Joel S. Demski

Managerial Uses of Accounting Information

Second Edition

Springer

MANAGERIAL USES OF ACCOUNTING INFORMATION
Second Edition

Springer Series in Accounting Scholarship

Series Editor:

Joel S. Demski
Fisher School of Accounting
University of Florida

Books in the series:

Christensen, Peter O., Feltham, Gerald A.
      Economics of Accounting - Volume I
           Information in Markets
Christensen, Peter O., Feltham, Gerald A.
      Economics of Accounting - Volume II
           Performance Evaluation
Ronen, Joshua, Yaari, Varda (Lewinstein)
      Earnings Management
           Emerging Insights in Theory, Practice, and Research
Demski, Joel S.
      Managerial Uses of Accounting Information, Second Edition

MANAGERIAL USES OF ACCOUNTING INFORMATION
Second Edition

by

Joel S. Demski
Fisher School of Accounting
University of Florida

Joel S. Demski
Fisher School of Accounting
University of Florida
333 Gerson Hall
Gainesville, FL 32611

to Millie

Contents

Preface

The second edition of *Managerial Uses of Accounting Information* reflects a decade of teaching this material to a variety of students, ranging from undergraduate sophomores to graduate students as well as a decade of growth in our understanding of information's role in an organization. While the spirit and intent of the first edition remain, the approach is (I hope) noticeably different. I have learned that a two pronged approach beginning with a stronger focus on fundamentals followed by a presentation of accounting as an artful rendering of those fundamentals is simultaneously clarifying and edifying. Staying slightly closer to the economics of cost, choice and contracting and removing minutia showcase the central ideas in a profoundly more clear fashion.

But as I said, the spirit and intent of the first edition remain. So it seems appropriate to repeat the message in the original Preface.

This book is an invitation to study managerial uses of accounting information. Three themes run throughout. First, the accounting system is profitably thought of as a library of financial statistics. Answers to a variety of questions are unlikely to be found in prefabricated format, but valuable information awaits those equipped to interrogate the library. Second, the information in the accounting library is most unlikely to be the only information at the manager's disposal. So knowing how to combine accounting and non-accounting bits of information is an important, indeed indispensable managerial skill. Finally, the role of a professional manager is emphasized. This is an individual with skill, talent, and imagination, an individual who brings professional quality skills to the task of managing.

This book also makes demands on the reader. It assumes the reader has had prior exposure to financial accounting, economics, statistics, and the economics of uncertainty (in the form of risk aversion and decision trees). A modest acquaintance with strategic, or equilibrium, modeling is also presumed, as is patience with abstract notation. The book does not make deep mathematical demands on the reader, but neither does it take mathematical shortcuts. An acquaintance with calculus and simple optimization is presumed. (Otherwise the many opportunities to use optimization software such as the Solver routine in Excel will be less than fully digested.) The major prerequisite is a tolerance for (if not a predisposition toward) abstract notation.

This style and list of prerequisites are not matters of taste or author imposition. The study of accounting is serious business; it demands an ability to place accounting in a large environment, complete with uncertainty, strategic considerations, and a fuzzy demarcation between the organization and its environment. A professional quality manager has this ability, and the study of accounting at the level of serious professional encounter demands no less. This is the nature of the subject. To ask less of the reader is to denigrate the art of professional management and to limit unjustly our exploration.

That said, as before, you will find the book purposely void of color and gratuitous photos. Our subject matter is too intellectually intriguing and too important to be treated otherwise.

Intellectual debt in any undertaking of this sort is enormous. You will find selected references at the end of each chapter, chosen to provide a sample of the breadth and depth of the debt in this particular case. I have also tried, in offering these references, to provide a sense of the historical development of this body of knowledge. More personally, I owe a deep intellectual debt to Jerry Feltham, Chuck Horngren, David Kreps, Carl Nelson, David Sappington and Bob Wilson. John Christensen, John Fellingham and especially Sybil Bartel, Haijin Lin and Rick Young provided invaluable encouragement, reading and guidance in the creation of this manuscript.

My largest debt, though, is to my wife Millie. Her constant encouragement, counsel and support have made my studies, my academic career and this project possible and enjoyable.

Joel S. Demski
Fernandina Beach, Florida

1

Introduction

This book invites the reader to study how accounting information is used in the management of an organization. It is a book that deals with accounting; yet the central feature is using the accounting, as opposed to doing the accounting. Stated differently, our study of accounting adopts a managerial perspective. It stresses use of the various accounting products, not their production. Emphasis is placed on use by a well-prepared and responsible manager.

Our study is inspired by the complexity and subtlety of two seemingly innocuous questions: *what might it cost*, and *did it cost too much*? Imagine some important decision the organization is facing, such as introduction of a new product. Following the underlying decisions through their natural cycle, the early phase would be concerned with, among other things, what various product design options might cost. Subsequently, with the chosen design in place and in production, we turn to evaluating the managers. Here, among other things, the evaluation would address how much the product actually cost. What might it cost and did it cost too much turn out to be natural, important, recurring questions, questions with a distinctly accounting flavor.

Reigning folklore suggests these are two ways of asking the same basic question, and that the answer is to be found in a well designed accounting system. Yet as comforting as the folklore is, it invites a serious misidentification of how accounting information is used.

What we mean by the cost of something is not a unique, knowable datum. Rather, it depends on the economic circumstances at hand and on how the underlying decision at hand has been framed. There simply is not

a straightforward or unique answer to the question "what might it cost?" Likewise, when asking whether it cost too much we are engaging in performance evaluation, an inherently retrospective activity. But information that is useful in making a decision is not necessarily useful in evaluating the manager who made that decision, and vice versa. Life, fortunately, is much more complex. And this is what opens the door for a serious study of managerial uses of accounting information.

This preamble is, in fact, code for a particular philosophy and approach to the study of accounting. Briefly, accounting is one of many information resources at the disposal of the professional manager. It is a highly useful, sophisticated, and adaptable resource. Used with skill, it can be of considerable value. Used without skill, it can lead to devastating and embarrassing errors. How to use the accounting resource is our focus. There is a temptation to think in terms of rules, recipes, and handy guidelines for this purpose. Yet this is the antithesis of the philosophy and approach expounded here.

Rules, recipes, and handy guidelines for how to use the accounting products are crutches for the less than well-prepared and not so responsible manager. Fortunately, managerial life is more interesting than that. The purposeful use of accounting is critically dependent on the circumstance at hand. The professional quality manager recognizes this and is prepared to add professional judgment to the exercise. Our study can help prepare the manager to make these judgments, but it cannot relieve the manager of their necessity.

The purpose of this introductory chapter is to expand on this theme and provide an overview of our study. Following a brief reminder of the typical array of accounting resources, we examine alternate approaches to the study of accounting. We then discuss the essential ingredients for the exercise to capture the crucial features of the implied managerial task. Finally, we outline the stages of our study.

1.1 Accounting Resources

The organization's accounting system provides a number of important resources. It provides a language. Accounting is often called the "language of business." Liabilities, net worth, bottom line, cost of goods sold, periodic income, and fund balance are all well-used, familiar terms. We often use the language of accounting to convey various facts about a corporation, a partnership, or proprietorship, a not-for-profit entity, a public sector entity, or an entire economy. The focus, in turn, might be the entity as a whole or some part thereof. The widespread use of accounting as a language should be abundantly clear from your prior study of financial accounting.

Accounting also provides a model of consequences. Every organization charts its progress, in part, with its financial statements. A personal debit card statement, GE's consolidated financial statements, and the University of Florida's current fund balances provide ready examples.

What might happen to earnings per share? What will opening the new plant do to our balance sheet? Is the (accounting) revenue less the (accounting) cost positive for this product? We often use accounting summarizations to help assess the financial progress of an organization. By implication, then, we often project what we think, expect, or even hope a future accounting summarization might look like if specified policies are pursued. Accounting provides a model of consequences.

The other side to this is that accounting provides a portrayal of the organization that others will see and use. To illustrate, competitors will be interested in one's public financial record, as will taxation agencies. The astute homeowner will inquire about the insurance company's financial health, just as the astute professor seeking greener pastures at a competitor will look into budget matters. Similarly, the astute competitor will study the financial strengths of its main competitors.

Finally, accounting is a repository of financial data. It is a well-maintained, structured, and defended financial library. The manager will often find useful information in the accounting library; and the accounting renderings of the manager's current activities will be deposited in the library.

This library metaphor pervades our study. We do not go to the usual library without an understanding of how the library is organized, nor do we expect to find off-the-shelf ready made answers to every inquiry we bring.[1] Similarly, we know of specialized libraries and have confronted the question of which library to query. Contrast the Law and Social Science Libraries at the University of Florida, for example.[2] We also know it is sometimes preferable to acquire information on personal account. Typically, we read our daily newspaper at home, without retrieving the newspaper from the library. Similarly, an efficient housing search would begin with a web-based inquiry of the posted options.

The same holds for the accounting library. The professional manager knows how this library is organized and maintained, and how to retrieve

---

[1]Emphatically, we don't simply Google the accounting library to learn what some product cost. The accounting library is not the sole source of insight; and it is constructed, as you will learn, in highly specific fashion, a fashion that may or may not speak to the decision at hand.

[2]An important advantage of the accounting library is its reliability. Serious effort is given to defending it against error, or worse. Care is taken to record events with considerable accuracy. Of course, this means some types of information are delayed (or not admitted). Revenue is not recognized when the customer announces an intent to purchase, even though this may be a remarkable, euphoric piece of good news. Rather, revenue is recognized at a later stage, at a time when the veracity of the claim can be better or more easily verified.

information from it. The professional manager also knows what types of information are likely to be found in the accounting library, and how to combine that information with information from other sources, including those sources that are personally maintained.

The professional manager is, among other things, a skilled user of the accounting library. This skill is the focus of our study.[3]

1.2   Modes of Study

This brings us to the question of how best to study the art of using the accounting library. One method might be labeled the "imperative." The idea is to decree or divine how the accounting should be performed and used. This is how revenue should be measured, this is how product cost should be measured, this is how performance relative to budget for the division manager should be measured. All are expressions of this philosophy.

While admittedly a red herring, it is worthwhile at the outset to dispense with the imperative theme. At one level it creeps in when financial reporting is encountered. This is a consequence of regulation. GAAP requiring this or that treatment is a common theme. This subtly shades into an imperative. After all, while accounting can be confusing, we can at least rely on GAAP to give it structure. GAAP is comfortable in this regard; it implies a widely applicable, correct answer to the question of how the accounting should be done.

At another level, I, personally, have found this imperative mode depressingly endemic to accounting. My students are usually frustrated and disappointed when "good" accounting is not identified. They seem to want a correct answer, an imperative. (It is one thing to identify a correct calculation of product cost, given an announced algorithm, and quite another to pick the algorithm.) After all, GAAP itself is all about learning and following the rules of a regulatory agency. Yet these same students would be sadly disappointed if their economics professor advocated the best allocation of a family's budget without reference to tastes, opportunities, and prices. Family budget allocations are influenced by economic forces, and the same goes for accounting.

So it should be made clear at the start: we will treat the accounting library as one among many resources at the manager's disposal. It is an economic resource. How best to construct it and how best to use it depend in such critical fashion on the circumstance at hand that general guidelines

---

[3]A corollary observation is the professional manager has a responsibility to help manage the accounting library. The acquisition policy at the public library is guided by consumer tastes, and we expect no less for the accounting library. From the manager's perspective, the accounting library is one more resource to be efficiently used and developed.

and rules of thumb are not available. Professional judgment is required, in the same way that it is required when a new product is launched, when an R&D project is contemplated, or when an evaluative conference with a subordinate is being planned.

If not adopting the "imperative school," then how are we to proceed? Another alternative is the codification approach. Here we document practice, including the latest consulting products, looking for commonalities and so-called best practices. Variety is to be expected. For example, municipalities tend to use recognition rules that formally record a purchase order as an expense. This is done to keep detailed track of commitments because spending limits are strictly enforced. The commercial organization uses a slower recognition rule but also keeps close track of purchases in its cash management operations. Similarly, hospitals tend to use elaborate product costing systems, while airlines do not think in terms of the cost of serving an individual customer. Of course, the hospital faces de facto cost-based pricing[4] while the airline adopts more of a system or network view of its products.

Here we run the risk of being overwhelmed by detail, and not taking care to identify and document what forces are shaping the accounting products. We invite a bias toward the status quo, and sidestep the question of what distinguishes a best from a less than best practice. Today's best practice is worthy of scrutiny and imitation. Yet our task extends from today to tomorrow to well beyond tomorrow.

The remaining interesting alternative is a conceptual approach. This emphasizes an image, a mental image, of the library and circumstance at hand. Several advantages follow. Our image must combine library and circumstance. We are therefore forced to provide a conceptual or generic description of a typical accounting library. This we will do in terms of aggregation, well chosen approximations to the organization's cost curve, and judicious use of cost allocation. We are also forced to provide a conceptual or generic description of circumstance. This we will do, in terms of other products, other activities, other sources of information and competitors in the product market, all of which impinge upon the managerial activity at hand.

This approach also allows us to treat the accounting library as an economic resource. We think of it, abstractly, as producing benefits for a cost. Yet this serves more to place words on an important managerial judgment than to inform that judgment.

---

[4]So-called DRG (diagnostic related group) categories have been used by Medicare to set reimbursement schedules. In turn, these prices are informed by product cost calculations; and negotiations with major commercial insurance carriers are informed by DRG prices and cost statistics. Hospitals did not install elaborate product costing schemes until the advent of these cost-based pricing procedures.

The conceptual approach also has its disadvantages. It forces us to mix accounting procedure and circumstance. Accounting procedure by itself could fill several books. It also forces us to think in terms of a small, parsimonious model of accounting and circumstance; otherwise we become overwhelmed with detail. It is also not easy. Studying methods of accounting is an inherently easier task. It is not open ended, and correct answers are readily verified (versus readily constructed).

Our approach, then, is conceptual.[5]   This forces us to focus on fundamentals, and offers the prospect of a clarifying perspective. Efficiently dealing with fundamentals, however, leads to a thematic approach that places demands on the reader. It demands patience and it demands tolerance for abstract notation. It also presumes familiarity with financial accounting (e.g., recording of transactions and the accrual process), economics (e.g., allocation of a budget in light of tastes and market prices and the profit maximizing view of firm behavior), and statistics (e.g., probability and regression). We will also make modest use of calculus and optimization and, as noted, abstract notation (in the form of sets and functions).[6]

1.3   Ingredients for an Interesting Stew

That said, this book mixes several essential ingredients to bring out central features of the accounting landscape. A first ingredient is *uncertainty*. We routinely admit uncertainty. The reason is we want the accounting measurements to tell us something. This implies there is something we don't know. Not knowing something is modeled as uncertainty. Where possible we will suppress uncertainty, but only to develop our theme as efficiently as possible. For example, uncertainty will not play a major role when we study the manner in which product costs are calculated. Subsequently, when we study how one might extract data from the accounting library to estimate a product cost, uncertainty will play a central role. Otherwise, by definition, we would have nothing to estimate.

---

[5] The conceptual orientation should be distinguished from a theoretical study. A theoretical study would begin with first principles and deduce various implications, such as the nature of a cost allocation scheme that has significant information content. Theory deals with underlying principles. It informs our study; indeed, references to the theoretical literature are provided at the end of various chapters. But our study is purposely structured to stay between the purely descriptive and the purely theoretical. The purely theoretical is too far removed from practice. The purely descriptive is too ephemeral.

[6] Luddites erroneously believed manufacturing machinery should be destroyed as it led to lower employment. A variation on this erroneous theory is that human capital in the form of economics, statistics, and so on should not be used in the study of accounting. To the contrary, economics, statistics, and so on make our study of accounting more productive and (to my mind) more exciting.

A second ingredient is *other sources of information*. It is important to understand and acknowledge that the accounting system does not have a monopoly on financial measurement or insight. We would not look to Homer for the answer to a mathematical question, just as we would not rely on our physician for insight into the market for satellite mapping services. Equally clear, we wouldn't look to the accounting system for something more readily available elsewhere. As humorous and as obvious as this appears, there is a deeper side. When multiple sources of information are available, they are often combined in highly unintuitive fashion. This will be particularly significant when we study performance evaluation in the light of various measures of performance.

A third ingredient is *multiple products or services*. A single product firm is just not a useful platform for our purpose. Literally, a single product story means the organization produced so many units of a good or service in a single time period and then closed down. The accounting is too easy. Accruals are irrelevant, as are interdependencies among products.[7]

A fourth ingredient is an assumed *model of behavior*. To put some structure on the idea that a manager is using the accounting measures, we are forced to say something about how the measures are used. For this purpose we will assume the manager is an economic agent. This means the manager's behavior is so consistent it can be described as if the manager had a utility function and selected from among alternatives so as to maximize that utility. Going a step further, we will assume this takes the form of expected utility maximization. This is done because the use of probabilities in the description allows us to say something about how information is used. In turn, this is critical to our venture, since we model accounting as providing information to and about the manager.

This behavior assumption, then, allows us to mix uncertainty, alternate sources of information, and the use of probabilities to govern the processing of information. This is useful and insightful. It is also costly. People are prone to systematic (and not so systematic) violations of the tenets of economic rationality, and we will invoke this at appropriate times in our study. Also, economic rationality is not too friendly to the view that one of the resources provided by accounting is a model of consequence. The economic actor comes ready equipped with a fully developed model of consequence. This schism, too, will be noted at appropriate points in our study.

---

[7]A personal computer manufactured in one period is a distinct economic product from the same personal computer manufactured in another period. The second exists at a different time, just as the resources used in its production were consumed at a different time. A single product firm has one product, in a single period setting. If we are worried about depreciation, for example, we have multiple time periods and therefore multiple products. This is why the economic theory of the single product firm has nothing to say about depreciation.

On the other hand, economic rationality has its advantages. Economic forces are hardly benign. Using them adds structure to our task; and, as the reader will discover, leads to significant, counterintuitive insights into informed professional use of accounting measurements.

A final assumption, nearly too obvious to mention, is that *accounting is not free*.[8] If accounting is costly, we should then expect its practice to reflect this fact; we should expect it to be less than perfect. The inevitable tensions between cost and quality should be controlling. Our study will routinely make use of less than perfect accounting measurements. This is reality. Accounting can always be improved, if one is willing to pay the price. Economic forces enter to stop us short of the best that is feasible. We will not explicitly dwell on this theme. It is implicit throughout the study.

1.4   Overview

Our study will focus on the two metaphorical questions of what might it cost and did it cost too much. We do this in four steps. Initially, in Chapters 2 through 7 we study product costing. In Chapters 8 through 12 we study managerial decision making with an emphasis on the "what might it cost" theme. In Chapters 13 through 18 we study managerial performance evaluation, with an emphasis on the "did it cost too much" theme. Chapter 19, the concluding chapter, provides a synthesis.

The pattern in each step along the way is to begin with fundamentals, and then introduce accounting, interpreted as an artful application of the fundamentals. In the product costing arena, then, we begin with the economic theory of the (single product) firm. This is the stepping-off point of our study. Many managerial concepts have their roots in economic theory. What we mean by product cost, for example, is rooted in the economic theory of the firm. Yet the single product orientation blinds us to interactions among products and across periods, so in Chapter 3 we extend the story to a multiproduct firm, where products might coexist in a given period or be phased across periods. Here we encounter an admonition that the only meaningful concept of product cost is marginal cost, an admonition that will guide us throughout our study. Also, as it is commonplace to refer to

---

[8]Literally, billions are expended each year on accounting for economic activity. Deeper, though, is the other side of the coin. Using accounting is costly. It takes skill, practice, and time. In addition, we humans are not expert at digesting large amounts of unstructured data. Predigested, codified, and summarized presentations are the norm. We should not make the mistake of presuming the best way to deal with accounting information is to collect and display as much as possible. Accounting aggregates data for a variety of reasons, one of which is our inability to process large amounts of data in an unstructured format.

these fundamentals as the theory of the firm, we will throughout our study refer to the organization of concern as a firm. This will serve as a gentle, recurring reminder of our fundamentals.

From here, in Chapter 4, we juxtapose accounting and the economic fundamentals. This portends a continuing theme of less than perfect measurement of economic concepts, thanks to less than perfectly functioning markets. Chapters 5 through 7 then bring the firm's financial data bank, or accounting library, into focus. Here the emphasis is on product costing. This is an important topic, and it serves as a vehicle to develop the library theme. We emphasize the typical accounting library makes judicious use of three building blocks: aggregation (as too much detail is overwhelming), cost curve approximation (as a more sophisticated cost expression is overbearing), and cost allocation. The same techniques are also used to measure cost incurred in a manager's department or division. We also emphasize the accountant's product costing as an artist's rendering of the fundamentals, moving, historically, from the impressionism school to the modernism school and its emphasis on activity based costing.

We then turn to the second step in our odyssey: managerial decision making, where the same approach emerges. In Chapters 8 through 10 we focus on the fundamentals of economic rationality, decision framing, and choice under uncertainty, followed by strategic choice. We then turn, in Chapters 11 and 12, to artful use of these fundamentals. Inherently small versus large decisions are stressed, where the distinction revolves around whether the decision is largely straightforward or highly nuanced in nature. Underlying these examinations are decision framing techniques that call for various expressions of product cost, our what might it cost theme. The accounting library is routinely helpful in these matters, but extracting its information, its clues, requires an understanding of how the library's data were put together (the above mentioned building blocks) and the particular decision frame we find comfortable.

This leads to the third step of evaluating the manager. Again we move from fundamentals to the accountant's palette. In Chapters 13 through 15 we examine a contracting setting where imperfect markets create a pay-for-performance arrangement between the firm and the manager. This, in turns, leads to an interest in what we mean by performance, and structures our theme of did it cost too much. We then turn to the accountant's rendering in Chapters 16 through 18, moving from single to multiple managers, coordination and divisionalized structures.

Chapter 19, as noted, concludes with an emphasis on synthesis.

1.5   Summary

This book offers an opportunity to study managerial uses of accounting in-
formation. Compared with financial accounting, the topic is inward looking;
it concerns managerial activities inside the firm. This is more pedagogical
than descriptive, however. The firm can hardly survive without paying close
attention to capital, labor, and product markets (not to mention govern-
mental activities). The study flows from product costing to decision making
to performance evaluation. This flow is designed to assemble all parts of
the puzzle in orderly fashion, and to emphasize the two thematic questions
of what might it cost and did it cost too much. The risk in the flow is that
the parts will be viewed more as separate entities than as building blocks
for a more delicate and interacting fabric.

   The study is also not separated from the realities of managerial life. We
readily assume a setting where multiple goods and services are available.
Uncertainty and multiple sources of information are also center pieces of our
study. We also assume the professional manager, the user of the accounting
information, responds to economic forces in a largely consistent fashion.

   Finally, the study is not separated from financial accounting. Exter-
nal and internal reporting activities share the same library. Management's
progress is, in part, judged by its financial reports; and governance of the
accounting library is influenced by the regulatory apparatus of financial
reporting.

   Read on!

1.6   Bibliographic Notes

It seems appropriate to begin with some historical perspective. Luca Pa-
cioli's *Suma de Arithmetica, Geometria, Proportioni, et Proportionalita*,
published in 1494, provided the first systematic description of the practice
of double entry record keeping, though accounting per se is much, much
older (Basu and Waymire [2006]). Cost accounting is largely the product of
the 19th century. For example, E. St. Elmo Lewis' third edition of *Efficient
Cost Keeping*, published in 1914, states the "... first edition ... was issued
in 1910 in response to what we believed to be a well-defined interest among
businessmen in cost finding." Clark [1923] provided the first comprehensive
treatment of costing. Solomons [1968] provides a delightful historical sur-
vey, while Cooper and Kaplan [1991] provide a more modern perspective.
Also, a helpful resource to keep in mind as we proceed is the two volume
handbook edited by Chapman, Hopwood and Shields [2007].

2

Economic Foundations: The Single Product Firm

The purpose of this chapter is to review several important ideas in economics. The firm operates in and is disciplined by markets, so we begin with the economist's notions of a market and market value. From there we move to the economist's portrayal of a firm as an institution that straddles factor and output markets. In this view, the firm uses market prices and its production function to decide what to produce and sell, and how to produce what it has chosen to produce. And it is at this point our study begins to take shape. Framing the firm's choice of output and input into revenue and cost components introduces us to the economic theory of cost. This is the foundation upon which the accountant's product costing art is built, a foundation that will guide us at every twist and turn of our journey. We extend this review in the next chapter to multiproduct firms.

This material is critical to our development. This is the foundation on which our notions of cost, revenue and income are rooted. You will find the going formal at times, and this is on purpose. Accounting is too important, too useful, and too intellectually fascinating to be shorted due to a shallow understanding of foundations.

2.1 Perfect Markets

A perfect market is a trade mechanism in which some fungible[1] item, such as a beverage, a transportation service, an hour of labor service, or an automobile, is tradable without restriction under known, constant terms of trade. This stylization is deceptively simple. Whatever the item, we know exactly what it is at the time of acquisition. We know the purity of the beverage, the reliability of the transportation service, the skill and motivation with which the hour of labor will be delivered, and the quality of the automobile. We also know the price of the item in question. We can purchase a fractional amount, no transaction costs of any kind are experienced, and no courts are necessary to enforce the terms of trade.

Some abstraction will drive the point home. Suppose trade is calibrated in a common currency, called dollars. Let q be the quantity of the item in question and P be the price expressed as dollars per unit. We know P; and q can be any real number. If $q > 0$ we pay Pq and receive q units. If $q < 0$, we receive $-Pq$ (Remember, the negative of a negative is positive!) and deliver q units. Naturally, we would not arrange to purchase $q > 0$ units if we did not have Pq dollars with which to pay the supplier, just as we would not promise to deliver $q < 0$ units if we did not have (or have access to) these units.[2]

Trade takes place without ambiguity or friction in a perfect market. If we have to ask what the price is, the market is not perfect. If the price per unit depends on how many units are involved, the market is not perfect. If we have to pay a broker to arrange the trade, the market is not perfect.

2.2 The Firm Straddles Markets

The firm now enters the story as an organization that stands between, that straddles, markets. The firm is more efficient than pure market arrangements at organizing production. The university is a ready example. Instead of daily prices for each and every class and graded assignment, we have a collection of policies and conventions, designed by those with decision rights and administered by bureaucrats, that govern such things as course offerings and schedules, degree requirements, and so on. A super market is another example, where we have a huge variety of products acquired, stocked and offered for sale, as opposed to a huge variety of individual vendors with a single spot market for each and every item.

---

[1]That is, freely exchangeable in whole or in part.

[2]This is one of the fictions of a perfect market. People actually pay their bills. Similarly, if you contract for the cable company to arrive at 10:00 am on Thursday, the technician actually arrives on the promised date at time. Fiction can be appealing.

We begin, however, with the fiction of a single product firm. This opens the door to our understanding of cost, at minimal complexity. Suppose, then, a firm is equipped to produce some good, say, pencils. Let $q \geq 0$ denote the quantity of pencils produced and sold. The quantity produced depends on what resources, called factor inputs, the firm uses and what production technology it possesses. Though factors come in endless variety, we develop the idea with but two, say labor and capital. Denote the two factor inputs by $z_1 \geq 0$ and $z_2 \geq 0$. (Since they are inputs, we require them to be non-negative, just as the output is required to be non-negative.)

In turn, how inputs can be transformed into outputs is catalogued in the firm's production function. Denote this function by $q = f(z_1, z_2)$. If inputs z_1 and z_2 are supplied, any output quantity between $q = 0$ and a maximum of $q = f(z_1, z_2)$ can be produced. We should think of the function $f(z_1, z_2)$ as providing a complete and reliable description of what the firm can produce. The following diagram emerges.

$$\boxed{inputs \ (z_1, z_2) \rightarrow \boxed{f(z_1, z_2)} \rightarrow output \ q}$$
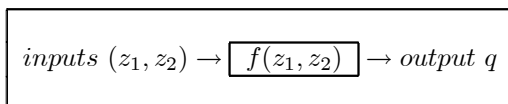
EXHIBIT 2.1

We naturally assume no free lunch, in the sense that zero input produces nothing other than zero output: $0 = f(0, 0)$.

Example 2.1 To illustrate, suppose $f(z_1, z_2) = \sqrt{z_1 z_2}$, again for $z_1, z_2 \geq 0$, and also that the technology must maintain $z_1 \leq 15$.[3] Thus, to produce, say, $q = 5$ units, the feasible possibilities consist of any pair of factors such that $5 \leq \sqrt{z_1 z_2}$, or $5^2 = 25 \leq z_1 z_2$ and $z_1 \leq 15$. So, for any feasible $z_1 > 0$ we require $z_2 \geq 25/z_1$.

What output and inputs does the firm choose? Recall that the firm straddles output and input markets; and it pays attention to the prices in those markets. Let $\widehat{P}$ denote the price per unit in the output market, P_1 the price per unit in the first input market and P_2 the price per unit in the second input market. All three markets are perfect.[4] The technical possibilities open to the firm are defined by the production function; and the market prices lead the firm to its profit maximizing choice.

Suppose the firm considers a production plan of q units of output, based on inputs of z_1 and z_2. Assume the plan is feasible, with $q \leq f(z_1, z_2)$. The firm will then receive $\widehat{P}q$ from customers in the product market and will

---

[3] Being a *function,* $f(z_1, z_2)$ assigns exactly one output quantity q to a given input list, $z_1 \geq 0$ and $z_2 \geq 0$. Also, this is an example of a Cobb-Douglas production function. The generalized version in the two factor case takes the form $q = z_1^\alpha z_2^\beta$ for non-negative exponents and factors of course. Notice the example uses $\alpha = \beta = \frac{1}{2}$.

[4] This is hardly necessary, but greatly simplifies our task.

pay a total of $P_1 z_1 + P_2 z_2$ to suppliers in the two factor markets. Its profit, or income, will be the net of receipts and payments: $\widehat{P}q - P_1 z_1 - P_2 z_2$. The firm chooses the feasible production plan with the largest profit.

Symbolically, we may describe its behavior as solving the following maximization problem.[5] Notice we denote the maximum profit by $\Pi(\widehat{P}, P)$, where $P = [P_1, P_2]$ is the listing of factor prices. The firm's maximum profit depends on the market prices it faces (as well as its presumably fixed technology).

$$\Pi(\widehat{P}, P) \equiv \max_{q \geq 0, z_1 \geq 0, z_2 \geq 0} \widehat{P}q - P_1 z_1 - P_2 z_2 \qquad (2.1)$$
$$\text{s.t.} \qquad q \leq f(z_1, z_2)$$

Viewed in this fashion, the firm possesses some exogenously specified technology that is recorded in its production function. It then takes price signals from the input and output markets and uses these signals to select the best production plan.

Example 2.2 To put words to this music, return to the above example and assume the selling price is $\widehat{P} = 40$ while the factor prices are $P_1 = 5$ and $P_2 = 20$. So we want to maximize $40q - 5z_1 - 20z_2$ subject to (1) $q \leq \sqrt{z_1 z_2}$ and (2) $z_1 \leq 15$, again for $q \geq 0$, $z_1 \geq 0$, and $z_2 \geq 0$. We readily find an optimal quantity of $q = q^* = 15$, along with respective factor choices of $z_1^* = 15$ and $z_2^* = 15$. Notice we designate optimal choices with an asterisk (*); this convention will be maintained throughout. The firm earns a profit of $40(15) - 5(15) - 20(15) = 225$.[6]

Two interpretive points will be important in subsequent developments. First, we have confined the exposition to two factors simply to avoid tedium. We should be thinking in terms of a large number of inputs, say $q = f(z_1, z_2, ..., z_m)$ where m is a large number. For example, imagine the different inputs in a modestly sized grocery store.

Second, the story we have sketched is a single period story. With more detail we would think in terms of units of output in each period, inputs of various kinds in each period, and profit defined via the present value of the resulting cash flow series. Many, many factors and a multiperiod

---

[5] One might ask whether this maximization actually has a solution. Here we side step various technical assumptions that would ensure a solution exists. These issues also extend to such niceties as the factors displaying diminishing marginal productivity. Likewise, if the production function is only defined for a given range of inputs, as in the first example, this is understood in our specification.

[6] The easiest way to verify this is indeed the solution is to use the optimization package in, say, Excel. Once you have verified the solution, try it again, but without the $z_1 \leq 15$ constraint. What happens? This is why we invoke $z_1 \leq 15$. Further notice our little firm earns strictly positive profit, implying it enjoys some type of market power. This particular assumption, of market power, is unnecessary but helps keep us focused on fundamentals.

orientation will turn out to be important elements in understanding the accountant's work. But that is getting ahead of the story.

A final point here concerns the nature of the maximization problem that we used to depict the firm's choice of output and inputs in (2.1). The essential ingredients in that exercise are the production function and the market prices. We completely solved the firm's problem without any reference to cost or revenue. This is an important lesson. Much of the data in any firm's financial data bank concerns the cost of various activities. It is possible to describe the firm's behavior economically with no explicit reference to cost. It is also possible to describe the firm's behavior with explicit reference to its cost. Different ways of framing a choice problem lead to different measures of cost. Cost is not a unique concept, either to the economist or the accountant.

2.3   The Economic Cost Function

To begin developing this theme, stay with the one output, two inputs story in (2.1). Fix the output at some feasible but otherwise arbitrary quantity q. Now define the cost of this output quantity to be the minimum factor payments that must be expended to produce q in light of the factor prices. Denote this minimum expenditure by $C(q; P)$ where, again, $P = [P_1, P_2]$ is the listing of factor prices[7]. We have the following construction.

$$C(q; P) \equiv \min_{z_1 \geq 0, z_2 \geq 0} P_1 z_1 + P_2 z_2 \qquad (2.2)$$
$$\text{s.t. } q \leq f(z_1, z_2)$$

Repeating this process for all possible output quantities gives us a cost function, denoted $C(q; P)$. Importantly, economic cost is the minimum expenditure on factors that will allow the firm to produce the quantity in question. We carry along the factor prices in the notation to remind ourselves that, in general, the mix of factors used to produce some level of output will depend on the factor prices. Moreover, with the factor prices given, output is the sole explanatory variable of the firm's cost.

A typical cost function based upon explicit factor prices might appear as displayed in Figure 2.1.[8]   Notice that cost is zero when quantity is zero, reflecting our earlier assumption that zero input implies zero output.

---

[7] You are probably wondering about the obsessive notation. But it will be important to remind ourselves that the mix of factors depends on their relative prices. Outsourcing is an example.

[8] The noted cost function in Figure 2.1 can be derived from technology and factor price specifications, but with a more complicated technology than presumed in our series of illustrations.

Beyond that, our graph depicts a situation where larger output always necessitates higher cost.
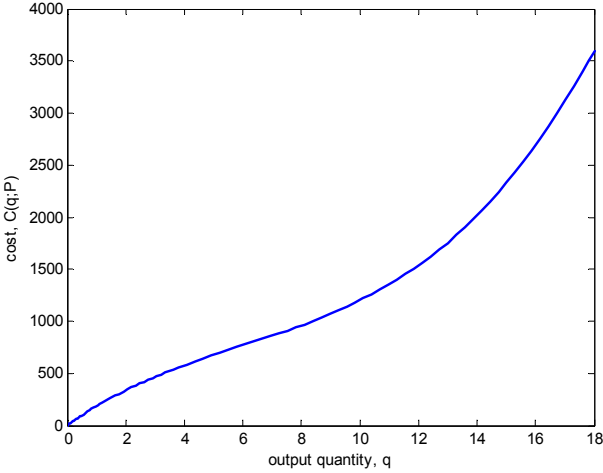
FIGURE 2.1. $C(q;P) = 200q - 18q^2 + q^3$

2.3.1   Cost Function Terminology

Various derivative (no pun) notions of cost are used at this juncture. Suppose we focus on some specific output level, say, q. $C(q;P)$ is the total cost of producing output quantity q, given the firm's fixed technology and given factor prices P. At this point, the important concepts of average, incremental and marginal cost surface. All deal with cost in relation to output quantity, holding factor prices constant.

Definition 1 *In the single product firm, average cost at output quantity $q > 0$ is total cost divided by quantity, or $C(q;P)/q$.*

Definition 2 *In the single product firm, the incremental cost of $\Delta > 0$ units at output quantity $q \geq 0$ is the difference between the cost of producing $q + \Delta$ units and q units, or $C(q + \Delta; P) - C(q; P)$.*

Definition 3 *In the single product firm, the marginal cost of output at output quantity $q \geq 0$, denoted $MC(q; P)$, is the rate at which cost changes with respect to change in quantity, or $MC(q; P) = \frac{\partial C(q;P)}{\partial q}$.*[9]

---

[9] Equivalently, marginal cost of quantity q is the slope of the line that is tangent to $C(q;P)$. Also notice that since we are holding P constant in the exercise we use a partial derivative in the definition of marginal cost.

Notice incremental and marginal cost are defined for any feasible quantity, while average cost (since we are dividing by quantity) requires a strictly positive quantity. Also, the incremental cost of one unit at output quantity q is $C(q + 1; P) - C(q; P)$. Many define this as marginal cost, and it is numerically "close" to marginal cost (and exactly equal to marginal cost if the cost curve is linear). While adequate for many purposes, this approximation will not be sustainable in all that follows, so we proceed from the beginning with the correct definition.

Example 2.3 Consider the example on which Figure 2.1 is based: $C(q; P) = 200q - 18q^2 + q^3$ (for some specific factor price specification). For $q > 0$, average cost is $C(q; P)/q = 200 - 18q + q^2$. Total and average cost are listed below in Table 2.1, for selected output quantities. Total cost, recall, is plotted in Figure 2.1. Now suppose $q = 9$. What is the incremental cost of one additional unit at this point? $C(10; P) - C(9; P) = 1,200 - 1,071 = 129$. Likewise, what is the incremental cost of $\Delta = 3$ more units, if $q = 5$? $C(5 + 3; P) - C(5; P) = 960 - 675 = 285$. Incremental cost, for $\Delta = 1$, is also listed in Table 2.1, along with marginal cost, calculated via $\frac{\partial C(q;P)}{\partial q} = 200 - 36q + 3q^2$.[10]

TABLE 2.1: $C(q; P) = 200q - 18q^2 + q^3$

output q	total cost $C(q; P)$	average cost $C(q; P)/q$	marginal cost $MC(q; P)$	incremental cost $C(q + 1; P)$ $-C(q; P)$
0	0	N/A	200	183
1	183	183	167	153
5	675	135	95	93
6	768	128	92	93
7	861	123	95	99
8	960	120	104	111
9	1,071	119	119	129
10	1,200	120	140	153

Average and marginal cost for this example are plotted in Figure 2.2. Notice how average cost is 183 for $q = 1$, declines to 119 (for $q = 9$), and then rises again. An economist interprets this as a region of economies of scale followed by diseconomies of scale. Further notice how marginal cost declines from 200 (at $q = 0$) to a minimum of 92 and remains below average

---

[10] A tormented person would now examine a larger Δ in the incremental cost construction and then ask you for the average incremental cost. Also notice that if you are handy with calculus you will be able to convince yourself that average cost is a minimum here when $q = 9$.
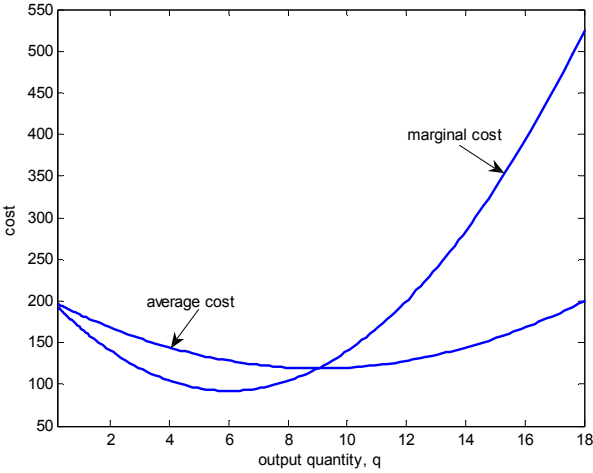
FIGURE 2.2. Average and Marginal Cost

cost until the two are equal at $q = 9$. When marginal cost is below average cost, average cost is declining. Average and marginal cost are equal when average cost is a minimum. Conversely, when marginal cost is above average cost, average cost is increasing.

2.3.2   A Closer Look at the Cost Function

This terminology is central to our work. However, in laying it out we worked with an illustrative cost function, without bothering to explicitly link the function to the firm's factor choices in (2.2) above. We now dig a bit deeper, because there is more to the story. For this purpose, glance back at the cost function definition (yes, in (2.2)). Intuitively, the factor choices depend upon how much output is to be produced as well as upon the market prices for those factors. Reflecting this in our notation, denote the optimal factor choices for quantity q (and factor prices P of course) by $z_1^*(q; P)$ and $z_2^*(q; P)$. Of course, cost is simply the total of what the firm spends on the *optimal* factor choices. So we have the equivalent cost expression of

$$C(q; P) = P_1 \cdot z_1^*(q; P) + P_2 \cdot z_2^*(q; P) \tag{2.3}$$

Two facts now emerge. First marginal cost is simply the sum of the factor prices multiplied by the rate at which optimal use of that factor changes with respect to output:

$$MC(q; P) = \sum_i P_i \frac{\partial z_i^*(q; P)}{\partial q} \tag{2.4}$$

Importantly, marginal cost is all about marginal factor consumption. Also, remember in this calculation that we are holding factor prices constant. This leads to the second important fact hidden in (2.3). The rate at which cost changes with respect to a factor price is nothing other than the optimal factor consumption:[11]

$$\frac{\partial C(q;P)}{\partial P_i} = z_i^*(q;P) \tag{2.5}$$

Intuitively, the rate at which cost increases with respect to some factor's price, say energy, depends on how much of that factor is being used.

Example 2.4 This is getting a bit out of hand, so let's look at an example. Here we use the technology in Example 2.1, but absent the extra constraint on the first factor. So technology is completely specified by $q \leq \sqrt{z_1 z_2}$. Cost, now, is defined by

$$C(q;P) \equiv \min_{z_1 \geq 0, z_2 \geq 0} P_1 z_1 + P_2 z_2$$
$$\text{s.t. } q \leq \sqrt{z_1 z_2}$$

Here it is easy to verify the optimal factor choices are given by $z_1^*(q;P) = \sqrt{P_2/P_1} \cdot q$ and $z_2^*(q;P) = \sqrt{P_1/P_2} \cdot q$.[12] As we would suspect, the factor choices depend on how many units are to be produced and their prices. If, for example, the first factor carries a relatively low price we use relatively more of it in producing the output. This is why we express the factor choices as depending on quantity and factor prices. And plugging this into (2.3) we have a cost curve of $C(q;P) = 2\sqrt{P_1 P_2} \cdot q$.[13] From here you should notice the cost function is linear, exhibiting a constant marginal cost of $2\sqrt{P_1 P_2}$. It is a classic case of constant returns to scale.

---

[11] This is so important it carries its own name: Shepard's Lemma. Notice you can derive it simply by differentiating (2.3) with respect to one of the factor prices. That said, this device is the beginning point for analyzing the potential impact of selected factor price changes.

[12] To verify this solution, concentrate on some $q > 0$ (as $q = 0$ clearly calls for zero inputs). Notice this requires strictly positive amounts of each factor. Moreover, we will have $q = \sqrt{z_1 z_2}$, as $q > \sqrt{z_1 z_2}$ is infeasible and $q < \sqrt{z_1 z_2}$ is wasteful. Now suppose the first factor is set at some positive amount, z_1. Then producing q units requires $z_2 = q^2/z_1$ (as this implies $\sqrt{z_1 z_2} = \sqrt{z_1(q^2/z_1)} = q$). With this substitution we want to minimize $P_1 z_1 + P_2 q^2/z_1$. Setting the derivative equal to zero yields $P_1 - P_2 q^2/z_1^2 = 0$, which implies $z_1^2 = P_2 q^2/P_1$, or $z_1 = \sqrt{P_2/P_1} \cdot q$. In turn, this implies $z_2 = q^2/z_1 = \sqrt{P_1/P_2} \cdot q$.

[13] Now try your hand at verifying (2.4) and (2.5):

$$MC(q;P) = \sum_i P_i \frac{\partial z_i^*(q;P)}{\partial q} = \sqrt{P_1 P_2} + \sqrt{P_1 P_2} = 2\sqrt{P_1 P_2}$$

and

$$\frac{\partial C(q;P)}{\partial P_1} = \sqrt{P_2/P_1} \cdot q$$

Example 2.5 It will prove useful in what follows to extend this example slightly. Keep everything as specified, except add an upper bound on availability of the first factor: $z_1 \leq \bar{z}_1$ where $\bar{z}_1$ is the presumed upper bound. Proceeding as before, cost is now defined by the following program:

$$C(q; P) \equiv \min_{z_1 \geq 0, z_2 \geq 0} P_1 z_1 + P_2 z_2 \tag{2.6}$$

$$\text{s.t. } q \leq \sqrt{z_1 z_2}$$

$$z_1 \leq \bar{z}_1$$

It turns out the optimal factor choices are as displayed in Table 2.2. (See the Appendix!) Notice that when output is sufficiently low (so the bound on the first factor is not an issue), we return to the setting of Example 2.4. And this pattern holds until the first factor is no longer productive or usable at the margin, i.e., until the upper bound of $\bar{z}_1$ is reached.[14] At that point we are forced to use an "inefficient" mix. Putting it all together, we then price these optimal factor choices at their respective factor prices, expression (2.3), to exhibit the firm's cost curve. See Table 2.2. From here, the marginal cost follows, as implied by (2.4). Finally, we also exhibit the effect of raising the bound on the first factor. This is, of course, nil when the bound does not matter; beyond that point, however, raising the bound lowers the cost, as a higher bound allows for more efficient mixing of the factors at that point.

TABLE 2.2: Cost Illustration ($\phi = \sqrt{P_2/P_1}$)

	$0 \leq q \leq \bar{z}_1/\phi$	$q > \bar{z}_1/\phi$
$z_1^*(q; P)$	$\phi \cdot q$	$\bar{z}_1$
$z_2^*(q; P)$	q/ϕ	$q^2/\bar{z}_1$
$C(q; P)$	$2\sqrt{P_1 P_2} \cdot q$	$P_1 \cdot \bar{z}_1 + P_2 \cdot \frac{q^2}{\bar{z}_1}$
$MC(q; P)$	$2\sqrt{P_1 P_2}$	$2P_2 q/\bar{z}_1$
$\frac{\partial C(q;P)}{\partial \bar{z}_1}$	0	$P_1 - P_2\left(\frac{q}{\bar{z}_1}\right)^2 < 0$

Example 2.6 Continuing, suppose the factor prices in Example 2.5 are $P_1 = 5$ and $P_2 = 20$, which implies $\phi = \sqrt{P_2/P_1} = \sqrt{20/5} = 2$. Further assume a capacity limitation on the first factor of $\bar{z}_1 = 15$ units. This provides the numerical version of Table 2.2 that is displayed in Table 2.3.

If, then, we set $q = 7$, marginal cost is 20 and total cost is 140. But if we set $q = 12$, where we are using an "inefficient" mix of factors, marginal cost increases from 20 to 32 while total cost increases from 140 to 267.

---

[14] This occurs at an output of $q = \sqrt{P_1/P_2} \cdot \bar{z}_1$ and can be verified by setting $z_1^*(q; P) = \sqrt{P_2/P_1} \cdot q = \bar{z}_1$.

TABLE 2.3: Cost Illustration with $P_1 = 5, P_2 = 20, \overline{z}_1 = 15$

	$0 \leq q \leq 7.5$	$q > 7.5$
$z_1^*(q; P)$	$2q$	15
$z_2^*(q; P)$	$q/2$	$q^2/15$
$C(q; P)$	$20q$	$75 + \frac{4}{3}q^2$
$MC(q; P)$	20	$\frac{8}{3}q$
$\frac{\partial C(q; P)}{\partial \overline{z}_1}$	0	$5 - \frac{20}{225}q^2$

2.4   Shadow Prices

Our exploration to this point opens the door to another important insight, one that will play heavily in our work that follows. This is the idea of a shadow price.

Suppose we are looking for the minimum of some objective function, say $f(x)$, but subject to a constraint that some other function of x, call it $g(x)$ not fall below some amount denoted b (for bound). So our optimization is limited by the constraint $g(x) \geq b$. Denote the optimal value of x by x^* and the optimal value of the objective function by $f(x^*)$. Now, the shadow price on the constraint measures the rate at which $f(x^*)$ changes with respect to b. Popular optimization packages, such as that in Excel, routinely report shadow prices.[15]   And in an approximate sense we can think of the shadow price as indicating how much the optimal objective function will change if we change the constraint by a single unit.

Shadow prices are important for pragmatic reasons and for the clarity they bring. On the pragmatic side, suppose we are facing a constrained optimization and one of the constraints has a relatively large shadow price. That suggests we look into whether the constraint can be relaxed, via subcontracting, redesigned workflow, or what have you. This pragmatic device will show up repeatedly in subsequent chapters.

On the clarity side, glance back at the cost function expression in (2.6). Notice we are minimizing subject to two constraints. Using the setting in Table 2.3, we find the shadow prices on the two constraints for representative output choices reported in Table 2.4. (You should verify this with your favorite optimization package.)

Notice the shadow price on the $z_1 \leq 15$ constraint is 0 when the upper bound of 15 is not constraining, but is -7.8 at an output of $q = 12$ units, where it is constraining. The rate at which cost changes here with respect

---

[15] The usual caveat of presuming suitable regularity conditions are satisifed applies here. Shadow prices are often called dual variables or Lagrange multipliers.

to that upper bound is -7.8. The rate is negative because *increasing* the upper bound *decreases* total cost. Increasing the upper bound one unit (i.e., from 15 to 16 units) will lower total cost by about 7.8.[16]

TABLE 2.4: Factor Details for $P_1 = 5, P_2 = 20, \overline{z}_1 = 15$

output	cost	z_1^*	z_2^*	shadow price on $z_1 \leq 15$	shadow price on $q \leq \sqrt{z_1 z_2}$
$q = 7$	140	14	3.5	0	20
$q = 12$	267	15	9.6	-7.8	32

More interesting is the shadow price on the $q \leq \sqrt{z_1 z_2}$ constraint. Notice it is 20 when $q = 7$ and 32 when $q = 12$. Now reread the last sentence in Example 2.6. Is this an artifact? Marginal cost is nothing other than the shadow price on the technology constraint, the rate at which the total of expenditures (on factors) changes with respect to output quantity.

TABLE 2.5: Shadow Prices for General Case

output	shadow price on $z_1 \leq \overline{z}_1$, or $-z_1 \geq -\overline{z}_1$	shadow price on $q \leq \sqrt{z_1 z_2}$, or $\sqrt{z_1 z_2} \geq q$
$0 \leq q \leq \frac{\overline{z}_1}{\phi}$	0	$2\sqrt{P_1 P_2}$
$\frac{\overline{z}_1}{\phi} < q$	$P_2 \left(\frac{q}{\overline{z}_1}\right)^2 - P_1$	$\frac{2P_2 q}{\overline{z}_1}$

Shadow prices and their important connection to marginal cost are explored further in the Appendix. There you will notice we frame the constraints so the shadow prices are non-negative. And for the record, Table 2.5 displays the shadow prices in general fashion, in non-negative frame, for our running illustration. Though awkward at times, being careful to frame the optimization so we we have non-nagative shadow prices will be an advantage in what follows.[17]

---

[16] If you work out the details, think Excel, you will find this unit increase in the upper bound lowers total cost at $q = 12$ from **267** to **260**, or by **7**. The world is not linear, and neither is our example.

[17] By convention, the shadow prices on the inequality constraints are presented in non-negative fashion. As developed in the Appendix, for a minimization problem this requires the constant be on the right had side and the inequality be stated in "greater than or equal to" format. So in Table 2.5, the first constraint is modeled as $-z_1 \geq -\overline{z}_1$ while the second is modeled as $\sqrt{z_1 z_2} \geq q$. As mentioned, this becomes awkward at times, but provides a consistent approach. Thus, the (non-negative) shadow price on the $\overline{z}_1$ constraint actually refers to the rate at which the objective function varies with

2.5   Cost and Revenue Framing

Now return to the original problem of locating the firm's best production plan, consisting of an output quantity and a set of inputs, as symbolized in (2.1). For a given quantity, q, the firm's revenue is $\widehat{P}q$, and its cost is given by the cost function, $C(q; P)$. Profit, of course, is revenue less cost. Maximizing profit now leads to the following depiction of the firm's problem:

$$\max_{q \geq 0} \widehat{P}q - C(q; P) \tag{2.7}$$

The same decision is made as in the original formulation in (2.1). All we have done is frame the analysis in revenue and cost terms. Moreover, differentiating the profit expression provides us with the time honored marginal revenue (MR) equals marginal cost (MC) condition for optimality:[18]

$$\frac{\partial \widehat{P}q}{\partial q} - \frac{\partial C(q; P)}{\partial q} = 0$$

The important observation here is we have decomposed the firm's problem into two stages. In the first stage we solve the input question by constructing the firm's cost curve, $C(q; P)$. This is the minimum expenditure, or most efficient set of inputs, necessary to produce a specified output. Second, we search over the output possibilities to locate the output level with maximal profit, or maximum of revenue less cost.

One way to frame the firm's problem calls for no explicit notion of product cost. We merely focus directly on inputs and outputs. This is the framing approach used in (2.1). Another way to frame the firm's problem is to focus on revenue and cost of output. This calls for an explicit cost function. Here we relegate input choice to the cost function construction exercise. This is the framing approach used in (2.7). The analyses in (2.1) and (2.7) are equivalent. They both locate the firm's best choice. Alternative, though equivalent, ways to frame a decision will be a recurring theme in our study. The marginal revenue equals marginal cost idiom rests on a particular way of framing the firm's profit maximization problem.

Example 2.7   This equivalence, in fact, is readily demonstrated in our running example (Examples 2.2 and 2.6) where the selling price is $\widehat{P} = 40$ and the respective factor prices are 5 and 20. In Example 2.2 we identified

---

respect to $-\overline{z}_1$. That is, increasing $\overline{z}_1$ is the opposite of increasing $-\overline{z}_1$. Again, see the Appendix.

[18] With output and factor prices given, the expression in (2.7) gives us profit as a function of output q. The first order condition for a maximum is that the first derivative with respect to q vanish, or that marginal revenue equal marginal cost. (Regularity conditions, in the form of smooth and well-behaved functions, are required to guarantee the first order condition tells the whole story.)

optimal choices of $q^* = z_1^* = z_2^* = 15$. In Example 2.6 we derived the firm's cost curve and associated marginal cost expression of

$$MC(q; P) = \begin{cases} 20 \text{ if } 0 \leq q \leq 7.5 \\ \frac{8}{3}q \text{ if } q > 7.5 \end{cases}$$

With a selling price of 40, revenue is given by $40q$ of course, so we now want to maximize

$$40q - C(q; P)$$

Differentiating this expression, and noting we will be in the range where $q > 7.5$ we have

$$40 - \frac{8}{3}q = 0$$

which implies $q^* = \frac{3}{8}(40) = 15$. At this point, revenue totals $40(15) = 600$, cost totals 375 (See Example 2.6), and we are back to a profit of 225 (and respective factor choices of 15 and 15 units).

The firm in Example 2.7 (or Example 2.2), then, earns a strictly positive profit, or *economic rent*. Put differently, economic rent arises when the firm earns more than necessary to compensate all factors of production, including capital. In a single period setting this reduces to having a strictly positive profit. In a multiperiod setting it is associated with a higher than required return on capital.

2.6   Short-Run versus Long-Run Cost

To this point the firm has been free to vary its factors, subject to whatever is allowed by its technology. This is a *long-run* setting, one where all factor choices are open. The central idea in such a setting is the firm is free to vary its factor inputs at will. This is why we insisted that zero input produces nothing other than zero output: $0 = f(0,0)$. And this, in turn, guaranteed that the cost of zero output is, well, zero: $C(0; P) = 0$.

In the short-run, the firm can only vary some of its inputs. We illustrate the effect by assuming the first factor is fixed at some amount $z_1 = \widehat{z}_1$. Of course, we could envision many versions of the short-run, depending on which factors are fixed and at what amounts.

With this input so fixed, the firm's technology is specified by the usual production function but with a constraint reflecting the fixed factor or factors. From this point, construction of the short-run version of the cost function proceeds in the usual fashion. We simply define the short-run cost as the minimum expenditure on resources that will make it possible to produce the output in question, given the technology and given the noted

factor constraint or constraints. For our simple story we have the following:

$$C^{SR}(q; P) \equiv \min_{z_1 \geq 0, z_2 \geq 0} P_1 z_1 + P_2 z_2 \tag{2.8}$$
$$\text{s.t. } q \leq f(z_1, z_2)$$
$$z_1 = \widehat{z}_1$$

Notice we denote the short-run version of the cost curve by $C^{SR}(q; P)$. It will be understood from the context which factor or factors are fixed. (This saves a notational assault on our senses.)

In short, $C^{SR}(q; P)$ is constructed in precisely the same fashion as its long-run counter part, $C(q; P)$, except we constrain (in this case) the first factor to its fixed amount of $z_1 = \widehat{z}_1$. This implies $C^{SR}(q; P) \geq C(q; P)$. The two cost curves are equal when (and if) q is such that the minimization solved to construct $C(q; P)$ selects $z_1 = \widehat{z}_1$.

Terminology enters at this point as well. We could, to be sure, beleaguer you with short-run average, short-run marginal, and short-run incremental cost. But these constructs follow directly from our earlier definitions. More important is the notion of fixed cost. In the long-run, we have zero cost at zero output, $C(0; P) = 0$. Not so in the short-run. Glace back at (2.8). What is the short-run cost when $q = 0$? We clearly set $z_2 = 0$, but by definition are stuck with and therefore paying for $z_1 = \widehat{z}_1$. At $q = 0$, the solution to (2.8) provides $C^{SR}(0, P) = P_1 \widehat{z}_1$. Naturally, the fixed cost depends on which factors are fixed and at what levels they are fixed.

We also speak of variable cost in this context. Total variable cost at output level q is total short-run cost less fixed cost, or $C^{SR}(q, P) - C^{SR}(0; P)$. Average variable cost may now be calculated. Fixed and variable are sufficiently important concepts that we add them to our list of formal definitions.[19]

Definition 4 *In the single (or multi-) product firm, fixed cost is the short-run cost of zero output, or $C^{SR}(0; P)$ for some defined short-run setting.*

Definition 5 *In the single (or multi-) product firm, variable cost of output q is the short-run cost of producing output q less the corresponding fixed cost, or $C^{SR}(q; P) - C^{SR}(0; P)$ for some defined short-run setting.*

Example 2.8   To illustrate, return to Example 2.3 (and Figure 2.1), where we presumed factor prices were such that the (long-run) cost curve was given by $C(q; P) = 200q - 18q^2 + q^3$. Now suppose one of the factors is irrevocably set in anticipation of producing $q = 9$ units, (this will become clear in a moment) and that the short-run cost curve is given by $C^{SR}(q; P) = 162 + 204.5q - 25q^2 + 1.5q^3$. So the fixed cost is, yes,

---

[19] Moreover, short-run marginal cost equals marginal variable cost, just as short-run incremental cost equals incremental variable cost.

$C^{SR}(0, P) = 162$. Total and average short-run cost are listed in Table 2.6 for selected output levels. Short-run and long-run total cost are plotted in Figure 2.3. Figure 2.4 provides a plot of long-run and short-run average cost.

Notice that $C^{SR}(q; P) > C(q; P)$ for all output levels except $q = 9$. This reflects our assumption that one of the factors is fixed in anticipation of producing $q = 9$ units. In turn, this implies average short-run cost exceeds average long-run cost at all points, except $q = 9$ where they are equal.

Example 2.9 To provide a second illustration, one where the factor choices are explicit, return to Example 2.6. Further suppose the firm has $z_1 = 10$ units in place, and this supply cannot be altered. This implies the firm can only vary the second factor, and the (hopefully familiar) technology thus implies $z_2 = q^2/10$ units are required to produce output quantity q. So we have a specific short run-cost curve of $C^{SR}(q; P) = 5z_1 + 20z_2 = 5(10) + 20(q^2/10) = 50 + 2q^2$. And the (short- run) fixed cost is $C^{SR}(0; P) = 50$. In Figure 2.5 we plot this particular short run cost curve and its long run parent, $C(q; p)$ derived in Example 2.6.

TABLE 2.6: $C^{SR}(q; P) = 162 + 204.5q - 25q^2 + 1.5q^3$

output q	total cost $C^{SR}(q; P)$	average cost $C^{SR}(q; P)/q$	marginal cost $MC^{SR}(q; P)$
0	162	N/A	204.5
1	343	343.0	159.0
5	747	149.4	67.0
6	813	135.5	66.5
7	883	126.1	75.0
8	966	120.8	92.5
9	1,071	119.0	119.0
10	1,207	120.7	154.5

A special case occurs when the short-run cost curve is linear: $C^{SR}(q; P) = F + vq$. In this case the fixed cost is F, the short-run marginal cost is the constant v, the average variable cost is v, and the incremental short-run cost of an additional unit (regardless of q) is v. We mention the linear case because accountants usually approximate the firm's cost curve with a linear cost function; and it is important to distinguish the economist's theory from the accountant's art.[20]

---

[20] Two additional points emerge here. First, a semantic qualification is in order. Common usage is to call a function of the form $F + vq$ linear, though strictly speaking it
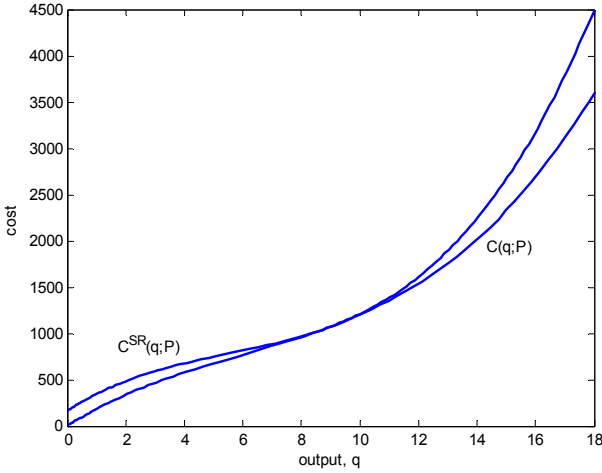
FIGURE 2.3. Short-Run and Long-Run Cost Curves

This completes our survey of important matters in the economic theory of the single product firm, including cost function terminology. Understanding the firm's cost function, or the behavior of its costs, is an important task. Specialized language has evolved to aid the manager in this task. It is important to remember, however, that a short-run cost analysis is idiosyncratic, as it depends on which specific factors are assumed unalterable. The long-run cost curve possesses no such ambiguity; all factors are alterable in the long-run. But then, how long is the long in long-run?

2.7 Summary

Our review of the economist's view of the single product firm stresses the notion that the firm has productive opportunities, catalogued in a production function. It exploits these opportunities by straddling input and output markets, to maximize its profit. The firm is a mechanical enterprise in this view. It has no control problems, no imagination, no entrepreneurial spirit, and no professional management. It has markets and a production function.

---

is linear only if $F = 0$. (Otherwise it is affine.) Second, if the cost curve is linear, we have a particular problem describing the firm's behavior. If $\widehat{P} > v$, the firm can make arbitrarily large profit by letting q be arbitrarily large. And if $\widehat{P} = v$, we have massive indifference. We resolve this dilemma by putting additional restrictions on the production function, as illustrated by the now familiar $z_1 \leq \overline{z}_1$ restriction.
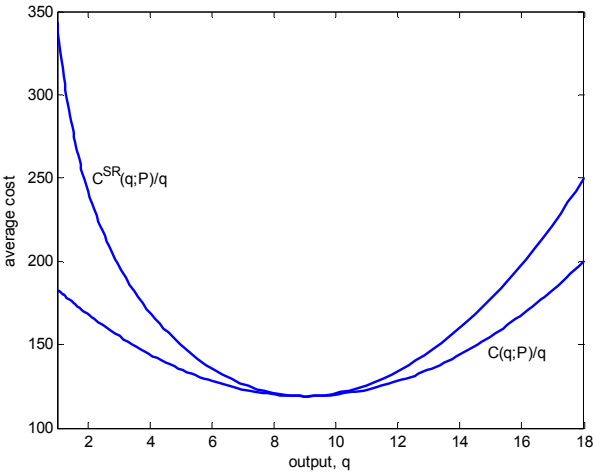
FIGURE 2.4. Short-Run and Long-Run Average Cost Curves

A theory, however, is designed to focus on central features. In our case, the central feature is cost. The firm might frame its problem of selecting an optimal production plan in many ways. Various ways of framing this choice lead to various notions of cost. These notions of cost and the associated terminology will be essential in subsequent development. Total cost, incremental cost, average cost, and marginal cost are important. Also important are distinctions between long-run and various short-run expressions of total cost. In turn, a particular short-run cost curve gives rise to fixed, variable, and average variable cost as well.

Two caveats should also be noted. One is our interest in cost extends to an interest in our competitor's costs whenever the output market is imperfect. The second is the fact an economist's notion of cost includes payments for all factors of production. An accountant's notion of cost excludes payments to residual claimants. Recall from the study of financial accounting that net income is revenue less expenses, including interest payments but excluding any transactions with the common stockholders. If a firm had no debt whatever, its capital would be supplied entirely by the common stockholders. In such a case the economist's cost curve would include the cost of capital, while the accountant's would exclude the cost of capital. This, by design, suggests the accountant's art is not a direct application of the economist's theory.

Cost is a subtle topic, and cost illiteracy is a serious virus.
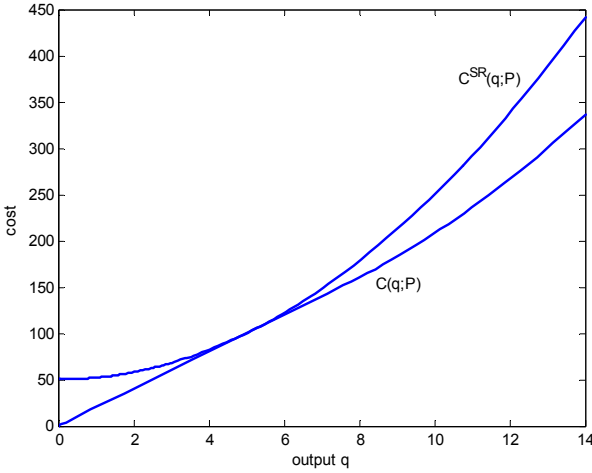
FIGURE 2.5. Short-Run and Long-Run Cost Curves

2.8   Appendix: Constrained Optimization and Shadow Prices

Additional insight into the theory of cost is available when we dig deeper into the cost expression in (2.2), with an eye on the shadow prices. A little background will set the stage. Suppose we want to find the value of some variable, call it x, that minimizes objective function $\omega(x)$ subject to the constraint that $g(x) \geq b$, where b is some constant. It now turns out that if x^* is a solution to this problem then there exists a shadow price, $\lambda \geq 0$, such that (1) $\omega'(x^*) - \lambda g'(x^*) = 0$, (2) $\lambda[g(x^*) - b] = 0$, and (3) $\frac{\partial \omega(x^*)}{\partial b} = \lambda$.[21] Notice the "formalism" has the constraint in "greater than or equal to" fashion, and with a constant on the right hand side.[22]

A useful characterization is the Lagrangian expression where we recast the problem in terms of the objective function less the constraint "priced"

---

[21] This is an application of the Kuhn-Tucker Theorem, where, as usual, we presume suitable regularity is present in the sense, casually, we have a well behaved objective function and constraint set.

[22] If, instead, we are dealing with a maximization problem, we represent the constraint in "less than or equal to" format: $\max_x f(x)$ subject to $g(x) \leq b$. Doing so keeps the shadow price sign non-negative, and we thereby wind up with the same characterization of an optimal solution. Further notice that with multiple constraints we have a shadow price for each, just as when the control variable, x, has multiple elements the noted differential condition applies to each single element. This will become clear in what follows.

at its shadow price:
$$\Psi = \omega(x) - \lambda[[g(x) - b]$$

Now return to Example 2.5 where we faced the following program, initially displayed in (2.6), for identifying the firm's cost function:

$$C(q; P) \equiv \min_{z_1 \geq 0, z_2 \geq 0} P_1 z_1 + P_2 z_2$$
$$\text{s.t. } q \leq \sqrt{z_1 z_2}$$
$$z_1 \leq \overline{z}_1$$

Let's recast our problem in the noted fashion where the constraints are in "greater than or equal to" fashion with a constant on the right hand side. Recognizing we have two constraints (and therefore will have two shadow prices) we have the following::

$$C(q; P) \equiv \min_{z_1 \geq 0, z_2 \geq 0} P_1 z_1 + P_2 z_2 \qquad (2.9)$$
$$\text{s.t. } \sqrt{z_1 z_2} \geq q$$
$$-z_1 \geq -\overline{z}_1$$

Two constraints imply we have two shadow prices. Denote them respectively as $\lambda \geq 0$ and $\overline{\lambda} \geq 0$. The Lagrangian is

$$\Psi = P_1 z_1 + P_2 z_2 - \lambda[\sqrt{z_1 z_2} - q] - \overline{\lambda}[-z_1 + \overline{z}_1] \qquad (2.10)$$

Notice $\frac{\partial \Psi}{\partial q} = \lambda$; marginal cost is a shadow price. Likewise, $\frac{\partial \Psi}{\partial \overline{z}_1} = -\overline{\lambda}$.

Now, if z_1^* and z_2^* are a solution to (2.9) we also have shadow prices denoted $\lambda \geq 0$ and $\overline{\lambda} \geq 0$ such that the following four equalities hold. (Notice we are simply paraphrasing the above conditions, for the case of two factors or choice variables and two constraints.)

$$P_1 - \frac{\lambda z_2^*}{2\sqrt{z_1^* z_2^*}} + \overline{\lambda} = 0 \qquad (2.11a)$$

$$P_2 - \frac{\lambda z_1^*}{2\sqrt{z_1^* z_2^*}} = 0 \qquad (2.11b)$$

$$\lambda[-\sqrt{z_1^* z_2^*} + q] = 0 \qquad (2.11c)$$

$$\overline{\lambda}[z_1^* - \overline{z}_1] = 0 \qquad (2.11d)$$

Initially assume output q is sufficiently low that the second constraint is not binding. This implies $z_1 < \overline{z}_1$ and (2.11d) therefore implies $\overline{\lambda} = 0$. Solving the remaining 3 equations in 3 unknowns provides:

$$z_1^* = \phi \cdot q$$

$$z_2^* = q/\phi$$

and

$$\lambda = 2\sqrt{P_1 P_2}$$

where $\phi = \sqrt{P_2/P_1}$. In turn, $z_1^* = \phi q \leq \overline{z}_1$ requires $q \leq \overline{z}_1/\phi$. Plugging these factor choices into (2.3) gives us a cost of $C(q; P) = P_1 z_1^* + P_2 z_2^* = 2\sqrt{P_1 P_2} \cdot q$, for $0 \leq q \leq \overline{z}_1/\phi$.

Conversely, suppose $q > \overline{z}_1/\phi$. This implies $z_1^* = \overline{z}_1$. Solving the remaining 3 equations in 3 unknowns now provides:

$$z_2^* = \frac{q^2}{\overline{z}_1}$$

$$\overline{\lambda} = P_2 \left(\frac{q}{\overline{z}_1} \right)^2 - P_1$$

and

$$\lambda = \frac{2 P_2 q}{\overline{z}_1}$$

Again plugging these factor choices into (2.3) gives us a cost of $C(q; P) = P_1 z_1^* + P_2 z_2^* = P_1 \overline{z}_1 + P_2 \frac{q^2}{\overline{z}_1}$, for $q > \phi \overline{z}_1$.

2.9   Bibliographic Notes

The best place to begin a review of the economic theory of the firm is a standard economics text. More sophisticated treatments can be found in Chambers [1988], Kreps [1990], and Spulbur [1989]. Stigler [1987] is a personal favorite in the intermediate category. Luenberger [1973] is a somewhat technical though favorite reference on shadow prices.

2.10   Problems and Exercises

1. The Chapter stresses the idea that the firm straddles input and output markets. Explain this notion. What role does the firm's cost function play as it straddles input and output markets?

2. Expressions (2.1) and (2.7) provide equivalent descriptions of the profit maximizing firm's behavior. Why is a cost function present in expression (2.7) but not in expression (2.1)? What purpose is served by the firm's cost function?

3. We insisted in notationally describing the firm's cost curve with $C(q; P)$ as opposed to simply $C(q)$. Explain.

4. Return to Example 2.1. Let $q = 7$ units. Plot all combinations of the two factors that will support production of the 7 units. Limit the first factor to $1 \leq z_1 \leq 40$. Repeat this for $q = 9$ units. What explains the shape of your curves?

5. *construction of cost function*
   Return to the setting of Example 2.6, where technology is specified by $q \leq \sqrt{z_1 z_2}$. Now, however, assume the factor prices are $P_1 = 2$ and $P_2 = 32$.

   (a) Initially assume no upper bound on either factor. What is the firm's cost curve, and what is the corresponding marginal cost curve?

   (b) Suppose the market price is 16 per unit. Can you determine the firm's optimal quantity? What is the source of the difficulty?

   (c) Now assume for the rest of the problem that the first factor cannot exceed an upper bound of 25, i.e., $0 \leq z_1 \leq 25$. Determine the firm's cost curve.

   (d) Suppose output sells for a price of 64 per unit. Determine the firm's optimal plan (output and associated factors). Do this two ways, by focusing directly on profit as in Example 2.2, and then by focusing on revenue less cost as in Example 2.7

   (e) Consider specific output levels of $q \in \{5, 15, 30\}$. For each of these output levels solve numerically for the firm's cost (Excel is recommended), and verify your answer with the above determined cost curve. What shadow prices emerge from your solution? How do you interpret them?

   (f) Suppose the first factor is fixed at $\widehat{z}_1 = 12$. Determine the associated short run cost curve. What is the fixed cost?

6. *long-run cost function*
   Ralph's firm produces a single product. Its long-run cost function is given by $C(q; P) = 900q - 40q^2 + q^3$.

   (a) Determine Ralph's total cost, average cost, marginal cost and incremental cost for each integer value of q between 0 and 30. (Let incremental cost be the cost of one additional unit.)

   (b) In a perfectly competitive market, what market price would you expect, and what output would you expect Ralph to produce? (Hint: under perfectly competitive markets the firm's profit will be precisely zero).

7. *short-run cost function*
   This is a continuation of problem 6 above. Ralph's short-run cost curve is given by $C^{SR}(q; P) = 1,200 + 860q - 45q^2 + 1.2q^3$.

(a) Determine Ralph's fixed, total variable and average variable cost.

(b) Determine Ralph's short-run total cost, average cost, marginal cost and incremental cost for each integer value of q between 0 and 30. (Again, let incremental cost be the cost of one additional unit.)

(c) Plot and interpret Ralph's long-run and short-run average cost.

(d) Plot and interpret Ralph's long-run and short-run marginal cost.

8. *short-run cost function*
Return to the setting of problems 6 and 7 above. The short-run cost curve depicts a case where some factors of production have been fixed at their levels at the efficient output point. Determine another short-run cost curve that could be interpreted as characterizing a case where a different set of factors of production has been fixed at the efficient point.

9. *monopolist's output*
This is a continuation of problem 6 above, where we focused on the long-run cost curve. Suppose Ralph is a monopolist. The market price, as a function of the quantity placed on the market, is $1,400 - 10q$, so Ralph's revenue is $(1400 - 10q)q$. Determine Ralph's optimal output and economic rent. (Rent is strictly positive profit.)

10. *construction of cost function*
Ralph must select the best combination of four factors of production, denoted $z_1 \geq 0, z_2 \geq 0, z_3 \geq 0$ and $z_4 \geq 0$, to produce q units of output. Technical requirements are defined by the following constraints (where you should notice the first two are perfect substitutes and the last two are also perfect substitutes):

$$z_1 + z_2 \geq q$$
$$z_3 + z_4 \geq q$$
$$z_1 \leq 5$$
$$z_3 \leq 6$$

The respective factor prices are $1, 2, 3$, and 4 per unit.

(a) Determine Ralph's cost curve, for $0 \leq q \leq 10$. Plot the cost curve. How do you interpret the shadow prices?

(b) Repeat for a particular short-run setting of $\widehat{z}_1 = 3$.

11. *optimal plan without cost*
Ralph has a one product firm that uses two factors of production to make some good or service. Call the two factors capital, denoted $K \geq 0$, and labor, denoted $L \geq 0$. Also denote the output quantity

by q. To make $q \geq 0$ units of output, Ralph must supply labor and capital in such quantities that the following constraint is satisfied:

$$q \leq \sqrt{KL}$$

Output sells for 15 per unit, while capital costs 9 per unit and labor costs 4 per unit. Ralph's profit, of course, is how much he receives from customers less how much he pays for capital and labor, or $15q - 9K - 4L$. The catch is Ralph cannot use capital in excess of 7 units. Putting everything together, your task is to solve the following maximization exercise:

$$\max_{q \geq 0, K \geq 0, L \geq 0} 15q - 9K - 4L$$
$$\text{s.t. } q \leq \sqrt{KL}; K \leq 7$$

12. *optimal plan with cost*
    This is a continuation of the immediately above problem. Now, however, we decompose Ralph's choice into a cost exercise followed by a quantity exercise.

    (a) Determine Ralph's cost of producing $q \geq 0$ units, $C(q; P)$.

    (b) Now that we have Ralph's cost curve fully specified, we turn to his revenue and profit. Revenue, of course, is $15q$, as each unit of output sells for 15. So Ralph's profit is $15q - C(q; P)$. Determine Ralph's profit maximizing choice of output and corresponding maximal profit.

    (c) What is the connection between this choice and the shadow prices that emerge in constructing Ralph's cost curve?

3
Economic Foundations: The Multiproduct Firm

The next step is to extend our exploration of the economic theory of cost to a multiproduct firm setting. All firms, all organizations for that matter, are multiproduct firms to an economist, meaning they produce more than one good or service. Acknowledging this fact, however, is not simply a nod to reality. Rather, cost here is not a straightforward extension of its single product firm counterpart. Understanding what remains and what does not survive the transition to a multiproduct firm is what opens the door to our understanding of the accountant's art.

Several issues emerge at this point. One is what we mean by a good or service. Here, as we shall see, and contrary to our intuition, a good or service produced in one time period is economically distinct from that "same" good or service produced in another time period. This leads to the insight that a firm marching through time, what we call a multiperiod firm, is just another version of a multiproduct firm.

Another issue is interperiod trade. Here, present value techniques come into play, and for this reason we return to the notion of perfect markets and develop present value as an intertemporal linkage across time periods. From there we lay out the basics of a multiproduct firm, concentrating on a two product firm which, in most cases, will be sufficient for our purpose. This leads to the firm's cost function, and issues of separability. From there we reinterpret the two product firm as a two period firm, where a single good or service is produced each period.

3.1   Back to Perfect Markets

Recall our depiction of a firm as an organization that straddles input and output markets. In a single product firm, the firm deals, by definition, with a single output market, and it has no life beyond the single good or service that it is producing. A multiproduct firm is a firm that deals, by definition, with more than one output market, in one or more time periods. This leads, naturally, to a richer set of market-based arrangements.

3.1.1   *What is a Good or Service*

But what is an output market, or what is a good or service? In economic terms, a good or service is defined by its *what*, *where* and *when*. A lecture on the economic theory of cost is a service (no comment necessary), but we must also specify where this lecture will be delivered and when. The same lecture delivered at another location or time is a distinct economic service. Similarly, a Corvette produced this month is economically distinct from a Corvette produced next month, just as a Corvette offered for sale in Los Angeles is distinct from one offered for sale in Miami.

   We naturally do not see distinct firms producing this month's and next month's Corvette, just as Corvette production is not completely independent of every other activity in General Motors' stable. This suggests synergy is rampant, that economic forces drive the firm to the bundle of goods and services that it produces.

3.1.2   *Present Value*

A special example of the *what*, *where* and *when* specification is your favorite currency. Think about dollars, in the bank, at a particular point in time. Time matters, as a dollar today is distinct from a dollar in 10 years time. It is now a short step to imagine claims to dollars in the bank at various points in time being traded in a perfect market.

   Suppose we want to purchase $100 that will be delivered in three years. Let P_3 be the price we pay today for delivery of $1 in three years. We must pay $100P_3$ in current dollars to arrange for delivery of $100 three years hence. Conversely, suppose we want to borrow $100, and repay the loan in one installment three years later. How much do we pay in three years? The price is P_3 per dollar. Let F be the amount we will pay back. The market demands the following: $100 = F \cdot P_3$. Thus, the trade is $100 today in exchange for $F = 100/P_3$ returned in three years.

   More generally, imagine a dated series of x_0 dollars at time $t = 0$, x_1 dollars at time $t = 1$ and so on through time $t = T$. We often display such a sequence as $[x_0, x_1, ..., x_T]$ or with a time line diagram:

EXHIBIT 3.1: Cash Flow Time Line

How much would we be required to pay at time $t = 0$ for this series of cash flows?   Equivalently, what is the time $t = 0$ value of this series of cash flows, its *present value*?   To answer the question we merely sum up the equivalent time $t = 0$ payments for each of the individual amounts. Let P_t denote the current (i.e., time $t = 0$) price of \$1 to be delivered at time t.   Of course, the current price of \$1 to be delivered at the current instant is $P_0 = 1$. The current price of the above cash flow series is therefore $x_0 + P_1 x_1 + \cdots + P_T x_T$.   As noted, we call this current price equivalent the *present value* of the noted cash flow series.   Thus, the present value of $[x_0, x_1, ..., x_T]$ is

$$PV = \sum_{t=0}^{T} x_t P_t \tag{3.1}$$

This likely strikes you as overly formal, but now is the time to get comfortable with the fact present value is a perfect market concept, it is nothing other than a market price.

To move to more familiar territory, notice we have made no assumption about whether the time intervals were in years, or were even of equal length. All we presumed was a perfect market in which we could exchange a dollar at time t for P_t current dollars. So now further assume (1) the periods are of equal length and (2) the ratio of adjacent prices, P_{t+1}/P_t is a constant, and express it as

$$P_{t+1}/P_t = (1 + r)^{-1}$$

where $r > 0$ is a constant. Of course, r is an interest rate and we have $P_1/P_0 = P_1 = (1 + r)^{-1}$. With the constant price ratio assumption we also have $P_2 = (1 + r)^{-2}$, and in general $P_t = (1 + r)^{-t}$.

This gives our present value construction a familiar flavor:

$$PV = \sum_{t=0}^{T} x_t (1 + r)^{-t} \tag{3.2}$$

The present value of cash flow series $[x_0, x_1, ..., x_T]$ is the discounted value of the series, using interest rate r. One often associates the idea of discounting with the common sense notion that we would prefer having a dollar today to waiting a year to receive the dollar. We emphasize, however,

that the discount rate is a market price and that the primitive idea is exchanging one series of cash flows for another in a perfect market.[1]

Present value will turn out to be essential in understanding the multiperiod firm.

3.2   The Multiproduct Firm

To lay out the basics of the multiproduct firm, suppose our firm now uses three inputs to produce two outputs. Denote the output quantities for the two products $q_1 \geq 0$ and $q_2 \geq 0$. Also denote the input quantities by $z_1 \geq 0, z_2 \geq 0$ and $z_3 \geq 0$. It will be useful at times to condense this notation and let q denote both quantities, so $q = [q_1, q_2]$, and z denote the entire list of inputs, $z = [z_1, z_2, z_3]$.[2]

Here, a production plan consists of q_1 and q_2 units of the two outputs being produced using inputs of z_1, z_2 and z_3. Feasible plans are catalogued with the production function, now denoted $[q_1, q_2] \in f(z_1, z_2, z_3)$, or $q \in f(z)$ for short. That is, $f(z)$ now describes the set of all combinations of the two outputs that can be produced by having inputs z. Notice that the production function relates the list of outputs to the list of inputs. We do not speak of one product and its inputs, combined with the other product and its inputs. The ability to speak in separable fashion is a specialized, uncommon situation. In general, we do not expect separability.

As in our earlier setting, we also avoid various annoyances by assuming all markets are perfect. The input prices are denoted, following our earlier setup, P_1, P_2 and P_3, or P for short. The prices in the respective output markets are denoted $\widehat{P}_1$ and $\widehat{P}_2$, or $\widehat{P}$ for short.

Yet again we describe the firm as straddling the input and output markets, subject to limitations imposed by its production function. Consider a feasible production plan to produce $q = [q_1, q_2]$ using $z = [z_1, z_2, z_3]$. The firm will receive $\widehat{P}_1 q_1 + \widehat{P}_2 q_2$ from customers in the product markets and will pay $\sum_{i=1}^{3} P_i z_i$ to suppliers in the factor markets. The firm chooses the feasible production plan with the largest profit, which we again denote $\Pi(\widehat{P}, P)$ – a total that depends on the output and input prices faced by the firm. This leads to a repetition of our earlier maximization problem in

---

[1] It should be evident how we move from this point to valuing more intricate arrangements, such as trading one series of cash flows for another or valuing the remaining portion of a particular series at some intermediate point in the future.

[2] Technically, q and z are now vectors. You will recall in Chapter 2 that when we grew tired of writing out the list of factor prices we resorted to the (vector) description of $P = [P_1, P_2]$.

(2.1):

$$\Pi(\widehat{P}, P) \equiv \max_{q_1, q_2, z_1, z_2, z_3 \geq 0} \widehat{P}_1 q_1 + \widehat{P}_2 q_2 - \sum_{i=1}^{3} P_i z_i \qquad (3.3)$$

$$\text{s.t.} \quad q \in f(z)$$

Nothing has changed, except we now think in terms of a list of products.

Example 3.1 To illustrate, we offer a slight variation on Example 2.6. Assume the selling prices are $\widehat{P}_1 = 40$ and $\widehat{P}_2 = 50$, while the factor prices are $P_1 = 20, P_2 = 15$ and $P_3 = 5$. The third factor is used in producing both products (think of it as a machine), while the first is specific to the first product and the second is specific to the second. The technology is specified by $q_1 \leq \sqrt{z_1 z_3}$ and $q_2 \leq \sqrt{z_2 z_3}$. We also assume this common third factor is limited to a total of $\overline{z}_3 = 15$. So we maximize profit by solving the following:

$$\Pi(\widehat{P}, P) \equiv \max_{q_1, q_2, z_1, z_2, z_3 \geq 0} 40 q_1 + 50 q_2 - 20 z_1 - 15 z_2 - 5 z_3$$

$$\text{s.t.} \quad q_1 \leq \sqrt{z_1 z_3}$$

$$q_2 \leq \sqrt{z_2 z_3}$$

$$z_3 \leq \overline{z}_3 = 15$$

We readily find optimal outputs of $q_1^* = 15$ and $q_2^* = 25$, along with respective factor choices of $z_1^* = 15, z_2^* = 41.667$ and $z_3^* = 15$.[3] And for later reference, the respective marginal costs at this point are 40 and 50. Profit totals 850. Parenthetically, if we are forced to set $q_2 = 0$, we return to the solution in Example 2.6 (where the factor indexing is adjusted to reflect one versus two products.)

3.3 The Multiproduct Cost Function

As before, we next divide the analysis into input and output components by constructing the firm's cost curve, and subsequently selecting the output levels that maximize revenue less cost. The only divergence from our earlier work is we now speak of the cost of a list of feasible outputs, $q = [q_1, q_2]$. This provides a direct parallel to the single product definition of cost in (2.2).

$$C(q; P) \equiv \min_{z_1, z_2, z_3 \geq 0} \sum_{i=1}^{3} P_i z_i \qquad (3.4)$$

$$\text{s.t.} \quad q \in f(z)$$

---

[3]Rounding is a recurring problem with this type of production function. Here $z_1^* = 625/15 = 41\frac{2}{3} \approx 41.667$.

Repeating this process for all possible output quantities gives us the firm's cost function, denoted $C(q; P)$. As usual we carry along explicit recognition of the factor prices to remind us that the mix of factors for any given output will, in general, depend on factor prices. Also, as noted in the single product case, once these factors are specified, output is the sole explanatory variable in understanding the firm's cost. Of course, output now is multidimensional. It also turns out that the factor mix used for one product will, in general, depend on how much of the other product is being produced.

This is best identified with a little more structure in the story, so we extend Example 3.1 to construction of the cost function.

Example 3.2 Return to Example 3.1, but now concentrate on the firm's cost function. Using (3.4) along with the factor prices and technology originally specified, we have the following construction.

$$C(q; P) \equiv \min_{z_1, z_2, z_3 \geq 0} 20z_1 + 15z_2 + 5z_3$$
$$\text{s.t. } q_1 \leq \sqrt{z_1 z_3}$$
$$q_2 \leq \sqrt{z_2 z_3}$$
$$z_3 \leq \overline{z}_3 = 15$$

Selected values are displayed in Table 3.1, and a plot is displayed in Figure 3.1. More will be said about this shortly. In the meantime, notice that if we set $q_2 = 0$, we are back to our old friend in Example 2.6.

TABLE 3.1: $C(q; P)$ for Example 3.2

q_1	q_2 0	2	4	6	8	10
0	0	34.64	69.28	103.92	138.56	175.00
2	40.00	52.92	80.00	111.36	144.22	180.33
4	80.00	87.18	105.83	131.15	160.33	196.33
6	120.00	124.90	138.56	159.00	187.00	223.00
8	160.33	164.33	176.33	196.33	224.33	260.33
10	208.33	212.33	224.33	244.33	272.33	308.33

3.3.1 Cost Function Terminology

Most of our earlier terminology extends to this setting, though we must be careful to accommodate the presence of multiple products.

Definition 6 *In the multiproduct firm, the incremental cost of $\Delta > 0$ units of product i at output quantity $q = [q_1, q_2]$, $q_1, q_2 \geq 0$ is the difference*
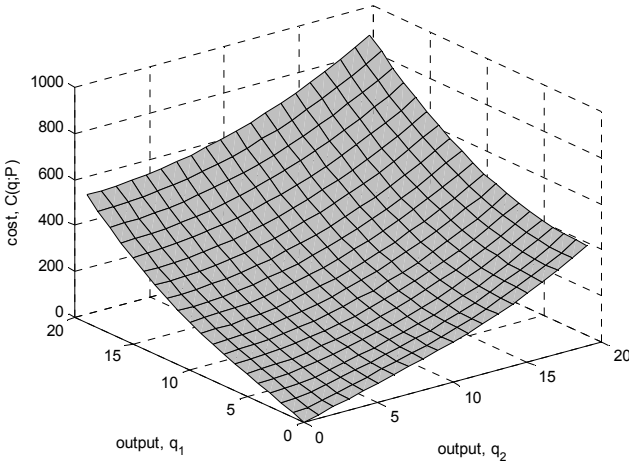
FIGURE 3.1. $C(q; P)$ for Example 3.2

between the cost of producing q_Δ units and q units, or $C(q_\Delta; P) - C(q; P)$ where $q_\Delta = [q_1 + \Delta, q_2]$ or $[q_1, q_2 + \Delta]$

Definition 7 *In the multiproduct firm, the marginal cost of output i at output quantity $q = [q_1, q_2]$, $q_1, q_2 \geq 0$, denoted $MC_i(q; P)$, is the rate at which cost changes with respect to change in quantity q_i, or $MC_i(q; P) = \frac{\partial C(q; P)}{\partial q_i}$.*

Moving to a short-run setting, our earlier definitions of fixed and variable cost apply equally as well to the single or the multiproduct firm. Average cost, however, must now be abandoned.

3.3.2 Cost Function Separability

To understand removal of average cost from our vocabulary, ask yourself: what is the total cost of producing but one of the products? In general there is no answer to this question. The difficulty is interaction. The mix of factors is likely to be affected by the totality of the output. After all, the firm chose to simultaneously produce this set of products, presumably due to an economic advantage (an advantage that is often called an economy of scope). For example, Toyota produces a number of models, just as a consulting firm offers a variety of services. We should expect some form of synergy among the models or services. And once this happens, there is no way to separate factor usage, and hence cost, in an unambiguous product-by-product fashion.

One way to see this is to focus on the marginal costs for the two products in Example 3.2. Below, in Table 3.2, we present selected values of each product's marginal cost. Notice how the marginal cost of either product depends on how much of the other product is being produced. For example, at the point $q_1 = 2$ units, the marginal cost of the first product is 20 if none of the second product is produced, but systematically declines when output of the second product is produced. The reason is the first and third factors are substitutes in producing the first product, but the third factor is also useful in producing the second product. Think of the third factor as a machine, and the first as labor. Producing the first product requires labor and a machine. But the machine is also used for the second product. So if we ramp up production of the second product, we use a larger machine; but a larger machine means less labor is required for the first product, and thus its marginal cost declines.

TABLE 3.2: Marginal Costs for Example 3.2

q_1	q_2	$MC_1(q; P)$	$MC_2(q; P)$
0	2	0	17.32
2	0	20.00	0
2	2	15.12	11.34
2	4	10.00	15.00
2	6	7.18	16.16
4	2	18.35	6.88
4	4	15.11	11.34
6	2	19.22	4.80
6	6	16.00	12.00

Of course we must also remember the machine can be only so large ($z_3 \leq 15$), and at that point this interaction disappears. This is most evident when we plot the marginal costs. We plot the first product's marginal cost in Figure 3.2, for the case where each output varies between .5 and 10 units. Notice for low levels of output that the first product's marginal cost decreases as the second product's output increases. This reflects the noted synergy phenomenon of sharing an ever increasing, larger machine (z_3).[4] Eventually, however, the upper limit on machine size takes over and

---

[4]Some care is required here because the cost function is not well behaved in the neighborhood of $q_1 = q_2 = 0$. The reason is the shared resource, the z_3 factor. If $q_2 = 0$, the firm's cost with only the first product being produced is $C([q_1, 0]; P) = \sqrt{2P_1 P_3} \cdot q_1$ (as long as q_1 is not so large the z_3 upper bound becomes an issue). And in this region the marginal cost of the first product is clearly $\sqrt{2P_1 P_3}$ (= 20 in the Example 3.2 setting). However, if just a tiny amount of the second product is being produced, the marginal cost of the first product is arbitrarily small when an arbitrarily small amount
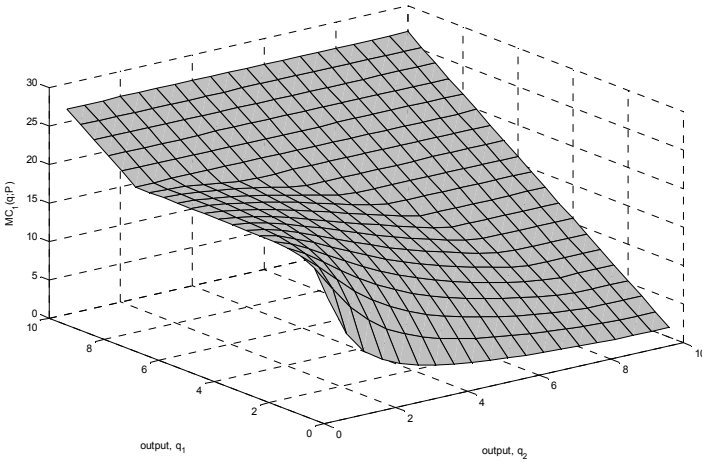
FIGURE 3.2. $MC_1(q; P)$ for Example 3.2

the interaction disappears. A plot of the second product's marginal cost would reveal a similar picture.

A second way to explore these interactions is to focus on the factor choices themselves. In Figure 3.3 we plot the optimal choice of the first factor (z_1^*) while holding output of the first product constant at $q_1 = 2$ units, but varying the output of the second product from .50 to 10 units. Notice how use of the first factor, a factor devoted exclusively to the first product, declines as output of the first product is held constant but output of the second product increases. This decline persists until the shared factor (the machine), which is increasing as output of the second product increases, reaches its upper limit. Interactions, as we keep saying, preclude a separate view of each product's cost. If, then, we interpret the first factor as labor supplied exclusively to the first product (and the third factor as a machine used by both products), we have a setting where expenditures on labor for the first product depend on how much of the second product is being produced.

An even more remarkable picture emerges when we plot consumption of the first factor as a function of the price of the second factor. Set output of each product at 2 units ($q_1 = q_2 = 2$). Further set the first and third factor prices at $P_1 = 20$ and $P_3 = 5$, as in Example 3.2. We now track how optimal use of the first factor (z_1^*) varies as we vary the price of the second factor between $P_1 = 2$ and $P_2 = 25$. This is the story in Figure 3.4.

---

of the first product is being produced. This shows up in Table 3.2, where we report $MC_1(q; P) = 0$ when $q_1 = 0$ and $q_2 = 2$, but $MC_1(q; P) = 20$ when $q_1 = 2$ and $q_2 = 0$.
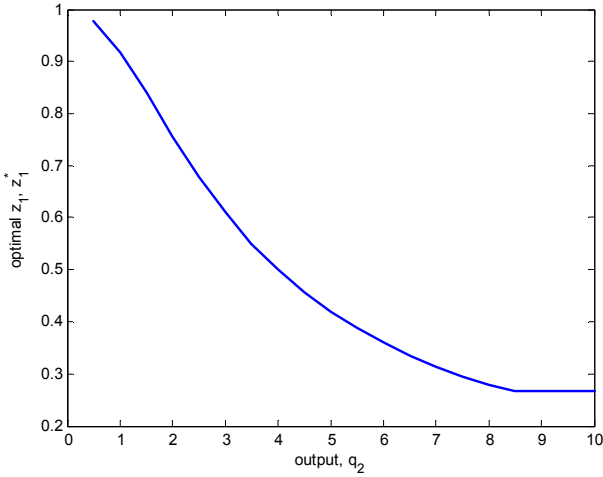
FIGURE 3.3. Factor z_1 as q_2 increases, $q_1 = 2$

Again it helps to interpret the first two factors as labor and the third as a machine. If labor for the second product is very inexpensive, we use a lot of labor and a small machine, and the small machine in turn necessitates a lot of labor for the first product. But if the labor for the second product is rather expensive, we use a large machine and little labor in the second product; but a large machine also results in less labor for the first product. Expenditures on labor for the first product depend on the price of labor for the second product.[5]

A third way to explore these interactions is to simply bite the bullet and look at the closed form expression for the cost curve. Though this is getting out of hand, the factor prices and technology in Example 3.2 imply the following cost curve:[6]

$$C(q; P) = \begin{cases} 2\sqrt{100q_1^2 + 75q_2^2} & \text{if } 4q_1^2 + 3q_2^2 \le 225 \\ 75 + \frac{4}{3}q_1^2 + q_2^2 & \text{otherwise} \end{cases} \tag{3.5}$$

This clearly conveys the impression of interaction!

---

[5] This is another instance where our insistence on carrying along the price symbolism, i.e., the P in $C(q; P)$, is important.

[6] This, as well as the general case, can be derived by extending the derivation in Chapter 2's Appendix. The upper bound on the third factor is not an issue as long as $\sqrt{\frac{P_1}{P_3}q_1^2 + \frac{P_2}{P_3}q_2^2} \le \overline{z}_3$. Below this point, cost is given by $C(q; P) = 2\sqrt{P_3(P_1q_1^2 + P_2q_2^2)}$. Beyond this point, cost is given by $C(q; P) = P_3\overline{z}_3 + P_1q_1^2/\overline{z}_3 + P_2q_2^2/\overline{z}_3$. Notice that if we set one of the outputs to zero the cost expression and upper bound on the third factor expressions both revert to the single product case explored in Chapter 2 (given proper adjustment for the indexing of course).
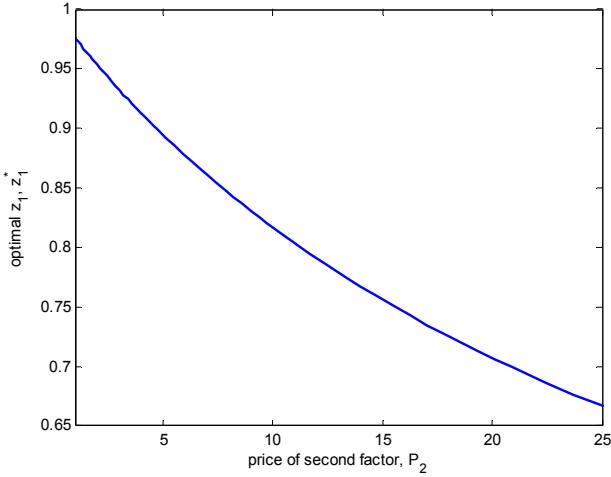
FIGURE 3.4. Factor z_1 as P_2 increases, $q_1 = q_2 = 2$

Example 3.3 This interaction phenomenon also shows up when we recast Example 3.1 in terms of maximizing revenue less cost. Revenue is given by $40q_1 + 50q_2$, and cost is displayed in (3.5). The rub is which region of the cost curve we are in depends on how much of each output is being produced, the $4q_1^2 + 3q_2^2 \leq 225$ qualification. At the optimum, we have $4q_1^2 + 3q_2^2 > 225$, and in this region, which you should verify, marginal revenue equals marginal cost for each product.

Stepping back from the details, suppose we could write the firm's cost curve expression in the following, separable fashion:

$$C(q; P) = G(q_1; P) + H(q_2; P) \tag{3.6}$$

Were this the case, our two product firm's cost would be the sum of two components, one that varied only with the first product and another that varied only with the second product (given factor prices, of course). We could then unambiguously identify the cost of each product and even speak of the average cost of each product.[7] This is the essence of a separable cost function.

Definition 8 *The multiproduct firm's cost function is separable if it can be expressed as the sum of single product cost functions.*

---

[7]In the related short-run case, we would further presume any fixed cost is separable as well. Otherwise, the short-run cost expression would appear as $C^{SR}(q; P) = F + G^{SR}(q_1; P) + H^{SR}(q_2; P)$. And now we could speak with conviction about the separable variable cost. But the joint or common fixed cost poses problems for our terminology.

But if the cost function is not separable we have some type of interaction among or between products. This means we cannot speak unambiguously about the cost of one of the products; and given this we cannot speak unambiguously about the average cost of one of the products. Lack of separability also means a product's marginal cost generally depends on the other products and a wide array of factor prices. Separability is the sheer, total absence of any interactions in the production process. And absent any such interaction one must wonder why the firm houses the combination of products in the first place.

In general, then, we are unable to express the multiproduct firm's cost curve in separable fashion. Products may share a common resource, such as machinery, location or workforce expertise. One product may consume resources that are in short supply and would otherwise be used for another product. Conversely, the products may be jointly produced as when we produce steak, ribs, and hot dogs.[8]

Parenthetically, the accountant usually calls any nonseparable accounting cost expression a setting in which joint products are produced. Our study, however, will distinguish economic from accounting cost, including nonseparable economic cost from nonseparable accounting cost.

It is also time to mention short-run matters. At this point, we could indulge ourselves and revisit the distinction between long-run and short-run cost. However, nothing new, with the exception of notation, would surface.

Regardless, the economic theory of cost is unrelenting. For a given set of factor prices, P, cost is completely explained by output, by volume of production so to speak. The same holds for marginal cost. This is why we keep writing the expression $C(q; P)$. Under separability, (3.6) above, cost and marginal cost are completely explained (given factor prices) by volume or quantity of the product in question. Absent separability, as in (3.5) above, cost and marginal cost are completely explained (given factor prices) by the the volume or quantity of all of the products.[9]

3.3.3   Ubiquity of Marginal Cost

This "preaching" I hope has caught your attention. In the single product case we introduced formal definitions of average, incremental and marginal cost, all based on the fundamental definition of the firm's cost function in

---

[8] In fact the separability described in (3.6) is equivalent to the condition that $\frac{\partial^2 C(q;P)}{\partial q_1 \partial q_2} = 0$.

[9] In later chapters we will confront the claim that cost is explained by variables other than volume or quantity of output (given prices). We will see, however, that this phenomenon occurs because we use approximate expressions for the firm's cost curve and aggregate a large variety of products into sets of bundles. Both are essential in the land of reality, and both lead to errors relative to economic theory.

expression (2.2). We also delved into the world of short-run cost, where at least one factor is fixed, giving rise to fixed and variable cost. We have also introduced formal definitions of incremental and marginal cost for the multiproduct firm; and short-run notions of fixed and variable cost could be pursued as well. But what about average cost?

The short answer is that it is time to remove the term "average cost" from our vocabulary. Absent cost function separability, average cost is a meaningless concept in the multiproduct firm setting. The reason, as noted above, is that without separability we simply do not know what to put in the numerator of the "cost"/"units" expression that is central to the average cost construction.

As you will learn, if you ask an accounting system what some particular product cost, it will nearly always provide an answer. The answer will arrive in the form that "the product cost so much per unit." How, then, are we to interpret this expression of cost per unit? To what economic fundamental does it correspond? It cannot be average cost, as average cost is a meaningless concept in such a setting. That leaves us with marginal cost. That's it. There is nothing left to say. To the extent the accountant's product costing art provides a useful measure of cost per unit, it can only be an estimate of the marginal cost of that product.

That said, we should take some comfort in marginal cost being the undeniable reference point for the accounting system's approach to product costing. After all, decisions about products, such as whether to drop, expand, or modify them, invariably lead to cost issues, which boil down to variations on the marginal cost theme.[10]

Example 3.4 Just to drive this important point a little deeper, suppose the firm's technology and factor prices are such that its cost is given by $C(q; P) = 10q_1 + 10q_2 + 5q_1q_2$. Further suppose the firm produces $q_1 = 4$ and $q_2 = 5$ units. Total cost is clearly 190. At this point the marginal cost of the first product is $10 + 5(5) = 35$, while that of the second is $10 + 5(4) = 30$. What are their respective average costs? There is no answer to this question, as average cost is not defined here. How are we to apportion the $5q_1q_2 = 5(4)(5) = 100$ component between the two products? If we assign it all to the first product, respective "average" costs are 35 and 10. If we assign it all to the second they are 10 and 30. This is silly.[11]

---

[10] As we said, the theory of cost is unrelenting. This focus on all products and efficient use of factors tells the entire story, though with slight modification once we admit strategic considerations for the firm, in either its product or factor markets. There, factor choices may have a strategic counterpart as well. Examples are how much to invest in learning more about your cost curve, or deliberately engineering a low marginal cost so as to influence a would-be competitor.

[11] It is also curious. Economically we do not speak of average cost in a multiproduct firm, except under highly specialized conditions. In financial reporting, though, we are committed to valuing inventory at "average cost."

3.4   A Multiperiod Interpretation

Now add time to the story. As we know, real firms produce multiple goods and services in multiple periods. Your university offers a variety of courses, semester after semester; Toyota offers a variety of models, year after year. Economically, we recognize this by treating the firm's activities as stretched out on a time line. In our two product excursion, we now assume q_1 denotes units of some product delivered in the first period and q_2 units of a product delivered in the second period. These might be identical products, so to speak, but being delivered in distinct time periods renders them distinct products to an economist.

Technology follows our familiar story, using factors of z_1, z_2 and z_3, except time now enters in an important way. It will be sufficient for our purpose to assume the technology is such that the first and third factors must be available at the start of the first period (e.g., labor for the first period product plus a machine), while the second factor must be available at the start of the second period (e.g., labor for the second period product, which also uses the already installed machine).

We further assume transactions take place in various spot markets according to the time line in Exhibit 3.2. (We term these spot prices because they are the market prices at the time the products or factors are delivered.)

$$-P_1 z_1 \qquad \widehat{P}_1 q_1 \qquad \widehat{P}_2 q_2$$
$$-P_3 z_3 \qquad -P_2 z_2$$
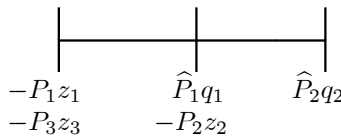
EXHIBIT 3.2: Cash Flow Assumptions

At the start of the story, time $t = 0$, the firm acquires the first and third factors, paying P_1 and P_3 per unit, respectively. At the end of the first period, time $t = 1$, first period output emerges and customers pay for the first period output, at a price of $\widehat{P}_1$ per unit. The firm also acquires the second factor at this point, paying P_2 per unit. Finally, at the end of the second period, time $t = 2$, second period output emerges and customers pay for the second period output, at a price of $\widehat{P}_2$ per unit.

The sole difference here is factor payments and customer payments occur at various points on the time line.[12] To no surprise, this leads us to a focus on present values.

---

[12] We might entertain different timing assumptions, but that is immaterial to what follows.

3.4.1   Present Value to the Rescue

Now suppose the firm is contemplating some specific production plan, based on factor acquisitions (z_1, z_2 and z_3) and outputs (q_1 and q_2). Further suppose the interest rate is r. Recalling the present value construction in (3.2), this plan's cash flows have a present value of

$$-P_1z_1 - P_3z_3 + (\widehat{P}_1q_1 - P_2z_2)(1+r)^{-1} + \widehat{P}_2q_2(1+r)^{-2} \qquad (3.7)$$

Notice the parallel to the profit expression in (3.3). In the single period case, we merely total customer payments and factor expenditures. In the multiperiod setting, we adjust the payments and expenditures for the time at which they take place, the present value calculus so to speak. This gives us a de facto time $t = 0$ profit calculation, fully adjusted for timing issues.

From here it is a short step to viewing the firm as selecting its plan, its input and output profiles through time, to maximize the present value of payments less expenditures. If you enjoy notation, here is the two period version of the original profit maximization problem in (3.3).[13]

$$
\begin{aligned}
PV(\widehat{P}, P) \quad \equiv \quad & \max_{q_1, q_2, z_1, z_2, z_3 \geq 0} -P_1z_1 - P_3z_3 + \qquad\qquad (3.8)\\
& (\widehat{P}_1q_1 - P_2z_2)(1+r)^{-1} + \widehat{P}_2q_2(1+r)^{-2}\\
\text{s.t.} \quad & q \in f(z)
\end{aligned}
$$

In particular, the solution to (3.8) is the production plan that maximizes the present value of customer receipts less factor payments. This provides the equivalent of a time $t = 0$ cash flow of $PV(\widehat{P}, P)$.

Example 3.5 To illustrate, suppose the spot product prices are $\widehat{P}_1 = 44$ and $\widehat{P}_2 = 60.5$, while the spot factor prices are $P_1 = 20, P_3 = 5$ and $P_2 = 16.5$. Also assume the interest rate is $r = 10\%$. First period output, output of $q_1 \geq 0$, is governed by $q_1 \leq \sqrt{z_1z_3}$ and second period output, output of $q_2 \geq 0$, is governed by $q_2 \leq \sqrt{z_2z_3}$. And, as usual, the common factor is limited to a maximum, here a maximum of $\overline{z}_3 = 15$, or $z_3 \leq 15$.

For some arbitrary (though feasible) sets of outputs and inputs, using the timing in Exhibit 3.2 and the present value construction in (3.2), the present value of the resulting cash flows is

$$-20z_1 - 5z_3 + (44q_1 - 16.5z_2)(1.1)^{-1} + 60.5q_2(1.1)^{-2}$$

---

[13]Two important, though subtle points are being ignored here. One is the firm is engaging in transactions today, at time $t = 0$, in anticipation of the prices that will prevail in later periods. Here where we are assuming no uncertainty (a nice double negative), this is not an issue, but it is once uncertainty is admitted. Second, we are not worried about time consistency. The firm does not actually commit itself to the z_2 factor until time $t = 1$, though we are pretending it makes this decision at time $t = 0$. Things of this sort can (and do) become an issue, and fall under the heading of the "time consistency problem."

But this reduces to

$$-20z_1 - 5z_3 + 40q_1 - 15z_2 + 50q_2 = 40q_1 + 50q_2 - 20z_1 - 15z_2 - 5z_3$$

So the firm's problem of maximizing the present value of the cash flows reduces to the following:

$$
\begin{aligned}
PV(\widehat{P}, P) &\equiv \max_{q_1, q_2, z_1, z_2, z_3 \geq 0} 40q_1 + 50q_2 - 20z_1 - 15z_2 - 5z_3 \\
\text{s.t.} \quad & q_1 \leq \sqrt{z_1 z_3} \\
& q_2 \leq \sqrt{z_2 z_3} \\
& z_3 \leq \overline{z}_3 = 15
\end{aligned}
$$

We readily find optimal outputs of $q_1^* = 15$ and $q_2^* = 25$, along with respective factor choices of $z_1^* = 15$, $z_2^* = 41.667$ and $z_3^* = 15$. The present value of the cash flows is 850.[14] Now glance back at Example 3.1. A multiperiod setting is essentially a multiproduct firm, with a time dimension added into the story.

3.4.2 The Multiperiod Cost Function

From here the next step, that of identifying the firm's multiperiod cost function, is profoundly important. The drill itself should be familiar. We decompose the firm's problem into factor choice depending on output, the cost function, followed by maximizing revenue less cost. It is the first half that concerns us. The additional nuance now is that time matters, as transactions and production take place at different points on the time line; invoking the old adage, "time is money." Invested funds now become a factor of production. Present value becomes essential.

Cost, then, is the minimum present value of factor expenditures that will allow production of the noted output schedule. Recalling our specific timing assumption in Exhibit 3.2, cost is defined in this case via

$$
\begin{aligned}
C(q; P) &\equiv \min_{z_1, z_2, z_3 \geq 0} P_1 z_1 + P_3 z_3 + P_2 z_2 (1+r)^{-1} \qquad (3.9) \\
\text{s.t.} \quad & q \in f(z)
\end{aligned}
$$

Note well, we do not generally speak of the cost of but one of the products in a multiperiod firm, and thus, in a multiperiod setting, we do not

---

[14] Now check the time consistency issue mentioned in the prior footnote. Also notice the interpretation of this 850 datum. In a single period setting, where there is no timing issue and no future, profit is cash from customers less cash for factors. So 850 is the firm's profit in Example 3.1. In Example 3.4, where we have used consistent prices, the present value of cash from customers less cash for factors is 850. That is, it is as if the firm's collections less payments netted 850 in time $t = 0$ dollars. Either way, our firm enjoys a rent of 850 in this story.

generally speak of the cost incurred in a single, specific period. Separability is generally absent, among products and through time.

Example 3.6 Return to the setting of Example 3.5. Recalling the assumed prices and timing conventions (Exhibit 3.2), the firm's cost is given by (3.5), though now with a distinct present value flavor. The firm will, under these circumstances, adopt the identical production plan identified earlier. The resulting cash receipts and expenditures are displayed in Table 3.3. For example, the third factor is acquired at time $t = 0$, by expending $5(15) = 75$, just as customers pay, at time $t = 2$, a total of $60.5(25) = 1,512.50$.

TABLE 3.3: Cash Flows for Example 3.6

	time $t = 0$	time $t = 1$	time $t = 2$
z_1 factor, 20(15)	-300		
z_3 factor, 5(15)	-75		
z_2 factor, 16.5(41.667)		-687.50	
sales of q_1, 44(15)		660.00	
sales of q_2, 60.5(25)			1,512.50
net cash flow	-375	-27.50	1,512.50

The present value of these cash flows is, as we know,

$$-375 - 27.5(1.10)^{-1} + 1,512.5(1.1)^{-2} = 850$$

Returning to the cost expression in (3.5), we also know production of $q_1 = 15$ and $q_2 = 25$ has a cost of

$$C(q, P) = 75 + \frac{4}{3}q_1^2 + q_2^2 = 75 + \frac{4}{3}(225) + 625 = 1,000$$

(as at this point the upper limit on the third factor has been reached, i.e., $4q_1^2 + 3q_2^2 > 225$.) Now glance back at Table 3.3. The factor expenditures total 375 at $t = 0$ and 687.50 at $t = 1$. The present value of these expenditures is

$$375 + 687.50(1.10)^{-1} = 375 + 625 = 1,000$$

Thus, the firm's cost reduces to a present value equivalent of $1,000$. Moreover, the respective marginal costs at this point are readily determined to be $MC_1(q; P) = 40$ and $MC_2(q; P) = 50$ And here you should not lose sight of the fact the first period's marginal cost now reflects anticipated production and factor prices for the second period's activities.

Cost, then, continues to be explained by expenditures on (the optimal mix of) factors. The nuance is that these expenditures now not only do not

separate by product, but they take place through time and, economically, also do not separate by time. Moreover, invested funds, implicit or explicit, now become a factor of production. In Example 3.6, notice, we have a cost of $1,000$, based on immediate expenditure of 375 followed by $t = 1$ expenditure of 687.5, for a total of $1,062.50$. This total expenditure is the equivalent of (1) 375 expended immediately, (2) 625 expended immediately and (3) interest of $r \cdot 625 = .1(625) = 62.5$, a total of $1,062.50$.

This reflects the fact we have collapsed the multiperiod activities into their present value equivalent, thereby implying actual expenditures differ from the present value cost construction by the price attached to invested funds. Another way to see this is to observe, again back to Example 3.6, that the marginal cost of the second product is 50. But this is the marginal cost of second period output, viewed economically at the start of the first period. At time $t = 1$ (now assuming the third factor is fixed), you can verify the marginal cost of the second product (at the optimal quantity of $q_2 = 25$) is $55 = (1 + r)50$! Time matters, and therefore the time value of money, so to speak, enters the costing calculus.

We continue this theme in Chapter 4, where we juxtapose accounting and economic renderings of this story.

3.5   Summary

The economic theory of cost in the multiproduct, multiperiod setting provides the foundation for our study of accounting inside the firm. Though it is tempting to quickly survey the single product case (Chapter 2) and move on to the accountant's art, it is important to juxtapose the single and multiproduct stories. The focus on efficient choice of factors applies to both stories. But once we acknowledge the firm chooses which products to produce in which periods, we acknowledge economic interaction among the products. And once we acknowledge interaction, the only concept of product cost with which we are left is marginal cost. Moreover, economic cost summarizes efficient use of all factors, including funds or investment capital.

This is the central, compelling theme we carry forward to our study of the accountant's art. Many mistakes have been and continue to be made because of cost illiteracy. This, I hope, will not apply to us.

3.6   Bibliographic Notes

Your finance text is the best place to review present value mechanics and applications; a personal favorite is Ross et al. [2006]. Hirshleifer [1970] is a superb source for a deeper treatment. Debreu [1959] is a long time

favorite for many reasons, including his elegance and his exposition of what a commodity is. Chambers [1988] and Nadiri [1987] provide exceptional treatments of the economic theory of the multiproduct firm (Nadiri also examines so called ray average cost where we attempt to peer through the cloud of nonseparability); the more adventuresome should also consult Fare and Primont [1995]. The separability theme is examined in Demski and Feltham [1976], Noreen [1991] and Christensen and Demski [1995]. Time consistency issues are flagged in Kydland and Prescott [1977].

3.7 Problems and Exercises

1. Present value is a key concept in linking the idea of a multiperiod firm with that of a multiproduct firm. Explain.

2. Why, in general, is average cost not defined, not meaningful, in a multiproduct setting?

3. Suppose a firm produces two products. This might be two distinct products, produced in a one-period setting. It also might be the same commodity produced in two different periods. In this sense the multi-period firm is a multiproduct firm whose activities are stretched out on the time line. Discuss.

4. Return to the setting of Example 3.1. What is the marginal cost of each product, given production is taking place at the optimal, profit maximizing solution. Comment on your findings.

5. In example 3.6 we calculated the firm's cost, given its production plan, as

$$C(q, P) = 75 + \frac{4}{3}q_1^2 + q_2^2 = 75 + \frac{4}{3}(225) + 625 = 1,000$$

   Is the firm's fixed cost 75? Explain.

6. *prices and present value*
   Suppose the interest rate is $r = 10\%$. What is the current price, at time $t = 0$, of a promise to deliver $1,000$ at the end of period 3? What would the price of this promise be at the end of the first period? Explain your reasoning.

7. *present value*
   Ralph is practicing present value mechanics. For this purpose, the following cash inflows are assumed. Each cash inflow occurs at the end of the indicated year. Whatever interest rate is present is constant throughout the horizon.

end of year	dollar amount
1	1,200
2	6,750
3	8,690
4	5,000
5	7,800
6	6,200
7	125

(a) Compute the present value, as of the start of the first year, assuming a discount rate of $r = 12\%$ per year.

(b) Repeat the above exercise for discount rates of $8\%, 10\%, 14\%$, and 16%.

(c) Now assume a discount rate of 11%. Compute the present value, as of the start of year t of the remaining cash flows. Do this for $t = 1, 2, ..., 7$. For example, the present value at the start of year $t = 6$ will be $6, 200(1.11)^{-1} + 125(1.11)^{-2}$.

8. *marginal cost*

Return to Example 3.1 but assume the price of the first product is $P_1 = 10$. If the second product is not present, what is the firm's optimal output of this product? Conversely, if both products are present, what is the firm's optimal output of this product? Explain.

9. *marginal cost*

Glance back at Figure 3.4 where consumption of the first factor varies with the price of the second factor, even though the two factors are used exclusively in the production of distinct products. For the same specification as used in Figure 3.4, plot the marginal cost of the first product (again, given $q_1 = q_2 = 2$) as a function of P_2. What is the intuition for your finding.

10. *cost in a multiproduct firm*[15]

Ralph is toying with concepts of cost. A two product firm, with quantities denoted q_1 and q_2, is being studied. Three distinct cost functions are being explored: (1) $C_1(q; P) = 10q_1 + 5q_2$; (2) $C_2(q; P) = 6q_1 + q_1^2 + 8q_2 + q_2^2$; and (3) $C_3(q; P) = 7q_1 + 9q_2 + q_1q_2$.

(a) As a warm-up exercise, Ralph decides you should fill in the following table for each of the cost functions. (Incremental cost refers to the incremental cost of one additional unit of output.)

| output | | total | average cost | | marginal cost | | inc. cost | |
q_1	q_2	cost	of q_1	of q_2	of q_1	of q_2	of q_1	of q_2
100	50							
60	50							
40	50							
30	10							
30	50							
30	70							

(b) Plot each of the functions.

(c) Write a brief paragraph about your observations on this exercise.

(d) Also write a brief paragraph on each of the following two questions: (i) what is the economic significance of nonlinearities in the cost functions; and (ii) what is the economic significance of interactions in a cost function (e.g., the q_1q_2 term in $C_3(q; P)$)?

11. *timing*
Return to Examples 3.5 and 3.6. Now suppose all factor payments and collections from customers take place at the end of the second period. Determine the spot prices such that the firm's optimal plan and rent are unaltered by this change in the time at which payments are made.

12. *interest rate*
Return (again) to Examples 3.5 and 3.6. Now assume the same spot prices but that the interest rate is $r = 18\%$. Determine an optimal plan, the firm's cost of producing the optimal output, and the marginal cost of each product at the optimal output. Explain any differences relative to the original example.

13. *multiproduct firm*
Ralph uses three inputs (denoted $z_1 \geq 0, z_2 \geq 0$ and $z_3 \geq 0$) to produce two products (denoted q_1 and q_2). Respective factor prices are $P_1 = 1, P_2 = 5$ and $P_3 = 2$. Technology requires $z_1 + z_3 \geq q_1$ and $z_2 + z_3 \geq q_2$.

(a) Write down the optimization program to determine Ralph's cost, $C(q; P)$.

(b) Determine the first product's marginal cost if $q_1 < q_2$.

(c) Determine the first product's marginal cost if $q_1 > q_2$.

(d) Explain your findings intuitively, and in terms of the shadow prices on the technology constraints.

14. *multiproduct firm*

Ralph has grown up and now manages a two product firm. The technology requires a mixture of capital and labor to produce each product. Capital is shared, while labor is specific to each of the two products. For the first product, capital ($K \geq 0$) and labor ($L_1 \geq 0$) must satisfy the following, in order to produce q_1 units: $q_1 \leq \sqrt{KL_1}$. Likewise, producing q_2 units of the second product requires capital (K) and labor (now $L_2 \geq 0$) such that: $q_2 \leq \sqrt{KL_2}$. In addition, total capital is limited to a maximum of 200 units. (So $K \leq 200$.) Naturally, K, L_1 and L_2 are all required to be non-negative. Capital costs 100 per unit, labor for the first product costs 150 per unit and labor for the second product costs 175 per unit. The first product sells for 275 per unit, and the second sells for 300 per unit.

(a) Initially suppose only the first product is present. Determine and interpret Ralph's optimal production plan.

(b) Next suppose only the second product is present. Determine and interpret Ralph's optimal production plan.

(c) Now assume both products are present. Determine and interpret Ralph's optimal production plan.

(d) Repeat $(a), (b)$ and (c) assuming the first product sells for 200 per unit.

(e) Fill in the following table

q_1	q_2	$MC_1(q; P)$	$MC_2(q; P)$
50	50		
50	100		
50	150		
100	50		
100	100		
100	150		
150	50		
150	100		
150	150		

(f) Write a short paragraph interpreting your findings.

15. *multiperiod firm*

Return to problem 14 above. Now assume this is a multiperiod setting, as in Exhibit 3.2. Further assume the interest rate is $r = 10\%$. Spot factor prices are 100 per unit for capital, 150 per unit for labor for the first product and 192.5 per unit for labor for the second product. The first product now sells for 302.5 per unit, and the second sells for 363 per unit.

(a) Answer part (c) of the original exercise.

(b) Answer part (e) of the original exercise. Explain your findings. Also identify the point on the time line at which these costs are being calculated. Explain.

4
Accounting versus Economics

We now turn to a juxtaposition of the accountant's art with the economist's theory. Two themes converge at this point, themes that will accompany us throughout our study. One is simply the fact that all of the detail presumed by the economist is tempered with pragmatism. Factor consumptions are typically treated in aggregate fashion, and are often recorded with a financial reporting bias.[1]

A second theme is the fact that the firm's financial records, even with this aggregation, form a data bank, or library. And as with any library, the holdings are carefully circumscribed and organized. The accounting library is purposely restricted. What is in the accounting library must have integrity. It must pass scrutiny. We do not record speculative, unverifiable events such as alleged capital gains for a nearly unmarketable fixed asset. This information might be useful, but it would not be found in the accounting records. The accounting records are maintained with a high degree of integrity. An audit trail must be present. This influences what we place in the records.

Thus, if a customer places an order but tenders no deposit or prepayment, conventional accounting will not record such an order at the time it is received. No consideration has been received, no asset is to be recognized, and no liability is to be recognized. Conversely, suppose the customer had accompanied the order with a down payment of 100 dollars. Conven-

---

[1] Depreciating a long-lived asset with straight line as opposed to economic depreciation is a ready example.

tional accounting will now record the order, to the extent of recognizing a liability for 100 dollars. Yet in either case the firm would take notice of the customer's order. Similarly, many firms collect market share and customer demographic information, employ forecasting services, and maintain detailed employee records. Most of this information is outside of the formal accounting records.[2]

Indeed, a popular myth is that financial and managerial accounting are separate worlds. This myth is pedagogically useful, if one seeks to learn procedures. However, it is pedagogically stifling if one seeks to learn how to use financial data. You will see, as the chapter unfolds, that we deal with the traditional financial accounting theme of determining a firm's income. We hope the early appearance of an arguably financial reporting issue will help you visualize accounting (all accounting) as a particular data bank, complete with advantages and disadvantages.

We proceed as follows. In the first section we return to the multiproduct, but single period firm initially presented in Chapter 3. Here, however, we focus on the accountant's art in determining the income associated with each of the products. From there we put considerable structure on this art, by defining such important items as direct and indirect cost pools. We then extend this structure, or recipe, to the multiperiod setting. We conclude with a brief reminder of the importance of accounting convention in these matters.

4.1   Back to the Multiproduct, Single Period Firm

We begin with the single period case, where the firm exists for one brief period of time, but produces multiple products. Two will, in fact, suffice. It will also prove convenient to have a modest amount of structure. So we return to the story in Example 3.1, but with generalized prices. The firm uses three factors, again denoted z_1, z_2 and z_3, to produce outputs denoted q_1 and q_2. Respective factor prices are denoted P_1, P_2 and P_3, and respective selling prices are denoted $\widehat{P}_1$ and $\widehat{P}_2$.

---

[2]Consider a mail order merchandiser. Item X is out of stock. One customer orders X, instructing the merchandiser to charge the item to a credit card. The credit card will be charged when X is shipped. For financial reporting purposes, no record is made until X is shipped. A second customer orders X, including a check for full payment. The merchandiser immediately records cash and a liability for financial reporting purposes. In both cases, however, detailed customer records are maintained. Indeed, capturing all data in an efficient and timely manner is the centerpiece of so-called enterprise software. These systems simultaneously deal with accounting, with human resources, with customers, with regulators, with investors, with production schedules, and with supply chains and suppliers.

4.1.1   The Economic Story

Economic cost is, by now, a familiar construct:

$$C(q; P) \quad \equiv \quad \min_{z_1 \geq 0, z_2 \geq 0, z_3 \geq 0} \quad P_1 z_1 + P_2 z_2 + P_3 z_3 \qquad (4.1)$$

$$\text{s.t.} \qquad q_1 \leq \sqrt{z_1 z_3}$$

$$q_2 \leq, \sqrt{z_2 z_3}$$

$$z_3 \leq \overline{z}_3$$

Here the third factor is common to the production of both goods or services, while the first is unique to the first product and the second is unique to the second product. Moreover, the common factor is limited to an upper bound of $\overline{z}_3$. This structure will be exploited momentarily.

Now suppose the firm produces strictly positive amounts of each product, say $\widetilde{q}_1 > 0$ and $\widetilde{q}_2 > 0$. This production is based on factor usages of $\widetilde{z}_1, \widetilde{z}_2$ and $\widetilde{z}_3$. These factor quantities are the solution to (4.1) when $q = \widetilde{q}$. So the firm incurs an economic cost total of [3]

$$C(\widetilde{q}; P) = P_1 \widetilde{z}_1 + P_2 \widetilde{z}_2 + P_3 \widetilde{z}_3 \qquad (4.2)$$

Likewise, customer payments total

$$\widetilde{R} = \widehat{P}_1 \widetilde{q}_1 + \widehat{P}_2 \widetilde{q}_2 \qquad (4.3)$$

And its profit, which is also its net cash flow as all transactions take place on the basis of immediate cash payment, is simply $\widetilde{R} - C(\widetilde{q}; P)$.

4.1.2   The Accounting Story

What about the accounting? Our work is considerably simplified by the fact this is a single period story. The firm acquires factors, produces, and delivers to customers, but everything takes place without significant lapse of time. On the revenue side, the accounting system will of course record the customer payments. But this is just the customer payments noted by the economist in (4.3). So accounting revenue agrees with the economist's revenue measure in this case, namely $\widetilde{R} = \widehat{P}_1 \widetilde{q}_1 + \widehat{P}_2 \widetilde{q}_2$.

Similarly, on the cost side, the accounting system will track the consumption of factors. In our transparent setting, the firm incurred a cost of $P_1 \widetilde{z}_1$ on the first factor, a cost of $P_2 \widetilde{z}_2$ on the second factor and a cost of $P_3 \widetilde{z}_3$ on the third. Payments were made upon acquisition. Don't miss the nomenclature. To the accountant, what was paid for something is its

---

[3] We assume, just to keep the story uncluttered, that the factor choices by the firm are indeed the efficient choices. As you will see, this assumption, while simplifying, has no effect on our rendering of the accountant's art.

cost, a sacrifice so to speak. Identification of factor consumptions and their respective costs is the building block used by the accountant to paint his portrayal of the firm's financial history. Though it seems unnecessarily elaborate in such a simple setting, imagine these factor costs being recorded in three "cost pools:"

TABLE 4.1: Cost Pools

pool	amount
#1: first factor	$P_1 \widetilde{z}_1$
#2: second factor	$P_2 \widetilde{z}_2$
#3: third factor	$P_3 \widetilde{z}_3$

Notice the total of the three cost pools is the total accounting cost, and this agrees with the economist's measure of cost in (4.2). As with the revenue measure, this is inevitable given we are in a one period story. In turn, accounting income of $\widetilde{R} - C(\widetilde{q}; P)$ agrees with the economist's measure of profit. Thus, in our single period setting, the economic and accounting renderings agree in aggregate.

From this point, we also want to construct the firm's income statement. We already know the total, but can we break it down on a product-by-product basis? We wind up with the construction in Table 4.2, where we exploit our knowledge of the firm's technology and the recording of factor costs in separate cost pools:[4]

TABLE 4.2: Tentative Income Display for Multiproduct Firm

	1st product	2nd product	total
revenue	$\widehat{P}_1 \widetilde{q}_1$	$\widehat{P}_2 \widetilde{q}_2$	$\widetilde{R}$
expenses			
first factor	$P_1 \widetilde{z}_1$		$P_1 \widetilde{z}_1$
second factor		$P_2 \widetilde{z}_2$	$P_2 \widetilde{z}_2$
third factor	?	?	$P_3 \widetilde{q}_3$
			$\widetilde{R} - C(\widetilde{q}; P)$

After all, we know the revenue generated by each of the products, and we also know the first factor is used exclusively in the production of the

---

[4]The important structure in the technology can be gleaned from (4.1). Now you know why we began the chapter in such a formal fashion; technology points us toward the way we measure product costs.

first product, just as the second factor is used exclusively in the production of the second product. The third factor, though, is problematic, as it is used in the production of both products. Lack of separability, as we have stressed, is commonplace.[5]

How to proceed? The accountant selects between two basic approaches (in a realistic setting a combination of the two would be used). One approach is to refuse to paint a picture of complete separability and simply show the third factor as associated, in total, with both products. This provides the following income statement display.

TABLE 4.3: Income Display with Unassigned Third Factor Cost

	1st product	2nd product	total
revenue	$\widehat{P}_1\widetilde{q}_1$	$\widehat{P}_2\widetilde{q}_2$	$\widetilde{R}$
expenses			
first factor	$P_1\widetilde{z}_1$		$P_1\widetilde{z}_1$
second factor		$P_2\widetilde{z}_2$	$P_2\widetilde{z}_2$
margin	$P_1\widetilde{q}_1 - P_1\widetilde{z}_1$	$P_2\widetilde{q}_2 - P_2\widetilde{z}_2$	
third factor			$P_3\widetilde{q}_3$
net income			$\widetilde{R} - C(\widetilde{q}; P)$

Notice we identify a margin for each of the products, but stop analyzing on a product-by-product basis beyond that point. Total margin less the cost of the problematic third factor gives us firm-wide income. The idea in this approach, then, is to work on a product-by-product basis as far as possible, and beyond that point convert to a focus on income of the firm as opposed to income associated with each of its products. Doing so, however, does not mean we have sufficiently addressed the separability demon. We know, from Example 3.2, for instance, that consumption of, say, the first factor depends on output of both products and on the price of all three factors.

The second approach the accountant might use here is to adorn the setting with an appearance of separability. In a sense, this entails pretending separability is present. The trick is to invent some way to apportion the cost of the third factor among the two products. We might assign half the total to each product, do it in proportion to their respective margins, respective revenues, or even physical units. We leave the choice to the reader, and simply assign some amount x to the first product and the remainder

---

[5] Stated differently, how much of the total income was produced by the first product? We know its revenue was $\widehat{P}_1\widetilde{q}_1$, but in the language of income measurement, what was its cost of goods sold? Surely the answer includes $P_1\widetilde{z}_1$, and beyond that we must confront the common factor's cost.

of $P_3\tilde{z}_3 - x$ to the second product. This provides the following rendering of the firm's income, one where we create the impression of two separate products living side by side.

TABLE 4.4: Income Display with Assigned Third Factor Cost

	1st product	2nd product	total
revenue	$\widehat{P}_1\tilde{q}_1$	$\widehat{P}_2\tilde{q}_2$	$\tilde{R}$
expenses			
first factor	$P_1\tilde{z}_1$		$P_1\tilde{z}_1$
second factor		$P_2\tilde{z}_2$	$P_2\tilde{z}_2$
third factor	x	$P_3\tilde{z}_3 - x$	$P_3\tilde{z}_3$
	$\widehat{P}_1\tilde{q}_1 - P_1\tilde{z}_1$	$\widehat{P}_2\tilde{q}_2 - P_2\tilde{z}_2$	
net income	$-x$	$-P_3\tilde{z}_3 + x$	$\tilde{R} - C(\tilde{q}; P)$

Notice there is no formally acknowledged interaction in the Table 4.4 income display. The income is presented as though part is derived from the first product and the remainder from the second. Indeed, income and cost averages can even be calculated for each product as well. Of course, this is silly, but one can do it. (Some even do.)

At this point you are probably suffering from terminal cynicism. But remember we are opening the door to the accountant's art. And one of the things we do is ask the accountant to put an accounting value on end of period inventory. Though there is no ending inventory here, suppose for the sake of argument that one unit of the first product was left in inventory at the end of the period. What inventory value would the accountant assign? It must be some expression of the cost of producing that unit. Well, in the first case it would simply be $P_1\tilde{z}_1/\tilde{q}_1$, the presumptive cost per unit; and in the second case would it be $(P_1\tilde{z}_1 + x)/\tilde{q}_1$.

Example 4.1 To illustrate this meandering we return to the setting of Example 3.1, which is consistent with the technology we have been assuming. The assumed selling prices are $\widehat{P}_1 = 40$ and $\widehat{P}_2 = 50$, while the factor prices are $P_1 = 20, P_2 = 15$ and $P_3 = 5$. Respective outputs are $\tilde{q}_1 = 15$ and $\tilde{q}_2 = 25$ units, while respective factor consumptions are $\tilde{z}_1 = 15, \tilde{z}_2 = 41.667$ and $\tilde{z}_3 = 15$. Revenue totals $\tilde{R} = 40(15) + 50(25) = 1,850$ and cost totals $20(15) + 15(41.667) + 5(15) = 1,000$. This implies the following cost pools.

pool	amount
#1: first factor	300
#2: second factor	625
#3: third factor	75

From here we examine the two approaches to dealing with the common factor cost of 75. The important fact is either way the accountant identifies a cost of each product. The difference is how many of the cost pools are treated as explicit costs of the products. Not assigning the common factor cost to individual products results in the following.

	1st product	2nd product	total
revenue	600	1,250	1,850
expenses			
first factor	300		300
second factor	___	625	625
margin	300	625	925
third factor			75
net income			850

Conversely, assigning the common factor cost to both products, here on the basis of the number of units produced, leads to the following display. (So the x in Table 4.4 is $x = \frac{15}{15+25}(75)$.)[6]

	1st product	2nd product	total
revenue	600	1,250	1,850
expenses			
first factor	300		300
second factor		625	625
third factor	28.13	46.87	75
net income	271.87	578.13	850

4.1.3   Per Unit Product Costs

You are probably wondering why we have gone down this road of trying to separate the firm's income into the amount due to or contributed by each product. One reason, of course, is we often find circumstances where we

---

[6] The rounding here is carried over to Table 4.5.

want to know, as best as possible, how much income is being contributed by various products. For example, do we want to keep the restaurant open for breakfast? Do we want to continue offering some consulting product, such as executive search?

A second reason is the accountant is often asked to value some inventory total, such as manufactured yet unsold electronic components. In this case we initially identify the costs associated with the components, and then convert them to per unit costs. For example, if $1,000$ units were manufactured, product costs totaled $5,000$ and 100 units remain in inventory, the accountant would value them (ignoring lower of cost or market issues) at $5,000/1,000 = 5$ per unit, which implies an inventory valuation of $5(100) = 500$. Pedantic as it seems, now is the time to learn that if you ask the accounting system what a product costs, it will almost surely give you an answer.

Unfortunately, the answer the accounting system gives is almost surely not the answer to your question. To illustrate, we return to Example 4.1 and tally the implied cost per unit in each of the income renderings in that example. See Table 4.5. All we have done is take the costs associated with the product in question and divide by the number of units. Glance back at the case in Example 4.1 where the common factor cost is not assigned to individual products. Notice the total cost reported for the first product is 300 (with $\widetilde{q}_1 = 15$) while that for the second is 625 (with $\widetilde{q}_2 = 25$).[7] In this case, the accountant contends the cost per unit is $300/15 = 20$ for the first product and $625/25 = 25$ for the second.

TABLE 4.5: Per Unit Product Costs for Example 4.1

	1st product	2nd product
common factor not assigned	20	25
common factor assigned	21.88	26.87
marginal cost, $MC_i(\widetilde{q}; P)$	40	50

Several comments are in order. First, as we know, marginal cost is, in general, the only meaningful measure of product cost in the multiproduct firm. Yet, while the economist stresses marginal cost, the accountant focuses on assigning observed factor costs to products. The accountant is not being a contrarian or a dolt here. No, marginal cost is knowable beyond a shadow of doubt (a good pun) only in a textbook; it is subjective in the world of affairs. And the accountant cannot routinely embrace subjectivity as he paints the firm's financial history. This would open the recorded history to serious waves of opportunism.

---

[7] These totals would be reported as the cost of goods sold for each product.

Second, the gulf between the accountant's and economist's approaches is driven in part by the fact the accountant must reconcile the accounting cost per unit to the total cost incurred. Returning yet again to Example 4.1, total cost incurred is, recall, $1,000$. Let α denote the accounting cost per unit of the first product and β the accounting cost per unit of the second product. For the sake of argument, suppose the accountant will assign all factor costs to the two products. This means $\alpha \geq 0$ and $\beta \geq 0$ must satisfy

$$15\alpha + 25\beta = 1,000$$

as 15 units of the first product are present along with 25 units of the second. Clearly, dreaming up some procedure that would provide accounting costs per unit equal to the respective marginal costs will not work here, as the respective marginal costs are 40 and 50:

$$15(40) + 25(50) = 1,850 > 1,000$$

Thus, in this example, whatever product cost per unit the accountant is reporting, it is clearly not a very good estimator of marginal cost here. Tweaking the accounting system to better measure marginal cost will be a recurring theme in our work.[8]

Finally, continuing to refer to the product cost per unit will grow weary, yet we also want a constant reminder that the accountant's art is an important force in producing these per unit product costs. We will, for this purpose, term the per unit product costs provided by the accountant *unit costs*. This is not very imaginative, but it will serve to remind us of the accountant's art and its conceptual distance from the economist's marginal cost.

4.2   The Underlying Recipe

With this introductory excursion behind us, it is time to be more specific about the accountant's costing art. The easiest way to visualize the costing recipe is to imagine all costs incurred during the period as being recorded in various cost pools, one for each distinct cost category. These are costs that must be "accounted for" during the period, covering all sorts of factors.

---

[8] To dig a bit deeper here, the firm's cost in this example, as originally laid out in expression (3.5), is $C(q; P) = 75 + \frac{4}{3}q_1^2 + q_2^2$ (presuming output is such that $4q_1^2 + 3q_2^2 > 225$, which it is here). This implies respective marginal costs of $MC_1(q; P) = \frac{8}{3}q_1$ and $MC_2(q; P) = 2q_2$. (So now try your hand at $q_1 = 15$ and $q_2 = 25$.) Importantly, now, we are in a region where marginal cost is increasing as we increase output, a point where we are encountering a diseconomy of scale. But with marginal costs behaving in this fashion, we will have $MC_1(q; P) \cdot q_1 + MC_2(q; P) \cdot q_2 > C(q; P)$, again remembering we also have $4q_1^2 + 3q_2^2 > 225$.

The general idea is to identify the factor consumptions and their respective factor costs. These factor costs will eventually be assigned to products or simply expensed (and thus assigned to the period). They must be assigned or expensed. So we have the following definition, where the word temporary reminds us these costs are to be assigned or expensed.[9]

Definition 9 *A cost pool is a temporary collection of factor costs.*

In Example 4.1 we had three cost pools, one for each of the factors. See Table 4.1.

Next is the grouping into categories, those cost pools associated with products and those that are expensed, meaning are associated with the period in which they were incurred.

Definition 10 *A product cost pool is a cost pool that is associated with the firm's products, as opposed to the time period in which it is incurred.*

Definition 11 *A period cost pool is a cost pool that is associated with the time period in which it is incurred, as opposed to the firm's products.*

In Example 4.1, the first two cost pools were product cost pools. The third cost pool was treated as a period cost pool in the construction in Table 4.3 but as a product cost pool in the construction in Table 4.4

Though we stress it is the cost pool itself that is associated with products or the period, it is also possible to have a hybrid cost pool, one where a fraction of the costs in the pool is assigned to products and the remainder to periods. But it is sufficient to focus on the basics at this point. That said, a product cost pool may be direct, in the sense we find it possible and convenient to associate the cost pool with a single product. Otherwise, we term it an indirect product cost pool. In turn, an indirect product cost pool, meaning it is not associated with a single product, may be associated with various products, with other cost pools or with various products and the period in which it is incurred, the noted hybrid case. Emphatically, it is not associated with a single product.

Definition 12 *A direct product cost pool is a product cost pool that is associated with a single product.*

Definition 13 *An indirect product cost pool is a product cost pool that is not associated with a single product.*

In Example 4.1, if the third cost pool is treated as a product cost pool, it is an indirect product cost pool as it is associated with both of the products.

---

[9] If you are handy with debits and credits and the closing process at the end of the accounting cycle, cost pools are temporary accounts, accounts that must be closed at the end of the accounting cycle.

Finally, a unit cost, as mentioned, is cost per unit that is implied by the recipe.[10]

Definition 14 *A specific product's unit cost is the total cost associated with that product divided by the number of units of that product.*

Unit costs for the two approaches to dealing with the third cost pool in Example 4.1 are displayed in Table 4.5

In short, the accountant's costing recipe begins by identifying the resources, the factors, consumed during some time period. The costs of these resources or factors, are recorded, explicitly or implicitly, in a number of cost pools. The cost pools themselves are designated as to whether they accumulate product costs or period costs. Period cost pools are expensed during the period. Product cost pools are assigned to products, directly (direct cost pools) or indirectly (indirect cost pools).

Of course, this begs the question of which of the myriad of factor consumptions fall into the product cost category as well as how indirect costs are assigned to products. These design elements will occupy a great deal of our study, but in each and every case you will recognize the basic recipe.

4.3   The Multiperiod Case

We now switch to the multiperiod case. As stressed in Chapter 3, the distinction between a multiproduct and a multiperiod firm revolves around the role of time. In the multiperiod case transactions need not be consummated immediately in cash, and factor payments may indeed take place in a time period distinct from the use of the factors. Moreover, common factors may now be common across periods, e.g., depreciation of a machine. This is where we encounter accrual accounting.

4.3.1   The Economic Story

To lay this out and tie into the costing recipe, we use the two product structure relied upon earlier in the chapter, but treat the first product as essentially the first period and the second product as the second period. Important timing assumptions are laid out in Exhibit 3.2. Briefly, recall, we have the first and third factors paid for at time $t = 0$, production and delivery of the first product taking place during the period followed by customer payments at time $t = 1$. The second factor is paid for at time $t = 1$, production and delivery of the second product take place during the second period, and those customers pay at time $t = 2$, at which point

---

[10]In offering this formalism we choose not to tempt fate and define unit cost as the average of the product costs associated with that product.

the story ends. The interest rate is denoted r. Given these assumptions, and continuing with our usual notation for spot prices, the firm's cost, in present value terms, is given by the following slight variation on (4.1):

$$C(q; P) \equiv \min_{z_1 \geq 0, z_2 \geq 0, z_3 \geq 0} P_1 z_1 + P_2 z_2 (1+r)^{-1} + P_3 z_3 \qquad (4.4)$$

$$\text{s.t.} \quad q_1 \leq \sqrt{z_1 z_3}$$
$$q_2 \leq, \sqrt{z_2 z_3}$$
$$z_3 \leq \overline{z}_3$$

Retracing our steps, now suppose the firm produces and sells $\widetilde{q}_1 > 0$ in the first period followed by $\widetilde{q}_2 > 0$ in the second period. This production is based on factor usages of $\widetilde{z}_1, \widetilde{z}_2$ and $\widetilde{z}_3$; and these factor quantities are the solution to (4.4) when $q = \widetilde{q}$. This leads to the cash flows depicted in Table 4.6, where you will notice we have introduced the shorthand notation of cash flow at time t being denoted CF_t.

TABLE 4.6: Cash Flows for Multiperiod Firm

	time $t = 0$	time $t = 1$	time $t = 2$
1st factor	$-P_1 \widetilde{z}_1$		
3rd factor	$-P_3 \widetilde{z}_3$		
2nd factor		$-P_2 \widetilde{z}_2$	
sale of 1st product		$\widehat{P}_1 \widetilde{q}_1$	
sale of 2nd product			$\widehat{P}_2 \widetilde{q}_2$
net cash flow, denoted CF_t	$-P_1 \widetilde{z}_1 - P_3 \widetilde{z}_3$	$\widehat{P}_1 \widetilde{q}_1 - P_2 \widetilde{z}_2$	$\widehat{P}_2 \widetilde{q}_2$

It is important to understand the cash flows, CF_t, in Table 4.6 are cash flows between the firm and its owners. The firm is assumed to maintain a zero cash balance. So the owners invest, initially, the amount $P_1 \widetilde{z}_1 + P_3 \widetilde{z}_3$, which is spent on the noted factors. At the end of the first period, at time $t = 1$, cash in the amount $\widehat{P}_1 \widetilde{q}_1 - P_2 \widetilde{z}_2$ flows between the firm and its owners, meaning further investment if $\widehat{P}_1 \widetilde{q}_1 - P_2 \widetilde{z}_2 < 0$ or a dividend if $\widehat{P}_1 \widetilde{q}_1 - P_2 \widetilde{z}_2 > 0$. Similarly, at the end of the second period a liquidating dividend of $\widehat{P}_2 \widetilde{q}_2$ is paid to the owners and the firm ceases to exist.[11]

Given these choices and cash flow consequences, the firm incurs an economic cost total, in present value terms, of

$$C(\widetilde{q}; P) = P_1 \widetilde{z}_1 + P_2 \widetilde{z}_2 (1+r)^{-1} + P_3 \widetilde{z}_3$$

---

[11] If the firm maintained a nonzero cash balance, interest on that balance would be accrued, and this would affect subsequent cash flows. Maintaining a zero cash balance avoids this complication.

Likewise, recognizing the time dimension to the spot markets, customer payments lead to a revenue total, again in present value terms of

$$\widetilde{R} = \widehat{P}_1 \widetilde{q}_1 (1 + r)^{-1} + \widehat{P}_2 \widetilde{q}_2 (1 + r)^{-2}$$

And its profit, or rent, which is the present value of the sequence of cash transactions, is $\widetilde{R} - C(\widetilde{q}; P)$. We can also amuse ourselves at this point by examining each product's marginal cost. For sure, this is taking on the appearance of redundancy.

At this point, then, we routinely speak of the firm's revenue, cost and profit as well as each product's marginal cost, all in present value terms. We also encounter the firm's economic income each period.[12] To sketch this, let PV_t denote the present value of the remaining cash flows as of time t:

$$PV_0 = CF_1(1 + r)^{-1} + CF_2(1 + r)^{-2}$$
$$PV_1 = CF_2(1 + r)^{-1}$$

and

$$PV_2 = 0$$

Notice the emphasis on the future: PV_t reflects the cash flows that occur beyond time t. At time t, then, the value of the remaining cash flows is simply PV_t.

Now define the firm's economic income as change in (present) value of the remaining cash flows plus the cash flow of the period. This gives us income at the inception, I_0, followed by first period income of I_1 and second period income of I_2, calculated in the following fashion:

$$I_0 = PV_0 + CF_0$$

$$I_1 = PV_1 - PV_0 + CF_1$$
$$I_2 = PV_2 - PV_1 + CF_2$$

Notice the sum of the three income numbers (as $PV_2 = 0$ because the firm ceases to exist at that point) is simply the sum of the cash flows: lifetime income equals lifetime cash flow.

$$I_0 + I_1 + I_2 = CF_0 + CF_1 + CF_2$$

---

[12] We have been casual in our use of the terms profit and income to this point. These two terms are often used interchangeably. We will treat profit as numerical gain, as when we speak of the firm's profit in a one period setting or the present value equivalent in a multiperiod setting. Economic rent in this sense is profit. However, we will reserve the term income for when we are attempting to identify "gain" each period (as when we speak of economic income) or when we resort to accounting-based calculations, such as the income of the period or the income attributable to a specific product. Though this likely strikes you as arcane, you are about to discover profit in this sense and total income in this sense are generally not the same.

Also notice, though the algebra is deferred to you, that $I_1/PV_0 = I_2/PV_1 = r$: economic income reports a constant rate of return on the respective period's initial asset base.

The time $t = 0$ construction, I_0, however, is a bit awkward. If markets are perfect, no firm earns rent. And if rent is absent, the initial investment is equal in value to the present value of the future cash flows, or $I_0 = PV_0 + CF_0 = 0$, and all is well. If markets are a little less forgiving, rent is present. The economist is not shy, and records the rent at the firm's inception, implying $I_0 = PV_0 + CF_0 > 0$.[13]

Thus, in addition to profit or rent and our ever present marginal costs, the economist speaks forcefully of the firm's economic income. In particular, economic income is simply the financial cost of the funds implicitly tied up in the enterprise. For example, with $I_1/PV_0 = I_2/PV_1 = r$ we have $I_1 = r \cdot PV_0$ and $I_2 = r \cdot PV_1$. In each period t, the remaining cash flows have a value at the beginning of the period equal to PV_{t-1}. Selling this claim to the cash flows and investing the proceeds would earn $r \cdot PV_{t-1}$. Emphatically, economic income is the factor cost of funds committed to the firm.

Now, ponder the fact we walked (quickly I admit) through this fantasy without ever mentioning when to recognize revenue, asking what a product cost, or asking what specific assets were on the balance sheet. It is simply a sequential rendering of the cash flow implications. The accountant is not so fortunate, and for good reason is denied the option of simply claiming some PV_t amount.

Example 4.2 Return to the setting of Examples 3.5 and 3.6. The product spot prices are $\widehat{P}_1 = 44$ and $\widehat{P}_2 = 60.5$, and the spot factor prices are $P_1 = 20, P_3 = 5$ and $P_2 = 16.5$. The interest rate is $r = 10\%$, and the common factor is limited to a maximum of $\overline{z}_3 = 15$. Further assume the firm follows the previously identified profit maximizing path by producing and selling $\widetilde{q}_1 = 15$ in the first period and $\widetilde{q}_2 = 25$ in the second period, while consuming factor quantities of $\widetilde{z}_1 = 15, \widetilde{z}_2 = 41.667$ and $\widetilde{z}_3 = 15$. In present value terms, revenue is

$$\widetilde{R} = 44(15)(1.10)^{-1} + 60.5(25)(1.10)^{-2} = 1,850$$

Likewise, in present value terms, the cost is

$$C(\widetilde{q}; P) = 20(15) + 16.5(41.667)(1.10)^{-1} + 5(15) = 1,000$$

So the firms profit is $\widetilde{R} - C(\widetilde{q}; P) = 1,850 - 1,000 = 850$, a familiar number for sure. And this leads to the cash flow summarization originally displayed in Table 3.3, and reproduced below. The cash flows total

---

[13] This becomes awkward because the use of present values is based on perfect markets, but perfect markets imply zero rent.

$$-375 - 27.5 + 1,512.5 = 1,110.$$

Now consider the continuation values. For $t = 0$ we have

$$PV_0 = -27.50(1.10)^{-1} + 1,512.50(1.10)^{-2} = 1,225$$

Similarly, for $t = 1$ we have

$$PV_1 = 1,512.50(1.10)^{-1} = 1,375$$

And of course $PV_2 = 0$. This gives

$$I_0 = PV_0 + CF_0 = 1,225 - 375 = 850;$$

$$I_1 = PV_1 - PV_0 + CF_1 = 1,375 - 1,225 - 27.5 = 122.50;$$

and

$$I_2 = PV_2 - PV_1 + CF_2 = 0 - 1,375 + 1,512.50 = 137.50$$

In turn, the lifetime income totals $1,110$, the lifetime cash flow:

$$850 + 122.50 + 137.50 = 1,110$$

Also notice the constant rate of return calculation and the fact $CF_0 + CF_1 + CF_2 = 1,110 = I_0 + I_1 + I_2$.

This is all summarized as follows.

	time $t = 0$	time $t = 1$	time $t = 2$
z_1 factor, 20(15)	-300		
z_3 factor, 5(15)	-75		
z_2 factor, 16.5(41.667)		-687.50	
sales of q_1, 44(15)		660.00	
sales of q_2, 60.5(25)			1,512.50
net cash flow, CF_t	-375	-27.50	1,512.50
continuation values, PV_t	1,225	1,375	0
economic income,			
$I_t = PV_t - PV_{t-1} + CF_t$	850	122.50	137.50
I_t/PV_{t-1}		.10	.10

Though this is wearing thin you should ponder the fact that even though the firm's activities are not separable, and that includes the firm's cost curve, the economist has no difficulty divining the period-by-period economic income. This is not witchcraft; no, it reflects the fact economic income is the factor cost of funds provided the enterprise. The market price of these funds is known, this is the interest rate r; and the amount of these funds is equally well known, this is the continuation present value, PV_t.

4.3.2   The Accounting Story

We now turn to the accountant's rendering of this multiperiod story. Here, in sharp contrast with the economist, the accountant's recipe comes to the fore. Initially we identify cost pools for each period. And at this point we encounter the problematic common factor. It is used in both periods and thus will be assigned in some manner to both periods. We no longer have the luxury of simply assigning it in total to the two periods, as we did in the single period case (Table 4.3). Rather, it must somehow be divided between the two periods. So the cost pools take on the following appearance.

TABLE 4.7: Cost Pools for Multiperiod Case

pool	1st period	2nd period
#1: first factor	$P_1\tilde{z}_1$	
#2: second factor		$P_2\tilde{z}_2$
#3: third factor	x	$P_3\tilde{z}_3 - x$

It is clear pool #1 is a direct product cost pool in the first period, just as pool #2 is a direct product cost pool in the second period.[14] Consistency requires pool #3 be treated as either an indirect product cost pool in both periods or as a period cost pool in both periods. For the sake of argument we adopt the former. The burning question, though, is how much of the expenditure on the common factor belongs in each of the periods. We know whatever answer we prescribe must sum to the factor cost incurred, $P_3\tilde{z}_3$. We leave the question unanswered, at least for the moment, and simply construct accounting income in generic fashion:

$$\widehat{I}_1 = \widehat{P}_1\tilde{q}_1 - P_1\tilde{z}_1 - x$$

and

$$\widehat{I}_2 = \widehat{P}_2\tilde{q}_2 - P_2\tilde{z}_2 - P_3\tilde{z}_3 + x$$

See Table 4.8.

Though the lack of closure is unsettling, you should expect some ambiguity here. After all, we have, for sound reason, stepped away from the economist's calculations, most importantly by not booking the up-front gain of I_0. Lack of separability further clouds the issue of measuring each

---

[14]Again, a cautionary note is in order. Being a direct product cost pool does not mean the factor consumption recorded in that pool is independent of other products or other factor prices.

period's accounting income.[15]   Regardless, we have universal agreement on the totals: lifetime accounting income equals lifetime economic income equals lifetime cash flow. Glance back at Table 4.6. You will see

$$\widehat{I}_1 + \widehat{I}_2 = \widehat{P}_1\widetilde{q}_1 - P_1\widetilde{z}_1 - x + \widehat{P}_2\widetilde{q}_2 - P_2\widetilde{z}_2 - P_3\widetilde{z}_3 + x = CF_0 + CF_1 + CF_2$$

This algebraic conclusion does not imply that in the long-run all methods of accounting accomplish the same thing. If we freeze the events, all methods of accounting will report the same cumulative income. Of course, accounting is likely to have an effect on what events take place.[16]

TABLE 4.8: Tentative Income Display for Multiperiod Firm

	1st period	2nd period
revenue	$\widehat{P}_1\widetilde{q}_1$	$\widehat{P}_2\widetilde{q}_2$
expenses		
direct cost pool	$P_1\widetilde{z}_1$	$P_2\widetilde{z}_2$
indirect cost pool	x	$P_3\widetilde{z}_3 - x$
income, $\widehat{I}_t$	$\widehat{P}_1\widetilde{q}_1 - P_1\widetilde{z}_1 - x$	$\widehat{P}_2\widetilde{q}_2 - P_2\widetilde{z}_2 - P_3\widetilde{z}_3 + x$

4.4   Accounting Conventions

Though tempting to call a halt at this point, it is important to step back and put a little more structure on the accountant's art. As we know, the firm's accounting system will report a balance sheet and an income statement each reporting period. It is here that the recognition rules of accrual reporting have their impact. The general idea is to depict a stock (the balance sheet) and a flow (the income statement) in a way that reflects the "economic substance" of the firm's activities. If the firm acquires plant that will be used for many periods, our common factor, this plant is providing services for many periods. This is an asset, just as cash on hand is an asset. Similarly, if it has manufactured but not sold some product, and if that product will be sold in a subsequent period, an asset exists.

A cash basis recognition rule takes an extremely conservative approach to asset recognition. Cash is the only asset that is recognized. This approach leads to a streamlined balance sheet, to say the least. It paints an unusual picture of the firm's financial health. For example, cash basis recognition in Table 4.8 would set $x = P_3\widetilde{z}_3$.

---

[15] A variation on this theme surfaces in a divisionalized firm where we attempt to associate some of the total income with each of the divisions as well as some of the total assets and even, at times, total liabilities.

[16] Were that not the case, this book would be extremely short.

Accrual procedures recognize noncash assets.[17]   The ways in which this might be accomplished are endless. In general terms, we think of an asset as something that will render service in the future. A liability is an obligation to render service or transfer assets in the future.   Cost is incurred when resources are consumed for some purpose. Revenue and expense are subdivisions of owners' equity. Revenue is an asset increase or liability decrease associated with serving the firm's customers. An expense is an asset decrease or liability increase associated with serving the firm's customers. Beyond these broad conceptions considerable detail, convention, and reporting regulation are combined to produce financial reporting practice.

Nevertheless, ambiguity is commonplace, as illustrated by our problematic x in Table 4.8. We know the total income for the two periods, but are perplexed about how to divide that income between the two periods. Truth is, regulation, judgment, convention, and pragmatic concerns combine to specify x.

Why dwell on this ambiguity? Suppose we look in the financial records and ask what was the cost of the product sold in period 1. We will find an answer. An answer will be there, with no hint of ambiguity. Now ask yourself to what question this answer is responding. $(P_1 \widetilde{z}_1 + x)/\widetilde{q}_1$ is the answer to the question: what was the (unit) cost for financial reporting purposes of the product sold in period 1? Is this the question we were asking?

Though it does not surface in our simple model, the accountant also deals with an overwhelming amount of data.[18]   No attempt is made to keep detailed track of every one of these inputs. They are lumped into categories: office supplies, labor of various type, materials of various types, and so on. This grouping is not perfect. Unlike items are bundled together. Detail that the economist exploits is purposely abandoned.

The accountant has neither the detail nor the market prices assumed in the economist's world. The accountant cannot replicate the economist's constructs. Yet the accountant uses the economist's language and ideas. This creates confusion. The accounting library contains many references to cost, revenue, and income. The intuitive meaning of these references is derived from economics. That intuition is inadequate. Ambiguity surrounds

---

[17]A common claim is that the accrual summation is superior to the cash based summation, as it at least attempts a proper matching of revenue and expense, or accomplishment and effort. We, however, adopt the perspective of a user of the accounting library. Both summations contain information. The two together are likely to be more useful than either standing alone. Related to this is another common claim that managers succumb to the temptation of short-run behavior by trying to maximize their accounting performance. We will analyze this at length in subsequent chapters. For now notice the incongruity. Accrual procedures are designed to provide a long-run perspective, while managers are unduly tempted to maximize these measures!

[18]Try listing the factors consumed by your favorite fast food enterprise, and that is the easy case.

these references, even when we understand the particular accounting procedures that have been employed. Procedures also vary from library to library. In addition, the procedures employed are influenced by financial reporting concerns.[19]

The professional manager understands what is in this library and how to decipher what is found to best suit the purpose at hand. The deciphering exercise is usefully thought of as identifying the revenue recognition and expense matching rules that govern the particular library.

4.5   Summary

This chapter has used our two product firm, in single and two period formats, to contrast the economist's and accountant's portrayals of the firm's activities. Both focus on products, customer payments, factors, and payments for the factors. The economist, taking full advantage of the presumed knowledge of the firm's technology, is able to identify economic rent, economic income and economic cost. Separability issues confine him to total cost (in present value terms) and marginal cost; and his periodic income is nothing other than the factor cost of the committed financial resources for the period.

The accountant, who operates without the mythical advantage of all-powerful knowledge, employs a constructive procedure to identify product and period costs, revenue and of course accounting income.

Period versus product cost is an important distinction. The economist, recall, does not deal with this distinction. The textbook economist, equipped with perfect markets, encounters no ambiguity. Yet the period versus product cost distinction is central to the accounting process. Some cost items are associated with the product, while others are associated with the period in question. The economist stresses knowledge of the firm's cost function. The accountant works with partial knowledge of this cost function, and even then in highly aggregated fashion.

---

[19] Financial reporting uses a two-step process to identify the accounting income for a particular period. The first step is to identify the revenue; the second is to match expenses with the identified revenue. The general guideline for recognizing revenue is that a substantive interaction in the product market must have occurred. Regulatory authorities call for recognition of revenue when it is "realized" or "realizable" and "earned." Stated differently, revenue is not to be recognized until the interaction between the firm and its customer is "virtually complete" and a "transaction" has occurred. With this in hand, we search for the cost of the factor consumptions associated with that revenue. As you will learn, convention has a heavy influence here.

4.6   Bibliographic Notes

Connections between accounting and economics have been explored in a variety of contexts. Paton [1923] is a classic. Whittington [1992] is an excellent introduction to accrual accounting and carries along numerous links to economics. Parker, Harcourt, and Whittington [1986] provide a collection of readings on income measurement drawn from the accounting and economics literature. Demski and Feltham [1976] stress the interplay between economic and accounting cost. Beaver [1998] emphasizes information content of the accounting measures. Christensen and Demski [2003] also emphasize information content, and build their analysis on a simplified economic model of the firm, similar to the approach taken here.

4.7   Problems and Exercises

1. Carefully contrast the concepts of economic income, economic rent, and accounting income.

2. In the story beginning with Tables 4.1 and 4.2, the firm's income totals $\widetilde{R} - C(\widetilde{q}; P)$. We also claim this is the firm's net cash flow. Explain.

3. In our multiperiod setting (e.g., Table 4.8), it turns out that over the two periods, cash flow, economic income and accounting income agree with one another. Verify this claim. Explain.

4. Define period and product cost pools. When is a period cost expensed? When is a product cost expensed? What does the economic theory of the firm say about the distinction between period and product costs?

5. Carefully contrast a product's marginal cost with its unit cost.

6. Are the factor consumptions reflected in a direct cost pool independent of the price of other factors or, for that matter, of the quantity of other products produced? Explain

7. Return to Tables 4.3 and 4.4. What is the cost of goods sold total for each of the products in each of the tables. Explain.

8. *product versus period cost*
   Suppose pool #3 in Table 4.7 is designated a period cost pool. Redo the income displays in Table 4.8. Explain your finding.

9. *marginal versus unit cost*
   Ralph manages a two product firm. The technology requires a mixture of capital and labor to produce each product. Capital is shared,

while labor is specific to each of the two products. For the first product, capital (K) and labor (L_1) must satisfy $q_1 \le \sqrt{KL_1}$ in order to produce q_1 units. Likewise, producing q_2 units of the second product requires capital (K) and labor (now L_2) such that $q_2 \le \sqrt{KL_2}$. Total capital is limited to a maximum of 200 units, so $K \le 200$. In addition, once either product is sold it is removed from inventory and packaged for shipping. This packaging requires B boxes such that $q_1 + q_2 \le B$. Ralph will produce and sell $q_1 = 200$ and $q_2 = 100$ units. Capital (K) costs 300 per unit, labor costs 250 per unit for the first product (L_1) and 400 per unit for the second product (L_2), and packaging (B) costs 50 per unit.

(a) Determine the optimal factor combinations, the total cost and the marginal cost of each product.

(b) Determine the shadow prices associated with each of the constraints in the optimization program in (a) above.

(c) Determine the unit cost of each product. For this purpose, treat all of the factor costs as product costs. The two labor cost pools will be direct cost pools of course. You should assign the capital cost to products on the basis of the relative direct labor costs.

(d) Write a brief paragraph discussing the connection between marginal and unit costs in this setting.

10. *accounting versus economic income*
    Return to the setting of Example 4.2, but now assume the firm must pay, at time $t = 0$, a franchise fee of 850 in order to legally produce and sell its products. Further assume it adopts precisely the same production plan. Determine its economic income and accounting income for each of the two periods. Also verify that this production plan is indeed optimal.

11. *accounting versus economic valuation*
    Ralph is pondering the difference between economic and accounting descriptions of financial life. Provide three explicit examples, one where a good guess is economic value exceeds accounting value, one where accounting value exceeds economic value, and one where they are about equal. In each case identify the accounting recognition rules that produce the particular bias, or lack of it, in the accounting valuation.

12. *accounting expense versus economic cost*
    Paton states "...the accountant's 'expense' for the particular business and the economist's 'cost of production' are two quite different things...[The] whole scheme of accounting is based upon the plan of

showing as costs or expense only the expirations of purchased com-
modities and services, not the economic value of the services con-
tributed by the business itself in furnishing capital and management"
(Paton [1923], pages 493-494). Carefully comment on the difference
between accounting expense and economic cost.

13. *accounting income*
We often see cases where a firm's accounting income is reported and
the price of the firm's equity, traded on an organized exchange, be-
haves in seemingly strange fashion. Provide a coherent story in which
a firm reports significantly higher income for a period and its value
(as determined by traders on an organized exchange) declines.

14. *accounting versus economic history*
Ralph forms a firm by investing $1,000$ dollars. This cash is immedi-
ately paid for a machine with a useful life of 3 years. The net cash
inflow from this machine will be 110 at the end of the first year, 0
at the end of the second year, and 1,197.90 at the end of the third
year. Net cash inflow is paid as a dividend immediately upon receipt.
Also the third year net cash flow of 1,197.90 consists of 1,000 from
customers and 197.90 salvage value received when the machine is re-
tired at that time. (The firm ceases to exist after the year 3 dividend
is paid.)

(a) Assume Ralph's accountant uses sum of the years' digits depre-
ciation. Tell Ralph's history with end-of-year balance sheets,
periodic income statements, and periodic cash flow statements.
The initial balance sheet should show an asset (call it P&E) of
$1,000$ and capital stock of $1,000$.

(b) Assume the interest rate is $r = 10\%$. Tell Ralph's history, again
with balance sheets, income statements, and cash flow state-
ments, but in terms of economic income.

(c) Construct a 3-year income statement for Ralph. Does the total
of the income numbers in your answer to (a) agree with this
answer? What about the total of the economic income numbers?

(d) Closely examine your accounting and economic income numbers
for the second year. What numbers in the overall history deter-
mine the economic income in the second period? (Hint: think
in terms of change in present value plus dividends.) What num-
bers in the overall history determine the accounting income in
the second period?

(e) To what extent is it correct to say accounting income is a back-
ward looking calculation, based on actual transactions, and eco-
nomic income is a forward looking calculation, based on antici-
pated transactions?

15. *the accountant's task*
    "Accounting...might best be defined as the art which attempts to break up the financial history of a business into specific units, a year or less in length. In other words, it is the business of the accountant to prepare valid statements of income and financial condition in terms of specific periods of time; and the propriety of a particular procedure cannot be judged fairly except in terms of its effect upon the integrity of the particular statement" (Paton [1922], page 469). Do you agree? Carefully explain your reasoning.

16. *accounting versus economic history*
    Ralph has designed a consumer product, and launched a manufacturing and sales organization. To keep the problem uncluttered, the firm has a life of exactly three years. It is incorporated at time $t = 0$, with Ralph acquiring all shares for $1,526.35$. (No apologies are offered for this obtuse amount.) You will also notice Ralph lives in a tax-free environment. The interest rate is $r = 9\%$. Subsequent to incorporation, the following cash transactions occur.

	$t = 0$	end yr 1	end yr 2	end yr 3
payment for				
equipment	1,526.35			
materials		200	100	0
labor		300	400	300
misc. factors		200	600	600
customer payments		1,400	1,600	1,500
dividends paid		700	500	600

(a) Determine Ralph's economic income for each period.

(b) Determine Ralph's accounting income for each period. You should use straight line depreciation for the equipment, and treat all of the other expenditures as direct cost pools associated with the respective period's output.

(c) Verify that lifetime income (economic or accounting) equals lifetime cash flow for the firm.

(d) Now assume no dividend is paid until the end of year 3. Any cash on hand is invested at $r = 9\%$. Determine Ralph's cash flows, economic income for each period, and accounting income for each period. Write a short paragraph detailing your observations.

5

A Closer Look at the Accountant's Art

We now take a closer look at the accountant's product costing art. We know accounting measures of various costs incurred during a period are recorded or assembled in a variety of cost pools. This reflects three significant departures from the underlying economics. First, nothing approaching the detail presumed by the economist is tolerated. Rather, cost pools reflect aggregate clusters of factor consumption. Second, while the economist does not concern himself with separability, the accountant literally separates activities and their costs by periods. Third, in contrast with the economist, the accountant does not regard the implicit cost of funds committed to the firm a component of cost as the firm marches through time. Rather, this is buried, if you will, in accounting income. It is below the line, so to speak.

But how does the accountant move from cost pools to product costs? The answer is found in the accountant's approach to so-called cost behavior. Each cost pool is viewed as a stand alone cost story, complete with its own approximate cost function. It is useful to think of each cost pool as accumulating the costs associated with some set of activities, with these costs being reasonably well described by a cost function, or subfunction (as we are dealing with but a portion of total cost when focusing on a specific pool). This subfunction almost surely will be a linear approximation, one that holds only over a restricted range of activities. We will use the acronym LLA to describe such a "local linear approximation." And it is this noted (appearance of) separability and linear approximation that the accountant exploits to move from cost pools to product costs.

We begin with an extended illustration, mechanically moving from cost pools to product costs. The building blocks of linear approximation and

aggregation are then formalized. We then combine these building blocks with cost allocation to provide a general description of how product costs are estimated. Even so, the details are close to numbing in effect. Keep in mind this is the accountant's world.

Yet it is the recipe that emerges that is important. No two firms measure product costs in identical fashion. But focusing on a firm's approach to aggregation, to the LLAs it employs, and to how aggressive it is in pursuing cost allocation will allow you to quickly identify the major details of its cost measurement system. This is why we stress the trio of aggregation, LLAs, and cost allocation. Do not lose sight of this recipe.

5.1   An Extended Illustration

Ralph, Ltd. is a management consulting organization. During its most recent year, Ralph provided service to two clients. One client (hereafter client A), a manufacturing firm, hired Ralph to design and install a new enterprise-wide information system. The other client (hereafter client B), a municipality, hired Ralph to study labor turnover in the city government. Both projects were completed just as the year came to a close.

Ralph, Ltd. is legally organized as a corporation. All common stock is owned by Ralph. Besides Ralph, the firm employs three associates, several technical specialists, and several nontechnical staff.

Financial records for the year in question show the various costs in Table 5.1, organized in a variety of cost pools, were incurred. Clearly, multiperiod issues are present, and various accruals have been used to approximate the costs of the underlying factor consumptions. But that is only part of the story.

Notice some costs are identified by client, i.e., subcontracting and reimbursable items. The salaries of the three associates are separately identified, while those of the others are grouped into technical and nontechnical totals. The data for our cost construction exercise arrive in aggregate form.

Professional development covers expenditures on technical materials and short courses that the employees use to maintain and increase their technical expertise. Transportation costs are due largely to leased automobiles that are used by Ralph, the three associates, and some technical employees. This is considered a routine cost of doing business and is not explicitly billed to clients. Air travel, on the other hand, is routinely billed as a reimbursable cost to specific clients. Depreciation is included in the equipment category. Lease amortization is included in the office space and transportation categories.

Importantly, then, explicit expenditures, such as payments to employees, are reflected in some of the cost pools, while interperiod allocations or assignments, such as depreciation, are used in identifying the costs in other

pools. The firm has designed its accounting library to capture expressions of factor consumptions, as detailed in Table 5.1.

TABLE 5.1: Costs Incurred by Ralph, Ltd.

cost pool	amount ($)
Ralph's salary	150,000
salary of first associate	120,000
salary of second associate	90,000
salary of third associate	80,000
salaries of technical employees	115,000
salaries of nontechnical employees	95,000
fringe benefits (insurance, pensions	
accrued vacations, employment taxes)	130,000
subcontracting	
client A	110,000
client B	15,000
other reimbursable costs	
client A	70,000
client B	35,000
advertising	15,000
supplies	48,000
transportation (other than reimbursable)	32,000
professional development	135,000
equipment	140,000
office space, heat, light, etc.	220,000
federal, state and local taxes	
(exclusive of employment taxes)	95,000
interest	25,000
total	1,720,000

Ralph and the three associates keep detailed records of how their time is spent. These records reveal the following.

	client A	client B	unbilled
Ralph	20%	30%	50%
first associate	75%		25%
second associate		70%	30%
third associate	70%	30%	

Unbilled time refers to time spent by the respective employee that cannot be directly associated with any of the client projects. Time spent on professional development, searching for and bidding on new projects, training

staff, and general administrative chores are all lumped into this category. Ralph expects the unbilled percentage to average about 25% of the total.

Bonuses were also paid to the various employees. The bonus pool was 250,000 dollars. It was shared among all employees in proportion to their salaries. The bonus was paid two months after the end of the year in question. Thus, it is not included in the above tally. Ralph, Ltd. also declared and paid a dividend of 120,000 at the time the bonuses were paid.

Exclusive of the bonuses and dividends, these costs total 1,720,000. What was the cost of client $A's$ project? What was the cost of client $B's$ project? (Do not prejudge the issue of whether the bonuses or dividends are costs.)

5.1.1   One Among Many Answers

In Table 5.2 we show how a typical accounting system might answer these questions. To avoid distraction, (000) has been omitted.

The cost construction begins with the salaries of Ralph and the three associates. We know their total salaries, and the breakdown of their time across the two clients and the unbilled category. This leads to respective assignments of 176, 132, and 132. Notice we are treating the unbilled category as a third product at this point. More will be said about this choice.

Now consider the salaries of the other employees. Here we chose to assign these salaries to the three products based on the above identified salary breakdowns.[1] Think of the 440 total salary of Ralph and the associates as labor input that we can specifically identify with the three products. We then assign the cost of the other labor inputs in proportion to how these directly identified inputs are used: 176/440, 132/440, and 132/440. These calculations provide us the noted total labor cost (exclusive of fringe and bonus) for the three products: 260, 195, and 195 for a total of 650.

Next we tackle the fringe and bonus. Both are assigned on the basis of total labor cost, exclusive of fringe and bonus. We are told this is how the bonus was determined. One might take the view that the bonus is a type of periodic profit sharing and should not be assigned to individual products. This view has merit. Our construction treats it as another conduit for delivering compensation for employee services.[2]

---

[1] We might want to use time rather than salary of Ralph and the associates here. We also might want to ask Ralph for a subjective estimate of how the other employees were used. At this juncture we are describing the basic philosophy of cost construction. Once this is well understood, we will turn our attention to the questions of how to adapt what we find in the accounting library to our purpose at hand and how to structure what is placed in the library in the first place.

[2] Notice we use the forthcoming bonus in the construction, an as yet unrecorded expenditure. Presumably last year's bonus was paid and therefore recorded this year, but we are purposely ignoring it on the presumption it relates to last year's production.

TABLE 5.2: Product Cost Construction for Ralph, Ltd.

cost pool	client A	client B	unbilled	total
labor costs				
Ralph				150
.2(150)	30			
.3(150)		45		
.5(150)			75	
first associate				120
.75(120)	90			
.25(120)			30	
second associate				90
.70(90)		63		
.30(90)			27	
third associate				80
.70(80)	56			
.30(80)		24		
subtotal	176	132	132	440
technical employees				115
(176/440)(115)	46			
(132/440)(115)		34.5		
(132/440)(115)			34.5	
nontechnical employees				95
(176/440)(95)	38			
(132/440)(95)		28.5		
(132/440)(95)			28.5	
total, excluding fringe, bonus	260	195	195	650
fringe (130/650 = 20%)	52	39	39	130
bonus (250/650 = 38.5%)	100	75	75	250
total labor cost	412	309	309	1,030

Fringe is likely to be a complicated affair. Younger employees have less vacation time than older employees. FICA taxes apply only up to a particular salary limit. Health insurance is a complicated package arrangement with the insurance vendor. Our construction deals with this in a nearly cavalier yet common fashion. We lump it all together and simply average!

In this way we assign total labor cost of 1,030 to the three products: 412, 309, and 309, respectively.

Materials, supplies, and so on are treated in a parallel manner. Subcontracting costs are identified by specific projects. We assign them accordingly. Other items are also identified by specific projects.

TABLE 5.2 Continued: Product Cost Construction for Ralph, Ltd.

cost pool	client A	client B	unbilled	total
materials, supplies, etc.				
subcontracting	110	15		125
other directly identified items	70	35		105
supplies (48/650 = 7.4%)	19.2	14.4	14.4	48
transportation (32/650 = 4.9%)	12.8	9.6	9.6	32
subtotal	212	74	24	310
unit costs (labor, materials, supplies)	624	383	333	1,340
unallocated costs				
advertising				15
development				135
equipment				140
office space				220
interest				25
taxes				95
subtotal				630
total				1,970

Supplies are not so identified. We chose to assign supplies the way we assigned fringe and bonus payments. This reflects the hunch that supplies are used with labor, and labor cost exclusive of fringe and bonus is a reasonable indicator of the manner in which supplies are consumed. Transportation is treated the same way. Since transportation largely consists of automobiles supplied to various employees, it seems reasonable to assign these costs in that fashion.

The remaining costs are not assigned to specific products.[3] We therefore wind up with a unit cost of 624 for client A, 383 for client B, and 333 for the "unbilled product."[4]

What about some of our choices in this construction? Were we wise in our handling of transportation? Should we turn around and assign the unbilled product? Surely equipment and office space could be assigned to products. Feelings of uneasiness are to be expected here. Cost construction is a matter of choice, choice in the presence of considerable ambiguity.

For example, we decided it was best to treat unbilled as a separate product. An important reason is that a major activity for our consulting firm

---

[3] You should recognize the pattern displayed in Tables 4.3 and 4.4.

[4] Recall we reserve the phrase "unit cost" for the accounting cost per unit, and in this case each client is treated as a single unit of some product.

is developing new clients and new skills. In this sense, one of today's products is getting ready or preparing to serve better tomorrow's clients. The current period cost of this preparation is reflected in the unbilled category.

Perhaps, then, we should have assigned all advertising (15) and all development (135) to the unbilled category. We decided against such an assignment because the unbilled category is in reality a murky joint product. We have lumped administrative items into this category as well, and cannot fully separate the two items. So it seemed best to adopt the construction presented. For the same reason we did not search into prior years' records to find an unbilled category to assign to the current projects.

Dividends present another dilemma. Our inclination, especially given training in financial accounting, is to keep it invisible in the cost construction. Remember, though, Ralph owns 100% of the capital stock. Is the dividend a payment to capital or a payment to labor or a return of capital? As you were warned, the details are close to (if not completely) numbing.

Cost construction is a matter of choice. Our choices are catalogued in Table 5.3. We reiterate that the product costing exercise is one of constructing expressions of cost. Countless choices are involved in any such construction. We will learn that these choices depend, in subtle ways, on the circumstances at hand. For the moment, the important point is to acknowledge the presence of choices in the algorithm.

5.1.2   Central Features of the Construction

Several features of our construction, and the setting in which it takes place, should be noted. First, the exercise began with a set of cost pools (in Table 5.1). These carry the initial recording of the identified factor consumptions.[5] The firm designed its accounting system to capture these consumptions in these cost pools. Yet, when it comes to estimating the cost of the products we may ignore some of these pools (we did not, however), and we may go in search of others: the as yet unrecorded bonus and dividends in our case. This seemingly innocuous move on our part belies an important message of always being willing to tailor what the library reports.

Second, as laid out in Chapter 4, each cost pool is designated as a product or period cost pool. For example, the various salary pools, the fringe pool and supplies pool were all designated as product cost pools, meaning they will be associated with the products. Conversely, advertising, equipment and interest, among others, were designated as period cost pools, meaning they are treated as costs of the period (in which they were incurred) and not associated with the products per se.

---

[5] Again, check the definition. These are temporary accounts, meaning they are accounts that will be closed by the end of the accounting cycle.

Third, the product cost pools were further subdivided into direct and indirect categories. Subcontracting and other reimbursable costs were directly identified with the respective products. In turn, some of the indirect product cost pools were treated in seemingly obvious fashion. This is the case for the salaries of Ralph and the associates, where time records were used to assign the respective salary totals to the products.[6] These time records, however, may be more or less exact. But the identification data were available, and we implicitly regarded them as sufficiently reliable to use in the construction exercise.[7] In addition, this identification may or may not line up with actual payments for the inputs. Consider payments to the three associates. We identify their salaries and prorate them over the products in question. But fringe is also a part of the payment for their services. We don't have explicit prices for fringe, only the total of accrued payments. So we average. A parallel comment applies to the as yet unpaid bonus.

Contrast this with the subcontracts. There, we presume each subcontract explicitly identifies the client projects and subcontractor payments. End of story, perhaps. What if there is a long term relationship between Ralph and a subcontractor, and it is implicitly understood payments are "smoothed" through time? For example, how do we interpret the case where the auto mechanic fixes a minor item and says, "I'll catch you next time"?

Of course, many inputs are not identified by product. For some inputs this is impossible. How much of the office space is used by each product? We might prorate cost of office space among the products, but this space is jointly used by all the firm's productive activities. This is why it was designated a period cost pool.

Other inputs are not identified by product, simply because to do so would be impractical. Supplies is an illustrative category. We don't know what is in this category, or even what each separate item cost. What we know is that a category called "everything else" is used to account for all materials that are not accounted for in more explicit fashion. Some items in this category are treated on an accrual basis, while others are treated on a cash basis. The firm does not keep track of separate items or their respective costs. Everything else, so to speak, is lumped together.

Regardless, our goal is to associate the cost of supplies with the products, so the supplies cost pool is designated an indirect product cost pool. But

---

[6] Even so, as each of these pools is not identified with a single product, they are indirect product cost pools.

[7] Quality of source data will not be taken for granted in our study. Just to illustrate, suppose one of the associates is new and eager to succeed. One client is "old hat." The firm has considerable experience with the client and its problems. The other client is new, and is calling on the firm to work in new and novel ways. Can we rely on our anxious associate to be consciously and subconsciously unbiased in estimating time spent on the two clients? The federal data bank on fishing stories has a similar problem.

absent any explicit association, we must find some method for assigning the cost to the products. Here we use labor cost (exclusive of fringe and bonus) as the basis on which to assign the costs in this pool to the products. (Glance back at Table 5.2.) The remaining indirect product cost pools are treated in parallel fashion.

TABLE 5.3: Choices in Product Cost Construction for Ralph, Ltd.

cost pool	type designation	assignment basis
salaries		
Ralph	indirect product	reported percentage
first associate	indirect product	reported percentage
second associate	indirect product	reported percentage
third associate	indirect product	reported percentage
technical employees	indirect product	Ralph, assoc. salaries
nontechnical employees	indirect product	Ralph, assoc. salaries
fringe benefits	indirect product	assigned salaries
subcontracting[8]	direct product	direct
other reimbursable costs	direct product	direct
advertising	period	
supplies	indirect product	assigned salaries
transportation	indirect product	assigned salaries
professional development	period	
equipment	period	
office space, etc.	period	
taxes	period	
interest	period	
bonus	"indirect product"	assigned salaries
dividends	"period"	

All of this is summarized in Table 5.3. Dwell on the designation of each cost pool using the direct product, indirect product and period cost pool categories introduced in Chapter 4.[9] We stress there is nothing sacrosanct in these designations and association techniques.

This is why we stressed the theme of choices in cost construction. We chose not to assign some pools to products, to treat them as period cost pools. We chose to assign other (indirect product cost) pools to products

---

[8] Though for display purposes the subcontracting and other reimbursable cost pools are displayed as single pools, being direct there is one for each product.

[9] Notice our treatment of the bonus and dividend payments. Neither is a technical liability as of the end of the year, and thus entered formally in the accounting records.

using particular assignment techniques. We chose to assign the directly identified categories to the respective products. Even this direct assignment may be ambiguous. For example, we often can associate some labor with a specific product. If the firm has a policy of full employment and we want to estimate the marginal cost of that product, it is not clear that we should assign the cost of the directly identified labor to the product.[10]

Product costing is a well-developed art. It is an art practiced in the face of considerable ambiguity. Our immediate aim is to place some structure on this art form, as an aid to documenting the choices that lead to typical cost constructions.

5.2   Unit Costing Art

The economist, recall, begins with all factors, all factor prices, and the production function in full view. In this luxurious state, the cost function is derived, and marginal cost is showcased. The accounting system attempts to emulate this derivation. But it begins with a handicap. Some factors are known, some market prices are known, and categories and averages of others are known. Relative to an economist, the accountant's product is by necessity delivered with ambiguity. To engage in allegory, the accountant is called upon to create nouvelle French cuisine in a high school cafeteria.[11] This is why we reserve the term unit cost for the cost per unit that is recorded in the accounting library.

We have a glimpse of this handicap in Table 5.2. Recall that accounting cost records systematically exclude capital cost. So to begin, we have items in economic cost that are not included in the accountant's tally of total cost for the period. This is the reason for our warning to think carefully about how to treat dividends in the Ralph, Ltd. example. Dividends might be return of capital, might be a payment to capital suppliers, or might even be a payment to Ralph for labor services.

Also recall that timing differences between economic and accounting income are likely. This implies we should anticipate timing differences on when particular cost items are recognized. Economic and accounting depreciation differ. Cash recognition procedures applied to miscellaneous

---

[10] The choices are even more subtle. They extend into designing the data gathering in the first place. Which items to group together, which ancillary data to collect, and which minor items to treat on a cash basis are all matters of choice.

[11] A less apocryphal analogy arises with price indices. There we take a basket of goods and track the market price of that basket of goods through time. Of course, the quality of the goods may change with time, the array of substitutes may change, relative prices of other goods may change in different fashion, and so on. We use the price index to give us an overall picture, recognizing its limitations. The same holds in accounting, even when we do not use constant dollar techniques.

supplies is another example. Various employee bonuses provide additional illustrations. In addition, the economic cost function presumes ruthless, mistake free production. Should we presume this to be the case in Ralph, Ltd.?

Finally, economic cost reflects an assumption that all inputs and all prices of inputs are known. The accountant, as we have discussed, does not have this knowledge. So, even if total accounting cost agreed with economic cost, we should expect slippage in the individual component constructions.

What, then, does the accountant do? Is there any pattern or systematic tendency that is employed in the art of product costing? The answer is yes. Three building blocks are fundamental to the accountant's art: aggregation, linear approximation and allocation. They were evident in the Ralph, Ltd. illustration and are discussed further below.

5.2.1   Aggregation

The aggregation building block is a familiar technique. Imagine all the transactions that occur in a grocery story. The checkouts electronically record all the items sold. Payroll records are extremely detailed. Supplier records are also detailed. The store manager has an incomprehensible array of data. These data are aggregated for obvious reasons. The same applies to accounting records in general.

The most vivid example of this aggregation in Ralph, Ltd. is the supplies category, with a total identified cost of 48,000. Imagine what might be grouped together in this category. Miscellaneous, immaterial (pun intended) office supplies are surely included. Paper, pencils, pens, miscellaneous laptop supplies such as replacement batteries, and so on are all included in this manner. They are also included on a cash basis. Inventory records are not maintained for such trivial items. Janitorial supplies and bulk paper are also included. The list goes on. Some individuals will use a different mix of supplies than others. One client project will entail use of a different set of supplies than another.

None of this detail is used in the cost construction exercise. Instead, we group somewhat like items together. These groupings manifest themselves in the cost pools with which we began the unit cost construction exercise for Ralph, Ltd. This is where the firm's approach to aggregation becomes apparent. For example, does it use a large or a small number of indirect product cost pools?

5.2.2   Linear Approximation

Once a cost pool is identified, the linear approximation building block comes into play. Consider the manner in which we assigned transportation cost to the three products in Ralph, Ltd. We took the total recorded in the transportation cost pool of 32,000, and divided it by the total salary cost of

650,000. Transportation cost averages 4.9% of salaries. We then used the 4.9% datum to assign transportation cost to each product, as a function of their respective (assigned) salary costs. Notice we treat the cost pool as though it had its own, separable cost function. And cost in this pool is not a function of output, as in economic theory, but is a function of some *synthetic*, aggregate "output" variable, total salary cost in this case.[12]

Suppose for the sake of argument that separability is not of overwhelming concern and that it is reasonable to view transportation cost as varying with labor input. More labor input necessitates more transportation. Further suppose total salaries is a good measure of labor input for this purpose. This means we should visualize transportation cost as some function of total salary cost.[13]

What might this function look like? A typical auto lease contract calls for a monthly payment that is independent of mileage, up to a limit. The parties usually commit for several years, but monthly, weekly, and daily arrangements are also possible. In addition, mileage charges are usually imposed once mileage goes beyond the specified limit. Gasoline, insurance, tolls, parking, and so on are usually paid by the lessee.

This suggests the function would look something like that depicted in Figure 5.1. If labor input drops to zero in the short-run, the lease arrangements can likely be scaled back so only a modest payment is made.[14] If labor input jumps dramatically, the lease arrangements can likely be expanded, and on favorable grounds. The nonlinear graph in Figure 5.1, labeled the presumed cost curve, is meant to be suggestive. We gloss over details such as when to increase the fleet by another unit, and so on. The important point is we do not expect the cost to be zero when labor input is zero, and we do not expect the cost to increase proportionately with labor input. For the sake of argument, our graph depicts the cost as increasing less than proportionately with labor input.

Contrast this with the manner in which we assigned the transportation cost. That assignment used the function

$$\text{transportation cost} = (32/650)\text{salaries} = (.049)\text{salaries}$$

A salary level of 195, for example, was assigned a transportation cost of $(.049)195 = 9.6$. The implied cost function is also plotted in Figure 5.1.

---

[12] In the general estimation literature, an explanatory variable is called an independent variable. Present-day jargon in the accounting, marketing, and management literatures calls them cost drivers. We will do our best to avoid the phrase, as it implies a level of separability and understanding that is unlikely to be present. For this reason we prefer the phrase synthetic variable.

[13] Indeed, we are saying the cost in one cost pool is reasonably thought of as a function of the costs in some other cost pools!

[14] Cancellation provisions are common features of automobile lease arrangements.
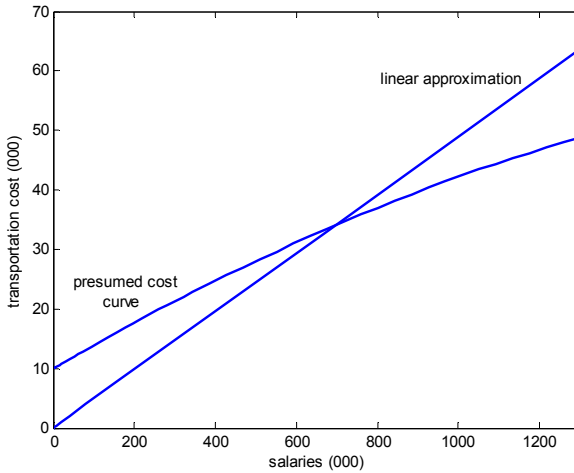
FIGURE 5.1. Presumed and Approximate Cost Curve for Transportation Cost Pool

Notice we have plotted both the presumed cost function and the "linear approximation" used in our cost construction.

How should this be interpreted? We began our thought exercise with some presumed cost curve, relating cost in the transportation cost pool to a synthetic variable of salaries. We then overlaid a linear approximation, and used it to assign the cost pool to products, by linking the synthetic variable to the products themselves. The approximation itself may or may not be close. Examine the graph in Figure 5.1 when salaries are close as opposed to far removed from 650.

Accounting procedures invariably begin with an approximation of various components of the firm's cost curve. These approximations, in turn, are usually linear approximations, where the costs in a cost pool are linearly related to output or synthetic variables. We use the phrase *local linear approximation*, or LLA for short, to remind ourselves of the use of this technique. The adjective local is carried along to remind us there is no guarantee the approximation is accurate over a wide range. The presumption is that it is sufficiently accurate over a restricted or local range.

In this way we construct product cost data, unit costs in our terminology, by working with subsets of factors aggregated into cost pools, by identifying approximate expressions for how the costs of these subsets of factors vary, and by linking these building blocks to the products themselves. The linkages may be direct or second-hand. The explanatory variables used in the component of the cost function might be the products themselves or some intermediate or synthetic explanatory variable. For example, we

linked the labor cost of the associates in explicit fashion, while we linked that of the technical employees in second hand fashion.

The importance of LLAs in the accountant's art is driven by three considerations. First, we simply do not know "the cost curve." We literally invented the presumed cost expression in Figure 5.1. All we really know is the single point of salaries $= 650$ and transportation cost $= 32$. Admittedly, we might collect other points from recent periods and use common sense and introspection to construct a few others. Regardless, the costing exercise begins with an absence of what the economist presumes to know.

Second, even if we knew the cost curve, pragmatic considerations would lead us to use approximations. Beyond a point (no pun), there are diminishing returns to keeping track of detail. (Minutia is the operative noun.) Backing off on detail leads to an approximation. A local linear approximation turns out to be the overwhelming choice.

Third, the accounting library must maintain its integrity. Verifiability is important. We must be able to verify source documents and calculations. Linear computations are simply easier to verify.

Use of LLAs also leads to some popular, far too common, misuse of terminology. The general equation for a line is $y = a + bx$, with intercept a and slope b. Think of this as a function of x:

$$y = f(x) = a + bx$$

In our world of cost pools, y is the cost total in some pool, and x is the variable, output or some synthetic variable, used to explain that cost total. So this has the appearance of a cost curve, or more formally subcost function.

Now, if the slope is nil, $b = 0$, common usage is to say the cost in this cost pool is a fixed cost. Likewise, if the intercept is nil, $a = 0$, common usage is to say the cost in this cost pool is strictly variable. And in the general case of $a \neq 0$ and $b \neq 0$, common usage calls a the fixed cost component and bx the variable cost component of the cost in the cost pool.

Of course, this terminology would be precise if we were discussing an economic cost curve that was linear. Moving to an LLA requires some diligence. The algebra is the same, but the economic content is not. This leads to confusion. And this is compounded by the fact we are dealing with a cost pool as opposed to the entire array of factors.

Fixed and variable cost have economic meaning when we begin with the economist's (short-run) cost curve. Conversely, the accountant uses a host of LLAs to construct an approximation to the economist's cost curve. Each of these LLAs relates some component of cost to some explanatory variable. This results in a functional description of $y = f(x) = a + bx$. But a is nothing other than the intercept of the LLA in question. b is nothing other than the slope of the LLA in question. Common usage, as we said, is to call a the fixed component of the cost in question and b the per unit variable cost component of the cost in question.

It would be imprudent to deny existence of this common usage. It also would be imprudent not to dwell on the fundamental difference between (economic) fixed cost and the intercept of an LLA.

Suppose separability is not a major concern and that we also constrain the explanatory variable to lay in some restricted range, say $x_L \leq x \leq x_H$. The interval from x_L to x_H is called the relevant range. It may then be the case that b is a reasonable approximation of the marginal cost of the cost component in question, when x is so restricted. May, however, is a statement of possibility, not of inevitability. This is why we stress the terminology of an LLA with intercept a and slope b.[15]

5.2.3   Cost Allocation

Using the chosen LLA to assign the costs in the cost pool to other cost pools or products is the very essence of cost allocation. Glancing back at Tables 5.2 and 5.3 we see cost allocation at work in virtually every indirect product cost pool. For example, the transportation cost pool had a total of 32 (000), and we decided upon an LLA of

$$\text{transportation cost}\ = y = .049x$$

where x is total salaries. In turn, the two clients and the unbilled categories tallied respective salary costs of $260, 195$ and 195. Thus, client A is allocated $.049(260) = 12.8$, etc.

Notice we take the accumulated cost in an indirect product cost pool and distribute it, allocate it, assign it, to activities on some hopefully reasonable basis. The chosen LLA provides the basis. But there is more to the story.

First, this is a procedure. It is an accounting procedure that has no explicit reference point in the economic theory of cost. The economist never allocates cost. The economist knows the cost curve; the accountant uses cost allocation to attempt to say something about a point on the unknown cost curve.[16]

---

[15] Continuing, in Chapter 2, expression (2.5) and the accompanying note 11, we noted Shepard's Lemma, that the partial derivative of economic cost with respect to a specific factor's price is the amount of that factor consumed in production of q units. In the world of accounting approximation, then, the partial derivative of accounting cost in some cost pool with respect to a well chosen price index for factors in that pool should approximate the value of the independent variable in the cost pool's LLA.

[16] On the other hand, the economist often worries about sharing the cost of a common facility, say, neighbors jointly sharing the cost of a neighborhood improvement. The economist uses the language of allocating the cost of the common facility among the individuals. This is a misnomer. Sharing the cost in this context is a euphemism for making payments in a way that the common facility is paid for. Do not confuse actual payments with the accountant's cost allocations. The latter always take place in the accounting library. They may be descriptive of resource transfers, but they are not resource transfers. They are calculations in a data bank.

Second, the cost pool's LLA might employ a single explanatory variable, or it might employ multiple explanatory variables. The usual story is the single explanatory variable case.

Third, the allocation may assign the cost pool's costs to virtually any combination of products, other cost pools, or the period. For example, human resource costs often wind up partially assigned to product, to various overhead pools such as maintenance, and to the period.

As cost allocation is an accounting phenomenon, and is central to the accountant's art, it behooves us to link it to our growing list of formalities. For this purpose, recall we defined a direct product cost pool as a product cost pool that is associated with a single product. Now suppose our firm produces a number of products, with quantities given by $q = [q_1, q_2, ..., q_n]$. A direct product cost pool's LLA must be of the form $y = a + bq_i$, otherwise it is not unambiguously associated with a single product. An indirect cost pool, in turn, is a product cost pool not associated with a single product. This means its LLA is of the form $y = a + \sum b_i q_i$ or of the form $y = a + bx$, where x is one or a set of synthetic variables. Thus, the very structure of the cost pool's LLA is intimately connected to its type designation. Cost allocation now arises when we use the independent variable or variables in the cost pool's LLA to assign the cost in that pool to other pools, products or the period.

Definition 15 *Given a cost pool with LLA based on output, $y = a + \sum b_i q_i$, or based on synthetic variables, $y = a + bx$, cost allocation is an accounting procedure whereby the total cost in the cost pool is assigned to some combination of the products, other cost pools, and the period, using the independent variable or variables in the pool's LLA.*

Several elements of this definition should be noted. Cost allocation is an accounting procedure, one that deals with the manner in which the cost pool, a temporary account recall, is closed. The procedure itself is tied to the pool's LLA. If the pool uses products as explanatory variables in its LLA, then the allocation is based on products. If the pool uses some synthetic explanatory variable in its LLA, then the allocation is based on that synthetic variable. Moreover, the allocation might treat the intercept of the LLA in one fashion, say by treating it as a period cost, and the remainder in another fashion.

In addition, while we stress cost allocation as a procedure for assigning the cost in a cost pool, that cost pool itself may well be the recipient of allocated costs. Indeed, allocation also arises in an interperiod context. Straight line depreciation is nothing other than allocation based on a synthetic variable of time. Cost incurred in one accounting period will, inevitably, include interperiod allocation. Thus, some of the period's cost

pools are themselves allocation recipients. This is the very essence of accrual accounting.[17]

5.3   The Constructive Procedure

We will learn to identify and engineer specific aggregations and *LLA*s in our study, as well as to be judicious in our cost allocation. We close this overview of the accountant's product costing art with a paraphrase of the constructive procedure used in Ralph, Ltd.

The cost construction begins with identification of the costs of various inputs. This identification comes in the form of various cost pools that aggregate the costs of the various factors into a manageable number of cost pools. From our study of financial accounting, we know the costs identified in each pool will likely be some combination of actual expenditures and accrual measures of resource consumption.

For the sake of discussion, suppose we have 7 such categories (with respective cost totals denoted $cost_j$) along with three products (with respective output totals denoted q_i). Think of this as beginning with all product and period costs for the period and cataloguing them into 7 cost pools. Extensive aggregation is the key.

cost pool	type
1	direct product (#1)
2	direct product (#2)
3	direct product (#3)
4	indirect product
5	indirect product
6	period
7	period

Next we select an *LLA* for each cost pool. Pool 1 consists of factor costs that can be directly identified with the first product. So we use an explanatory variable of units of the first product, implying an *LLA* of $cost_1 = a + bq_1$. Parallel choices are made for categories 2 and 3. Subcontracting and other directly identifiable items were handled in this fashion in Ralph, Ltd.

Pool 4 consists of factor costs that we have grouped together and will assign to the three products in some indirect fashion. We must now select some synthetic (explanatory) variable that can be used to relate $cost_4$ to products. To illustrate, it may be possible to readily identify the labor input of some employees with specific products. These employees are often

---

[17]In dynamic terms, then, we would envision, say, a depreciable asset as hosting a sequence of cost pools.

called direct labor. In turn, we may find it reasonable to view pool 4 costs as well explained by the synthetic explanatory variable of dollars of readily identifiable labor, denoted $L\$$. This implies an LLA of $\text{cost}_4 = a + bx$, where $x = L\$$. Salaries of the technical and nontechnical employees were handled in this fashion in Ralph, Ltd. where we used assigned salaries of Ralph and the associates as the synthetic explanatory variable in our LLA for these costs. Similarly, in dealing with the fringe benefits we used total assigned labor cost as a synthetic explanatory variable.

Pool 5 also consists of factor costs that we have grouped together and will assign to the three products. Here we might use a different synthetic variable in the LLA. Perhaps this is a manufacturing firm and we know how much manufacturing capacity, measured by hours of machine time, was used by each product. If we find it reasonable to view cost_5 as well explained by machine hours, we would use machine hours to assign the cost to the products. Alternative stories could surely be told here, but the important point is to focus on the aggregation, LLA, allocation trio.

On the latter point, we will learn that the allocation procedure for dealing with the indirect product cost pools can be one of two types. In pool 4, for example, we have an LLA of $\text{cost}_4 = a + bx$ and, presumably, both the slope and intercept are nonzero. One approach to allocation allocates the total cost in the pool to products, on the basis of the synthetic variable. A second approach allocates the intercept amount to the period and the remainder to the products. Indeed, we will further learn that given the approximate nature of these allocations we generally work with estimated allocation rates. But this is getting ahead of our story.

Finally, pools 6 and 7 are particular aggregations of period costs. They will be assigned to the period, not the products. Of course this does not imply we are uninterested in their economic structure. It just means they are not part of the unit cost calculation.

This, then, is the general way in which product costs, unit costs to us, are constructed. We will encounter many variations on this theme. For example, we may use more than one explanatory variable for a particular cost pool. We may even use a nonlinear approximation on rare occasion. The extent of the aggregation will vary from situation to situation. In each case, though, the recipe is the same. We combine aggregations, approximations and allocation to mold product cost statistics.

5.4   Short-Run versus Long-Run Marginal Cost

A remaining issue in this overview is the connection between the accountant's unit cost measure and the economist's marginal cost measure. Is it reasonable to treat the unit cost as a readily available approximation

to marginal cost? If so, would this be long-run or some specific short-run marginal cost?

The answers are (1) perhaps it is reasonable and (2) whether we are approximating long- or some specific short-run datum depends on the circumstances at hand. Code these answers, for short, as "perhaps" and "it all depends." At the risk of feeding some deep seated cynicism, the intellectual point should not be missed. The only product cost measure of merit in a multiproduct firm is marginal cost; the accountant's art, in turn, may or may not well approximate marginal cost of some product. It all depends on the economic forces at hand as well as the countless choices made in constructing the accounting library.

Though this will concern us in subsequent chapters, it is important to confront the basic idea at this juncture. To keep things as uncluttered as possible, we illustrate with extensions of two previous examples.

Example 5.1 In Examples 2.3 and 2.8 we postulated a single product firm with long-run cost curve $C(q; P) = 200q - 18q^2 + q^3$ along with specific short-run cost curve $C^{SR}(q; P) = 162 + 204.5q - 25q^2 + 1.5q^3$. This provides a long-run marginal cost of $MC(q; P) = 200 - 36q + 3q^2$ and a short-run marginal cost of $MC^{SR}(q; P) = 204.5 - 50q + 4.5q^2$. Now focus on output of $q = 7$. This implies a short-run cost of $C^{SR}(7; P) = 883$ and accompanying marginal cost of $MC^{SR}(7; P) = 75$. (See Table 2.6.) Now imagine an LLA of $y = 358 + 75q$. This has the virtue of reporting the correct marginal cost (of 75) and of agreeing with total cost at $q = 7$: $358 + 75(7) = 883$. Think of this as a single cost pool, where we treat the intercept (of $a = 358$) as a period cost. Conversely, if we treat the intercept as a product cost, our LLA would be $y = 126.14q$, as $883/7 = 126.14$. Arguably, now, the first LLA has a short-run flavor, as that is how we constructed it, while the second has a long-run flavor, as it implies no fixed cost.

In marginal cost terms, the first LLA implies a marginal cost of 75 while the second implies a marginal cost of 126.14. In Figure 5.2 we plot these two marginal cost estimates along side their correct long-run and short-run counterparts. Notice, 126.14 is closer to either marginal cost curve for "low" or "high" output, while in between, 75 is closer. As you were warned, the viability of the accountant's unit cost as an estimate of marginal cost depends on the economic forces at hand as well as the choices made in constructing the accounting library.

This illustrates our answers of "perhaps" and "it all depends" to the rhetorical questions at the start of this section.

Example 5.2 For a second illustration, we return to Example 4.1, where three factors are used to produce two products, subject to an upper bound on the third factor. Denote the output by $q = [q_1, q_2]$. We assume output is sufficiently high that the third factor is at its upper bound, implying a specific region of the long-run cost curve or a specific version of the short-run cost curve with the third factor fixed at its upper bound.
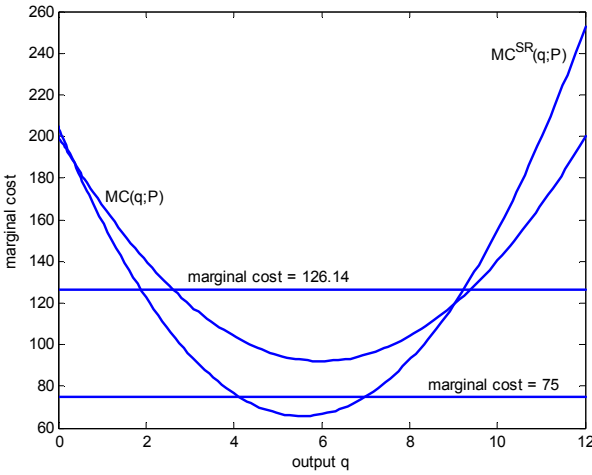
FIGURE 5.2. Marginal Cost Approximations

Either way, the three factors are naturally catalogued in three distinct cost pools, and this leads to the following cost pool expressions[18]

$$\text{cost}_1 = \frac{4}{3}q_1^2$$
$$\text{cost}_2 = q_2^2$$
$$\text{cost}_3 = 75$$

which reflects the economic cost curve in this region of $C(q; P) = 75 + \frac{4}{3}q_1^2 + q_2^2$, along with respective marginal costs of $MC_1(q; P) = \frac{8}{3}q_1$ and $MC_2(q; P) = 2q_2$. The accounting library will treat the first two cost pools as direct product cost pools, while the third will be treated as a period or a product cost pool.

Now suppose, as in the original example, that $q_1 = 15$ and $q_2 = 25$ are produced and we have cost pool totals of $\text{cost}_1 = 300$, $\text{cost}_2 = 625$ and $\text{cost}_3 = 75$. This gives us an LLA of $\text{cost}_1 = 20q_1$ (as $\text{cost}_1/q_1 = 300/15 = 20$) for the first pool, and $\text{cost}_2 = 25q_2$ (as $\text{cost}_2/q_2 = 625/25 = 25$). Treating the third pool as a period cost thus implies respective unit costs of 20 and 25 for the two products.

In turn, treating the third pool as an indirect product cost pool requires we specify some synthetic variable. We'll use $\widehat{q} = q_1 + q_2$; and this implies an LLA for the third pool of $\text{cost}_3 = 1.875\widehat{q}$ (as $\text{cost}_3/\widehat{q} = 75/40 = 1.875$).

---

[18] Given the assumed technology and factor prices, the third factor is at its upper bound (of $\overline{z}_3 = 15$) whenever $4q_1^2 + 3q_2^2 \geq 225$. See Example 4.1, and the predecessor details in Chapter 3, surrounding equation (3.5).

This leads to respective unit costs of $20 + 1.875 = 21.875$ and $25 + 1.875 = 26.875$.

The respective marginal costs at this point are $MC_1(q; P) = \frac{8}{3}(15) = 40$ and $MC_2(q; P) = 2(25) = 50$. With output sufficiently large that the third factor is at its upper bound, the firm is producing in a region where its marginal costs are increasing in output. This leads, as flagged in Table 4.5, to the accountant's systematic assignment of actual costs to products leading to unit costs that are systematically below marginal costs.

Now glance back at Table 4.5. All we have done here is reconstruct that display, but now using the explicit costing mix of aggregation, LLAs and cost allocation.

Stepping back, any application of the accountant's product costing art leads to a product cost statistic. Suppose we decide to treat some product cost statistic as an approximation of marginal cost. Is this best thought of as an approximation to long-run or to some short-run marginal cost? No general answer is possible. The accountant's art produces a number. This number may be close to or far removed from the portion of the long-run marginal cost curve we had in mind. Similarly, this number may be close to or far removed from the portion of some short-run marginal cost curve we had in mind.

This may appear curious or even cynical. Yet it is the natural manifestation of treating accounting as a library. Various choices go into design of the library. The resulting choices may produce something that is close to what we want or not so close. We and our immediate curiosity are just one of many users of that library. How we use the library depends on how the library was constructed and on our context. Rules and recipes stand in the way of professional quality interrogation of the accounting library. Professional judgment is an essential ingredient in the use of the accounting library.

5.5   Summary

The accounting library routinely collects various product cost measures, or statistics. These cost statistics are accounting constructions. Initially, various aggregation choices are made. Costs recognized by the accounting library are aggregated into a variety of cost pools. Some of these pools are treated as period costs and expensed in the period in which they arise.

The remaining pools are assigned to products in some systematic, verifiable fashion. The assignment procedure for each category centers on identification of an approximation to the cost curve, or cost subfunction, associated with the cost pool in question. These cost pool specific cost expressions may use output or synthetic output variables as explanatory variables. In any event, they are almost always linear in appearance, and

presumptively valid over a limited range of activity, hence our acronym LLA.[19] Movement from the indirect product cost pools to the products themselves is accomplished by the uniquely accounting phenomenon of cost allocation.

Taken together, we emphasize the three library design parameters of aggregation, LLAs and cost allocation. The resulting unit costs are, at best, reasonable approximations to marginal costs. This follows from the joint facts that, relative to the economist, the accountant deals with approximations and the total of his product and period costs for the period must sum to the costs recognized in the period.

The accounting library, by nature, catalogues a wealth of information. Unfortunately, information in one circumstance is noise in another. This means we must learn to filter, to extract what is in the library. The professional manager understands the accounting library and how to extract whatever it contains that is useful to the purpose at hand.

5.6   Bibliographic Notes

The history of cost accounting is traced in Solomons [1968] and in Johnson and Kaplan [1987]. The connection between economic and accounting cost was emphasized by Clark [1923], and his famous expression of different costs for different purposes. Demski and Feltham [1976] continue this theme, with emphasis on the idea of approximation. Cost allocation has been explored in a variety of contexts, too numerous to mention. Decision making connections, with an emphasis on marginal cost, are examined in, say, Demski and Feltham [1976] and Zimmerman [1979] as well as in Kaplan [1973] and Baker and Taylor [1979] where a duality connection is emphasized. Demski [1981] and Verrecchia [1982] explore cost allocation criteria, where we attempt to design cost allocation procedures with a less than fully specified context.

5.7   Problems and Exercises

1. The cost construction illustration in Table 5.1 treats interest but not dividends as a cost. Give one set of circumstances in which dividends

---

[19]Terminology enters here. The LLA takes the form $y = a + bx$, and common usage, recall, refers to the intercept as fixed cost and the slope as variable cost per unit of the explanatory variable. We caution that the intercept is simply the intercept of the LLA, nothing is said or implied about what cost might be incurred at $x = 0$. So we insist on calling a the intercept of the LLA. But to exhibit limits to our purity, we continue to term $y = a + bx$ a linear function though it is actually an affine function; a linear function has a zero intercept: $y = bx$.

would not be treated as a component of economic cost and another in which they would.

2. Add another cost pool to the list in Table 5.1: employee bonuses, with a total of $95,000$. This total was paid to employees early in the year, based on their and the firm's performance in the prior year. How should this alter the construction of each product's cost in Table 5.2? Explain.

3. A major university has launched a pollution reduction campaign that, among other things, will tally miles flown by employees on commercial carriers, presumably to "cost" each such mile in terms of carbon emissions. Though the details are vague, the analogy to product costing is apt. Comment on this particular aspect of the campaign.

4. Return to Example 5.2, where we explored treating the third pool as a period or a product cost pool. Can you find an LLA and allocation scheme for the third pool that makes the resulting unit cost for the first product as close as possible to that product's marginal cost? What happens to the ability of the second product's unit cost to approximate that product's marginal cost? Explain.

5. A so-called "step cost" arises when some factor of production is acquired in specific, integer units. To illustrate, it might be possible to lease machine time at the rate of $5,000$ per unit, where units are measured in thousands of hours. So any number of hours of machine time strictly above zero and below $1,000$ will cost $5,000$; any number between $1,000$ and strictly below $2,000$ will cost $10,000$, and so on.

   (a) Plot the implied cost curve.

   (b) In such a situation we often hear someone say "If we expand output, our fixed costs will increase." Carefully analyze this statement, in economic terms and in accounting terms.

6. *product costing*
   Return to the product cost construction illustration in Table 5.1. Numerous assumptions were used in the costing exercise, reflecting period versus product cost distinctions and the LLAs used to allocate product costs among the products. Now find two other sets of assumptions, one set that maximizes the product cost for client B, the municipal client, and another set that minimizes the product cost for the municipal client. Present your calculations in a format comparable to that in Table 5.2. Also, be certain to identify the LLAs and cost allocation in each step of each construction, and provide an adequate defense of your choices.

7. *product costing*
Various nonprofit organizations report the total funds raised, the amount spent on various social services, and the amount spent on administration and fund raising. We might think of such an organization as having n products in any given period; $n-1$ of the products are the various social services provided by the organization and the nth is the internally consumed fund raising and administration product. What pressures might this disaggregate reporting place on the product costing apparatus?

8. *long-run marginal cost*
Suppose Ralph has a single product firm with long-run economic cost given by $C(q; P) = 300q - 20q^2 + q^3$.

   (a) Suppose $q = 10$ units are produced. What total cost will be incurred? What cost per unit will the accounting system report? What is the marginal cost of production at this point?

   (b) Repeat part (a) for the case $q = 7$.

   (c) Repeat part (a) for the case $q = 15$.

   (d) Write a short paragraph explaining your results.

   (e) Write a second short paragraph explaining what would happen here were this a multiproduct firm.

9. *long-run versus short-run economic cost*
Suppose Ralph's long-run economic cost curve is again given by

$$C(q; P) = 300q - 20q^2 + q^3$$

where q denotes output in this conveniently single product firm.

   (a) Tabulate total and marginal cost for $q \in \{0, 1, 2, 3, ..., 20\}$. Also plot marginal cost for $0 \leq q \leq 20$; and notice the efficient point is $q = 10$ units, where average cost for the single product firm is a minimum.

   (b) Next consider a particular short-run cost curve given by

$$C^{SR}(q; P) = F + 290q - 21q^2 + 1.1q^3$$

   Determine F if we are to interpret $C^{SR}(q; P)$ as some short-run cost curve consistent with Ralph's long-run cost curve and an efficient scale of $q = 10$ units, so $C(10; P) = C^{SR}(10; P)$.

   (c) Plot the resulting marginal short-run cost, for $0 \leq q \leq 20$. Contrast this with their long-run counterparts. The best way to do this is to plot all four curves on the same graph.

10. *accounting LLA*

    Return to the above problem dealing with Ralph's cost curve. Now suppose Ralph's accountant approximates $C^{SR}(q; P)$ with an *LLA* of $C^{SR}(q; P) \approx a + bq$. Further suppose the accountant does this by setting the slope of the *LLA* equal to the marginal cost at $q = 10$ and so that the total cost at $q = 10$ equals the approximate cost at that point: $C^{SR}(10; P) = a + b(10)$. Graph the *LLA*. What is its slope? What is its intercept? Over what range does this strike you as a reasonable approximation to the underlying short-run and long-run cost curves? Is the intercept a fixed cost?

11. *accounting LLA*

    Repeat the *LLA* construction in problem 10 above, but with the "anchoring point" at $q = 12$. (So you want the *LLA* to agree with $C^{SR}(12; P)$ and its slope to equal marginal cost at $q = 12$.) What is its slope? What is its intercept? Over what range does this strike you as a reasonable approximation to the underlying short-run and long-run cost curves? Is the intercept a fixed cost?

12. *product costing*

    Ralph's Service provides consulting expertise to not-for-profit entities. Several partners lead various consulting teams that provide the services on a contract basis. Each team consists of the aforementioned partner and a group of professional people drawn from Ralph's stable of professional labor. Ralph employs what we will learn to call a job order costing system to document the cost of each consulting engagement. Each engagement, or client, is costed out based on (i) actual partner time, which averages 120 per hour; (ii) specific identifiable costs (such as for specialized materials); and (iii) allocated professional staff, indirect labor, and miscellaneous supplies. During a recent month the following events occurred:

	client A	client B	client C
partner time (hours)	100	450	250
professional staff time (hours)	1,200	900	800
specific costs (dollars)	18,000	12,000	145,000

    In addition, the following support costs were incurred, each catalogued in a separate cost pool:

professional staff labor cost	55,000
indirect labor cost	45,000
cost of misc. supplies	24,000

    Determine the (unit) cost of each of the engagements.

13. *product costing*
    Ralph's Firm manufactures and sells two products, code named A and
    B. The manufacturing process is relatively simple, with each product
    passing through the same set of machines and using the same labor
    force. For convenience, we might think of this as a manufacturing
    facility with a single department. The accounting system uses three
    types of cost pools: (i) direct labor, where the cost of any labor eas-
    ily identified with a specific product is recorded; (ii) direct material,
    where the cost of any materials readily identified with a specific prod-
    uct is recorded; and (iii) overhead, where all other product costs are
    recorded in a single, aggregate pool.

    The accounting for direct materials uses conventional inventory ac-
    counting procedures. For convenience, we assume the price paid
    suppliers for these items does not vary. In that way, we need not
    worry about LIFO, FIFO, or whatever in the direct material inven-
    tory accounts. (Alternatively, we could assume Ralph's Firm uses
    "just-in-time" inventory procedures with its suppliers.) In a similar
    vein, proper accrual procedures are used in recording the direct labor
    cost. Thus, direct labor cost in a particular period corresponds to
    direct labor input during that period, regardless of any lags in paying
    the employees. Finally, proper accrual procedures are also used in
    determining manufacturing overhead for any particular period. In-
    cluded in this category are such things as insurance, property taxes,
    supervision, indirect manufacturing labor, fringe benefits for labor,
    miscellaneous materials, depreciation, and energy costs – all properly
    concerned with manufacturing operations.

    During a recent period, the following was observed (and recorded):

	product A	product B
units produced	1,200	4,800
direct material cost	6,500	3,500
direct labor cost	3,000	9,000
manufacturing overhead cost	45,000	

    What is the per unit manufacturing cost of A and of B? Determine
    your answer by allocating total manufacturing overhead on the basis
    of (i) units produced; (ii) direct material cost; (iii) direct labor cost;
    and (iv) the total of direct material and direct labor cost.

14. *unit costs*
    Ralph manufactures two products. Total manufacturing cost (TMC)
    is described by an LLA of $TMC = 40,000 + 10q_1 + 5q_2$, where q_i
    denotes units of product i.

(a) Ralph is contemplating two possible production plans. Plan #1 calls for $q_1 = 2,500$ and $q_2 = 2,500$ units, while plan #2 calls for $q_1 = 3,500$ and $q_2 = 1,400$ units. Determine total manufacturing cost for each plan.

(b) Suppose Ralph employs a unit costing procedure in which the "fixed" cost, the intercept in the LLA, is allocated to the products on the basis of total physical units. Determine the unit cost for each product under both plans.

(c) Conversely, suppose Ralph allocates the intercept component on the basis of relative separable cost incurred. Determine the unit cost for each product under both plans.

(d) Repeat the above, for the cases where plan #1 calls for $q_1 = 3,000$ and $q_2 = 1,000$ units, while plan #2 calls for $q_1 = 1,000$ and $q_2 = 3,000$ units.

(e) Carefully explain your unit cost results.[20]

---

[20] The phenomenon illustrated in the example is called Simpson's Reversal Paradox. In general terms, consider conditional probabilities and various events. It is possible to have probabilities $\pi(A|B) > \pi(A|B')$, yet also have $\pi(A|B \text{ and } D) < \pi(A|B' \text{ and } D)$ and $\pi(A|B \text{ and } D') < \pi(A|B' \text{ and } D')$, where the primes denote complements. What happens in probabilistic terms is the conditioning events combine in unintuitive yet logically possible ways. What happens in this unit cost exercise is the production quantities combine in unintuitive yet logically possible ways as they are passed through the unit cost calculation. It is the lack of meaningfulness of average cost in a multiproduct setting that allows us to construct this example. Sunder [1983] is an important reference.

6
The Impressionism School

We now take an extended look at settings where the firm produces a variety of heterogeneous products. Consider the private sector. A tool and die manufacturing firm, a consulting firm, and an internet merchandiser of computer equipment illustrate the genre. Alternatively, consider the not-for-profit sector. Research at a private university, CPR courses offered by the Red Cross, and religious material merchandising by a church illustrate the genre. Finally, consider the public sector. Operation of a regional exhibition hall, law enforcement, and sale of surplus materials illustrate the genre.

This list is not random. We have covered private, not-for-profit, and public sectors. In each sector we have illustrated manufacturing, service, and merchandising operations.

That said, you should have noticed the Chapter's title. Historically, product costing has relied on considerable aggregation in the cost pools and a relatively modest variety of synthetic variables in their respective LLAs. As technology continued to change, the costing art lagged behind, and arguably, though it turns out not necessarily, began providing less and less accurate estimates of marginal cost. Regardless, we refer to the art form in this area as the impressionism school, as the reliance on extensive aggregation leads to an emphasis on "immediate aspects of objects or actions without attention to details." Yet the approach remains important in historical and contemporary terms, and for that matter alone warrants our attention. Moreover, this allows us to introduce additional nuances in product costing art, and to deepen our understanding of the aggregation-LLA-allocation theme.

We begin with some additional terminology. From there we specify a set of cost pools and *LLA*s, and then explore a variety of methods for allocating the indirect product cost pools.

6.1   More Terminology

As we know, the accountant records the (accounting) cost of factors consumed during a period in cost pools and, as we also know, categorizes these cost pools into product and period categories.[1] That said, it is important to remember period costs, by definition, are not assigned to products. So when we ask the accounting library what a product cost, its answer will exclude any reference to period costs. This is one of the reasons we use the phrase *unit cost* to label whatever it is the accounting library reports as a product's cost.

This is not capricious behavior on the accountant's part. Rather, it reflects a concern for inventory valuation. Consider a manufacturing firm. The cost of factor consumptions to complete the manufacturing process is what defines the product cost pools. Any completed but unsold units would be held in finished goods inventory, and valued, for accounting purposes, at their unit costs. In turn, factor consumptions associated with, say, shipping and handling are incurred in the delivery process and, traditionally, not assigned to products themselves. Thus, in the usual GAAP-style income display, cost of goods sold will report the unit costs of items sold,[2] with period costs expensed elsewhere in the income display. *Gross margin* is the term reserved for revenue less cost of goods sold.

The product cost pools also lead us to some additional terminology. A direct product cost pool that records consumption of labor is, not surprisingly, called *direct labor*, just as a direct product cost pool that records consumption of material is called *direct material*. Any indirect product cost pool is often called *overhead*. *Prime cost* is the total of direct labor cost and direct material cost. *Conversion cost* is the total of direct labor and overhead cost.

As we have stressed, pragmatic considerations are present here. Some detail is purposely clouded; some distinctions are purposely not made. For example, a large firm will employ many people just to oversee and manage purchases of materials. Where will we find the cost of these "overhead" items? They may be in a separate overhead cost category, they may be

---

[1] Recall the definitions in Chapter 5.

[2] More precisely, cost of goods sold is the total of the unit costs of products whose revenue is recognized during the period. You will learn, shortly, that there is an additional nuance to this statement, one that is the source of considerable and unacceptable confusion.

in a separate period cost category, or they may be commingled with other items.

In this respect, it is useful to dwell on the fact direct labor is the cost of labor factors that the firm is able to and finds useful to associate with products, just as direct material is the cost of material factors that the firm is able to and finds useful to associate with products. (And remember that directly identified does not necessarily imply separable, as we saw in Figures 3.3 and 3.4.) All other labor factor, and all other material factor, consumptions are recorded in the indirect product or period pools

6.2   Data for an Extended Illustration

We now focus on a specific firm or organization, Ralph's Venture for short, and examine its product costing practices. To keep clutter to a minimum, two products are present, which we will imaginatively call job 1 and job 2. Think of this as two custom orders, one unit each. These were the only products present during the accounting period in question. In addition, the cost identified during this period totals 314 (000). What is the cost of each product?

Specific cost data are displayed in Table 6.1. Labor, for example, cost a total of $12 + 20 + 10 + 36 + 42 = 120$. (Notice we scale the story in terms of thousands (000).) Of this amount, only 12 and 20 are in direct cost pools, while 10 and 36 are in two distinct overhead pools, and the remainder is in a period cost pool. Aggregation is the order of the day.

TABLE 6.1: Cost Construction Data for Ralph's Venture

description	job 1	job 2	overhead A (OV_A)	overhead B (OV_B)	period	total
labor	12	20	10	36	42	120
materials	30	20	5	10	10	75
energy/space			25		10	35
depreciation			16	14	10	40
fringe			22			22
misc.			10		12	22
total	42	40	88	60	84	314

This leads to the 7 cost pools displayed in Table 6.2: direct labor for each job, direct material for each job, two overhead pools, and one period pool. The underlying data in Table 6.1 should provide a hint of the aggregation. Labor is present in every pool, as are various materials. Fringe is aggregated in overhead A, and so on.

TABLE 6.2: Ralph's Cost Pools

pool	designation	total
1	direct labor (job 1)	12
2	direct labor (job 2)	20
3	direct material (job 1)	30
4	direct material (job 2)	20
5	overhead A (OV_A)	88
6	overhead B (OV_B)	60
7	period	84

Next we specify the associated LLAs. These are obvious for the direct labor and direct material pools. For the first overhead pool, the total of direct labor cost, denoted $DL\$$, is used as a synthetic variable. We have the following specification[3]

$$OV_A = a + b \cdot DL\$ = 40 + 2 \cdot DL\$ \qquad (6.1)$$

That is, the OV_A pool uses $x = DL\$$ as a synthetic variable, and has an intercept of 40(000) and a slope of 2. The second overhead pool uses a synthetic variable of total direct material cost, denoted $DM\$$

$$OV_B = a + b \cdot DM\$ = 25 + 0.5 \cdot DM\$ \qquad (6.2)$$

We also might, and eventually will, specify an LLA for the period cost category. For now, however, we concentrate on the product cost pools.

With the cost pools and product cost pool LLAs in hand, then, we have only to specify the cost allocation to complete the costing exercise. Below we work through a number of approaches to the allocation leg of the costing recipe.

6.3   Assignment of Actual Overhead Totals

One possibility here, following our work in Chapter 5, is to use the synthetic variables in the respective LLAs to allocate the actual overhead totals. The finer details of the noted LLAs are ignored, as we allocate actual overhead on the basis of the synthetic variable's actual total. This leads to the construction displayed in Table 6.3.

---

[3] To avoid clutter, we take the synthetic variable, slope and intercept as givens at this point. Informed subjective assessment and statistical analysis are important guides in these choices.

TABLE 6.3: Product Costs based on Actual Overhead Totals

pool	job 1	job 2	total
direct labor	12	20	32
direct material	30	20	50
overhead A	$(88/32)(12) = 33$	$(88/32)(20) = 55$	88
overhead B	$(60/50)(30) = \underline{36}$	$(60/50)(20) = \underline{24}$	$\underline{60}$
unit cost	111	119	230

You should notice two features here. First, just as in the case of Ralph, Ltd. in Chapter 5, we allocate each indirect product cost pool, each over-head pool, on the basis of the specified synthetic variable. Consider the overhead A pool, where the synthetic variable is total direct labor hours, as given in (6.1). The two direct labor pools display a total direct labor cost of $12 + 20 = 32$. In turn, the cost in the overhead A pool totals 88, and we thus have an allocation rate of [4]

$$f_A = \frac{88}{32} \text{ dollars of overhead A per dollar of direct labor}$$

Job 1 incurred direct labor cost of 12, so it is allocated overhead A in the amount $f_A(12) = \frac{88}{32}(12) = 33$. Likewise, job 2 incurred direct labor cost of 20, leading to allocated overhead A in the amount $f_A(20) = \frac{88}{32}(20) = 55$. In parallel fashion, overhead B uses a synthetic variable of total direct material cost, as given in (6.2). With direct material cost totaling $30 + 20 = 50$ and overhead B totaling 60, we have an allocation rate of

$$f_B = \frac{60}{50} \text{ dollars of overhead B per dollar of direct material}$$

And this implies respective allocations of $\frac{60}{50}(30) = 36$ and $\frac{60}{50}(20) = 24$.

Second, also notice we have accounted for the total in each product cost pool in the construction. All of the direct labor and direct material costs are referenced in the construction; total allocated overhead A equals the total in the overhead A cost pool; and total allocated overhead B equals the total in the overhead B cost pool. As we have stressed, cost pools are temporary accounts in the accounting library, and they must be closed out at the end of the period. Here, the sum of product cost pools winds up as the sum of the product costs, just as the total of the period cost pools winds up as the period cost for the period.[5]

---

[4] Many call such an allocation rate a burden rate. We don't.

[5] In our continuing effort to stress fundamentals, we purposely sidestep the debit and credit depiction of these procedures. Technically, in the case of Table 6.3, the costs in each product cost pool are "moved" to a work in process account, and from there,

It will also be useful in what follows to extend the cost construction to reporting the firm's income. For this purpose suppose the first product, job 1, is completed and delivered to the customer, who pays the agreed upon price of 700. The resulting income calculation and format are displayed in Table 6.4.[6] Notice the cost of goods sold total is simply the cost of the product sold, the product whose revenue has been recognized. Similarly, the cost of the other product, 119 in this case, provides the end of period inventory valuation for that product.

TABLE 6.4: Ralph's Income (via Table 6.3)

revenue	700
cost of goods sold	111
gross margin	589
period costs	84
net income	505

6.4   Assignment of Estimated Overhead Totals

The allocation procedure in Table 6.3 assigns all of the indirect product cost totals, all of the overhead, to the products. This is possible because we wait until those overhead totals are known before performing the allocations. But this creates a troublesome delay, especially if we are dealing with a large number of products, as the overhead totals will not be completely identified until the end of the period. Imagine asking the accountant what job 1 cost and being told "I cannot tell you until the end of the period."

Once we admit the overhead assignments are approximations, we are led to a simple and commonplace modification of the procedure. We use an estimated (instead of the actual) overhead allocation rate to make the overhead assignments. This allows us to assign overhead to products as they are finished and avoids the troublesome delay. Pragmatism has many influences.

6.4.1   Normal, Full Costing

The procedure is straightforward, at least on the surface. Glance back at (6.1) where we displayed the presumed LLA for the overhead A pool.

---

presumably, to a finished goods account. You might want to try your hand at the procedures. But the important point is the very nature of the constructive procedure.

[6]The second product, job 2, is not yet complete and thus does not enter the current period's income calculation. Of course, were this some type of long term construction story we would be dealing with job 2 as well.

FIGURE 6.1. Original vs. Allocated LLA for Overhead A

We plot this LLA in Figure 6.1: $OV_A = 40 + 2x$ with synthetic variable $x = DL\$$. Further suppose we estimate the synthetic variable will total $DL\$ = 40(000)$. This estimate is called a *normal volume*, so we denote it $x_N = 40(000).$[7] It implies an estimated total cost in the overhead A pool of $OV_A = a + b \cdot x_N = a + b(40) = 40 + 2(40) = 120$. And this implies an allocation rate of $f_A = 120/40 = 3$ dollars of overhead A per dollar of direct labor. So instead of allocating overhead A at a rate of $\frac{88}{32}$ as we did in Table 6.3, we allocate it at a rate of 3 dollars of overhead A per dollar of direct labor. Note well: we are merely replicating the allocation method in Table 6.3, but on the assumption the total cost in the pool is 120 and direct labor dollars total 40. Further note that the allocation de facto creates an LLA given by $OV_A = 3 \cdot DL\$$, as opposed to the originally presumed $OV_A = 40 + 2 \cdot DL\$$ in expression (6.1). See Figure 6.1.
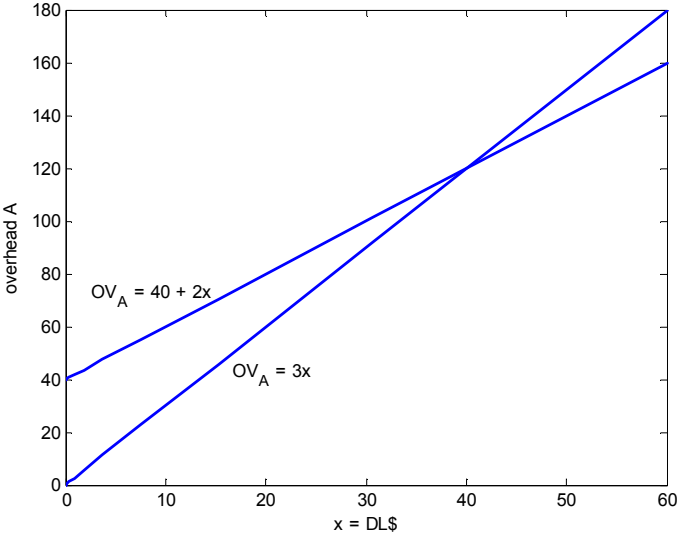
To round out the picture, suppose we also assume a normal volume of $x_N = 50(000)$ direct material dollars for the overhead B pool. This implies an overhead B total of $OV_B = a + b \cdot x_N = a + b(50) = 25 + 0.5(50) = 50$ and an allocation rate of $f_B = 50/50 = 1$ dollar of overhead B per dollar

---

[7] Various approaches to defining this so-called normal volume can be found: what we expect during the period, what we expect to average over the next several periods, or what our capacity will allow for example. The important point is the implied LLA that enters the product cost calculation.

of direct material. And from here a companion to Figure 6.1 based on the overhead B pool is readily envisioned.

This approach to allocation is called a normal, full (or absorption) cost system. It is called "full" costing since the goal is to fully assign the overhead to products, and it is called "normal" because the allocation rate is an estimated rate, estimated on the basis of a conjectured or otherwise specified normal volume. The resulting cost constructions are displayed in Table 6.5.

TABLE 6.5: Product Costs based on Normal, Full Costing

pool	job 1	job 2	total
direct labor	12	20	32
direct material	30	20	50
overhead A	$(3)(12) = 36$	$(3)(20) = 60$	96
overhead B	$(1)(30) = 30$	$(1)(20) = 20$	50
unit cost	108	120	228

Notice, however, that we have some residual amounts to deal with, as the overhead assigned to the products does not equal the overhead totals themselves. Using data in Tables 6.2 and 6.5 we have the reconciliation displayed in Table 6.6. Overhead A totals 88 but our procedure has allocated a total of $36 + 60 = 96$ to the products. Similarly, overhead B totals 60 but we have allocated a total of $30 + 20 = 50$ to products. Many use the terminology that overhead A is over-absorbed (by $96 - 88 = 8$) and overhead B is under-absorbed (by $60 - 50 = 10$) here. This is a bit too quixotic for us so we simply label this "over" or "under" amount the *plug* to the overhead pool. In this way, the total amount in the overhead pool equals the total of the allocations from the pool, including the plug. So the plug in this case equals cost in the pool less allocations from the pool to the two products. See Table 6.6.[8]

TABLE 6.6: Overhead Allocations under Normal, Full Costing

pool	allocated to product		allocated to cgs	total
	job 1	job 2	(plug)	
overhead A	36	60	$(8) = 88 - 36 - 60$	88
overhead B	30	20	$10 = 60 - 30 - 20$	60

[8]The plug is easy to identify if you display the cost pool total and allocations in a "T" account.

These errors (yes, these plugs) are inevitable. Surely our estimate of normal volume is in error, just as surely as our original LLA is in error.[9] But what to do with these errors? The answer is simple, include them in (yes, allocate them to) the period's cost of goods sold total.[10]   Enough said!

All of this leads to the income display in Table 6.7. Notice the cost of goods sold is calculated as the product cost of job 1 (Table 6.5) plus the sum of the overhead allocation errors or plugs (Table 6.6). It is imperative you remember this nuance of the plug being included in cost of goods sold and therefore in the period's income.

TABLE 6.7: Ralph's Income (Normal, Full Costing)

revenue	700
cost of goods sold $(108 + (8) + 10)$	110
gross margin	590
period costs	84
net income	506

You should also notice the connection between income and inventory effects when we explore different costing recipes. Our switch from allocating overhead based on actual overhead to a normal, full costing approach increased Ralph's income from 505 (Table 6.4) to 506 (Table 6.7). Now remember only job 1 has been sold. The inventory value of job 2 is 119 when allocating based on actual overhead (Table 6.3) versus 120 when using a normal, full costing approach (Table 6.5). The new, improved income number is 1 dollar higher because its associated ending inventory is 1 dollar higher.

In sum, we have the following formalism.[11]

Definition 16 *A normal, full cost costing system allocates each indirect product cost pool with LLA given by cost $= a + bx$ using a rate of $f = \frac{a}{x_N} + b$ per unit of synthetic variable x, where x_N is the normal volume for the synthetic variable. Any difference between actual and thus allocated cost is allocated to cost of goods sold.*

---

[9] Indeed, we should express the LLA as $y = a + bx + \epsilon$, where ϵ is the proverbial error term.

[10] The only concern with this answer is when the errors are "large" and not correcting them significantly affects the firm's ending inventory balance. In such a case we would redo the allocations at the end of the accounting period, in effect converting to allocations based on the actual overhead totals, as in Table 6.3. This is an unlikely occurrence.

[11] $f = \frac{a + b \cdot x_N}{x_N} = \frac{a}{x_N} + b.$

Example 6.1 To dig a bit deeper, suppose a special customer approaches Ralph about producing a third product. The offered price is 23. Ralph estimates direct labor will cost 5 and direct material will cost 2. Using the full cost allocation rates in Table 6.5 to estimate the overhead costs, we find a unit cost of 24, constructed as follows.

direct labor	5
direct material	2
overhead A	$3(5) = 15$
overhead B	$1(2) = \underline{2}$
unit cost	24

With an estimated cost of 24 and an offered price of 23, Ralph is not tempted.

But before declaring victory, ask yourself what will happen to Ralph's income if this offer is indeed accepted? Revenue, of course, will increase by 23, but what about cost of goods sold? Presumably direct labor will increase by 5 and direct material by 2. The overhead *LLA*s, in turn, suggest that overhead A increases at the rate of 2 per dollar of direct labor and that overhead B increases at the rate of 0.5 per dollar of direct material. (Recall (6.1) and (6.2).) Glancing back at the cost pool totals in Table 6.2, this suggests accepting this product would increase overhead A from 88 to $88 + 2(5) = 98$ and would increase overhead B from 60 to $60 + 0.5(2) = 61.$[12] With this insight, overhead A increases by 10, but we allocate 15 to the new product, so the overhead A plug (Table 6.6) changes from (8) to $(8) + 10 - 15 = (8) - 5 = (13)$. Similarly, overhead B increases by 1 but we allocate 2 to the new product, so the overhead B plug changes from 10 to $10 + 1 - 2 = 10 - 1 = 9$.

This implies accepting the offer increases Ralph's (normal, full costing) cost of goods sold from 110 (Table 6.7) to $110 + 24 - 5 - 1 = 128.$[13] This increase of $128 - 110 = 18$ is less than the proffered price of 23. Importantly, accepting the offer would imply a unit cost of 24, but would simultaneously lower the plug to cost of goods sold.

Think a little more about the allocation procedure for the overhead A pool. If you do not include the change in the plug, you are using the steep *LLA* in Figure 6.1; if you include the plug in your calculation, you are removing the intercept bias in the full cost analysis, and thereby using the original, less steep *LLA* in Figure 6.1. A parallel comment applies to the overhead B pool.

---

[12] Notice that for estimation purposes we assume the *LLA*s are reasonably accurate. So incremental overhead A is estimated to be $2(5)$.

[13] Notice that cost of goods sold would now be the sum of the two unit costs and the two plugs, or $108 + 24 + (13) + 9 = 128$.

As we said, it is imperative you remember the nuance of the plug to cost of goods sold. The overall cost effect of accepting such an offer will be found in the new product's unit cost *and* in the overhead plugs. The unit costs and plugs are generally not separable in a normal, full cost system.[14]

6.4.2   Normal, Variable Cost

Glance back (again) at Figure 6.1. The fundamental idea in full costing is to assign "all" the overhead to the products, de facto re-engineering the cost pool's presumed LLA as depicted in the Figure. As the LLA itself is an approximation, this re-engineering may, or may not, be helpful. For example, suppose the cost function is separable, in terms of products and in terms of the components recorded in separate cost pools. This suggests, in terms of this indirect product cost pool with LLA given by cost $= a + bx$, that the marginal effect of increasing the synthetic variable would be the slope b, not the slope of $\frac{a}{x_N} + b$ in the normal, full cost approach. Of course we must worry about all of this presumed separability, and so it is not at all obvious we have a strong argument. But keying on the slope in the original LLA is the central feature of what is called variable costing.

The idea is to allocate overhead to products using the slopes of the identified LLAs and to treat the intercepts of the overhead LLAs as period rather than product costs. After all, if the firm's cost curve really were linear, and therefore expressible in the single product case as $C^{SR}(q; P) = a + bq$, intercept a would be the (short-run) fixed cost and slope b would be the (short-run) marginal cost as well as the variable cost per unit. Arguing by analogy, then, in a variable costing system we assign overhead using the slopes of the identified LLAs as the allocation rate. The remaining overhead is not assigned to products. It is allocated to a period cost category.

As noted, product costing procedures of this sort are called variable costing systems. We append the adjective normal to remind ourselves we are using estimated allocation rates, based on the slopes of the respective LLAs. This explains the label *normal, variable cost.*

In comparison with a normal, full cost regime, then, we begin with the same cost pools and distinction between period and product costs.[15] Beyond that point, however, three differences surface. One difference is the manner in which the product costs, the unit costs, are calculated. In a vari-

---

[14]To be sure, our little demonstration presumes the overhead LLAs are reasonably accurate. Regardless, the unit cost and plug calculations are not generally separable in a full cost system.

[15]Notice the sentence refers to the same cost pools, but in comparison with a full cost regime we would also have the same totals in the pools only if the firm's behavior were unaffected by how it measured its product costs, an unlikely story for sure.

able cost regime the intercepts of the product cost pool LLAs are allocated to period cost categories.[16]

To see this in our running example, the direct labor and direct material tallies remain as before (in Table 6.5). Turning to the overhead allocations, we exploit the slopes in the assumed overheads LLAs. Glancing back at expressions (6.1) and (6.2), the overhead A allocation rate is now $f_A = 2$ per direct labor dollar while that for Overhead B is $f_B = 0.5$ per direct material dollar. Using these rates we now simply repeat the constructive procedure in Table 6.5 to derive the unit costs. Details are displayed in Table 6.8. Notice the respective unit costs have declined from 108 and 120 to 81 and 90. This reflects the systematic shift of some of the overhead pool costs into the period cost category.

TABLE 6.8: Product Costs based on Normal, Variable Costing

pool	job 1	job 2	total
direct labor	12	20	32
direct material	30	20	50
overhead A	$2(12) = 24$	$2(20) = 40$	64
overhead B	$0.5(30) = 15$	$0.5(20) = 10$	25
unit cost	81	90	171

The second difference between the normal full and variable cost systems is the residual errors, the plugs, in the overhead pools. Again glancing back at the LLA expressions (6.1) and (6.2), we know the respective intercepts of 40 and 25 are allocated to the period cost categories. And in Table 6.8 we have identified the amounts allocated to products. This provides, in Table 6.9, the parallel to the "plug" calculations in Table 6.6. The errors, which will again be allocated to cost of goods sold, equal their full cost counterparts only when normal and actual volume of the synthetic variable are identical (as is the case for overhead B here).[17]

---

[16] This idea, in principle, extends to the direct cost pools as well. It is conceivable the LLAs for direct costs might have nonzero intercepts. If so, we would assign the direct costs to products using the slopes of the respective LLAs. The intercept amounts would be expensed. For example, it may be impossible or impractical to alter the labor supply in the short-run. Simply because the firm finds it possible and convenient to identify particular labor costs with specific products does not imply these direct costs are variable costs. In effect, this would entail reclassifying the direct cost pool as an indirect product cost pool. That said, however, as a practical matter we generally treat the direct cost category LLAs as having zero intercepts.

[17] Here we have respective plugs of $(16) = 88 - 24 - 40 - 40$ and $10 = 60 - 15 - 10 - 25$. Of course, a less pure approach would simply allocate the intercept and the plug to the period cost category, de facto eliminating their calculation and allocation to cost of goods sold.

TABLE 6.9: Overhead Allocations under Normal, Variable Costing

pool	allocated to product job 1	allocated to product job 2	allocated to period	allocated to cgs (plug)	total
overhead A	24	40	40	(16)	88
overhead B	15	10	25	10	60

This leads to the third difference between the normal, full and variable cost systems. In the variable cost case we often, especially for internal reporting purposes, highlight the "variable" portion of the period costs. This is done by focusing on each product's *contribution margin*, defined as price less variable unit cost less variable period cost. To illustrate, suppose the period cost pool in our story contains a variety of items including shipping, that shipping the first product cost 24, and that this shipping cost is the only variable portion of the period cost pool. We now have Ralph's income calculated and displayed in normal, variable costing format.

TABLE 6.10: Ralph's Income (Normal, Variable Costing)

revenue	700
variable cost of goods sold $(81 + (16) + 10)$	75
variable period cost	24
contribution margin	601
period costs $(84 - 24)$	60
allocated overhead intercepts $(40 + 25)$	65
net income	476

If you are still awake, you will notice the income is lower in the variable cost regime. Recall our earlier point that the difference in income between any two regimes will equal the difference in change in ending inventory between the two regimes. Job 2 remains unsold, and thus in inventory. No other product inventory is in place. Hence, relative to the normal, full cost regime, its inventory valuation has dropped from 120 (Table 6.5) to 90 (Table 6.8), a decline of 30. And relative to the normal, full cost regime, income has fallen from 506 (Table 6.7) to 476 (Table 6.10), a decline of 30.

We formalize normal, variable costing as follows.

Definition 17 *A normal, variable cost costing system allocates each indirect product cost pool with LLA given by cost $= a + bx$ using a rate of $f = b$ per unit of synthetic variable x. Intercept a is allocated to a period cost*

pool. Any difference between actual and thus allocated cost is allocated to cost of goods sold.

Example 6.2 Return to Example 6.1, but now assume normal, variable costing is in place. Recall the offered price for the special product is 23, and Ralph estimates direct labor will cost 5 and direct material will cost 1. Using the variable cost allocation rates in Table 6.8, this suggests a unit cost of 18:

direct labor		5
direct material		2
overhead A	$2(5) =$	10
overhead B	$0.5(2) =$	<u>1</u>
unit cost		18

This implies income would increase by $23 - 18 = 5$, precisely the conclusion we reach in Example 6.1 when we are careful to pick up the change in the overhead plugs. Doing so, as we stressed, removes the allocation of a portion of the LLA intercepts, and results in an analysis identical to that under normal, variable costing.[18]

6.4.3 Remarks

Several remarks round out this comparison of full and variable cost systems. First, the economic interpretation of the two approaches is subtle and ambiguous. Presuming the intercepts in the overhead LLAs are positive, unit costs under a full cost system will exceed their counterparts under a variable cost system. But which is a better estimate of marginal cost? As developed in Chapters 4 and 5, there is no general answer.

Second, and in a related vein, many contend unit costs based on a full cost system are better estimates of long-run marginal cost, while those based on a variable cost system are better estimates of short-run marginal cost. This reflects an unstated assumption (beyond that of the usual separability concern) that the LLAs are accurate, even to the point of their intercepts well-measuring the appropriate short-run fixed costs. As we have stressed, this is not likely to be the case, and we are back to the prior paragraph: there is no general answer.

Third, institutional matters come into play at this point as well. GAAP requires full costing, as does the U.S. federal tax system. But this is only one set of demands placed on the accounting library (and it is easy, in

---

[18] Here there is no change in the overhead plugs because we assume the incremental direct labor and direct material do not affect the intercepts of any of the LLAs.

statistical terms, to convert ending inventory balances – and thus income – from variable to full cost format). So it is not surprising we find significant variety in the library approaches firms choose, even within the same industry. Moreover, one need not commit to a single approach. A common technique is to separate the overhead allocation rates into "fixed" and "variable" components. This is code for identifying separately the intercept and slope components of the unit cost assignments. To illustrate, let's combine the calculations in Tables 6.6 and 6.8. We use normal costing, but separately identify what the overhead allocations would have been under variable costing. See Table 6.11.

TABLE 6.11: Product Costs based on Normal Costing

pool	job 1	job 2
direct labor	12	20
direct material	30	20
allocated slope components		
overhead A	$2(12) = 24$	$2(20) = 40$
overhead B	$0.5(30) = 15$	$0.5(20) = 10$
variable unit cost	81	90
allocated intercept components		
overhead A	$1(12) = 12$	$1(20) = 20$
overhead B	$0.5(30) = 15$	$0.5(20) = 10$
full unit cost	108	120

6.5 Standard Cost Systems

Another approach to moving from cost pools to unit costs is to use estimated amounts for each product cost pool; and this can be done in either full or variable costing format.[19] These are called standard cost systems. For example, Intel's annual report states "Inventory is computed on a currently adjusted standard basis (which approximates actual cost on an average or first-in, first-out basis)." What Intel is telling us is some of their inventory balances are on an average cost model while others are on a FIFO model, but in both cases the balances are estimated from the underlying standard costs.

---

[19] If you are keeping track, we now have full or variable formats, in normal mode or in standard cost mode. We also have our initial variant of full cost in actual cost mode. We draw the line, however, and eschew actual cost in variable format. The reason is (beyond mere exhaustion) that this would entail separate cost pools for the intercept components of each overhead category, a presumption that runs counter to the *LLA* philosophy in the first place.

The procedure itself is a simple extension of normal costing. As usual, period and product costs are aggregated into various pools. For each cost pool, we record actual cost incurred. We then construct the unit costs using estimated quantities and prices for each factor of production.[20] Most surely, every product cost pool will have an actual amount that differs from its estimated counterpart. With a little luck, these errors will sum to a small amount. We then expense the errors, just as we expensed the errors, the plugs, in the overhead pools in a normal costing system.

Why bother? There are several reasons. First, with many cost pools and products, substantial bookkeeping economy is available with a standard costing system. LIFO is easier to implement with such a system. Transfers of partially completed products from one location to another (e.g., from manufacturing to regional warehouses) are also easier to record with standard costs.

Second, the juxtaposition of actual and standard costs is often a useful exercise. This allows the manager routinely to compare actual with estimated results. Large deviations are signals that the actual or estimated costs have been compromised. Moving this juxtaposition of actual and standard into the accounting library makes these comparisons more routine. It places them within the firm's formal reporting process.

Finally, we often evaluate a manager's performance using, among other things, a comparison of results achieved with resources consumed. Resources consumed are usually measured by the cost of resources consumed. Many resources are supplied by other managerial units within the firm. For example, maintenance may be done by a maintenance group. Subcomponents may be manufactured in a separate facility. Security may be provided by a security group. One division's students may take courses in another division of the university.

In each instance an important question arises. Do we want to cost these imported services at actual or at estimated amounts per unit? The answer is subtle and varied. For example, costing imported services at their actual cost imposes supplier inefficiencies on the importing manager's evaluation. Conversely, costing them at their estimated cost shields the importing manager's evaluation from factor price changes in the supplier department. In addition, a single firm may want to treat different managerial units differently on this score. This means we want an accounting system with the flexibility to cost these imported services as the situation demands. Standard costing is the answer.

---

[20] Naturally we work with aggregates here. We don't, literally, estimate prices and quantities for every single factor.

6.6   Summary

We have stressed the importance of viewing the accounting library as a collection of accounting constructions. The accountant's product costing art uses various aggregations and LLAs coupled with allocation procedures to construct product cost statistics. Product costing in a setting of heterogeneous products is quintessential costing art. Indeed, every product cost construction is a variation on the theme developed in this chapter.

We have also emphasized the importance of choice in constructing and thus in understanding the accounting library. Initially we must select the level of aggregation, which determines the structure of the cost pools themselves. Then we must identify those pools in terms of direct product, indirect product and period in nature. Special terminology enters at this point: direct labor pools reflect labor costs that the firm can and wants to identify directly with products; direct material pools reflect material costs that the firm can and wants to identify directly with products; overhead is the usual name for the remaining (and thus indirect) product cost pools.

Once the cost pools are identified and categorized, we turn to the question of selecting LLAs that well describe each category's costs. By definition, the explanatory variable for a direct product cost pool is units of the product in question. Overhead is ambiguous, and there we rely on synthetic variables.[21] This is where allocation enters, driven by pragmatic considerations. Cost allocations in an actual costing system must wait until the end of the accounting period when the cost totals are known. Under normal costing, we assign the overhead using estimated allocation rates. Any error, our plug, is simply allocated to cost of goods sold at the end of the accounting period.

Of course, if we have bothered to construct an estimated allocation rate for some overhead pool, we have given considerable thought to the nature of costs in that pool. This raises the specter of variable costing. If we have a reasonable estimate of the slope and intercept of the underlying LLA, we can treat the intercept as a period rather than a product cost. This is the essence of variable costing. Likewise, if we are reasonably good at these estimation exercises, we may want to implement a standard instead of a normal costing system.

Either way, the economic interpretation of full and variable cost statistics is far from straightforward. If the LLAs have nonnegative intercepts, we know the full costing unit cost is larger than the variable costing unit cost. Which statistic, which unit cost, is a better estimator of marginal cost, or a better estimator of performance? This depends on our purpose and on our circumstance. All we can say of a general nature is that a costing

---

[21] As we hinted, we also often deal with LLAs for period cost pools.

system that catalogues both statistics (as in Table 6.11) provides more information.

The contrast between this costing exercise and that of the economist should be pondered, again. The accountant's art has the flavor of a practical, first cut at estimating marginal cost. Subtle aspects of the underlying technology and economic forces are unlikely to be given much force in this highly aggregate approach. This is why we call this the impressionism school.[22]

6.7   Bibliographic Notes

Where to draw the line between product and period costs remains a controversial issue. Expensing of R&D under GAAP is illustrative. Choice between full and variable costing is part of this larger issue. GAAP stresses the importance of costs that are "clearly related" to production in identifying product costs; internally, of course, the firm faces no such constraint in designing its library. A large literature debates and analyzes this issue. Green [1960], Sorter and Horngren [1962], and Fremgren [1964] provide an excellent introduction to this literature. Miller and Buckman [1987] offer a dynamic perspective. Ties to pricing can be found in, say, Balakrishnan and Sivaramakrishnan [2002] and Banker and Hansen [2002]. Standard costing is also a long-standing subject. Indeed, Solomons [1968] provides links to the 19th century; and it certainly has a close association with the "scientific management" school (e.g., F. W. Taylor). Issues of attainability or tightness of the standards arise, as does the question of participation in setting the standards. Becker and Green [1962] provide a good entry to these themes.

6.8   Problems and Exercises

1. The accounting library uses aggregation, *LLA*s and cost allocation in assembling and presenting cost information. Carefully discuss the connection among these building blocks and the product cost terminology of direct labor, direct material, and overhead.

---

[22] The limit points of the impressionism school should also be mentioned. One is where we have joint products such as petroleum refining. Here, there are limits to the mix of products, and the impressionism school is hard pressed to deliver a useful estimator of marginal cost. The other limit point is where we have roughly continuous production of the same product, such as a brewery. This is called a process cost system, though it boils down (no pun) to a one product story intermixed with stages of completion. We defer exploration of the joint product and process cost settings to other sources.

2. Suppose we have a single product firm. The firm uses normal, full costing with a normal volume equal to its efficient scale or output level (where average economic cost is a minimum). The LLA is constructed by setting the slope equal to marginal cost at the efficient output level and the intercept so the two total cost expressions agree at that point. Why is there no difference between full and variable costing in this instance?

3. Assume in our extended illustration (Ralph's Venture) that both products are sold, with the selling price for the second being 250. Redo Tables 6.4, 6.7 and 6.10. Explain the pattern that emerges.

4. Suppose a firm tends to hold its production constant, and uses inventory to buffer the effects of random demand. (So it depletes inventory in good times and builds inventory in bad times.) Will the statistical variance of its net income under full costing be higher or lower than the statistical variance of its net income under variable costing? Explain.

5. *overhead pools*
Return again to the extended illustration (Ralph's Venture), but now assume all overhead is aggregated into a single pool with LLA estimated by $OV = 60 + 3 \cdot DL\$$. Redo Table 6.11, and comment on your observations. Continue to assume a normal volume of $DL\$ = 40(000)$.

6. *unit costs*
Ralph is dealing with two products, with respective quantities denoted q_1 and q_2. The cost structure is described by the following LLAs:

direct labor:	$DL = 400q_1 + 700q_2$
direct material:	$DM = 200q_1 + 800q_2$
overhead:	$OV = 50,000 + 1 \cdot DM$
selling and administrative:	$S\&A = 40q_1 + 120q_2.$

Determine the unit cost of each product if (1) normal, variable costing is used and if (2) normal, full costing is used. (For full costing, assume a normal volume of $DM_N = 50,000$ direct material dollars.)

7. *standard costs*
Return yet again to the extended illustration, but now assume a standard cost system is in use. The LLAs for the two overhead pools remain as originally specified. The standard (the estimated) direct labor and direct material costs are as follows:

	job 1	job 2
direct labor	10	23
direct material	34	18

Determine Ralph's income assuming standard, full costing is used and assuming standard, variable costing is used. That is, replicate Tables 6.7 and 6.10 for the standard costing cases.

8. *actual, full costing*
Return to problem 3−14 where Ralph manages a two product firm. Demand now limits him to producing $q_1 = 100$ units of the first product and $q_2 = 150$ units of the second. Labor used exclusively for the first product cost a total of 7, 500 and labor used exclusively for the second product cost a total of 19, 687.50. In addition, the machine used for both products cost 20, 000. This gives us three cost pools, direct labor for each product and an overhead pool. You should verify that these factor expenditures are consistent with the technology and factor prices specified in the original problem. Having done that, now determine the unit cost of each product. Do this by allocating overhead equally between the products, by allocating overhead on the basis of physical units of output, and by allocating overhead on the basis of direct labor dollars. Contrast your unit cost constructions with marginal cost, presuming of course $q_1 = 100$ and $q_2 = 150$.

9. *allocation of overhead plug to cost of goods sold*
Simple Manufacturing Company manufactures and distributes a single product. It records manufacturing costs using a normal, full costing system with overhead allocated on the basis of direct labor hours, using a normal volume of 20, 000 direct labor hours. In the most recent period, Simple expected to incur 100, 000 of manufacturing overhead. There was no work-in-process inventory at the beginning of the period.[23] Actual product costs turned out to be 50, 000 for direct materials, 270, 000 for direct labor (reflecting 30, 000 hours of direct labor), and 125, 000 for manufacturing overhead. Determine the balance in the overhead cost pool, the plug, that is allocated to cost of goods sold. Conversely, suppose 40% of the overhead allocation rate is the averaged intercept in the overhead LLA. What would be the allocated plug to cost of goods sold if the firm used normal, variable costing?

---

[23] Work-in-process inventory consists of partially completed products as the firm passes from one to an adjoining accounting period. It would be costed in terms of factors consumed to date.

10. *normal, full costing and income effects*

    Ralph produces and sells a wide variety of products. A new product proposal is under review. The tentative plan calls for production of 100 units of this new product. It will sell for 300 per unit. Ralph anticipates the following direct production costs: direct labor: 20 per unit; and direct material: 40 per unit. Overhead is described by an LLA of $OV = F + 2DC$, where DC denotes direct cost dollars (i.e., the sum of direct labor and direct material cost). Ralph uses normal, full costing with an allocation rate of 500% of direct cost, i.e., $5DC$. Finally, if this new product proposal is accepted, Ralph's period costs will increase, for this period only, by a total of 5,000.

    (a) What is the estimated unit cost of this new product?

    (b) By how much will Ralph's accounting income change if this new product proposal is accepted?

    (c) Suppose Ralph accepts this new product proposal, produces 100 units and incurs costs as described above. However, only 80 units sell this period. The remainder are sold next period, also at a price of 300 per unit. Determine the incremental effect on Ralph's accounting income in each of the two periods.

11. *normal, variable costing and income effects*

    Redo problem 10 above on the assumption normal, variable costing is used. Explain your finding.

12. *actual versus normal costing*

    Return to the setting of Ralph's Firm, problem 13 in Chapter 5. After reflection and analysis, Ralph concludes that total manufacturing overhead (OV) is best described with a linear model of the following form: $OV_t = \alpha + \beta y_t + \epsilon_t$. α and β are constants, y_t is the total of direct labor cost plus direct material cost in period t, ϵ_t is a zero mean random error term in period t (arising from such things as weather, shop floor congestion, and so on), and OV_t is total manufacturing overhead in period t.

    Ralph speculates that $\alpha = 20,000$ and $\beta = 1.00$. Ralph also speculates that manufacturing during the period in question will result in $y_t = 20,000$; i.e., direct labor and direct material will total 20,000. Using the output and cost data in the original problem, consider the following.

    (a) Suppose Ralph uses this analysis and speculation to implement a normal, full costing procedure. Determine the unit cost for each product. Further suppose half of the current period production of A and B has been sold. Determine ending finished goods inventory and cost of goods sold.

(b) Carefully discuss how Ralph's specification of the overhead LLA removes the ambiguity encountered in the original problem.

(c) Repeat part (a), assuming Ralph uses a normal, variable costing system.

13. *unit costs*

Ralph is exploring his understanding of various product costing models. For this purpose he envisions a single product firm, along with three cost pools: direct labor, direct material, and overhead. Ralph writes down the following LLAs to describe these three pools: $DL = a \cdot q$, $DM = b \cdot q$ and $OV = c + d \cdot DL$. DL denotes direct labor, DM direct material, and OV overhead. q denotes the units produced. Finally, a, b, c and d are constants. So the slope of the DL LLA is a, the intercept of the OV LLA is c, etc.

Ralph also recognizes the actual costs will differ from those described by the noted LLAs. To accommodate this reality, Ralph next envisions the actual cost in each of the pools as

$$\text{actual } DL = \widehat{DL} = a \cdot q + \varepsilon_1$$
$$\text{actual } DM = \widehat{DM} = b \cdot q + \varepsilon_2$$
$$\text{actual } OV = \widehat{OV} = c + d \cdot \widehat{DL} + \varepsilon_3$$

ε_1, then, is simply the difference between actual direct labor cost and that amount predicted by the LLA (given q units were actually produced).

Now suppose $\widehat{q}$ units are produced, and the actual costs accumulated in each cost pool are given by

$$\text{actual } DL = \widehat{DL} = a \cdot \widehat{q} + \varepsilon_1$$
$$\text{actual } DM = \widehat{DM} = b \cdot \widehat{q} + \varepsilon_2$$
$$\text{actual } OV = \widehat{OV} = c + d \cdot \widehat{DL} + \varepsilon_3$$

Further suppose normal volume is N units of output (i.e., $q = N$); and when standard costing is used the standard cost is based on the originally specified LLAs. Determine Ralph's unit cost (i.e., accounting cost per unit produced) assuming (1) standard, full costing; (2) full, normal costing; (3) variable, normal costing; and (4) actual, full costing. Hint: the unit cost under standard variable costing is $a + b + d \cdot a$.

14. *incremental effects*

Ralph is considering whether to respond to a customer's appeal for production of a special product. The offered price is $6,000$ and Ralph estimates incremental direct labor will cost 300 and incremental direct

material will cost 200. Ralph uses a single overhead pool with LLA given by $OV = 150,000 + 2 \cdot DL\$$, where $DL\$$ denotes direct labor dollars. Normal, full costing is used, with a normal volume of $10,000$ direct labor dollars.

(a) Determine the unit cost of this special product.

(b) What is the net change in Ralph's cost of goods sold as a result of this product being produced and sold. Explain.

(c) Repeat both parts above on the assumption normal, variable costing is used. Explain your finding.

15. *comparison of methods*
Ralph's Job deals with a small custom fabricator of display cabinets. The accounting system separately accumulates direct labor cost, direct material cost, and two overhead pools. The overhead pools are denoted, respectively, OV_A and OV_B. A recent reporting period begins with no work-in-process inventory. During the period three jobs (a, b and c) were worked on. The first two have been completed, and delivered to their customers while the third (job c) remains partially complete at the end of the period. Various overhead and period costs incurred are as follows. (The data are scaled in what follows for presentation purposes.):

	OV_A	OV_B	period
hourly labor	2,000	1,000	1,000
salary labor	4,000	5,000	6,000
various materials	4,000	10,000	12,000
heat and light	8,000		1,000
depreciation	6,000	2,000	2,000
misc.	9,000	5,000	3,000

Direct labor and direct material activities are summarized as follows:

	job a	job b	job c
direct labor	2,200	2,500	3,500
direct material	1,800	5,000	4,000

In addition, the overhead LLAs are given by $OV_A = 22,000 + 1.00 \cdot DL\$$ and $OV_B = 20,000 + .50 \cdot DM\$$ (where $DL\$$ denotes direct labor dollars and $DM\$$ denotes direct material dollars).

(a) Suppose an actual, full costing system is used. Determine the unit cost of each job as well as the period's cost of goods sold.

(b) Repeat (*a*) for the case where a normal, full costing system is used. Assume respective normal volumes of $DL\$ = 10,000$ and $DM\$ = 10,000$ for the two overhead pools.

(c) Repeat (a) for the case where normal, variable costing is used.

16. *variable versus full costing, income effects*
Consider a single product firm with the following LLAs, where q denotes units manufactured and selling and administrative is, of course, a period cost.

direct labor	$DL = 10q$
direct material	$DM = 10q$
overhead	$OV = 90,000 + 2DL$
selling and administrative	$SA = 120,000$

The product sells for 100 per unit. Initially no inventory is present. Production and sales quantities for five consecutive years are noted below. At no time is there any ending work-in-process inventory.

	prd 1	prd 2	prd 3	prd 4	prd 5
production	4,500	4,500	4,500	4,500	4,500
sales	3,000	5,000	4,500	4,000	6,000

Assume the various LLAs are completely accurate. Determine the income and ending finished goods inventory for each period, using normal, full costing and using normal, variable costing. Assume a normal volume of $q = 4,500$ units. How do you explain the period-by-period differences between full and variable cost income?

17. *full costing*
Ralph's Venture finds Ralph in a startup company. Ralph has prepared a business plan and a venture capitalist has agreed to provide the necessary funds. Ralph's business plan rests on the following cost and revenue structures, where q_m denotes units manufactured and q_s denotes units sold.

manufacturing cost	$TMC = 400,000 + 80q_m$
selling and administrative cost	$S\&A = 200,000 + 20q_s$
total revenue	$TR = 700q_s$

The business plan called for production and sale of 1,000 units in the first period, with steadily growing sales thereafter. The venture capitalist also required that Ralph present an audited financial statement at the end of each period. This statement was to be produced according to GAAP, using actual, full costing. During the first period, the estimated cost and revenue structures turned out to be exact. Ralph's Venture produced $q_m = 1,200$ units and sold $q_s = 900$ units. (So manufacturing cost totaled 496,000, $S\&A$ totaled 218,000; and revenue totaled 630,000.) The venture capitalist now examines the

financial statement prepared according to GAAP. It is much better than anticipated, and the venture capitalist turns to the telephone to call some friends and boast about Ralph's Venture.

Determine the income that was projected for the first year in Ralph's business plan and the income that was actually reported to the venture capitalist. Critically comment on the report and the venture capitalist's enthusiasm.

7
The Modernism School

As we have stressed, the impressionism school relies on considerable aggregation and a small number of synthetic variables in its approach to product costing art. The modernism school stresses what it calls *activity based costing, ABC* for short (and its companion managerial mind-set of activity based management). This approach began with dissatisfaction over the impressionism school's approach as technology advances led to more complex, integrated and capital intensive production processes. Just as the modernism movement was a rebellion against traditions, aimed at confronting a changed world, activity based costing can be viewed as a rebellion against traditional approaches to costing art, aimed at improving the estimation of product costs.[1]

That said, the coster's palette remains the same: aggregate cost pools, *LLA*s, and allocation. The distinctive features are envisioning the firm as a collection of micro-firms, called activities, and a willingness to work with remarkably less aggregation coupled with an equally remarkable set of synthetic variables, all designed to connect the costing art more realistically to the production technology. This also leads, in many cases, to more (if not all) of the cost pools being confined to the product as opposed to period

---

[1]It is also important to acknowledge modernism is considerably more dated than activity based costing, having its roots in the early part of the last century. Of course, it has given way to post modernism, and I am beginning to suspect the same is true of product costing art, as we now have such phrases as time-driven activity based costing.

category.[2] Of course, just as good art is in the eye of the beholder, good product costing is in the eye of the beholder. We have some work to do.

We begin with an overview and two stylized examples, and then formalize the central idea of partitioning the firm's technology into a set of more or less independent activities, each complete with local cost structure. We then return to an explicit technology specification, paralleling earlier work in Chapters 2 and 3, and concentrate on the issue of marginal cost estimation. This leads to a portfolio perspective, one where the firm hosts a number of products and any application of costing art will produce an entire portfolio of estimation errors. It then follows that there is no universal answer to which version of the costing recipe produces the least troublesome portfolio of errors.

7.1 Variations on a Theme

To be sure, the last two decades have witnessed a renaissance in product costing art, complete with consultants, software, and, at important junctures, further separation from the economic theory of cost. The basic idea, as we have hinted, is to better match the cost constructions with the technology, with the productive activity itself. The quest is more realism in the cost construction, as a venue for improved cost estimation. Once fashioned, however, we remain with cost pools, LLAs and allocation. The difference is the willingness to confront detail.

Example 7.1 We begin with a popular, albeit apocryphal example. Ralph (Who else?) invites a friend, Sally, to dinner. Their bill tallied as follows:

	appetizer	entree	total
Sally	6	14	20
Ralph	14	36	50
wine			80
tax			15
tip			35
grand total			200

Ralph is not cheap! It also turns out he is a wine snob, and insisted on a serious bottle of wine (most of which he consumed), and is also a big tipper.[3] Now think of this as two products, Ralph's and Sally's dinner. Total cost was 200, but how much did each product cost? The impressionism school would simply treat this as two dinners, costing 100 each or, treating the

---

[2] The pools may be re-engineered as well.

[3] The wine was a half bottle of Heitz Cellar's 2001 Martha's Vineyard Cabernet Sauvignon.

appetizers and entrees as direct costs, allocate the wine, tax and tip using the direct costs. So Sally's dinner cost $20 + (20/70)(80 + 15 + 35) = 57$ and Ralph's dinner cost $50 + (50/70)(80 + 15 + 35) = 143$. The modernism school, on the other hand, would exploit the available details, taking into consideration the fact Ralph is a wine snob and a big tipper. This leads to something like the following:

	Sally	Ralph
appetizer	6	14
entree	14	36
wine	20	60
subtotal	40	110
tax (10%)	4	11
tip	6	29
unit cost	50	150

Here we make a subjective judgement that 25% of the wine's enjoyment was consumed by Sally and the remainder by Ralph. And given Ralph's big tipper status, we costed Sally's tip at the conventional 15% (of 40), and assigned the remainder to Ralph. The tax, of course, is not aggregated and simply assigned at the noted rate of 10% of taxable items.

So what did each dinner cost? The impressionism school approaches the question in "quick and dirty" fashion, while the modernism school exploits various details and takes a more nuanced approach.[4] Notice, however, that either way we wind up with a total of 200, and that either way we have ignored some resource consumptions, such as transportation and Ralph and Sally's time. Most important, we have sketched a separable story. While both parties enjoyed the ambiance, the service, the wine, it is also fairly likely that had they eaten alone they would have ordered differently. Separability is a fiction.

Summing up our little story, the two approaches can lead to significantly different answers, the answers always add up, and we remain perplexed. That is the nature of the game.

Example 7.2 Now consider a two product story, where $q_1 = 100$ units of one product and $q_2 = 20$ units of a second product have been produced. To add some flavor to the story, think of the products as high-end imaging equipment, with the first being a "standard" specification and the second being an "advanced" more or less custom specification. Nine cost pools

---

[4]Both approaches inject themselves into the social fabric of the situation, and this shows up when firms engage in wholesale restructuring of their costing apparatus. By analogy, imagine your instructor announcing in the middle of the course that the grading system will be changed!

are present, direct labor and direct material for each product along with four overhead pools and a single period cost pool. Details (000 omitted) are presented in Table 7.1.[5] You will also notice in what follows that we focus on an actual costing system simply to avoid clutter and distracting minutia.

TABLE 7.1: Cost Pools

pool	designation	total
1	direct labor (q_1)	2,000
2	direct labor (q_2)	100
3	direct material (q_1)	3,000
4	direct material (q_2)	1,400
5	overhead A (OV_A)	3,100
6	overhead B (OV_B)	15,200
7	overhead C (OV_C)	3,000
8	overhead D (OV_D)	2,200
9	period	5,000
		35,000

Now suppose we invoke an extreme form of the impressionism school, aggregate all four overhead pools, and use direct labor cost as the synthetic variable for the aggregate overhead pool's LLA. This leads to an allocation rate of $f = (3,100 + 15,200 + 3,000 + 2,200)/(2,000 + 100) = 23,500/2,100 = 11.190$. And we wind up with the following unit costs.

TABLE 7.2: Unit Costs based on Direct Labor Allocation

pool	q_1 (100 units)	q_2 (20 units)
direct labor	2,000	100
direct material	3,000	1,400
overhead	22,381	1,119
total product cost	27,381	2,619
unit cost	274	132

Recalling the period costs total 5,000, this suggests the firm's cost curve is approximated by

$$\text{cost} \cong 5,000 + 274q_1 + 132q_2 \tag{7.1}$$

---

[5] We skip over the underlying details of what types of resources are included in the various pools. Glance back at Tables 6.1 and 6.2, for example.

The concerns here are, among other things, that we have said nothing about the underlying technology, the indirect product costs are a large percentage of total cost, and the synthetic variable of direct labor cost is, correspondingly, a small percentage of total cost.[6]

On closer analysis, the OV_A pool contains a variety of resources, largely labor support services, that closely follow direct labor, so we use direct labor cost for this pool's synthetic variable.

OV_B, on the other hand, reflects resources associated with materials, including receiving, handling and, it turns out, preparation for assembly. So we use direct material cost for this pool's synthetic variable.

OV_C reflects setup and inspection activities, with each unit of the second product requiring a unique setup while the first product is produced in batches of 50 units each. So we use number of setups as the synthetic variable here (and note we have a total of $100/50 + 20 = 22$ setups, as $q_1 = 100$ and $q_2 = 20$).

OV_D is problematic. It reflects various resources, such as floating personnel, routine maintenance, and so on, all essential but difficult to characterize. A consensus among the various managers emerges that these resources are associated with the resource consumptions in the other three overhead pools in roughly a ratio of 1 to 6 to 3. Think of this as defining a synthetic variable of "service units," where 10 units of service were provided, 1 unit to OV_A, 6 units to OV_B and 3 units to OV_C.

Putting all of this together, the original four overhead pools are now envisioned as distinct activities, and we approach the allocation of each one's cost in terms of the resources or services it provides other activities or products. This is summarized by the synthetic variables and allocation rates displayed in Table 7.3. For example, the OV_B pool begins with an initial tally of $15,200$ (Table 7.1). To this amount we add an allocation of the OV_D total of $(6/10)(2,200) = 1,320$, resulting in a revised total of $16,520$. Using the noted synthetic variable of direct material cost, we have an allocation base of $DM = 3,000 + 1,400 = 4,400$ (again, see Table 7.1), and thus an allocation rate of $16,520/4,400 = 3.755$ per dollar of direct material cost.

From here it is routine to construct the unit costs: we merely tally the direct product costs and allocate the indirect product costs, using the calculated allocation rates. This leads to the respective total product costs of $19,759$ and $10,241$ and unit costs of $19,759/100 \approx 198$ and $10,241/20 \approx 512$, as displayed in Table 7.4.

---

[6] Moreover, use of direct material cost as a synthetic variable leads to parallel concerns. For the record, this would imply respective unit costs of 210 and 449.

TABLE 7.3: Allocation Rates for ABC System

	OV_A	OV_B	OV_C
original amount	3,100	15,200	3,000
allocate $OV_D = 2,200$ $(1:6:3)$	220	1,320	660
total	3,320	16,520	3,660
synthetic variable	DL	DM	setups
base	2,100	4,400	22
rate	1.581	3.755	166.364

TABLE 7.4: Unit Costs based on ABC System

pool	q_1 (100 units)	q_2 (20 units)	total
direct labor	2,000	100	2,100
direct material	3,000	1,400	4,400
$OV_A = 1.581DL$	3,162	158	3,320
$OV_B = 3.755DM$	11,264	5,256	16,520
$OV_C = 166.364$	333	3,327	3,660
total product cost	19,759	10,241	30,000
unit cost	198	512	

In contrast with the impressionism construction, we rely on considerably more detail, less aggregation, and a variety of synthetic variables, all drawn from closer affiliation to the production process, the technology itself. The treatment of OV_D should also be noted, as those resources were tracked in indirect fashion to the products. Yet at each and every step the underlying algebra is one of imposing linearity, and everything adds up yet again. This manifests itself in the following implied cost expression:

$$\text{cost} \cong 5,000 + 198q_1 + 512q_2 \qquad (7.2)$$

We could, at this point, explore normal costing variations as well as the full versus variable costing distinction. However, ABC systems tend to emphasize a full costing approach, and may even include what would otherwise be treated as a period cost in the calculations. Regardless, the question at this point is whether expression (7.1) or (7.2) provides a "better" rendering of the firm's cost structure. The presumptive marginal cost of the first product has dropped, while that of the second, the customized product, has increased as we moved from the impressionism to modernism approach. But which answer is to be preferred?

Our reticence on this matter will become clear as we proceed.

7.2   The Underlying Structure

Reflecting on our work in Chapters 2 and 3, the firm's economic cost function, $C(q; P)$, concerns itself with all of the firm's resources and all of its products. It rests on the firm's technology; and given this technology, economic cost depends on output and factor prices. With the latter held constant, economic cost is fully explained by output. Moreover, marginal cost is the center of attention. Emphatically, separability is not presumed and synthetic variables are strictly outside the theory.[7]

Accounting's portrayal of economic cost, on the other hand, is remarkably pragmatic. For example, our work in Tables 7.1 through 7.4 focuses on only the products produced during a specific period, so separability has been introduced. It also relies heavily on a host of aggregate expressions, LLAs, and synthetic variables.

In so doing, the modernism school stresses the idea of an activity.

Definition 18 *An activity is a sphere of productive action that is modeled as if it is a separable form of production, with its own output measure and cost structure.*

Note well, modeled "as if" does not mean indeed separable, it merely means we choose to invoke an assumption of separability.

This leads, inexorably, to a collection of or variety of components of the firm's cost function, cost subfunctions in a sense. Moreover, the output of an activity may be something that benefits the firm's output per se or it may benefit other activities. Each activity, in this sense, is a micro-firm, complete with synthetic output measure.

We saw this in Table 7.3 where four activities are identified, with respective output measures of direct labor cost, direct material cost, setups, and

---

[7]Synthetic variables come into play once we step away from ruthless specification of the technology and factor consumptions. In addition to aggregating factors into cost pools, we aggregate output into output totals. In Example 7.2, for instance, we dealt with setup costs, yet in theory each single batch of output is a distinct output. This casual approach to measuring factors and output is what leads some in the modernism school to claim they have discovered non-volume cost drivers, or a cost expression that with factor prices given is not completely explained by output. Rather, the non-volume cost drivers are picking up measurement error.

Recall note 9 in Chapter 3: "In later chapters we will confront the claim cost is explained by variables other than volume or quantity of output (given prices). We will see, however, that this phenomenon is driven by the fact we use approximate expressions for the firm's cost curve and aggregate a large variety of products into sets of bundles. Both are essential in the land of reality, and both lead to errors relative to economic theory."

To push a bit further, suppose output is produced in batches of 50 units, and each batch necessitates a costly setup. If each batch is a distinct product, setup costs are linked to products; but if the batches are aggregated, setup costs are linked to an activity called setup.

service units. We also saw it in Table 7.2 where these four were collapsed
into a single activity. And we also saw it in the Ralph, Ltd. odyssey in
Table 5.3. The innovation here is to stress careful identification of activi-
ties, to work to understand their cost structures, and to target them for
managerial innovation.[8]

Definition 19 *An activity based costing (ABC) system partitions the firm's
technology into a set of activities, approximates each activity's cost struc-
ture with an LLA given by cost = a + bx and derives unit costs by allocating
activity costs to products via each LLA's synthetic variable, x.*

Naturally, the *ABC* system can be approached in terms of actual or
normal costing, and in full or variable (or both) formats, but these are
implementation details. The important distinction is the reliance on less
aggregation and a wider variety of synthetic variables; and the important
question is whether this approach leads to improved marginal cost esti-
mates. Exploration of this question, however, necessitates a more detailed
specification of the firm's technology.

7.3   Back to the Firm's Technology

We now posit an unusually explicit setting, where the firm produces two
products, with quantities denoted q_1 and q_2. Direct labor is present, but we
assume no other direct factors are present, simply to keep things focused.
Three overhead pools are also present, and each such pool has the natural
interpretation of an activity. Moreover, this is a single period story, so
we sidestep the important issue of separating one period's cost and output
from another's. You will also notice in what follows that we have no period
costs as well, thereby sidestepping the product versus period distinction.

Eight factors are present, with the first six being purchased in factor
markets and the remaining two internally produced. z_j denotes the quan-
tity of factor j and the market prices of the purchased factors are denoted
$P_1, ..., P_6$. So the firm's total expenditure on factors is simply $\sum_{j=1}^{6} P_i z_i$.

The first factor is direct labor supplied to the first product. We scale
the price so one unit of this factor corresponds to the required direct labor
for the first product. So to produce q_1 units we require $z_1 \geq q_1$ units of
direct labor for the first product. A similar scaling is used for the second
product, thus necessitating $z_2 \geq q_2$.

---

[8] Again, each cost pool in the impressionism world can be thought of as an activity,
but it is the managerial focus that is key. And naturally this managerial focus can
lead to a different set of cost pools. Hence the companion phrase of activity based
management. This leads to questioning whether each activity "adds value" and using
the resulting insights to re-engineer the technology.

As noted, factors $z_1, ...z_6$ are purchased and factors z_7 and z_8 are internally produced. This leads to the following program for determining the firm's cost of producing output $q = [q_1, q_2]$ in the face of price vector P, where expressions (7.3b) through (7.3f) characterize the firm's technology:

$$C(q; P) \equiv \min_{z_1 \geq 0, ..., z_8 \geq 0} \sum_{j=1}^{6} P_j z_j \tag{7.3a}$$

$$\text{s.t. } z_1 \geq q_1 \tag{7.3b}$$

$$z_2 \geq q_2 \tag{7.3c}$$

$$\sqrt{z_3 z_7} \geq A_{11} q_1 + A_{12} q_2 \tag{7.3d}$$

$$\sqrt{z_4 z_8} \geq A_{21} q_1 + A_{22} q_2 \tag{7.3e}$$

$$\sqrt{z_5 z_6} - z_7 - z_8 \geq 0 \tag{7.3f}$$

It is important we take some time to digest this story. (7.3a) is simply the total of expenditures on purchased factors. (7.3b) and (7.3c) require the firm supply adequate direct labor for each product. (7.3d) requires productive services delivered by factors z_3 and z_7 be sufficient to support production of q_1 and q_2. Think of this as some type of subassembly or service item. Each unit of the first product requires A_{11} units of the item, while each unit of the second product requires A_{12} units of the item. In turn, the total number of such items available depends on the supply of factors z_3 and z_7, and is governed by $\sqrt{z_3 z_7}$. Importantly, external factor z_3 and internal factor z_7 are substitutes in this regard.

TABLE 7.5: Cost Pools for Extended Illustration

	pool	amount
1	direct labor for q_1	$P_1 z_1$
2	direct labor for q_2	$P_2 z_2$
3	overhead A (OV_A)	$P_3 z_3$
4	overhead B (OV_B)	$P_4 z_4$
5	overhead C (OV_C)	$P_5 z_5 + P_6 z_6$

(7.3e) is a parallel story, related to another and distinct subassembly or service item. Notice that here factors z_4 and z_8 are substitutes. Moreover, while factors z_3 and z_4 are purchased in factor markets (e.g., some labor or subassembly item), factors z_7 and z_8 are internally produced according to the technology specified in (7.3f). Further notice purchased factors z_5 and z_6 are used to produce this supply of internal factors.

Each of the constraints corresponds to what we might term an activity. We will further assume that the various factor expenditures are recorded in distinct cost pools as described in Table 7.5.[9]

7.3.1   Marginal Costs

We know from our earlier work that the prize in this game is estimating each product's marginal cost. The exact expression follows from our earlier work on shadow prices. Notice program (7.3) contains 5 constraints, so label their respective shadow prices $\lambda_1, ..., \lambda_5$. From here it is routine to verify that the marginal cost of product i, $i = 1, 2$, given price vector P and given output of $q = [q_1, q_2]$, is the following linear expression:[10]

$$MC_i(q; P) = \lambda_i + \lambda_3 A_{1i} + \lambda_4 A_{2i} = P_i + \lambda_3 A_{1i} + \lambda_4 A_{2i} \qquad (7.4)$$

Notice the marginal cost is the sum of three terms. The first term, λ_i, is the shadow price on the direct labor constraint, so we have a term corresponding to direct labor. Indeed, our scaling ensures the shadow price on the first (second) constraint is simply the price of the first (second) factor. The second term is the shadow price on the third constraint, λ_3, multiplied by the number of units of the activity described in (7.3d) that the product requires. So we have a term consisting of cost per unit of activity multiplied by units of that activity, A_{1i}. And the third term is a parallel story: the shadow price on the fourth constraint, λ_4, multiplied by the number of units of the activity described in (7.3e) that the product requires, A_{2i}.

This is not magic. Rather, the technology was specified so we would have such an additive structure for the marginal cost expressions. Its importance will become clear as we turn to using product costing art to estimate these marginal costs.

7.3.2   Impressionism's Answer

Let's begin with the impressionism school, and assume all indirect product costs are aggregated into a single overhead pool, and that the LLA for

---

[9]Remember, factors z_7 and z_8 are internally produced, and thus are not explicitly priced. Rather, the expenditures on factors (z_5 and z_6) used to produce them are recorded in the overhead C pool.

[10]The Lagranian for program (7.3) is given by $\Psi = \sum_{j=1}^{6} P_j z_j - \lambda_1(z_1 - q_1) - \lambda_2(z_2 - q_2) - \lambda_3(\sqrt{z_3 z_7} - A_{11} q_1 - A_{12} q_2) - \lambda_4(\sqrt{z_4 z_8} - A_{21} q_1 - A_{22} q_2) - \lambda_5(\sqrt{z_5 z_6} - z_7 - z_8)$. Differentiating with respect to q_1 gives the first product's marginal cost expression, and differentiating with respect to q_2 gives its counterpart for the second expression. Likewise, differentiation with respect to factor z_1 will convince you that $\lambda_1 = P_1$, a direct (pun) consequence of our scaling. Similarly, you should be able to convince yourself that $\lambda_2 = P_2$.

this pool uses direct labor cost as a synthetic variable. Glancing back at Table 7.5, we see that total overhead will be $OV = OV_A + OV_B + OV_C = \sum_{j=3}^{6} P_j z_j$. Likewise, total direct labor cost will be $DL = P_1 z_1 + P_2 z_2$. Thus the allocation rate will be

$$f = OV/DL = \sum_{j=3}^{6} P_j z_j / (P_1 z_1 + P_2 z_2) \qquad (7.5)$$

In our streamlined setting, unit cost will consist of direct labor cost (per unit), P_1 or P_2, coupled with allocated overhead of $f \cdot P_1$ or $f \cdot P_2$. This provides the following unit cost expressions for the impressionism school:

$$c_i = P_i + f \cdot P_i \qquad (7.6)$$

Notice it is the sum of two terms, in contrast to the three terms in (7.4). Also do not miss the fact this provides a cost curve approximation of $C(q; P) \cong c_1 q_1 + c_2 q_2$, reflecting presumed separability and constant returns to scale.

7.3.3   ABC's Answer

In contrast, an ABC system will divide the technology into activities, and use activity consumptions and costs to guide the allocations. We begin with the same cost pools in Table 7.5. Glancing back at (7.3) we treat (7.3b) through (7.3f) as activities. (7.3b) and (7.3c) are simply direct labor cost stories, and are treated just as in the impressionism school.

The remaining three, however, are indirect product cost stories. Let's begin with (7.3f) where we have the expression $\sqrt{z_5 z_6} \geq z_7 + z_8$. Factors z_5 and z_6 are used to produce factors z_7 and z_8. The cost of producing these latter factors is recorded in the overhead C cost pool of $OV_C = P_5 z_5 + P_6 z_6$. In turn, a natural synthetic variable to use for this activity is the units of "service" it provides, i.e., $x_C = z_7 + z_8$. This provides an allocation rate for this activity of

$$f_C = OV_C / x_C = OV_C / (z_7 + z_8) \qquad (7.7)$$

Stated differently, this activity's LLA is given by $OV_C = f_C \cdot x_C$. Looking back at our definition of an activity based costing system, we are proceeding with a zero intercept in the LLA and thereby working with a full costing approach just as we did in examining the impressionism school's approach here.

Allocation of OV_C, in turn, is based on consumption as measured by the synthetic variable, x_C. Notice, however, that the synthetic variable is traced to the activities described by (7.3d) and (7.3e), in effect to the two other overhead pools. This provides revised overhead totals for the two

activities and their associated overhead OV_A and OV_B pools of $\widehat{OV}_A = OV_A + f_C \cdot z_7$ and $\widehat{OV}_B = OV_B + f_C \cdot z_8$.

Now look more closely at the activity described by (7.3d): $\sqrt{z_3 z_7} \geq A_{11}q_1 + A_{12}q_2$. A natural synthetic variable here is once again units of service provided by the activity, or $x_A = A_{11}q_1 + A_{12}q_2$. We therefore have an allocation rate of

$$f_A = \widehat{OV}_A/x_A = (OV_A + f_C \cdot z_7)/(A_{11}q_1 + A_{12}q_2) \qquad (7.8)$$

So we have an LLA for this activity of $\widehat{OV}_A = f_A \cdot x_A$.

In parallel fashion, the activity defined by (7.3e) requires $\sqrt{z_4 z_8} \geq A_{21}q_1 + A_{22}q_2$; and the natural synthetic variable here is $x_B = A_{21}q_1 + A_{22}q_2$. This leads to an allocation rate of

$$f_B = \widehat{OV}_B/x_B = (OV_B + f_C \cdot z_8)/(A_{21}q_1 + A_{22}q_2) \qquad (7.9)$$

and an activity LLA of $\widehat{OV}_B = f_B \cdot x_B$.

From these expressions, these cost subfunctions, we construct the unit costs. As in the impressionism setting, unit cost consists of direct labor cost coupled with allocated overhead. Keep in mind that the A and B activities consume the services provided by activity C. Each unit of the first product, for example, consumes A_{11} units of service from activity A and A_{21} units of service from activity B. This provides the following unit cost expressions for the ABC school, where we have, for product i, the direct (labor) cost of P_i coupled with allocations from the $\widehat{OV}_A$ and $\widehat{OV}_B$ pools whose synthetic variables track to the outputs themselves:

$$\widehat{c}_i = P_i + f_A \cdot A_{1i} + f_B \cdot A_{2i} \qquad (7.10)$$

Notice $\widehat{c}_i$ is the sum of three terms, just as in the marginal cost expression of (7.4), and in contrast to the two terms in the impressionism school's calculation. This suggests the ABC approach is on a stronger footing when it comes to estimating marginal cost. Regardless, do not lose sight of the fact the ABC approach suggests the firm's cost is well expressed by $C(q; P) \cong \widehat{c}_1 q_1 + \widehat{c}_2 q_2$, yet another separable, constant returns to scale story.

Example 7.3 An example is surely overdue. Assume factor prices are given by $P_1 = 20, P_2 = 10, P_3 = 1, P_4 = 2, P_5 = 3$ and $P_6 = 4$. Further assume, glancing back at (7.3d) and (7.3e), that $A_{11} = 1, A_{12} = 3, A_{21} = 3$ and $A_{22} = 1$. The important pattern here is the first product consumes more direct labor and more of the (7.3e) service while the second consumes less direct labor and more of the (7.3d) service. Now suppose $q_1 = 7$ and $q_2 = 9$ units are produced and the factor consumptions are the solution to program (7.3).[11] This implies a total cost of 632.33. (As usual, you should

---

[11] The factor choices are $z = [7, 9, 89.493, 55.836, 33.528, 25.146, 12.917, 16.119]$. You should verify this claim.

think in terms of scaled numbers.) Paralleling Table 7.5, we have the cost pool totals displayed in Table 7.6.

TABLE 7.6: Cost Pools for $q = [7, 9]$

	pool	amount	dollars
1	direct labor for q_1	$P_1 z_1$	140
2	direct labor for q_2	$P_2 z_2$	90
3	overhead A (OV_A)	$P_3 z_3$	89.49
4	overhead B (OV_B)	$P_4 z_4$	111.67
5	overhead C (OV_C)	$P_5 z_5 + P_6 z_6$	201.17

From here we construct the impressionism school's rendering. Recall in this particular setting that we assume the overhead is aggregated into a single pool, and that direct labor cost is the synthetic variable of choice. This implies, following (7.5), an allocation rate of $f = (89.49 + 111.67 + 201.17)/(140 + 90) = 1.749$. And we wind up, following (7.6), with the unit cost constructions in Table 7.7. Recalling the absence of any period costs in the story, we thus have a cost curve approximation of $C(q; P) \approx 54.99q_1 + 27.49q_2$.

TABLE 7.7: Unit Costs based on Direct Labor Allocation

pool	q_1 (7 units)	q_2 (9 units)
direct labor	140	90
overhead	244.90	157.43
total product cost	384.90	247.43
unit cost	54.99	27.49

Turning to the modernism school, we treat each overhead pool in the story as a distinct activity. Following (7.7), the synthetic variable for the OV_C pool is units of service (in this case service provided the other two overhead activities), measured by factor consumptions z_7 and z_8. Of course we must know these amounts, and it turns out we have $z_7 = 12.917$ and $z_8 = 16.119$. With $OV_C = 201.17$ we thus have an allocation rate of $f_C = 201.17/(12.917 + 16.119) = 6.928$. From here we allocate OV_C to the other two overhead pools, and then, following (7.8) and (7.9), construct the allocation rates as laid out in Table 7.8: $f_A = 178.99/34 = 5.264$ and $f_B = 223.34/30 = 7.445$.

TABLE 7.8: f_A and f_B Allocation Rates

	OV_A	OV_B
original amount	89.50	111.67
allocate OV_C, $f_C = 6.928$	$f_C \cdot 12.917 = $ 89.49	$f_C \cdot 16.119 = $ 111.67
total	178.99	223.34
synthetic variable	$A_{11}q_1 + A_{12}q_2$	$A_{21}q_1 + A_{22}q_2$
base	34	30
rate	$f_A = 5.264$	$f_B = 7.445$

With the allocation rates in place it is routine to construct the unit costs: following (7.10), we merely tally the direct product costs and allocate the indirect product costs, using the calculated allocation rates. This leads, in Table 7.9, to respective unit costs of 47.60 and 33.24. We thus wind up with a cost curve approximation of $C(q; P) \cong 47.60q_1 + 33.24q_2$.

Notice in either the impressionism or ABC approach that we begin with the same total product cost incurred and develop unit costs such that the sum of the unit costs is equal to the total product cost incurred: $632.33 = 54.99(7) + 27.49(9) = 47.60(7) + 33.24(9)$.[12]

TABLE 7.9: Unit Costs based on ABC System

pool	q_1 (7 units)	q_2 (9 units)	total
direct labor	140	90	230
$OV_A = 5.264A_{1i}q_i$	36.85	142.13	178.98
$OV_B = 7.445A_{2i}q_i$	156.35	67.01	223.36
total product cost	333.20	299.14	632.34
unit cost	47.60	33.24	

7.3.4 Back to Marginal Costs

But what about marginal cost? It turns out in this setting characterized by program (7.3) that the firm's cost curve is given by

$$C(q; P) = \widehat{c}_1 q_1 + \widehat{c}_2 q_2 \qquad (7.11)$$

where $\widehat{c}_1$ and $\widehat{c}_2$ are the ABC system's unit costs, as derived in expression (7.10), as well as Table 7.9. The technology is such that the ABC system

---

[12] Well, we have a slight rounding error! Beyond that, and more importantly, you should notice this is a static exercise. In reality, would the firm's behavior be affected by the way it measured its product costs? If not, we are wasting our time.

provides error free marginal cost estimates. This follows from the separable components of the technology (glance back at (7.3)) and from the "square root" structure of the technology coupled with the additive structure of the service consumptions. So you should not assume this is a general phenomenon. But you should assume the setting is designed to show the logic and intuition of the modernism school in its most pure form.[13]

Our setting in (7.3), then, is modernism's nirvana. The firm's cost structure is separable and linear, and unit costs constructed with the *ABC* approach faithfully measure each product's marginal cost. In turn, the impressionism school's approach is a laggard by comparison. In Example 7.3 we see that it overstates the first product's marginal cost ($54.99 > 47.60$) and understates that of the second ($27.49 < 33.24$). This precise pattern, though, is driven by the fact the first product is a relatively high consumer of direct labor, and thus is allocated a relatively large amount of the aggregate overhead. More broadly, except in a razor edge case, in this setting the

---

[13] Though a thorough explanation leads to a slight detour, we begin, following Chapter 2's Appendix, with the Lagrangian expression for program (7.3):

$\Psi = \sum_{j=1}^{6} P_j z_j - \lambda_1(z_1 - q_1) - \lambda_2(z_2 - q_2) - \lambda_3(\sqrt{z_3 z_7} - A_{11}q_1 - A_{12}q_2) - \lambda_4(\sqrt{z_4 z_8} - A_{21}q_1 - A_{22}q_2) - \lambda_5(\sqrt{z_5 z_6} - z_7 - z_8)$, where λ_k denotes the shadow price on the kth constraint. In turn, differentiating with respect to each of the factors, and setting each such expression to 0 provides the following 8 equations.

$$\frac{\partial \Psi}{\partial z_1} = P_1 - \lambda_1 = 0$$

$$\frac{\partial \Psi}{\partial z_2} = P_2 - \lambda_2 = 0$$

$$\frac{\partial \Psi}{\partial z_3} = P_3 - \lambda_3 z_7/2\sqrt{z_3 z_7} = 0$$

$$\frac{\partial \Psi}{\partial z_4} = P_4 - \lambda_4 z_8/2\sqrt{z_4 z_8} = 0$$

$$\frac{\partial \Psi}{\partial z_5} = P_5 - \lambda_5 z_6/2\sqrt{z_5 z_6} = 0$$

$$\frac{\partial \Psi}{\partial z_6} = P_6 - \lambda_5 z_5/2\sqrt{z_5 z_6} = 0$$

$$\frac{\partial \Psi}{\partial z_7} = -\lambda_3 z_3/2\sqrt{z_3 z_7} + \lambda_5 = 0$$

$$\frac{\partial \Psi}{\partial z_8} = -\lambda_4 z_4/2\sqrt{z_4 z_8} + \lambda_5 = 0$$

The first one implies, as claimed in (7.4) that $\lambda_1 = P_1$, just as the second implies $\lambda_2 = P_2$. Next use the 5th and 6th expressions to determine $\lambda_5 = 2\sqrt{P_5 P_6}$. From here, the 3rd and 7th can be used to determine $\lambda_3 = 2\sqrt{\lambda_5 P_3}$, and then use the 4th and 8th to determine $\lambda_4 = 2\sqrt{\lambda_5 P_4}$.

From here, we use the fact the various constraints will be binding, so we know that $\sqrt{z_3 z_7} = A_{11}q_1 - A_{12}q_2$, $\sqrt{z_4 z_8} = A_{21}q_1 - A_{22}q_2$ and $\sqrt{z_5 z_6} = z_7 + z_8$. And using these equalities and the shadow prices, you can now determine each product's marginal cost, using (7.4), each activity's allocation rate, using (7.7), (7.9) and (7.8), and each product's unit cost, using (7.10).

impressionism school will always overstate the marginal cost of one product and understate that of the other product, though the magnitude of these errors depends on finer details. For example, if overhead is a small fraction of direct labor cost, the errors are likely to be trivial, just as the opposite is likely to be the case when overhead is vastly larger in magnitude than the direct labor cost. Again, however, bold claims of this sort depend on the specified technology in (7.3).

7.4   Numerical Explorations

To explore further the horse race between these two schools, we make two small but significant changes in the technology of (7.3). First, the local production function in (7.3d) is changed from $\sqrt{z_3 z_7}$ to $z_3^\alpha z_7^\alpha$. $\alpha = .5$ is the original story, and exhibits constant returns to scale. $0 < \alpha < .5$ is a decreasing returns to scale story, e.g., additional production causes congestion of some sort and thus marginal cost increases. Likewise, $\alpha > .5$ is an increasing returns to scale story, e.g. additional production benefits from learning and thus marginal cost decreases. Second, the local production function in (7.3e) is changed from $\sqrt{z_4 z_8}$ to $z_4^\beta z_8^\beta$. β below, above or equal to .5 is a parallel decreasing, increasing or constant returns to scale story, but now in the $z_4^\beta z_8^\beta$ region of the production function. Beyond this we retain all of the price and A_{ij} specifications used in Example 7.3.

Notice, however, that the marginal cost expression itself, (7.4), is unaffected by these changes, though the shadow prices will be. Likewise, the impressionism and modernism unit cost calculations continue to be calculated precisely as described in (7.6) and (7.10). Moreover, the unit cost calculations are constrained by the requirement that total (product) cost incurred must equal the sum of unit cost times output over the two products. This tidiness is always maintained.

7.4.1   Decreasing Returns

Initially consider the case where $\alpha = \beta = .45$. This is a setting where each product's marginal cost is increasing. Below we plot the marginal cost of each product, $MC_i(q; P)$ along with the percentage error in each school's estimation of that marginal cost. The error measure, in other words, is

$$\frac{100(\text{marginal cost - unit cost})}{\text{marginal cost}}$$

Thus, the measure is positive if the unit cost is a downward biased estimate, is below marginal cost, and is negative if it is an upward biased estimate.

The first thing to notice is that each product's marginal cost increases with production of either product, as displayed in Figures 7.1 and 7.2. This
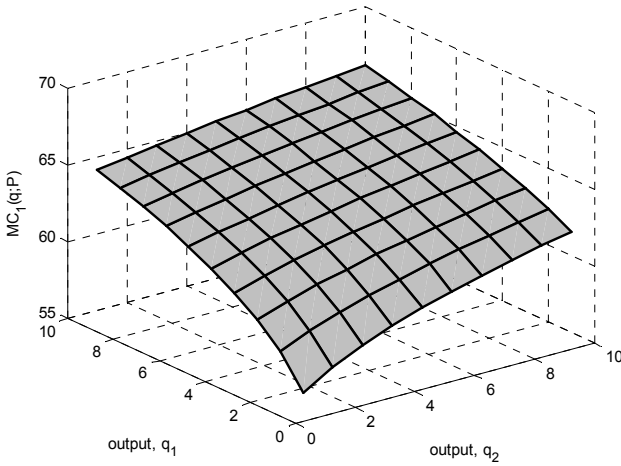
FIGURE 7.1. $MC_1(q; P), \alpha = \beta = .45$

is a story of steadily increasing marginal cost, reflecting the fact we now have decreasing returns to scale in local production functions (7.3d) and (7.3e). Keep this increasing marginal cost pattern in mind as we proceed.

Now consider the impressionism school's ability to reasonably well estimate these marginal costs. For the first product we see in Figure 7.3 that the error can be rather small or rather large, and when large it is an upward biased estimate. Conversely, for the second product, displayed in Figure 7.4, it provides a downward biased estimate, and is in most cases not that accurate an estimate.

The *ABC* approach on the other hand provides consistently downward biased estimates. It tends to provide a better estimate than the impressionism school for the second product, but not so for the first, as becomes clear in Figures 7.5 and 7.6. Further notice the *ABC* error for either product increases with output, as the strain imposed by everything adding up to the total cost incurred becomes more troublesome.[14]

7.4.2   Increasing Returns

Now let's try the opposite case of steadily decreasing marginal costs. Let $\alpha = \beta = .55$. We now have the error patterns displayed in Figures 7.7 through 7.10. Here, in the face of decreasing marginal costs, we see the *ABC* approach, Figures 7.9 and 7.10, consistently overstates marginal cost.

---

[14]Where have you seen this phenomenon before? Glance back at Table 4.5 as well as note 8 in Chapter 4.

FIGURE 7.2. $MC_2(q; P), \alpha = \beta = .45$

FIGURE 7.3. Impressionism Error for First Product, $\alpha = \beta = .45$

FIGURE 7.4. Impressionism Error for Second Product, $\alpha = \beta = .45$

FIGURE 7.5. ABC Error for First Product, $\alpha = \beta = .45$
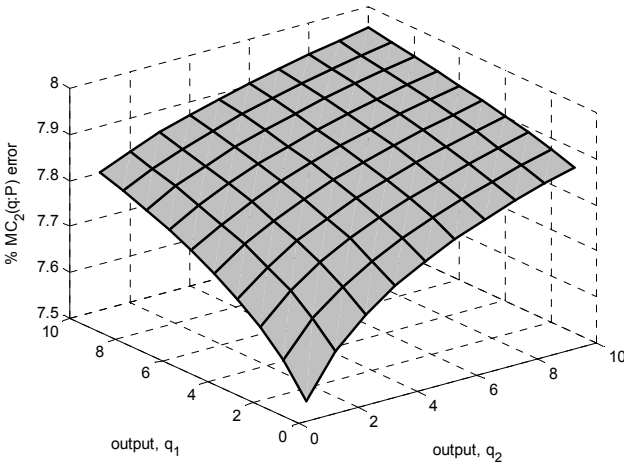
FIGURE 7.6. *ABC* Error for Second Product, $\alpha = \beta = .45$

This reflects the friendly somewhat separable technology in (7.3) and the fact that everything must add up to total cost incurred, making it impossible to correctly estimate both marginal costs. The impressionism school, Figures 7.8 and 7.9, with its slightly freer hand offers less of a defined pattern and does, on occasion, outperform the modernism calculations.

7.4.3   Mixed Case

Let's wrap this up by looking at a mixed case of decreasing returns in one activity and increasing returns in another. $\alpha = .55$ and $\beta = .45$ will accomplish this. In particular this leads to a case where the second product's marginal cost, $MC_2(q; P)$, decreases with q_2 but increases with q_1. The error patterns in Figures 7.11 through 7.14 emerge. Again, neither approach is dominating in its performance.

7.5   Portfolio of Errors

Economic theory is unrelenting in its depiction of the forces that determine a firm's cost structure and that at the product level the only meaningful concept of cost per unit is marginal cost. Staying close to theory, then, it is natural we emphasize product costing art as providing estimates of marginal cost. But the world of affairs, as opposed to the world of theory, has a pragmatic side. This implies, regardless of the costing system's sophistication, that we should, and the wise manager surely will, expect

FIGURE 7.7. Impressionism Error for First Product, $\alpha = \beta = .55$
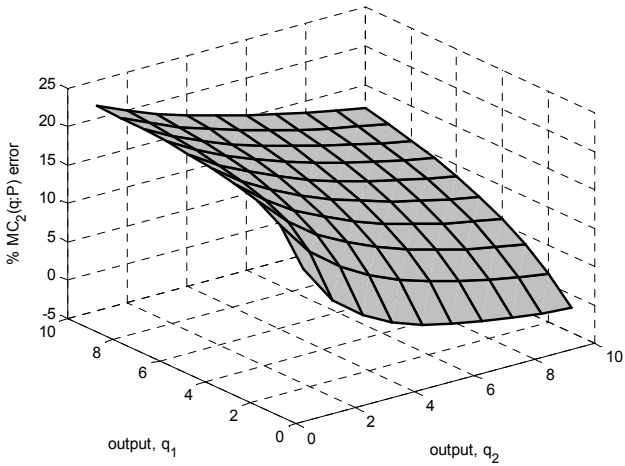
FIGURE 7.8. Impressionism Error for Second Product, $\alpha = \beta = .55$
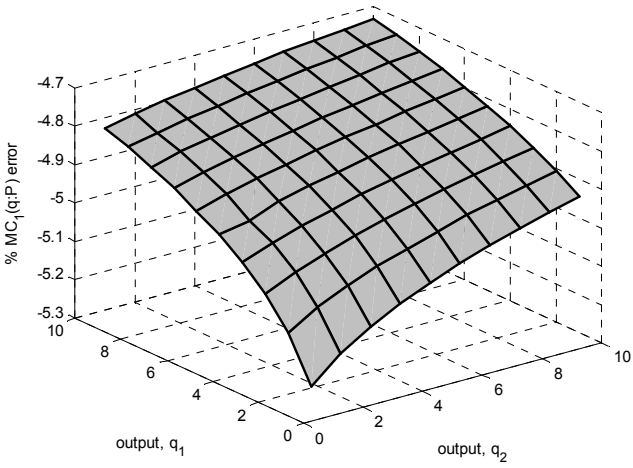
FIGURE 7.9. *ABC* Error for First Product, $\alpha = \beta = .55$
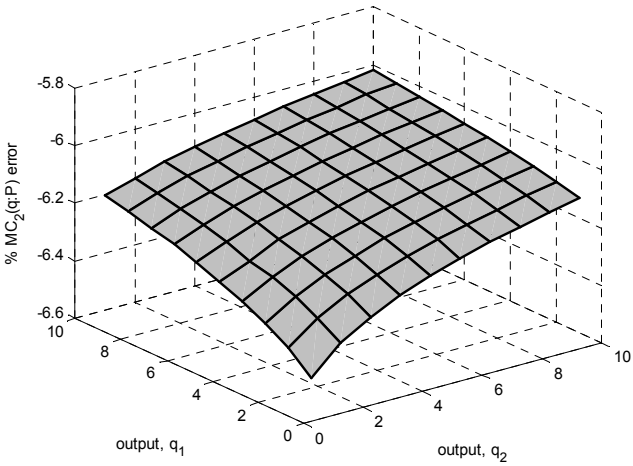
FIGURE 7.10. *ABC* Error for Second Product, $\alpha = \beta = .55$
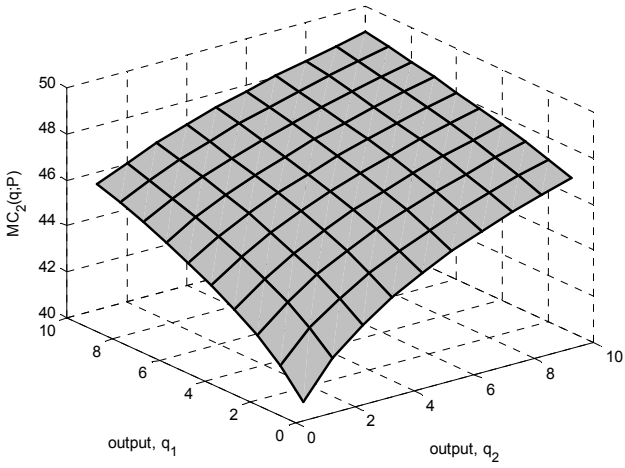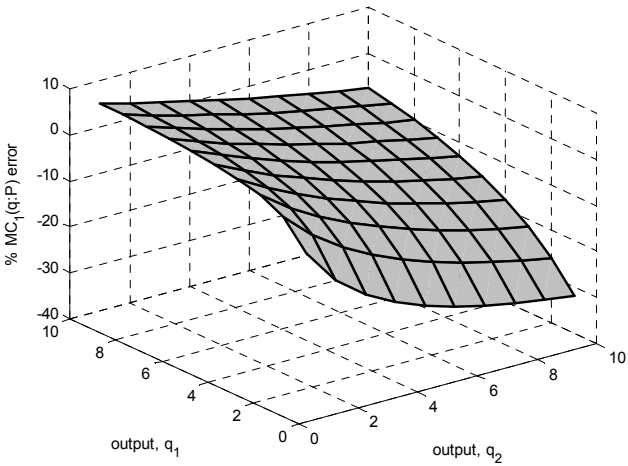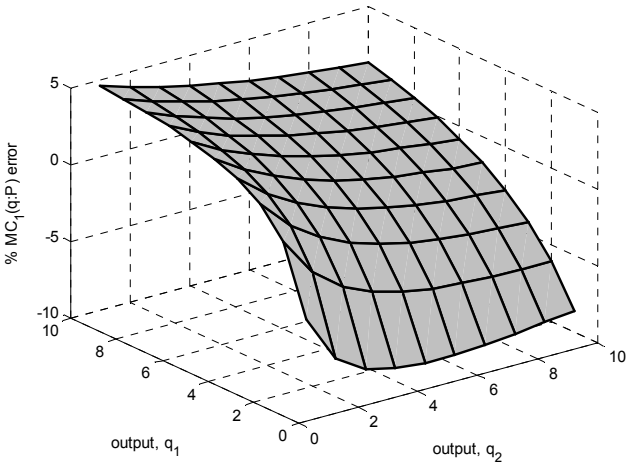
FIGURE 7.11. Impressionism Error for First Product, $\alpha = .55, \beta = .45$
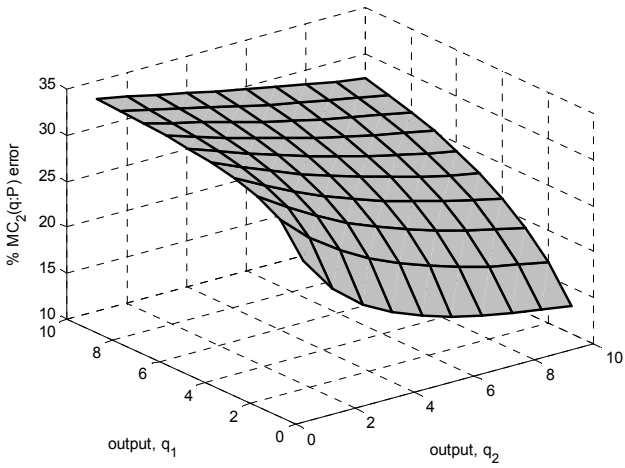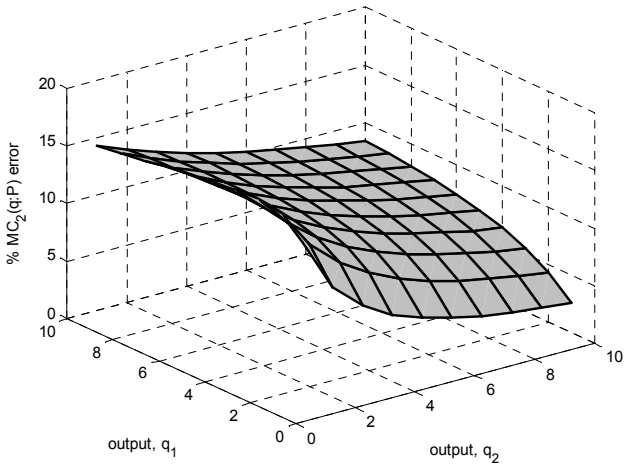

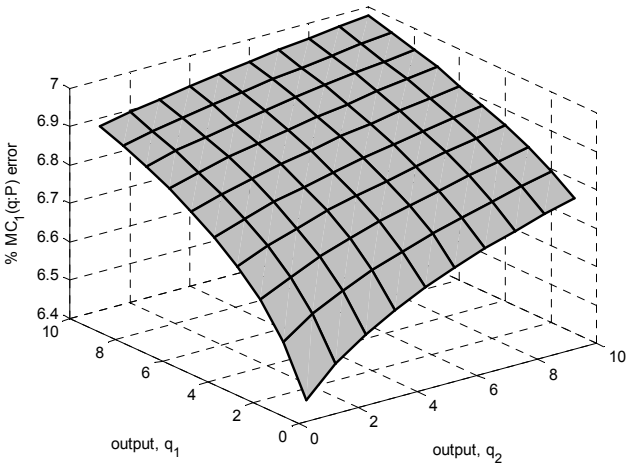FIGURE 7.12. Impressionism Error for Second Product, $\alpha = .55, \beta = .45$
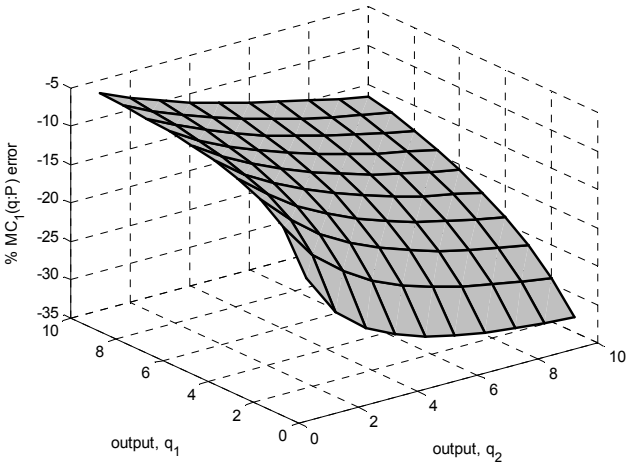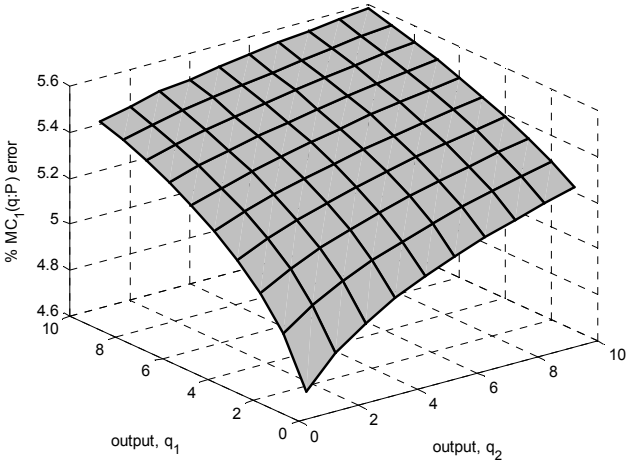
FIGURE 7.13. *ABC* Error for First Product, $\alpha = .55, \beta = .45$
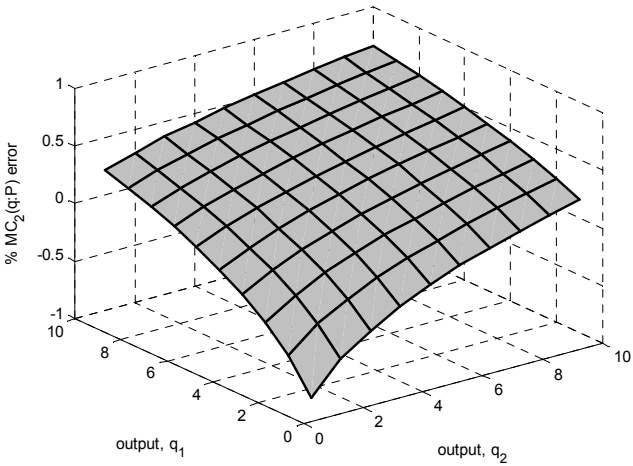
FIGURE 7.14. *ABC* Error for Second Product, $\alpha = .55, \beta = .45$

and anticipate errors in costing art's estimates of marginal cost. This is precisely what Figures 7.3 through 7.14 are all about.

Two facts emerge from this odyssey. One is that we can spend countless hours varying technology, prices, products and the specific rendition of costing art, and continue to wind up with various error patterns. Error is the norm, period. Second is that these errors are distributed across all of the firm's products, including the prior period, the current period, and any future period.

Indeed, a more realistic view of the error possibilities would introduce explicit multiperiod considerations, the possibility of randomly placing some cost items in incorrect periods and pools, and sheer randomness in production itself. Likewise, it would take a less benign view of separability issues than we imbedded in expression (7.3).[15]

This implies we have a portfolio of errors. In relative terms, we should expect winners and losers in the portfolio. Some products will be well-served by the costing art in place while others will not. To think all the errors can be driven to inconsequential amounts is to display a dysfunctional lack of understanding. Likewise, to think the portfolio of errors is uniformly reduced by moving the costing art closer to technology, by going down the modernism road, is simply naive. We can never, in reality, tightly connect to the technology. We remain constrained by presumptive separability, the inherent linearity of costing art, and the requirement that all of the product costs add up at the end of the period.

Indeed, this delightful (at least to me) tension between the impressionism and modernism schools is an example of what economists call the classical theory of the second best. Suppose we are working with a process where several variables must be carefully set in order to attain the optimal configuration. It is then possible that if one or more of these variables is not at its optimal level, the best setting for the other variable may be other than its optimal level. Mistakes or misalignments interact. In the world of product costing, moving but some of the interacting variables closer to the technology may, as we have seen, actually cause a deterioration in the performance of the system.

7.6   Summary

Product costing procedures and the accounting library are fine-tuned to the firm's production environment. We expect a social science library to

---

[15] Glance back at Figures 3.3 and 3.4 where we have a direct cost for one product that varies with both output of a different product and with the direct cost of that different product, simply because of the possibility of substituting factors in the direct category for factors in the indirect category. Also do not forget that lack of interperiod separability implies today's marginal cost depends on tomorrow's anticipated production.

differ from a physical science library; and we should expect the accounting library in an auto assembly factory to differ from the accounting library in a public high school. In each instance we deal with aggregation, *LLA*s and allocation, though the choices vary from firm to firm.

In making these choices the impressionism school stresses a pragmatic, almost "quick and dirty" approach while the modernism school stresses vastly more detail and a closer connection to the technology itself. This leads to a metaphorical horse race between the two schools. Yet the modernism approach offers the possibility of less error in estimating marginal cost along with the possibility of additional sources of error, and thus there is no clear winner in the race. After all, both retain an aggregate cost function approximation that is inherently linear, and both struggle with interperiod considerations.

An apt closing returns us to Sally and Ralph in Example 7.1: The two approaches can lead to significantly different answers, the answers always add up, and we remain perplexed. That is the nature of the game.

## 7.7	Bibliographic Notes

The linkage between the production environment and the accounting library's procedures is well-explored. Kaplan [1973], for example, stresses allocation and marginal cost estimation in a setting of interconnected services while Weil [1968] examines the relationship between economic structure and accounting joint products. Johnson and Kaplan [1987] emphasize connecting accounting procedures to the firm's technology, while Cooper and Kaplan [1991] provide an extensive examination of the modernism school's activity based costing approach. Explicit ties to technology are explored in Noreen [1991], Noreen and Soderstrom [1994, 1997] and Christensen and Demski [1995], where the connection between synthetic variables (or cost drivers) and output measurement is explored. Errors in product cost measurement are explored in Gupta [1993], Hwang, Evans and Hegde [1993] and Christensen and Demski [1997]. Rogerson [1992] and Christensen and Demski [2003a] stress perverse incentive consequences in the world of estimation errors. Implementation difficulties are documented in Anderson and Young [1995] and Anderson, Hesford and Young [2002], and stressed in Kaplan and Anderson [2007]. Our particular emphasis on technology and marginal cost estimation is patterned after Christensen and Demski [1997, 2003]. Lipsey and Lancaster [1956] identify the second best phenomenon.

7.8   Problems and Exercises

1. How much, in Example 7.1, should Sally pay for dinner?

2. Our comparison of the impressionism and modernism schools stressed estimation of marginal cost. Is this an appropriate focus of comparison? Explain.

3. We have stressed the role of presumptive separability in both the impressionism and modernism schools. What is meant by separability, and how does lack of separability lead to error in the unit cost measures?

4. A popular claim is that the modernism school provides more accurate unit cost measures because it concentrates on cause and effect relationships, well thought out synthetic variables, as opposed to the use of broad, arbitrary allocation rates in the impressionism school. Is the claim accurate? Explain.

5. *managerial implications*
   Suppose you encounter a setting such as Example 7.2 where a highly aggregate, impressionism approach is in place and it appears most of the firm's profit is due to one of the product lines. For instance, suppose in the Example that the first product sells for 275 per unit while the second sells for 450 per unit. A consultant performs a more detailed analysis, based on the modernism school, and reports the unit costs in Table 7.4. This dramatically alters the unit costs, and leads to the conclusion the firm is actually losing money on the product line that originally appeared to be the source of its profit. What is an appropriate response in such a setting? Will profit improve if the second product is abandoned or at least de-emphasized?

6. *no difference*
   Given a specific numerical version of the prices and technology in (7.3) such that the impressionism and modernism schools report the same unit costs, and thus perfectly estimate marginal cost.

7. *changing technology*
   Return to the setting of Example 7.3 and Tables 7.7, 7.8 and 7.9. Now parameterize the direct labor cost by $\gamma \cdot 140$ for the first product and $\gamma \cdot 90$ for the second product. (So $\gamma = 1$ is the original story.)

   (a) Determine each product's marginal cost, again given $q_1 = 7$ and $q_2 = 9$, as γ varies from $\gamma = .25$ to $\gamma = 5$.

   (b) Determine each product's unit cost, via the impressionism school, as γ varies from $\gamma = .25$ to $\gamma = 5$.

(c) Determine each product's unit cost, via the modernism school, as γ varies from $\gamma = .25$ to $\gamma = 5$.

(d) What happens to the respective errors in estimating marginal cost as γ varies from $\gamma = .25$ to $\gamma = 5$? Explain.

8. *unit versus marginal cost*
Return to Example 7.3 and focus on the constant returns to scale case of $\alpha = \beta = .5$. Determine total cost, marginal cost of each product, unit cost of each product under the impressionism school, and unit cost of each product under the modernism school for each of the following output pairs: $q = [3, 8], q = [8, 3]$ and $q = [10, 10]$. Also determine the corresponding percentage estimation errors.

9. *unit versus marginal cost*
Repeat the above for the mixed case of $\alpha = .55$ and $\beta = .45$.

10. *unit versus marginal cost*
Repeat the above for the mixed case of $\alpha = .45$ and $\beta = .55$. Explain whatever differences you see in the error patterns here versus in the above 2 exercises.

11. *unit versus marginal cost*
Return to problem 3-14, but now assume Ralph wants to produce $q_1 = 100$ and $q_2 = 200$.

(a) Determine Ralph's best combination of capital and labor, as well as the total cost of production.

(b) Determine the marginal cost of each product (given the above output schedule).

(c) Provide an accounting method such that the resulting unit cost of the first product well approximates its marginal cost. For this purpose, notice three "cost pools" are present: direct cost for the first product (i.e., $150L_1$), direct cost for the second product (i.e., $175L_2$) and an indirect, product cost pool, or overhead (i.e., $100K$). Does the second product's unit cost provide a reasonable estimate of that product's marginal cost? Explain.

12. *shadow prices*
Return to Example 7.3. Determine, using program (7.3), the optimal factor consumptions to produce the noted output of $q = [7, 9]$. Also determine the shadow prices for each constraint, and use them to verify the respective marginal costs of 47.60 and 33.24. Also verify the unit cost calculations for both schools.

13. *shadow prices*
This is a continuation of problem 12 above. Using your shadow prices

and factor choices, verify that all of the constraints in (7.3) are satisfied, as well as all of the first order conditions displayed in note 13.

14. *second best*
    Suppose we want to maximize $f(x, y, z) = 6x - x^2 + 9y - y^2 + 5z - z^2 - \theta xyz$, subject to $x, y, z \geq 0$.

    (a) Let $\theta = 1$. Determine the values of x, y and z that maximize $f(x, y, z)$.

    (b) Now suppose $x = 3, y = 1$ and $z = 0$. Is it an improvement to move z from $z = 0$ to $z = 1$, holding the other two variables constant? Why is it useful to move z away from its globally optimal setting?

    (c) Now suppose $x = 1, y = 4.5$ and $z = 2$. Is it an improvement to move x from $x = 1$ to $x = 0$.

    (d) Repeat (a) (b) and (c) above for the case of $\theta = 0$.

    (e) Write a short paragraph detailing your findings and what this implies about connecting the costing system closer to the firm's technology.

15. *lots of details*
    Numbing Ralph is a three product firm, complete with lots of details. The products vary in terms of direct labor and direct material requirements, but also in terms of how many units are manufactured in a given batch (which determines the number of costly setups), their "complexity," and in terms of their material handling transactions. Direct labor (DL) and direct material (DM) costs are displayed below, along with the setup, complexity and handling measures.

	q_1	q_2	q_3
direct labor (DL)	40	100	240
direct material (DM)	18	250	480
setups (S)	$q_1/500$	$q_2/500$	$q_3/400$
complexity units (U)	1	1	2
handling transactions (T)	12	5	20

    In turn, four overhead pools are present. Their *LLA*s are given by: $OV_1 = 750,000 + 0.4DL, OV_2 = 200,000 + 0.2DM + 1.5T, OV_3 = 750,000 + 0.0U$, and $OV_4 = 1,500S$. OV_1 includes the various direct labor-related costs, such as fringe benefits and supervision. OV_2 contains various direct-material related costs, including purchasing, receiving, inventory control and material handling. Notice the synthetic variables are direct material cost and an index of the number of

material transactions, T. OV_3 contains various product and process engineering costs. These costs are thought to be related to product complexity, as measured by the above noted complexity units tally. And OV_4 collects various setup costs.

(a) Suppose $q_1 = 2,500$, $q_2 = 2,500$ and $q_3 = 2,400$ units are produced, i.e., $q = [2{,}500,\ 2{,}500,\ 2{,}400]$, and costs turn out to be precisely as detailed above, and thus total $5,342,550$. Determine the unit cost of each product, assuming the overhead pools are aggregated and allocated on the basis of direct labor cost.

(b) Repeat using separate overhead pools and the noted synthetic variables. For the OV_2 pool where two such variables are present, allocate the pool on the basis of variable cost in that pool, i.e., $0.2DM + 1.5T$

(c) Repeat (a) and (b) for
$q = [2{,}500,\ 500,\ 2400]$,
$q = [5{,}000,\ 500,\ 5{,}200]$ and
$q = [2{,}500,\ 0,\ 0]$.

(d) We implicitly used a full costing approach in the above. What, using the modernism approach, is the variable cost per unit?

8
Consistent Decision Framing

Our focus now turns to the topic of managerial behavior. Studying managerial uses of accounting information requires we combine accounting and managerial behavior. This, in turn, demands we adopt some image, implicit or explicit, of managerial behavior. Just as we traced costing art's foundations to the economic theory of cost, we now trace decision making art's foundation to economic rationality.

We begin with a brief review of economic rationality and its central idea of consistent behavior. From there we examine the important topic of framing a decision. Framing refers to the description of a decision problem that we construct. It is a description or representation. It is also personal. We construct it. The framing exercise is also an application of managerial art. The gifted manager can balance detail and abstraction, the quantitative and the qualitative, inclusion and exclusion in describing or framing a decision problem.

Framing is important for a variety of reasons. We will see, for example, that various frames or descriptions call for various measures of cost and benefit. For example, in the preceding chapters we were careful to introduce the firm's problem as simultaneously selecting factors and products to maximize its profit, resulting in a problem statement that contained no measure of cost. We then decomposed the problem into first determining the firm's cost function and second selecting its outputs to maximize revenue less cost, resulting in a problem statement that contained an explicit measure of cost.

This little adventure, however, is only the beginning. There are countless ways to frame a decision, and each leads to a different measure of cost. What we mean by the cost of something is highly contextual. In fact, what we mean by the term cost in a decision setting depends on the economic forces at play in that setting and on the manner in which we have framed the decision. Many find this awkward and unintuitive (if not outright false). Yet cost is the very glue that connects the explicit and the implicit consumption of resources in a decision frame.[1] Drawing the line between what is explicit and implicit at different places leaves us with different measures of cost.

Initially, as noted, we review the tenets of economic rationality. We then present and examine three principles of consistent framing. In the following chapters we relate these framing principles to uncertainty and to various cost constructs. This provides an intimate connection between accounting and decision making.

Keep in mind that this is our initial foray into these matters. Important questions of framing uncertainty and strategic considerations, and of framing in a way that control considerations are included, are all deferred for the moment, as are explicit connections to the accounting library.

8.1   Economic Rationality

The label of rational behavior calls to mind someone who is intelligent, wise, and enlightened. In an economic setting this colloquialism is often refined to someone who pursues self-interest and wealth with an unrelenting, even unhealthy vigor. Yet, as often happens, there is both more and less to the popular conception.

The underlying idea is straightforward. Suppose we face the problem of selecting one choice from an available set of alternatives.[2] Think of the alternatives as contained in set A. Also denote one such alternative by a. Choice is confined to some $a \in A$. This is the first half of the setup. We have an exogenously specified problem of selecting one element from set A.

The second half of the setup introduces a criterion function, or preference measure. In particular, we presume the individual's choice behavior can

---

[1] For example, our ubiquitous marginal cost expression, $MC_i(q; P)$, rests on a frame where we are simultaneously and explicitly analyzing all of the firm's products. Likewise, in managing inventories, as in a retail setting, we worry about the cost of stockouts. This cost of being unable to meet a current customer's demand is simply a way of using cost to surrogate for an extended analysis and description of what happens when the customer's demand cannot be satisfied.

[2] We explore the tenets of economic rationality in terms of an individual facing a well-defined choice problem. Naturally this holds for any economic entity, including the behavior of a firm.

be described as though some criterion function is present, denoted, $\omega(a)$, and the best choice is the available one that produces the largest value of the criterion function. Symbolically, we have the following portrayal of the individual's choice process:

$$\max \omega(a) \qquad (8.1)$$
$$\text{s.t. } a \in A$$

Don't pass over the subtlety. The assumption is not that the individual has or uses such a function. It is that the individual's behavior can be described *as though* such a function were present and used in the noted fashion. The assumption is that choice behavior can be described by, can be modeled by, can be represented by, maximization of some function.[3]

Studying household behavior with a budget line and indifference curves is a case in point. Examine Figure 8.1, where we deal with the best feasible choice of two goods with respective quantities denoted q_1 and q_2. The straight line depicts the budget line. Any combination on or below the line is feasible (provided $q_1 \geq 0$ and $q_2 \geq 0$). The curved lines are indifference curves. The individual is indifferent among all combinations of q_1 and q_2 on any such curve. Moving in the northeast direction is preferred. Thus, consumption of any combination on the lower indifference curve can be improved. Better choices are available. Consumption of any combination on the highest indifference curve would be nice, but no such combination is feasible. This indifference curve is uniformly above the budget line. The middle curve is critical. It is tangent to the budget line. Anything on a lower indifference curve can be improved, while anything on a higher indifference curve is infeasible.

Nothing is said here about whether the individual walks around with these indifference curves. Rather, the story is one of describing the individual's behavior in terms of indifference curves and a budget line. The individual can identify combinations of q_1 and q_2 that are equivalent in terms of preference, and can identify combinations that are strictly better or worse, in terms of preference. This story is equivalent to one in which the individual's tastes are represented by a criterion function, or utility function.

Lest we doubt, let $a = (q_1, q_2)$ denote a particular choice of q_1 and q_2, or consumption bundle. The story in Figure 8.1 was generated with a budget line of $q_1 + q_2 = 10$ (reflecting equal prices of one dollar per unit and a budget of 10), and a utility function of $\omega(q_1, q_2) = q_1 q_2$. The tangency

---

[3]The individual thus comes equipped with considerable skill and self-insight. The fact a decision opportunity is present is known, and all of the alternatives have been identified. Set A is exogenously specified; and the individual behaves as though these alternatives are evaluated with the $\omega(a)$ function.
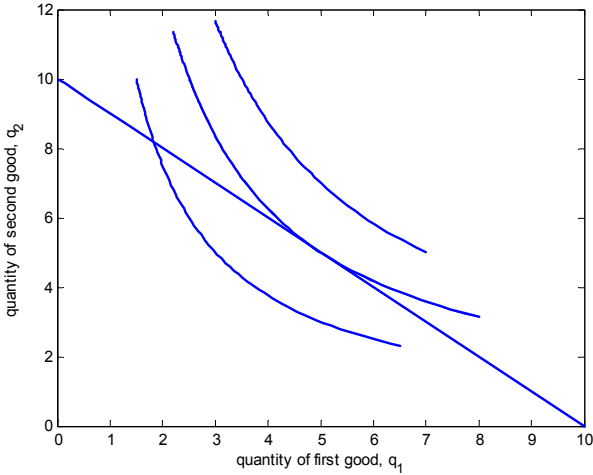
FIGURE 8.1. Budget Line and Indifference Curves

point is located by maximizing $\omega(q_1, q_2) = q_1 q_2$ subject to (1) $q_1 \geq 0$; (2) $q_2 \geq 0$; and (3) $q_1 + q_2 \leq 10$. The solution is $q_1^* = q_2^* = 5$.[4]

The idea of economic rationality, then, is that preferences are so well defined they can be described by a criterion function, yes a utility function. There is no claim, no requirement, that the individual possess and literally uses a utility function. The claim is the individual's behavior is so well defined that it can be described *as though* such a function were present. This leads to the question of what it means for behavior to be so well defined.

8.1.1   Consistency

The central feature here is consistency. Suppose we must select from some set $A = \{w, x, y, z\}$. Further suppose that we rank the choices in the order of w, x, y and z. Notice two things. Our ranking is complete. Take any two options from A. Either we are indifferent (e.g., w is as good as itself, w) or one is better than the other (e.g., w is better than z). Our ranking is also transitive. For example, w ranks above x and x ranks above y; and then w ranks above y.

Complete and transitive ranking is the hallmark of consistency. If our ranking is not complete, we are saying there are some comparisons that we find confusing; we cannot choose between them. If our rankings are

---

[4]To connect this to the conceptualization in (8.1), $a = [q_1, q_2], A = \{[q_1, q_2]|q_1 \geq 0, q_2 \geq 0, q_1 + q_2 \leq 10\}$ and $\omega(a) = \omega(q_1, q_2) = q_1 q_2$.

intransitive, we open ourself to foolishness, or worse. Suppose we say z beats w and w beats x and x beats z. Further suppose we currently possess x. w beats x, and we pay a dollar to switch to w. But z beats w and we pay a second dollar to switch to z. Finally, x beats z and we pay a third dollar to switch to x. We are now at the beginning point of the cycle, holding x but less $3. Not good!

It turns out that if the set A is finite, the following two statements are logically equivalent. First, we have a ranking of the elements in A that is complete and transitive. Second, there exists a function on A, say, $\omega(a)$, such that for any two $\grave{a}, \acute{a} \in A, \omega(\grave{a}) \geq \omega(\acute{a})$ only when $\grave{a}$ is ranked as good as $\acute{a}$. In this sense we say a function on set A that represents some ranking of the elements in A exists if (and only if) that ranking is complete and transitive.[5]

This probably strikes you as well beyond anything of interest in the study of accounting. The most important feature of economic behavior, though, is consistent tastes, consistent in the sense they are complete (we know what we like) and transitive (we don't cycle). Emphatically, this does not say greed or self-interest. It says complete and transitive.

8.1.2 Smoothness

You may have noticed Figure 8.1 uses an uncountable set of possible choices, while our digression on existence of a utility function used a finite set of possible choices. This leads to a second, more technical condition. If we are to have a utility function, we must have a complete and transitive ranking. But there are cases where this is not enough. These cases take the form of rankings that are not sufficiently smooth.[6] Both consistency and smoothness are required for existence of a criterion or utility function. In

---

[5] This is a paraphrase of an important result in the theory of measurement, due to Cantor and published over a century ago. That said, we are being a little casual here. Let B be the set of conceivable choices and A be any nonempty subset of B. $\omega(a)$ is everywhere defined; it is a function on B. A particular choice problem then arises when we encounter some nonempty subset of B, the set A. We treat sets A and B as the same sets in our narrative, hoping to convey the central idea without burdening the discussion with details better reserved for a thorough inquiry.

[6] A lexicographic ordering is a case in point. Let $a = [x, y]$, with x and y denoting quantities of two goods. Let set A be any combination of x and y with $0 \leq x \leq 1$ and $0 \leq y \leq 1$. Suppose in comparing a and a' you always look to the first good. Take the option with the largest amount of the first good. If a tie is present, take the one with the largest amount of the second good. Notice how the second good is important only when the consumption bundles have the same amount of the first good. Though these preferences are complete and transitive, no utility function exists. The preferences are not sufficiently smooth. The technical requirement is we be able to find a subset of A that is both dense (in the sense we can use it to bracket the other elements) and countable (a trivial issue when A is finite). This is hardly intuitive, so we just invoke the requirement of smoothness.

general terms, then, choice behavior can be described in terms of maximizing a criterion function when the underlying preferences are consistent and smooth.

Return to our initial characterization of choice behavior in expression (8.1). This characterization amounts to an assumption that (1) the individual has identified a choice that must be made; (2) has identified the feasible options, i.e., set A; and (3) brings consistent and smooth preferences to the exercise. This allows us to describe his behavior as selecting the best option in A, where best means the one that maximizes the criterion or utility score. Consistency is the critical feature. We require the individual not cycle (i.e., be transitive) and not be confused (i.e., be complete).[7] Annoying pathological cases are ruled out by the added requirement that these preferences be smooth.

8.1.3 *Consistent Framing*

Whenever we describe a decision problem, implicitly or explicitly, a framing exercise has been engaged. Whether to introduce a new product, whether to study this chapter seriously, what to eat for dinner, how to boost the morale of our work group all imply some framing of a decision problem.

We have, as noted, already experienced this framing exercise. In Chapters 2 and 3 we encountered competing frames for determining the firm's production plan. One frame focused on input and output prices, and technology. Another focused on output prices (or revenue) and cost. The cost function, $C(q; P)$, was central in this latter frame. That frame highlighted the revenue and cost of both products.

Again using our formalism of a choice problem in expression (8.1), suppose a^* is a solution to this problem. This means a^* is feasible; it is among the listed alternatives. $a^* \in A$. This also means $\omega(a^*) \geq \omega(a)$ for every $a \in A$. Stated succinctly, a^* is optimal.[8] It is the solution to an optimization problem.

Consistent framing is now easily introduced. It refers to ways to transform this optimization problem, but always so an optimal solution is identified. This is why we speak of consistent framing. There are countless

---

[7] Would we want a manager who is confused and known to cycle through the available options?

[8] The solution need not be unique. We may have more than one element of A that produces the maximal value of $\omega(a)$. This is why we used the phrase of "a solution" rather than "the solution." Also, there is no guarantee a solution even exists. For example, what is the solution to: maximize $\omega(a) = a$, subject to $0 \leq a < 1$. If you claim some feasible a is optimal, we can always retort by suggesting that you try $\widehat{a} = a + (1-a)/2$. But $\widehat{a} > a$, and $\widehat{a} < 1$! The point is not profound. Care should be exercised in making certain an optimization problem actually has a solution. Our discussion always presumes a solution exists. In turn, when presenting concrete optimization problems we are careful to make certain they are sufficiently well crafted to have a solution.

ways to describe or transform an optimization problem, without losing our ability to locate a solution. They are consistent in the sense they lead to an optimal choice.

It turns out there are three principles at work here, what we call the three principles of consistent framing. They are discussed in turn.

8.2   Irrelevance of Increasing Transformations

The first principle of consistent framing addresses the ability to transform the criterion function, $\omega(a)$. What is the best choice of a for the following

$$\max \omega(a) = a$$
$$\text{s.t. } 0 \leq a \leq 1$$

Trivially, the answer is $a^* = 1$.[9] We can do no better than set $a = 1$

In contrast, what is the solution to

$$\max \widehat{\omega}(a) = 50 + a$$
$$\text{s.t. } 0 \leq a \leq 1$$

Surely the answer is also $a^* = 1$. The former achieves a maximum of $\omega(a^*) = a^* = 1$, while the latter achieves a maximum of $\widehat{\omega}(a^*) = 50 + a^* = 51$. But the choice of a remains the same. Adding the constant, 50, to $\omega(a)$ does not affect the choice of $a \in A$. It simply shifts the criterion function by a constant amount, keeping every choice of a in the same relative position.

Suppose α is an arbitrary constant. Does it matter whether we maximize $\omega(a)$ or $\alpha + \omega(a)$? Suppose $\beta > 0$ is an arbitrary though strictly positive constant. Does it matter whether we maximize $\omega(a)$ or $\beta \cdot \omega(a)$? Surely not. Given set A, maximizing $\omega(a)$ and maximizing $\alpha + \beta \cdot \omega(a)$ will identify the same $a^* \in A$, for any α (whether positive or negative) and any $\beta > 0$. We use this simple idea so often its use usually goes unnoticed.

Examine Figure 8.2. Four functions are plotted over the range $3 \leq a \leq 7$. The maximum occurs in each case at the point $a^* = 5$. Irrespective of the given function used to evaluate the choice of a (respecting $3 \leq a \leq 7$), we always locate $a^* = 5$. Is this an accident?

The four functions were constructed as follows:

graph 1: $\omega_1(a) = 10a - a^2 - 20$;
graph 2: $\omega_2(a) = \omega_1(a) + 20 = 10a - a^2$;
graph 3: $\omega_3(a) = 1 + [\omega_1(a)]^2 = 1 + [10a - a^2 - 20]^2$; and
graph 4: $\omega_4(a) = \ln[\omega_2(a)] = \ln[\omega_1(a) + 20]$

---

[9]Is there a conflict with what was said in the prior footnote? $0 \leq a < 1$ and $0 \leq a \leq 1$ are different intervals. The former allows us to get arbitrarily close to $a = 1$, but never achieve $a = 1$. The latter allows us to achieve $a = 1$. Further observe that we have a unique solution in this case.
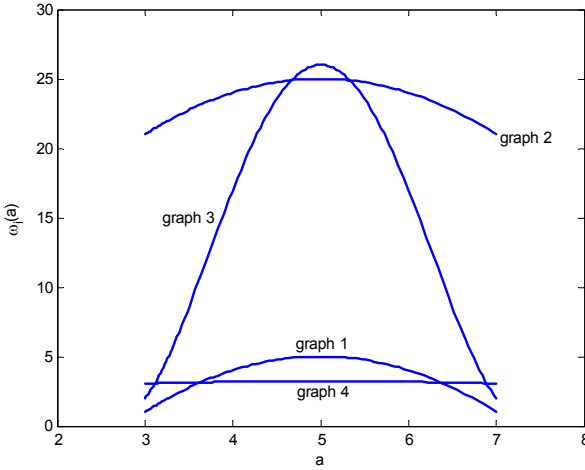
FIGURE 8.2. Increasing Transformations

Representative points are also displayed in Table 8.1.

<div>
TABLE 8.1: Various Functions Defined on $3 \leq a \leq 7$

a	$\omega_1(a)$	$\omega_2(a)$	$\omega_3(a)$	$\omega_4(a)$
3	1	21	2	3.0445
4	4	24	17	3.1781
5	5	25	26	3.2189
6	4	24	17	3.1781
7	1	21	2	3.0445
</div>

Graphs 2, 3, and 4 are all judiciously chosen transformations of $\omega_1(a)$. Each transformation is chosen in a way that leaves the point at which the function reaches a maximum undisturbed.[10]   On the other hand, graph 3 is visually helpful, while graph 4 is close to opaque. This suggests some transformations are more helpful than others.

But how much freedom do we have in this transformation game? The explanation is somewhat technical but we should not attribute it to magic.

---

[10] Notice the derivative of each function passes through zero at the same point, $a = 5$:
$\omega_1'(a) = 10 - 2a$;
$\omega_2'(a) = 10 - 2a$;
$\omega_3'(a) = [2\omega_1(a)][10 - 2a]$; and
$\omega_4'(a) = [10 - 2a]/\omega_2(a)$.

The key is what is called an increasing transformation, a transformation that always preserves order. If one item is larger than another before transformation, it must remain larger after transformation. For example, $a > b$ if and only if $a^3 > b^3$. Similarly, $a > b$ if and only if $2.45a - 20 > 2.45b - 20$. Also, if $a \geq 0$ and $b \geq 0$, $a > b$ if and only if $\sqrt{a} > \sqrt{b}$.

Definition 20 *Function T is an increasing transformation of function $\omega(a)$ if $\omega(a) > \omega(\widehat{a})$ if and only if $T[\omega(a)] > T[\omega(\widehat{a})]$ for every a and $\widehat{a}$ in the domain of the original function.*[11]

Return to Figure 8.1 and consider $a = 5.5$ and $\widehat{a} = 4$. For the first graph we find

$$\omega_1(5.5) = 10(5.5) - (5.5)^2 - 20 = 4.75 > \omega_1(4) = 10(4) - (4)^2 - 20 = 4$$

For the second we have

$$\omega_2(5.5) = 20 + 4.75 = 24.75 > \omega_2(4) = 20 + 4 = 24$$

For the third we find

$$\omega_3(5.5) = 1 + (4.75)^2 = 23.5625 > \omega_3(4) = 1 + (4)^2 = 17$$

Finally, for the fourth function we have

$$\omega_4(5.5) = \ln(24.75) = 3.2088 > \omega_4(4) = \ln(24) = 3.1781$$

In each instance we find $\omega_i(a) > \omega_i(\widehat{a})$. Continuing, for any feasible choices of $3 \leq a \leq 7$ and $3 \leq \widehat{a} \leq 7$, $\omega_1(a) > \omega_1(\widehat{a})$ is logically equivalent to the claim $\omega_i(a) > \omega_i(\widehat{a})$ for any of the noted transformations. This is apparent in Figure 8.1. One of the functions increases from one point to another if and only if all the other functions do as well.

The following fact emerges. Let a^* maximize $\omega(a)$ subject to $a \in A$. Also let T be an increasing transformation. Then a^* also maximizes $T[\omega(a)]$ subject to $a \in A$.[12] This is the first principle of consistent framing.

This fact is useful. Judicious use of increasing transformations may simplify what we are looking at. Add the constant 20 to $\omega_1(a) = 10a - a^2 - 20$. This is surely an increasing transformation. It simply removes an irrelevant constant from view. We do this every time we ignore fixed cost in a short-run maximization problem.

---

[11] The domain of $\omega(a)$ is the set of points over which the function is defined. In our example in Figure 8.1, the domain of the function is $3 \leq a \leq 7$. This is apparent from looking at the horizontal axis in Figure 8.1.

[12] Suppose a^* maximizes $\omega(a)$ subject to $a \in A$. Let T be an increasing transformation. Suppose a^* does not maximize $T[\omega(a)]$ subject to $a \in A$. We then have some $\widehat{a} \in A$ such that $T[\omega(\widehat{a})] > T[\omega(a^*)]$. Since T is an increasing transformation, this means $\omega(\widehat{a}) > \omega(a^*)$. And this implies a^* cannot maximize $\omega(a)$ subject to $a \in A$. Contradiction.

Of course, some art is involved here as well. Multiplying $\omega(a)$ by .2 does no harm, but it doesn't appear particularly useful. Cubing $\omega(z)$ does no harm, but it certainly appears noxious. Graphs 1 and 2 in Figure 8.1 do no harm. Graph 3 provides a more apparent picture. Graph 4 clouds the picture.

Example 8.1 Suppose we have a single product firm, currently producing $q = 100$ units that sell for $\widehat{P} = 10$. Cost is given by $C(q; P) = 8q$. A special customer has surfaced and offers to purchase two units by paying 9 per unit. Profit without the customer totals $10(100) - 8(100) = 200$, while profit with the special customer included would be $10(100) + 9(2) - 8(102) = 202$. Alternatively, frame the question in incremental terms. Incremental revenue would be $9(2) = 18$, while incremental cost would be $8(2) = 16$, implying an incremental profit of $2 = 202 - 200$.

When we focus on the difference in profit all we do is subtract the status quo profit. (Likewise, revisit our definition of incremental cost, e.g., $C(q + \Delta; P) - C(q; P)$ in the single product case in Chapter 2.) This is an increasing transformation. Similarly, ignoring fixed cost in a short-run setting simply amounts to transforming the profit function by adding a constant equal to the fixed cost. This simple transformation is illustrated by $\omega_2(a)$ (i.e., graph 2) in Figure 8.1, but you should also glance back at Example 2.8.

The first principle of consistent framing is that optimization problems are unaffected by increasing transformations. Simple transformations, adding a constant or multiplying by a strictly positive constant, often give a more friendly appearance. More deeply, these are but particular classes of increasing transformations.[13]

8.3   Local Searches are Possible

The second principle of consistent framing addresses our ability to search in smaller regions of set A for the maximizer of $\omega(a)$. The classic example is where a search committee sorts among numerous dean candidates and then submits the best three to the university president. This amounts to selecting the candidate from a pre-screened set of alternatives. Naturally, the trick is to make certain we have not erred in the pre-screening.

---

[13]It is important to remember we are transforming $\omega(a)$, and not its individual components. This admonition will become apparent when we introduce uncertainty. There we will discover transforming a choice problem is a fairly difficult problem unless risk neutrality is present. You should also notice the transformation applies to the range of $\omega(a)$, to all points generated by $\omega(a), a \in A$. In Figure 8.1, for example, we are unconcerned with any points generated by $\omega(a)$ with a outside of $3 \leq a \leq 7$.

Return to the exercise in Figure 8.1. We want to maximize $\omega_2(a) = 10a - a^2$ subject to $3 \leq a \leq 7$. (Notice how we have dropped the irrelevant constant!) This calls for us to search over a values between 3 and 7. Suppose instead we search over a smaller domain, say, by maximizing $\omega_2(a) = 10a - a^2$ subject to $4 \leq a \leq 7$. Consulting Figure 8.1 should convince us that the maximum occurs at $a^* = 5$, and our limited search has done no harm.

In a sense we have analyzed a smaller problem here. Our search was confined to the smaller region of $4 \leq a \leq 7$. We nevertheless located an optimal solution to the original, "larger" problem. No harm was done. But how do we convince ourselves no harm was done?

The answer is simple. Take the best of what we ignored, and test it against what we found. We ignored the alternatives in $3 \leq a \leq 4$. What is the best choice from among the alternatives we want to ignore? What is the maximum of $\omega_2(a) = 10a - a^2$ subject to $3 \leq a \leq 4$? As is obvious from Figure 8.1, the maximum over this limited range occurs at $a = 4$, and provides $\omega_2(4) = 24$.[14]

Our search over $4 \leq a \leq 7$ located $a^* = 5$, with $\omega_2(5) = 25$. In locating $a^* = 5$ we did not search all the alternatives. None of the choices we did not explicitly examine is better than $a = 4$, with $\omega_2(4) = 24$. Clearly, the best choice overall is $a^* = 5$. It's the best we found, and it beats everything in the subset we did not examine.

This illustrates the second principle of consistent framing. Suppose we tentatively select the best choice from a reduced or pre-screened set of alternatives. This tentative choice is best overall if it is better than the best of those not considered.

The terminology of opportunity cost is used to convey this principle. Suppose we face the problem of selecting the best action from some set of available actions. As usual, we invoke expression (8.1) and portray this task as maximizing $\omega(a)$ subject to $a \in A$.

Now take this set of available actions and divide or split it into two parts. Call the two parts A_1 and A_2. For example, divide the interval $3 \leq a \leq 7$ into (1) $3 \leq a \leq 4$ and (2) $4 \leq a \leq 7$.[15] Also, denote the best choice from set A_1 by a_1^*, with associated criterion function value $\omega(a_1^*)$. Similarly, denote the best choice from set A_2 by a_2^*, with associated criterion function value

---

[14] The point $a = 4$ is contained in both regions. This was done to avoid dealing with a problem formulated as maximize $\omega_2(a)$ subject to $3 \leq a < 4$. See the following note.

[15] Thus, the combination of the two intervals returns us to the original specification of $3 \leq a \leq 7$. In set terminology, we have $A_1 \cup A_2 = A$. It also might seem logical to make these two subsets disjoint. For example, why not use (1) $3 \leq a < 4$; and (2) $4 \leq a \leq 7$? The answer is that it is now awkward to talk about the maximum of $\omega_2(a)$ over the first region. Of course, we might avoid this by cleverly using (1) $3 \leq a \leq 4$; and (2) $4 < a \leq 7$. It seems easier just to allow $a = 4$ to be included in both subsets in this instance.

$\omega(a_2^*)$. Put differently, let a_1^* be a solution to maximize $\omega(a)$ subject to $a \in A_1$. Also let a_2^* be a solution to maximize $\omega(a)$ subject to $a \in A_2$. We now have a definition of opportunity cost.[16]

Definition 21 *Given the choice problem in (8.1), the opportunity cost of confining our search to subset A_1 is $\omega(a_2^*)$, the best we could do by selecting from among those alternatives in set A_2.*

The second principle of consistent framing is now easily stated. The best choice from set A_1 is best overall if its evaluation exceeds its opportunity cost.[17] a_1^* is the best choice from set A_1. a_2^* is the best choice from set A_2. a_1^* is best overall if $\omega(a_1^*) \geq \omega(a_2^*)$. The opportunity cost of a_1^* is the criterion function's score, the evaluation, of the best alternative among those not explicitly searched. Opportunity cost thus refers to something both foregone and not explicitly considered. It is stated in units of the criterion function, $\omega(a)$. The opportunity cost of selecting the best alternative from among those explicitly considered (those in A_1) is the criterion function's evaluation of the best alternative among those alternatives not explicitly searched (those in A_2).

TABLE 8.2: Opportunity Cost Calculations with
$\omega(v) = 1, \omega(w) = 2, \omega(x) = 3, \omega(y) = 4$, and $\omega(z) = 5$

included set A_1	excluded set A_2	best in A_1, a_1^*	best in A_2, a_2^*	$\omega(a_1^*)$	$\omega(a_2^*)$
$\{v, x, y\}$	$\{w, z\}$	y	z	4	5
$\{v, w, z\}$	$\{x, y\}$	z	y	5	4
$\{w, z\}$	$\{v, x, y\}$	z	y	5	4
$\{v, y, z\}$	$\{w, x\}$	z	x	5	3
$\{w, y, z\}$	$\{v, x\}$	z	x	5	3
$\{w, x, y, z\}$	$\{v\}$	z	v	5	1
$\{v, w, x, y, z\}$	null	z	N/A	5	N/A
$\{v\}$	$\{w, x, y, z\}$	v	z	1	5
$\{w\}$	$\{v, x, y, z\}$	w	z	2	5
$\{x\}$	$\{v, w, y, z\}$	x	z	3	5
$\{y\}$	$\{v, w, x, z\}$	y	z	4	5
$\{z\}$	$\{v, w, x, y\}$	z	y	5	4

---

[16] The definition requires A_1 and A_2 to be subsets of A, and to have their union equal A: $A_1 \subset A, A_2 \subset A$, and $A_1 \cup A_2 = A$.
[17] If $\omega(a_1^*) > \omega(a_2^*)$, a_1^* is best overall. If $\omega(a_1^*) < \omega(a_2^*)$, a_2^* is best overall. If $\omega(a_1^*) = \omega(a_2^*)$, both a_1^* and a_2^* are best overall.

Example 8.2 Suppose we have five alternatives. Describe them by set $A = \{v, w, x, y, z\}$. Further suppose the evaluation function is: $\omega(v) = 1, \omega(w) = 2, \omega(x) = 3, \omega(y) = 4$, and $\omega(z) = 5$. So $z \in A$ is the best choice.

Some possible ways to define the pre-screened or "included" set A_1 are noted in Table 8.2. Set A_2, the "excluded" set, contains all elements not in set A_1. In each instance, the opportunity cost of the best choice from A_1 is the evaluation of the best choice among those options excluded. Opportunity cost is used to control for those options not considered. If all options are in A_1, there is no opportunity cost. Be certain to verify the $\omega(a)$ constructions in Table 8.2.

It is important to understand opportunity cost depends on what is excluded from primary, explicit consideration. To dramatize, suppose we include option z in set A_1. Then we always have $a_1^* = z$. What is the opportunity cost of searching in set A_1? Symbolically, it is $\omega(a_2^*)$. But $\omega(a_2^*)$ might be $1, 2, 3$, or 4.[18] It all depends on what is excluded from A_1.

8.3.1   The Economist's Approach

Contrast this definition of opportunity cost with the colloquialism that opportunity cost refers to "what could have been achieved had a particular decision not been taken." Under this usage, opportunity cost refers to the best alternative foregone, the sacrifice associated with a particular action. For example, attending class commits your time to class, and thereby forecloses alternative use of that time.

To link this to our definition, return to the example in Table 8.2. Suppose the included set A_1 contains a single option. As usual, A_2 contains everything else. Let A_1 contain only the first option, v. What is the opportunity cost of searching only in A_1? It is the evaluation of the best among those in A_2. Clearly this is choice z; the opportunity cost is $\omega(z) = 5$.

In this sense, the colloquialism comports with our definition. If A_1 is a single choice, the opportunity cost refers to the best alternative foregone. This is because opportunity cost in our approach refers to the best alternative not considered. If only one option is considered, all others are not. The best of all others, the best alternative foregone, then gives us the opportunity cost. Equivalently, if (and when) we select a specific action, we have foregone all other possibilities, so the implicit sacrifice is best of those foregone, measured by the criterion function.[19]

---

[18] It could also be 5 if we happened to include z in both A_1 and A_2!

[19] Suppose A_1 contains a single option. A_2 then contains all others. Now reverse the two sets. This is equivalent to selecting among all but the one option, and then comparing the tentative selection with the single one excluded. If we search by placing a single element in A_1, the best choice – when it is the single element in A_1 – will be the one with the minimum opportunity cost. This is what Coase [1968, page 118] means when he says, "To cover costs and to maximize profits are essentially two ways of expressing

Opportunity cost arises to control for opportunities that have not entered the formal analysis. As such, it is an important framing device. A good manager will, among other things, be good at specifying set A_1. Intuition, common sense and experience are important inputs to this pre-screening or identification exercise. In the end, though, we always ask whether something intriguing was left out of the analysis. In a formal sense, this is the concept of opportunity cost. It naturally depends on the problem we face and on how we pre-screen the available choices for purposes of analysis. It is also measured in terms of the criterion function we are trying to optimize.

Thus, the second principle of consistent framing allows us to confine our search for the best choice to a reduced set of alternatives. Local searches are possible. We control for the remaining options by comparing the tentative choice with its opportunity cost. In turn, this process of "divide and conquer" should be thought of as an application of managerial art. Knowing which options to consider seriously serves to pre-screen the task. Judgment is essential. In a technical sense, we envision the manager as (subjectively) assessing the opportunity cost and proceeding with the analysis.[20]

Do you see a connection to shadow prices?

8.3.2 Shadow Prices

Consider the following exercise, where any action has two components, denoted $a = [x, y]$, and the set A is defined by a set of constraints.

$$\max_{x \geq 0, y \geq 0} \omega(x, y) \quad = \quad 10x + 12y \qquad (8.2)$$
$$\text{s.t.} \qquad x + y \leq 8$$
$$x + 2y \leq 12$$

The optimal solution, as reported by a typical software package, is $x^* = 4, y^* = 4$ and $\omega(x^*, y^*) = 88$, along with respective shadow prices for the two constraints of 8 and 2.

Now recall the shadow prices report the rate at which the optimal objective (or criterion) function will change as we alter the constraint in question. For example, what will the solution be if we change the first constraint from $x + y \leq 8$ to $x + y \leq 9$? The answer is $x^* = 6, y^* = 3$, and $\omega(x^*, y^*) = 96$. So we have improved our lot from 88 to 96. Notice that $96 - 88 = 8$. The increase in the objective or criterion function from 88 to 96 is no accident. It is the change in the constraint multiplied by the constraint's shadow price of 8.

_____

the same phenomenon." This citation draws from a reprint of some of Coase's writings on cost measurement, originally published in 1938.

[20] Another interpretation, based on bounded rationality, is satisficing. If the search over A_1 yields a sufficiently attractive (i.e., satisfactory) alternative, the search stops. Otherwise, we look further.

To dig a bit deeper, rewrite the first constraint in (8.2) as $x + y \leq \theta$. The shadow price on this constraint is given by

$$\frac{\partial \omega(x^*, y^*)}{\partial \theta}|_{\theta=8} = \lambda = 8$$

This suggests increasing the first constraint from $x + y \leq 8$ to $x + y \leq 9$ will produce a gain of approximately $\lambda(9-8) = 8(1) = 8$. In fact, since the criterion function and the constraints in (8.2) are all linear, and we are thus dealing with a *linear program*, we know the shadow prices are constant over well-defined regions (and indeed the shadow price on the first constraint when we impose $x + y \leq 9$ continues to be 8).[21]

Importantly, now, the shadow price is a stylized opportunity cost. Where have we searched for our optimal solution? Within the noted constraints. This is set A_1. What does the shadow price tell us? It tells us how the optimal objective function will change as we change the constraint, as we move to include alternatives not in A_1. If we increase the constraint parameter, we add to the list of options. The shadow price speaks to the change in the objective function associated with expanding the options allowed. It provides an indication of returns that are available with options that were excluded from the analysis.

Intuitively, a large shadow price tells us it may be worthwhile to alter the constraint in question, if possible. For example, the $x + y \leq 8$ constraint might refer to units of capacity in a manufacturing department. The shadow price of 8 raises the question of expanding this capacity. Suppose equipment can be leased on a short term basis for less than the shadow price. This suggests we have not yet found the optimal solution. The opportunity cost of confining ourselves to the stated constraints is "too large." Some interesting options remain unexplored.[22]

---

[21] The shadow price in a linear program remains constant as long as the optimal basis does not change. We use here an illustration in which a unit change in the noted constraint leaves the basis unchanged and thus leaves the shadow prices unchanged. Also notice, relative to our discussion of shadow prices in Chapter 2's Appendix, that (8.2) is framed as a maximization as opposed to a minimization exercise.

[22] To complete the story, suppose it is possible to alter the constraint from $x + y \leq 8$ to $x + y \leq 9$, at a cost of C. Think of A_1 as the set of alternatives defined by the original constraints and A_2 as the set of alternatives defined by the perturbed constraints but not in A_1. Then the best choice among the alternatives excluded from the initial formulation is $x^* = 6$ and $y^* = 3$. The associated objective function evaluation is $10(6) + 12(3) - C = 96 - C$. In our language, $96 - C$ is the opportunity cost of searching within the confines of the original constraint in (8.2). In turn, $96 - 88 = 8(1)$ is in this case also equal to the shadow price multiplied by the change in the constraint. The incremental gain from expanding the constraint is $8 - C$. Shadow prices are stylized opportunity costs stated in incremental terms in that they measure the rate at which the criterion function changes with respect to change in the constraint (exclusive of the cost of changing the constraint). Also, we should not lose sight of the fact that the shadow prices are local measures of rates of change. They are constant in this specific

Opportunity cost refers to the best among those choices not considered. Shadow prices report the rate at which the maximal objective function value will change as we change the respective constraints. This informs us about the potential returns to altering our formulation of the problem. Altering the formulation means looking beyond the alternatives allowed by the constraints, as formulated. It means looking outside set A_1.

8.4 Component Searches are Possible

The third principle of consistent framing concerns the ability to reduce the explicit dimensionality of a decision problem. This exploits the idea that it is often easier to work on a problem in sequential format.[23] We have, in fact, used extensively this technique in our study of the firm's cost function. Though the idea is so straightforward we often fail to recognize it when it is used, laying out its bare bones structure is notationally awkward. So we begin with an example.

Example 8.3 Return to Examples 2.5 and 2.6 where we were concerned with a single product firm's cost function. The choice problem we faced was the following:[24]

$$C(q; P) \equiv \min_{z_1 \geq 0, z_2 \geq 0} \omega(z_1, z_2) = 5z_1 + 20z_2 \tag{8.3}$$
$$\text{s.t. } q \leq \sqrt{z_1 z_2}$$
$$z_1 \leq 15$$

where we have inserted the additional notation to remind us the criterion function is $\omega(z_1, z_2) = 5z_1 + 20z_2$. Let's also agree that $q > 0$ (as otherwise the solution is trivial). Now think about the first constraint, $q \leq \sqrt{z_1 z_2}$. If $z = [z_1 z_2]$ is such that $q > \sqrt{z_1 z_2}$, our solution is not feasible; and if it is such that $q < \sqrt{z_1 z_2}$, our solution is hardly optimal because we have an excess supply of factors. Thus we know the solution will entail $q = \sqrt{z_1 z_2}$, or $q^2 = z_1 z_2$. Moreover, with $q > 0$ and $z_1, z_2 \geq 0$ we must have $z_1, z_2 > 0$. But this implies any potential choice of $z_1 > 0$ is matched with

$$z_2 = \frac{q^2}{z_1} \tag{8.4}$$

---

case because of the assumed linearity and because we are not perturbing the constraint an amount that would actually change the shadow price.

[23] This "multidimensionality" theme will resurface in later chapters when we worry about motivating proper balance among a variety of tasks, a so-called multitasking exercise.

[24] Notice this is in minimization format. However, $\max_{a \in A} \omega(a)$ is identical to $\min_{a \in A} -\omega(a)$. This is what allows us to invoke the same framing principles regardless of whether the problem is stated in maximization or minimization format.

Substitute this expression, (8.4), into (8.3). This gives us

$$C(q; P) \equiv \min_{z_1 > 0} \widehat{\omega}(z_1) = 5z_1 + 20\frac{q^2}{z_1}$$
$$\text{s.t. } z_1 \leq 15$$

Notice this is a one variable problem, thanks to the substitution laid out in (8.4). From here you should be able to verify all of the details displayed in Table 2.3.[25]

8.4.1 Cost Functions

Perhaps the most vivid illustration of this component search technique is construction of a cost curve. Rather than frame the firm's question in terms of simultaneously selecting inputs and outputs, we break it into stages. Input choices are initially formalized in the cost function. Output is then chosen by juxtaposing revenue and cost, with cost effectively surrogating for the myriad input choices.

Glance back at expression (2.1), where we formulated the (single product) firm's problem in terms of simultaneously selecting its inputs and output. With one output, q, and two inputs, $z = [z_1, z_2]$, think of the objective function, with specified prices, as simply a function of these three variables, or $\omega(q, z_1, z_2)$. Now glance back at expression (2.7)[26] What we have done is express inputs as a function of output, the cost function $C(q; P)$ in expression (2.2); and with this substitution the profit maximizing frame in (2.7) is a one variable problem with objective function $\widehat{\omega}(q)$. You might enjoy revisiting Examples 2.2 and 2.7.

In short, we can solve for inputs and outputs in one fell swoop or we can approach the problem in stages. Cost, in the guise of $C(q; P)$, carries all the factor input choices when we use a revenue less cost frame. And in the process we have engaged in a component search, a search over output quantity, by expressing the other components of the choice problem as depending on the explicit component, output. (That is, in the reduced frame, with prices given, revenue depends on variable q and cost depends on variable q.)

8.4.2 The General Idea

To smother this in notation, suppose we want to find the maximum of function $\omega(x, y)$, subject to the restrictions $x \in X$ and $y \in Y$. We might

---

[25] For example, ignore the constraint and differentiate: $\widehat{\omega}'(z_1) = 5 - 20q^2/z_1^2$. At the optimal solution, we know $\widehat{\omega}'(z_1) = 0$, or $z_1 = 2q$. And the constraint is not binding unless $q > 7.5$.

[26] Don't miss the subsection title, "Cost and Revenue Framing."

write this abstract problem as:[27]

$$\max_{x \in X, y \in Y} \omega(x, y) \tag{8.5}$$

The imperative is to search over combinations of $x \in X$ and $y \in Y$ to find the choices that give the maximum feasible value of $\omega(x, y)$.

Now rewrite the formulation in slightly different fashion:

$$\max_{x \in X} \{ \max_{y \in Y} \omega(x, y) \} \tag{8.6}$$

Concentrate on the portion included in brackets. For any tentative choice of x, this is a one variable problem. Suppose we tentatively specify $\widehat{x} \in X$. The portion in brackets now directs us to find the value of $y \in Y$ that maximizes $\omega(\widehat{x}, y)$. Denote the choice of y in this circumstance by $y = g(\widehat{x})$. (Glance back at (8.4)).

Now repeat this procedure for every possible $\widehat{x} \in X$. In this way we construct the function $y = g(x)$. That is, function $g(x)$ gives a best choice of y to match with each possible choice of x.

$$\max_{x \in X} \{ \max_{y \in Y} \omega(x, y) \} = \max_{x \in X} \omega(x, g(x)) = \max_{x \in X} \widehat{\omega}(x) \tag{8.7}$$

In short, we have re-expressed the problem as one of selecting the value of $x \in X$ that makes the function $\omega(x, g(x)) = \widehat{\omega}(x)$ as large as possible. Our task has taken on the appearance of a single variable problem.

Of course, this is not uninvolved. (A double negative seems appropriate.) We had to do the work to solve the inner maximization. The point, however, is valid. It is logically (and conveniently) possible to reduce the apparent dimensionality of a choice problem by "maximizing out" some choices.[28]

8.4.3   Interactions

In exploring this component search technique we have, however, focused on the case where the feasible choices, $x \in X$ and $y \in Y$ do not interact.

---

[27] So we have $a = [x, y]$ and $A = \{[x, y] | x \in X, y \in Y\}$.

[28] This is not a sleight of hand exercise. Assume the choice problem is well formulated, so the maximization problem has a solution. Also assume the inner maximization problem has a solution for every possible $x \in X$. Let x^* and y^* denote a solution to the problem as originally stated. Now suppose our rewritten problem identifies an optimal solution of $x^{**} \in X$ and $y^{**} \in Y$. Further suppose $\omega(x^{**}, y^{**}) > \omega(x^*, y^*)$. This implies we didn't have the correct solution in the first place, and is a contradiction. What if $\omega(x^{**}, y^{**}) < \omega(x^*, y^*)$? This means, using $y = g(x)$, we have $\omega(x^{**}, g(x^{**})) < \omega(x^*, y^*)$. But x^* is feasible, and $\omega(x^*, g(x^*)) = \omega(x^*, y^*)$. Otherwise, we did the inner maximization incorrectly. Hence, the point x^{**} and y^{**} is not a solution to the rewritten problem and we have another contradiction. Thus, the only possibility is $\omega(x^{**}, y^{**}) = \omega(x^*, y^*)$.
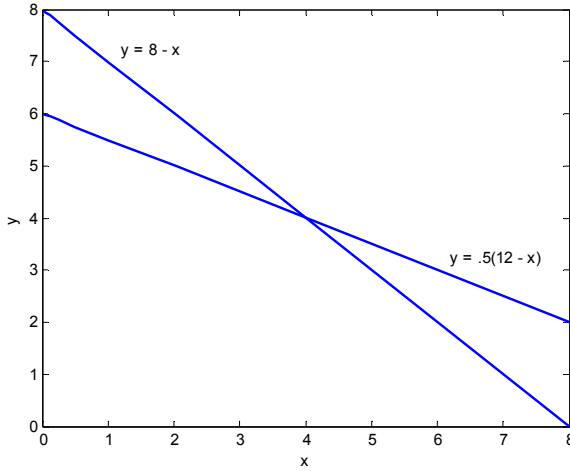
FIGURE 8.3. Constraints on Choice of y

Such interactions, fortunately, also lend themselves to component searches. The trick is to be careful in solving the inner maximization.

To see this, return to the two variable setting in (8.2), but now frame it as a search for the optimal x. To build intuition, think of this as a firm with two products and limited capacity, as specified by the constraints. (Limited capacity implies this is a short-run story, but we have used the first principle of consistent framing to jettison the fixed costs).

With this preamble, notice the constraints limit us to $0 \leq x \leq 8$. Now ask yourself, for any tentative value of x in this region, what is the best choice of y? The answer is simple. Each unit of y increases the objective function by 12 units, so we want y to be as large as possible. The first constraint in (8.2) tells us that $y \leq 8 - x$. The second constraint tells us that $2y \leq 12 - x$, or $y \leq .5(12 - x)$. Remember we want y as large as possible, given $0 \leq x \leq 8$. Hence, our preferred y is the largest value of y that satisfies both of these conditions.

See Figure 8.3 where we plot these two conditions.[29]   The largest value of y that satisfies both conditions is given by the lower of the two lines. This, in short, is our $y = g(x)$ function:

$$y = g(x) = \min\{8 - x; .5(12 - x)\}$$

---

[29] The essence of Figure 8.3 is we cannot write the constraints in separable fashion, i.e. as $x \in X$ and $y \in Y$. Rather, they take the form $[x, y] \in \{[x, y] | x + y \leq 8, x + 2y \leq 12, x \geq 0, y \geq 0\}$.

Examine this function more closely. Notice that for small x, the second constraint is binding while the converse is true for larger x. Also notice the constraints intersect at $x = 4$.[30]

Now substitute these tentative y choices into the original objective function of $\omega(x, y) = 10x + 12y$:

$$
\begin{aligned}
\omega(x, g(x)) \;=\;\; & \widehat{\omega}(x) = 10x + 12 \cdot \min\{8 - x; .5(12 - x)\} \qquad (8.8) \\
=\;\; & \begin{cases} 10x + 6(12 - x) = 72 + 4x, & \text{if } 0 \leq x \leq 4 \\ 10x + 12(8 - x) = 96 - 2x, & \text{if } 4 \leq x \leq 8 \end{cases}
\end{aligned}
$$

We now have an objective function that depends only on x. What is the maximum? The maximum occurs at $x = 4$. The slope of $\widehat{\omega}(x)$ is positive if $x \leq 4$; beyond $x = 4$ it is negative. We can do no better than set $x = 4$, which implies $\widehat{\omega}(4) = 88$ and $y = g(4) = 4$. We are back to our original solution in (8.2).

This is the third principle of consistent framing. It is possible to frame portions of a decision problem in implicit fashion, provided we are careful to make certain the explicit and implicit parts of our frame articulate. Consistent framing allows us to reduce the apparent dimensionality of a choice problem.

This leads to a pithy comment. In expression (8.8) we analyze x in terms of $10x + 12g(x)$. Is $12g(x)$ an opportunity cost? Technically, the answer is no. We are working with a frame in which $A_1 = A$; we have not circumscribed the choice of x. We are not limiting our search. We are only doing it in stages. The term is a type of externality cost.

This should give a hint of things to come. Altering the way we frame a choice problem often leads to an alteration in what we regard as the cost of some activity or product. For example, in the original formulation in (8.2) of the above illustration x had a profit margin of 10 and y had a profit margin of 12. But when we transformed or framed the choice to have the appearance of depending only on x, we concluded in expression (8.8) that x had a profit margin of 4 or -2. Apparently, what we mean by "profit margin" depends on how we have framed the choice problem. (This is the noted externality cost at work.)

This is why you were warned in the introduction that what we mean by the cost of some object depends on the economic context and on the way we have framed the decision. And this often leads to a decision frame in which the cost in question is far removed from some expenditure. Cost, that is, becomes more and more distant from expenditure.

---

[30] $8 - x = .5(12 - x)$, or $2 = .5x$, which implies $x = 4$.

8.5   Consistent Framing

We have referred to these framing exercises as consistent framing. The consistent adjective is code for an important assumption. We assume we enter the exercise with a well defined optimization problem. Moreover, in the presence of economic rationality this takes the form of maximizing $\omega(a)$ subject to $a \in A$. This is a given. It is exogenous. Our exploration begins with the choice problem in place.

Given this beginning, we may transform the objective function, search in limited domains, or reduce the apparent dimensionality of the problem. With care, these techniques, mixed in various ways, will lead us to identify an optimal solution.

These principles are based on optimization, on locating the maximum of some function over some defined region. In this sense, and to this degree, they are grounded in theory. Which frame is best is outside the theory. Also, where the problem statement comes from in the first place is outside the theory.

For that matter, we also might entertain some specification of the problem that is easier to analyze, even if this leads us to analyze a misspecified though easier to analyze problem. These latter concerns take us beyond our theory. This is where the theory of managerial action ends and the art begins.

8.6   Summary

Economic rationality rests on consistency in the art of making choices. We expect a professional manager to aspire to consistency, and thus ground our study on this assumption. Consistency (coupled with the technical smoothness requirement) means we can model the manager's choice behavior as though he were solving a well-defined optimization problem. From here the richness and importance of our study comes into view, as a given optimization problem can be transformed, or framed, in a variety of manners. Some seem to enhance the art of decision making, while others seem to hinder it.

This is why we stress decision framing is a mixture of art and theory. The theory side of the recipe uses three ingredients: the ability to transform an objective function, to engage in local searches, and to reduce the apparent dimensionality of a decision problem. Consistently done, nothing is lost by using these ingredients.

The local search idea relies on opportunity cost as the countervailing force. We stress opportunity cost is the evaluation measure's score of the best alternative not explicitly searched. The dimensionality reduction idea relies on "maximizing out" some choices. The economist's classical cost

function is the reigning example. You will learn that continued use of this idea creates a notion of cost in a decision setting that removes us further and further from expenditures on associated factors of production. This will be a recurring theme of our study.

8.7   Bibliographic Notes

Economic rationality is a deeply studied and controversial subject. The central issue is whether preferences can be measured, i.e., whether a utility or criterion function exists and the interplay between this existence question and patterns of human behavior. On the existence side, Demski [1980] provides an introduction, with deeper treatments in Fishburn [1970], Krantz and associates [1971], and Kreps [1988]. In the next chapter we will move further into this topic by introducing uncertainty and preference measures based on probabilities. On the human behavior side, Sargent [1993] explores "bounded rationality," as do Dawes [1988], and Nisbett and Ross [1990].

   Framing is explored in a variety of directions. Buchanan [1969] provides an extensive discussion of opportunity cost, linking it to the preferences that govern a decision problem. Demski and Feltham [1976] link various transformations of a decision problem, based on the principles of consistent framing, to concepts of cost. Naturally this interacts with framing, a point explored by Bonner [1999].

8.8   Problems and Exercises

1. What does it mean when we say consistency is the central feature of economic rationality? Might an individual characterized by undivided pursuit of wealth be economically rational? Might an individual characterized by undivided pursuit of social justice be economically rational? Explain.

2. The three principles of consistent framing were presented in terms of locating an element in a given set, $a \in A$, that make a given criterion function, $\omega(a)$, as large as possible. Carefully discuss the role of economic rationality in identifying and using these principles.

3. *nonlinear shadow price*
   Return to our discussion of shadow prices, and the maximization in (8.2). Now suppose the criterion function is $\omega(x, y) = 10x^2 + 12y^2$. Determine an optimal solution, along with the shadow prices on the two constraints. Repeat for the case where the first constraint is $x + y \leq 9$. Is the gain from relaxing the constraint numerically equal

to the original shadow price multiplied by the one unit increase in the constraint? Explain.

4. *shadow price under component search*
In Example 8.3 we worked through a reduced dimensionality version of Example 2.6. An important constraint is the requirement $z_1 \leq 15$. Determine the shadow price on this constraint for $q \in \{5, 10, 15\}$. Compare you shadow price with that in the original setting of Example 2.6. Explain your finding.

5. *increasing transformations*
Suppose we want to maximize $\omega(a) = 12a - a^2$, over $0 \leq a \leq 8$. Why does the first principle of consistent framing apply to transforming the entire function and not its individual components? Hint: what is the maximum of $[12a - a^2]^3$, subject to the noted constraint? Contrast this with the maximum of $[12a]^3 - [a^2]^3$, again subject to the noted constraint.

6. *incremental analysis*
Suppose a firm seeks to maximize its profit. It is presently producing and selling $\hat{q}$ units. It has an opportunity to produce and sell $\hat{q} + 1$ units. Carefully explain the use of the first principle of consistent framing when we analyze this in terms of the incremental revenue and incremental cost of the additional unit.

7. *opportunity cost*
Suppose you are going to the movie. The choices are a mystery, a high adventure story, a musical, or a documentary. Further suppose you absolutely cannot stand musicals. Use the concept of opportunity cost to frame the choice by pre-screening (pun) the musical.

8. *opportunity cost*
A retailer often frames a product stocking and placement decision by thinking (and analyzing) in terms of the opportunity cost per unit of shelf space. Is this a proper application of the principles of consistent framing? Explain

9. *shadow prices*
We find Ralph studying cost, and how cost depends on the way a choice problem is framed. Ralph now produces two products. Let x and y, respectively, denote the quantities of the two products that are produced and sold. Any nonnegative quantities satisfying the following constraints can be produced: (1) $x + y \leq 400$; and (2) $x + 2y \leq 500$. Ralph's revenue is given by $40x + 42y$, and his cost is given by $30x + 30y$. (Though this is clearly a short-run story, as capacity is fixed, we suppress the fixed cost and measure "profit" as $10x + 12y$.)

    (a) Determine an optimal solution for Ralph.

    (b) Determine the shadow prices on the two constraints.

    (c) In what sense are the shadow prices on the two constraints opportunity costs?

10. *component searches and product cost*
Return to problem 9 above. Now suppose Ralph likes to think in terms of how many units of the first product, x, to produce and sell. Clearly we require $0 \leq x \leq 400$. Within this range, it should also be clear Ralph would produce as many units of the second product as possible. This implies, for any such x, the corresponding choice of y would be $y = g(x) = \min\{400 - x; .5(500 - x)\}$. This implies profit as a function of x is given by $10x + 12g(x) = 10x + 12 \cdot \min\{400 - x; .5(500 - x)\}$.

    (a) Plot this expression, for $0 \leq x \leq 400$. Determine the optimal choice of x.

    (b) Next, observe (but verify that) this function simplifies to $10x + 3,000 - 6x$ if $0 \leq x \leq 300$ and $10x + 4,800 - 12x$ if $300 \leq x \leq 400$. Concentrate on the first range. What is the implied incremental or marginal cost of the first product in this range? Carefully explain your answer, in light of the fact this product was previously viewed as providing revenue of 40 per unit less cost of 30 per unit.

    (c) Why does the cost of the product depend on the decision frame?

11. *combinations of the framing principles*
Suppose we want to maximize $w(x, y) = 12x - x^2 + 18y - 3y^2 - 10$, subject to $x + y \leq 8$, $x \geq 0$ and $y \geq 0$. You should verify the solution has $x = 5.25$ and $y = 2.75$. Now consider the following. (i) Initially drop the constant of -10. (ii) Notice that if the constraint were not present, we would never set x above 6 or y above 3. Doing so lowers the objective function. Similarly, we would never set x below 6 or y below 3. A slight increase whenever the variables are below the noted targets will increase the objective function. (iii) This insight implies, with the constraint present, we would never set x below 5 (because y would never be set above 3). (iv) Together, then, we can locate the best choice of x by maximizing $12x - x^2 + 18(8 - x) - 3(8 - x)^2$, subject to the constraint $5 \leq x \leq 6$.

    (a) Try it.

    (b) Carefully document the use of the three principles of consistent framing in this exercise.

12. *framing and approximations*
    This problem works through a sequence of framing exercises.

    (a) Ralph produces a single product, with quantity denoted x. Profit is given by the expression $x(10-.5x)$, and capacity is constrained so $0 \leq x \leq 10$. Determine Ralph's optimal output.

    (b) A new customer arrives on the scene. Let y denote the quantity of output Ralph produces for this second customer. This customer is a mirror image of the first, so Ralph's problem is now to select quantities x and y to maximize profit of $x(10 - .5x) + y(10 - .5y)$, but now subject to a capacity constraint of $0 \leq x + y \leq 10$ (and nonnegative quantities of course). Determine Ralph's optimal output of each product, i.e., x and y. You should find $x = y = 5$.

    (c) Ralph likes to keep things simple, and enjoys working with single product decision frames. It turns out that the optimal x can be located in this case by maximizing any of the following functions: (1) $x(10 - .5x) + [50 - .5x^2]$; (2) $x(10 - .5x) + [-.5x^2]$; or (3) $x(10 - .5x) + [-5x]$. Verify this claim. Then carefully explain why each function allows us to identify the optimal choice of y.

    (d) Now suppose Ralph must immediately decide on the quantity of the first product (x); after this decision has been implemented, Ralph will learn whether demand for the second product materializes. If it does, and if Ralph supplies y units of the second product, total profit will be $x(10-.5x)+y(10-.5y)$. Naturally, we require $x+y \leq 10$. Let α denote the probability demand for the second product materializes. So Ralph's problem is now to maximize expected profit of $x(10 - .5x) + \alpha y(10 - .5y)$, subject to a capacity constraint of $0 \leq x + y \leq 10$. The solution is $x = 10/(1 + \alpha)$ and $y = 10 - x$. (x now denotes the immediate choice of first product quantity, and y the choice of second product quantity provided demand materializes.) How do you interpret this solution?

    (e) Finally, go back to Ralph's penchant for keeping things simple. It turns out the optimal x can be located here by maximizing any of the following functions: (1) $x(10-.5x)+\alpha[50-.5x^2]$; (2) $x(10-.5x)+\alpha[-.5x^2]$; or (3) $x(10-.5x)+[-10\alpha x/(1+\alpha)]$. Verify this claim. Then carefully relate each function to its counterpart in the initial story (where $\alpha = 1$).

13. *decision making*
    Consider a three product firm facing a constrained linear technology. The firm is organized into two departments, machining and assembly. Machine hours are constraining in the first department and labor

hours are constraining in the second department. The required machine and labor times for each product are listed below:

	#1	#2	#3
hours of machine time in dept. 1	1	2	3
hours of direct labor in dept. 2	2	4	5

Thus, each unit of product #1 requires 1 machine hour in department #1 and 2 direct labor hours in department #2, and so on. Total capacity is 12,000 machine hours in department #1 and 15,000 direct labor hours in department #2. In turn, total manufacturing cost, for any feasible production plan $q = [q_1, q_2, q_3]$ is given by $TMC = 200,000 + 18q_1 + 24q_2 + 45q_3$. Respective selling prices are 130, 145, and 185 per unit. Finally, the only period cost is specialized shipping "foam" that protects each of the products. This "foam" is purchased from a local supplier at a cost of 100 per pound. Each unit of product #1 requires 0.3 pounds of foam, each unit of product #2 requires 0.5 pounds of foam, and each unit of product #3 requires 0.7 pounds of foam.

(a) Formulate a program to maximize the firm's profit. Use four decision variables in your formulation, q_1, q_2, q_3, and F (the total quantity of foam purchased). Your program should have three capacity constraints, dealing with total machine hours in department #1, total direct labor hours in department #2, and total foam consumed.

(b) Without solving the program, what is the shadow price on the foam constraint? Carefully explain your reasoning. Then solve your program and verify your conjecture.

(c) Next formulate a program to maximize the firm's profit using but three decision variables, q_1, q_2, and q_3. Carefully explain the relationship between your two programs.

14. *decision making with interactions*
Ralph is now managing a firm with interdependent service centers. Two such centers are involved, say, power and maintenance. For each such center, 80% of total output goes to manufacturing and 20% goes to the other service center. The variable cost of power is 10 per unit while the variable cost of maintenance is 15 per unit. Production requires 800 units of each service. Hence, 1,000 (gross) units of each are produced at a total variable cost of 25,000. At this point Ralph starts to consider an opportunity to purchase power (all or some) from an outside vendor, and the following questions are to be answered with this in mind.

(a) What is the cost per unit of power? (Hint: simultaneous equations are necessary, and 13.5417 is an important number.)

(b) Now formulate and solve a program to determine the minimum cost activity levels for power and maintenance in order to provide manufacturing with at least 800 units of each service. You can infer from the noted production plan that each unit of power requires .2 units of maintenance and vice versa. So the technical constraints in your program should be $P - .2M \geq 800$ and $M - .2P \geq 800$, where P denotes units of power and M denotes units of maintenance.

(c) Compare your cost per unit of power in (a) above with the shadow price on the power constraint in (b) above. Carefully explain why they differ, if the differ, or why they are the same if they are the same.

(d) Suppose the outside source will sell power at 12 per unit. Should this offer be accepted? If so, determine the total saving.

(e) Next formulate and solve a program related to that in (b) above but with a third variable, x, units of power purchased from the outside source at 12 per unit.

(f) In part (d) above you used a cost of internal power of 13.5417 to answer the sourcing question, but in the program in (e) you used a cost of internal power of 10 per unit to answer the sourcing question. Carefully explain. (Hint: decision framing is at work.)

15. *framing with interactions*[31]
Ralph uses many factors to produce two products, and has framed his analysis to focus on the two products (denoted q_1 and q_2) and four explicit factors. The first product sells for 38 per unit, and the second for 33 per unit. The four factors are direct labor (DL, with a price of 13 per hour), direct material (DM, with a price of 5 per pound), service one (x, with a cost of .90 per unit of service), and service two (y, with a cost of 2.6 per unit of service). These two services provide essential service to the two products and to each other. This is also a short-run story, and Ralph has in-place two machines with respective capacities of 150 and 50 machine hours. Ralph, always clever, has transformed the problem to ignore the fixed costs associated with his fixed factors. The factor requirements are specified by the following technology constraints:

---

[31] Contributed by Rick Young.

direct labor	$DL \geq 2q_1 + q_2$
direct material	$DM \geq q_1 + 2q_2$
service one	$x \geq q_1 + 2q_2 + .1x + .2y$
service two	$y \geq 2q_1 + q_2 + .3x$
machine 1's capacity	$q_1 + q_2 \leq 150$
machine 2's capacity	$q_2 \leq 50$

Thus, each product uses one hour on the first machine, while only the second uses the second machine. Service one serves both products, itself and service two. Service two serves both products and service one.

(a) Determine Ralph's optimal production plan, consisting of the two products and four factors. Your program should explicitly solve for all six variables, and include all of the noted constraints along with the usual non-negativity requirements. Having done this, what is the cost per unit of each product used in your program. Do these costs comport with what is likely to be reported in the accounting library? Explain.

(b) Repeat (a) above, but now for the case where your program solves explicitly only for q_1, q_2, x and y.

(c) Repeat (a) above, but now for the case where your program solves explicitly only for q_1 and q_2.

(d) Suppose we add another constraint: $q_2 \leq 45$. What is the shadow price on this constraint in each of the above three frames. Explain.

9

Consistent Framing under Uncertainty

We now expand our work on consistent framing to include explicit recognition of uncertainty. This is important for several reasons. First, viewing accounting as a source of information naturally presumes information is valuable or useful. It must be able to tell us something we do not know. This implies uncertainty must be present. Second, and in parallel fashion, evaluating an agent's performance is a gratuitous exercise absent uncertainty, as certainty implies we already know how the agent has performed. Third, risk may well be an important consideration in a decision. But risk has no meaning in the absence of uncertainty. For that matter, our work to date on estimating marginal cost is pointless unless we don't know marginal cost. For pedagogical purposes we developed this theme without formally acknowledging uncertainty. But our work from this point forward does not allow such a convenient approach.

We begin by extending the earlier notion of consistency and smoothness to yet additional structure. This will allow us to describe choice behavior as though the *expected value* of a utility function were being maximized. We then exploit the structure in this model of choice to examine a powerful framing device of certainty equivalents, to examine risk aversion, and to examine information's arrival and use in a choice setting.

Keep in mind we are anchoring the art of managerial choice on economic foundations. This provides a parsimonious description of managerial behavior, one that emphasizes consistency in the face of economic forces. It focuses our study and leads to important insights. It is not, however, a universal description of behavior. At appropriate junctures we will intro-

duce the idea of systematic variations from economic rationality. This, of
course, presumes we understand economic rationality in the first place.

9.1   Explicit Uncertainty

The key to modeling choice in the face of uncertainty is to place some
structure on that uncertainty. To do this, think of a choice as producing
a consequence or outcome, e.g., you go to the movie and enjoy the movie.
With uncertainty, though, the outcome is not guaranteed, e.g., you go to
the movie and either enjoy the movie or not. Probabilities now enter the
story, as a measure of how likely the various outcomes might be.

   To be a bit less vague, suppose we must choose between two alternatives.
Call them, imaginatively, a_1 and a_2. So $A = \{a_1, a_2\}$. Either choice will
lead to some consequence or outcome. The possible outcomes, assessed in
dollar terms, are a gain or net cash inflow of 100, 240 or 400.[1]

9.1.1   Choices as Lotteries

One way to introduce the probabilities is to view each choice as essentially
a probability specification. We do this with the display in Table 9.1. a_1
produces (a good verb here) the 100 outcome with probability $0 \leq \alpha \leq 1$,
the 240 outcome with probability zero, and the 400 outcome with proba-
bility $1 - \alpha$. Conversely, a_2 produces the 240 outcome with probability one
(and the other two outcomes with probability zero).

TABLE 9.1: Probabilities on Outcomes

alternative	dollar outcome		
	100	240	400
a_1	α	0	$1 - \alpha$
a_2	0	1	0

   Each possible choice, then, provides a probabilistic description of the
possible outcomes. Indeed, we might even describe the choices by their
probabilities: $a_1 = [\alpha, 0, 1 - \alpha]$ and $a_2 = [0, 1, 0]$ is the natural notation.
Further notice that choice a_2 is a "safe" choice, as it will lead to the 240 for
certain. Choice a_1 is a "risky" choice, as it will lead to a dollar outcome of

---

[1]The outcome need not be monetary, and might even be a bundle of consequences,
perhaps even spread out through time. What follows, though, is less cluttered if we
proceed with a consequence assessed in monetary terms.

100 with probability α or a dollar outcome of 400 with probability $1 - \alpha$.[2] Stated differently, each possible choice is a lottery, a gamble over (dollar) outcomes. (And if you enjoy pithy comments, a_2 is a degenerate gamble in that it has probability mass on a single outcome.)

9.1.2 Choices as State Dependent Outcomes

An equivalent way to introduce the probabilities is to think of the outcome as being jointly produced by the choice and a random state. Probabilities are now assigned to states, and each choice literally maps states into outcomes. This is illustrated in Table 9.2 for our running choice problem, where we invoke two possible states, denoted s_1 and s_2.

TABLE 9.2: Probabilities on States

	state s_1	state s_2
probability, $\pi(s)$	α	$1 - \alpha$
outcomes		
under a_1	100	400
under a_2	240	240

To decipher this, we begin with probabilities assigned to states: state s_1 obtains with probability α, and state s_2 with probability $1 - \alpha$. For later reference we denote the probabilities $\pi(s_1)$ and $\pi(s_2)$. Further notice that the outcome of $100, 240$ or 400 depends on the state and on the choice. For example, if a_1 is selected and if state s_2 then obtains, the outcome will be a gain of 400. And from here we use the state probabilities to assign probabilities to outcomes. So choice a_2 produces an outcome of 240 regardless of the state, etc.

To reiterate, we have two equivalent approaches for dealing with the probabilities: directly assign them to each choice's possible outcomes (Table 9.1) or assign them to each possible state and from there infer the outcome probabilities (Table 9.2). On the surface, the former appears more intuitive; but it will turn out when we introduce information that the latter is more intuitive. So in the spirit of consistent framing, we begin with two equivalent probability frames.[3]

---

[2] For example, suppose $\alpha = .5$. Would you rather (1) flip a coin and receive 100 if it is heads and 400 if it is tails or (2) receive 240 for certain? If this doesn't catch your interest, add, say, 4 zeros to each outcome.

[3] The state approach is also used in extending the economic theory of cost to uncertainty. To give an example, albeit cryptic, in Chapter 2 we expressed the one-product firm's technology via $q \leq f(z_1, z_2)$, expression (2.1). Under uncertainty, output and factor supplies are uncertain and thus depend on the state as well. So we express the technology as $q^s \leq f(z_1^s, z_2^s)$ for each state $s \in S$.

9.2   Consistent Choice with Probabilities

Once the outcomes and probabilities are in place, we exploit the structure with an expected utility approach. The mechanics are deceptively simple. We assign, yes another assignment, a utility score to each outcome. And we then evaluate each choice via the expected value of the possible utility scores it induces.[4]

   Looking ahead, we want to specify these utilities in a particular manner. For this purpose, let w denote wealth and $U(w)$ denote the utility associated with wealth level w. Further assume that, prior to encountering the decision problem described in Tables 9.1 and 9.2, our individual has an initial wealth level of w_i. As the outcomes in the two Tables are expressed in terms of gains, our friend's wealth will turn out to be $w_i + 100$, $w_i + 240$ or $w_i + 400$. This leads to respective utility assignments of $U(w_i + 100)$, $U(w_i + 240)$ and $U(w_i + 400)$.

   Now let $E[U|a]$ denote the expected utility that follows from choice of $a \in A$. In our running illustration we then have the following criterion measures:

$$E[U|a_1] = \alpha \cdot U(w_i + 100) + 0 \cdot U(w_i + 240) + (1 - \alpha) \cdot U(w_i + 400) \quad (9.1\mathrm{a})$$

$$E[U|a_2] = 0 \cdot U(w_i + 100) + 1 \cdot U(w_i + 240) + 0 \cdot U(w_i + 400) \quad (9.1\mathrm{b})$$

And choice a_1 is best if $E[U|a_1] \geq E[U|a_2]$.

   Notice the role of the utility function, $U(w)$. It is defined on outcomes, not on the alternatives. The criterion function that applies to the choices, $E[U|a]$, is the expected value of the utility of outcomes that follows from choice a. Indeed, what we have accomplished is restructuring the $\omega(a)$ criterion function in the prior chapter (expression (8.1)) so that it exploits a probabilistic description of uncertainty. That is, we now work with $\omega(a) \equiv E[U|a]$.

   A decision tree display is given in Figure 9.1, where we show only the positive probability outcome possibilities. Each alternative leads, in general, to an array of possible outcomes. And each possible outcome is assessed in terms of its respective utility. Each initial branch, corresponding to a possible choice, is then "rolled back" to construct the expected

---

[4]Suppose some real valued variable can take on one of n possible vales, $x_1, x_2, ..., x_n$ with respective probabilities $\pi_1, \pi_2, ..., \pi_n$. Then the expected value of this variable is

$$E[x] = \sum_{j=1}^{n} \pi_j x_j.$$

Stated differently, then, we have a utility measure defined on outcomes. As the outcome is uncertain, the resulting utility is itself a random variable. And we measure the preferences for the alternatives by constructing the expected value of the utility for each such alternative.
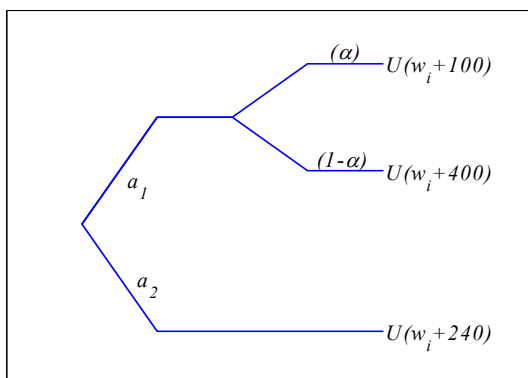
FIGURE 9.1. Decision Tree for Tables 9.1 and 9.2

value of the utilities. The rolled back calculation is the decision criterion, $\omega(a) \equiv E[U|a]$.

Of course this begs the question of what these utility assignments might be or reflect. Below, in Table 9.3, we display a modest variety of illustrative utility functions, all based on the wealth interpretation. The first, $U(w) = w$, is linear while the other two, being square root or negative exponential functions, are nonlinear. All three are increasing in w, implying more is preferred to less. The latter two, however, increase at a decreasing rate, implying diminishing marginal returns.

TABLE 9.3: Illustrative Utility Assignments

case	utility for wealth w
1	$U(w) = w$
2	$U(w) = \sqrt{w}, w \geq 0$
3	$U(w) = -\exp(-\rho \cdot w), \rho > 0$

Some relief is in order.

Example 9.1 Suppose $\alpha = .5$ and $U(w) = w$ in the setting described in Table 9.1 or 9.2. Following expression (9.1) we have the following evaluations:

$$\omega(a_1) = E[U|a_1] = .5(w_i + 100) + .5(w_i + 400) = w_i + 250$$

$$\omega(a_2) = E[U|a_2] = w_i + 240$$

and we see that $\omega(a_1) > \omega(a_2)$ regardless of initial wealth. This is probably the format in which you first encountered expected utility analysis. The

utility of the outcome is the outcome itself (here adjusted for initial wealth). This linearity implies risk neutrality, but that is getting ahead of the story.

Example 9.2 We now repeat the above, but with the second utility measure in Table 9.3: $U(w) = \sqrt{w}$. Here, however, we don't have the luxury of remaining silent on initial wealth, so let's suppose it is zero, $w_i = 0$. We then have the following:

$$\omega(a_1) = E[U|a_1] = .5\sqrt{w_i + 100} + .5\sqrt{w_i + 400} = .5\sqrt{100} + .5\sqrt{400} = 15$$

$$\omega(a_2) = E[U|a_2] = \sqrt{w_i + 240} = \sqrt{240} = 15.49$$

and we now see that $\omega(a_2) > \omega(a_1)$. But initial wealth matters. Repeating this for, say, $w_i = 500$ we see that the preference is reversed: $\omega(a_1) = 27.25 > \omega(a_2) = 27.20$. This reversal, as we will see, is caused by risk aversion coupled with initial wealth affecting the attitude toward risk.

Example 9.3 Continuing the saga, now consider the third utility measure in Table 9.3: $U(w) = -\exp(-\rho \cdot w)$, along with $\rho = 0.001$. Using zero initial wealth we find the following:

$$
\begin{aligned}
\omega(a_1) &= E[U|a_1] = -.5\exp(-.001(100)) - .5\exp(-.001(400)) \\
&= -.5\exp(-.1) - .5\exp(-.4) = -.78758
\end{aligned}
$$

$$\omega(a_2) = E[U|a_2] - \exp(-.001(240)) = -\exp(-.24) = -.78663$$

and conclude $\omega(a_2) > \omega(a_1)$. Repeating for the case $w_i = 500$, we do not alter the preference for a_2 as we have $\omega(a_2) = -.47711 > \omega(a_1) = -.47769$.

9.2.1 Scaling

It is important to remember, especially as we march through the above three examples, that choice between a_1 and a_2 is modeled in terms of asking whether we have $E[U|a_1] > E[U|a_2]$, the converse, or a tie (implying indifference). Nothing is said or implied about whether the $E[U|a]$ measure is "large," "small," or even negative. The reason is simple. The underlying utility measure, $U(w)$, is unique only up to admissible rescaling. If $U(w)$ is indeed our utility measure in the world of expected utility, then so is $\beta + \gamma \cdot U(w)$ for any arbitrary β and for any strictly positive γ.

Glance back at our evaluations in Example 9.2 where, with zero initial wealth, we calculated $\omega(a_2) > \omega(a_1)$. Multiplying $U(w) = \sqrt{w}$ by a positive constant will not alter this conclusion, nor will adding an arbitrary constant.[5]

---

[5] Rare is the measure that cannot be rescaled. Here, the utility measure is unique to what is called a positive affine transformation. Casual interlopers say it is unique to linear rescaling.

9.2.2   Consistency, Smoothness and Independence

This, then, is the apparatus we use for exhibiting economic forces in the presence of uncertainty. Recall, in our initial foray into rational choice, that we laid out the basic idea that the individual was required to be consistent and in an admittedly vague sense to also exhibit smooth preferences. Otherwise, we could not represent his choice behavior as though a criterion function, or measure, were being maximized, as in expression (8.1).

Here we continue with this maximization motif, so we continue to invoke the consistency and smoothness requirements. But we have also given specific structure to the measure being maximized, as is evident in the expectation structure in (9.1) and the above examples.

Doing so comes with two additional prices, or caveats. First, we must have probabilities in the story, otherwise taking an expected value is an empty concept. Second, these probabilities cannot interact with the utility measure. A form of independence must be maintained.

To provide some flavor of this additional requirement, suppose in our setting in Tables 9.1 and 9.2 we add a third alternative (a_3). This alternative consists of selecting the original a_1 with probability θ and the original a_2 with probability $1-\theta$. Under expected utility representation, its evaluation is:

$$
\begin{aligned}
E[U|a_3] &= \theta\{\alpha U(w_i + 100) + (1 - \alpha)U(w_i + 400)\} + (1 - \theta)U(w_i + 240) \\
&= \theta E[U|a_1] + (1 - \theta)E[U|a_2]
\end{aligned}
$$

That is, the evaluation of compound or sequential gambles is the expected value of the underlying evaluations. This forces the expected value structure that is central to the story.[6]

9.3   Certainty Equivalents

The expected value of the utility of the outcomes is an awkward, long-winded expression. Fortunately, restatement of this machinery in terms of certainty equivalents provides a useful interpretive device. Formally, alternative a's certainty equivalent is that certain wealth (or consequence), denoted CE_a, such that the individual is indifferent between choice a and a new choice that guarantees wealth in the precise amount CE_a. Since CE_a is a certain amount of wealth (i.e., occurs with probability one), its utility evaluation is given by $U(CE_a)$. And since we are indifferent between CE_a

---

[6]It also forces a form of independence in the taste for gambles. Let a, a' and a'' be three alternative choices. If we prefer a to a' we then prefer a compound gamble of a and a'' to a compound gamble of a' and a''(presuming a and a' are both engaged with the same probability).

and a we require

$$U(CE_a) = E[U|a] \tag{9.2}$$

Whatever alternative a entails, the individual is indifferent between it and its certainty equivalent.[7]

Returning to the three utility functions (defined on wealth) displayed in Table 9.3, we have the following generic certainty equivalents. In each case these are derived from the definition of $U(w)$ and the certainty equivalent definition in expression (9.2).

TABLE 9.4: CEs for Table 9.3

case	utility for wealth w	CE_a for $E[U|a]$
1	$U(w) = w$	$CE_a = E[U|a]$
2	$U(w) = \sqrt{w}, w \geq 0$	$CE_a = \{E[U|a]\}^2$
3	$U(w) = -\exp(-\rho \cdot w), \rho > 0$	$CE_a = -\ln\{-E[U|a]\}/\rho$

9.3.1 A Convenient Transformation

Notice, now, that the certainty equivalents provide an admissible transformation, in the sense of the first principle of consistent framing (Chapter 8). One choice is better than another if its expected utility is higher, and this is equivalent to its certainty equivalent being higher. If, for example, choice a_1 is strictly preferred to choice a_2, we have $E[U|a_1] > E[U|a_2]$. And this is equivalent to the statement that $CE_1 > CE_2$:

$$U(CE_1) = E[U|a_1] > E[U|a_2] = U(CE_2)$$

An alternative's certainty equivalent, then, is a guaranteed or certain amount, CE, such that the individual is indifferent between it and the alternative in question. This provides an intuitive expression of preference. Though it is a trivial transformation in the linear utility case of $U(w) = w$, where a choice's CE numerically equals its expected utility, it comes into its own otherwise.[8]

---

[7]The initial wealth remains an important part of the story. CE_a is that certain stock of wealth that is equivalent, in terms of preference, to the initial stock (w_i) plus whatever gains or losses accrue from selecting alternative a. The reason for carrying along the initial wealth complication will become clear as we proceed.

[8]We should be a little careful here. When $U(x) = x$, the expected utility is simply the expected value of x, and our CE_a expression is precisely as claimed:

$$U(CE_a) = CE_a = E[U|a] = E[x|a]$$

But what if we rescale this utility function via $\widehat{U}(x) = \beta + \gamma \cdot U(x) = \beta + \gamma \cdot x$? Our CE_a expression now becomes

$$\beta + \gamma \cdot CE_a = E[\widehat{U}|a] = \beta + \gamma \cdot E[x|a]$$

Example 9.4 To illustrate, return to the $U(w) = \sqrt{w_i + w}$ case of Example 9.2. The a_2 choice is an anomaly. It offers guaranteed wealth of $w_i + 240$, so without further ado expression (9.2) provides

$$\sqrt{CE_2} = E[U|a_2] = \sqrt{w_i + 240}$$

and we therefore have $CE_2 = w_i + 240$. Turning to a_1, (9.2) requires in the $w_i = 0$ story that

$$\sqrt{CE_1} = E[U|a_1] = .5\sqrt{100} + .5\sqrt{400} = 15$$

which implies $CE_1 = (15)^2 = 225$. We thus have $CE_2 = 240 > CE_1 = 225$. a_1, then, is equivalent to 225 in wealth, which does not stack up against a_2. Turning to the $w_i = 500$ story, $CE_2 = 500 + 240 = 740$. CE_1 is the solution to

$$\sqrt{CE_1} = E[U|a_1] = .5\sqrt{500 + 100} + .5\sqrt{500 + 400}$$

or $CE_1 = 742.42 > CE_2 = 740$.

Example 9.5 Now consider Example 9.3 where we assume a utility measure of $U(w) = -\exp(-.001w)$. Of course we continue to have $CE_2 = w_i + 240$. For the $w_i = 0$ story, CE_1 is, again following expression (9.2), the solution to

$$-\exp(-.001CE_1) = -.5\exp(-.1) - .5\exp(-.4)$$

which provides $CE_1 = 238.79.$[9] In turn, if you check the $w_i = 500$ story you will find $CE_1 = 500 + 238.79 = 738.79$, and, indeed, for an arbitrary initial wealth we would have $CE_1 = w_i + 238.79$. This will be explained shortly.

9.3.2 A Special Case

A highly special but insightful version of this certainty equivalent framing device arises when the utility function is the negative exponential version of $U(w) = -\exp(-\rho w)$, and w itself is normally distributed about a mean

---

Now we must adjust for the rescaling in moving between the (thus scaled) expected utility score and the corresponding certainty equivalent. Enough!

[9] Notice we want to solve

$$\exp(-.001CE_1) = .5\exp(-.1) + .5\exp(-.4) = .787579$$

Recalling that $\ln(\exp(z)) = z$, we have

$$\ln(\exp(-.001CE_1)) = -.001CE_1 = \ln(.787579) = -.238792$$

which implies $CE_1 = 238.79$.

of μ and with a variance of σ^2. The density function for such a normally distributed random variable, something you should have encountered in your introduction to statistics, is

$$f(w) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-(w-\mu)^2/2\sigma^2)$$

Now use our certainty equivalent definition in (9.2) to identify the certainty equivalent of such a wealth lottery:

$$- \exp(-\rho CE)) = E[U(w)] = - \int_{-\infty}^{\infty} \exp(-\rho w)f(w)dw$$

It turns out we have a simple and intuitive expression:[10]

$$CE = \mu - \frac{1}{2}\rho\sigma^2 \qquad (9.3)$$

The certainty equivalent is the expected value of w less one half its variance multiplied by the utility function parameter ρ.

This leads us into the language of risk and risk premia.

9.4   Risk Aversion

Suppose we are offered a choice of (1) flip a fair coin, receive $1,000$ dollars if heads and nothing if tails or (2) receive 500 dollars for certain. Most people would jump at the second, sure amount. This is the intuitive idea of risk aversion. We would gladly trade a risky alternative for a certain amount equal to its expected value.

We formalize this by juxtaposing an individual's certainty equivalent for a choice, CE_a, with the expected value of wealth associated with that choice, $E[w|a]$. In particular, we say an individual is *risk averse* whenever (1) $E[w|a] \geq CE_a$ for all a; and (2) the inequality is strict whenever a has strictly positive probability on at least two different wealth levels. If wealth

---

[10] This is not magic. We must evaluate the $- \int_{-\infty}^{\infty} \exp(-\rho w)f(w)dw$ integral, and fortunately $f(w)$ is another exponential term. Substituting in the density function we have

$$- \int_{-\infty}^{\infty} \exp(-\rho w)f(w)dw = \frac{-1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp(-\rho w)\exp(-(w-\mu)^2/2\sigma^2)dw$$

A little algebra and some heartburn now give us the equivalent expression

$$E[U(w)] = \frac{-1}{\sqrt{2\pi}\sigma} \exp(-\rho(\mu - \frac{1}{2}\rho\sigma^2)) \int_{-\infty}^{\infty} \exp(-(w-(\mu-\rho\sigma^2))^2/2\sigma^2)dw$$

But $\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp(-(w-(\mu-\rho\sigma^2))^2/2\sigma^2)dw = 1$, as we are dealing with a normal density. And we wind up with $-\exp(-\rho CE) = -\exp(-\rho(\mu - \frac{1}{2}\rho\sigma^2))$, or $CE = \mu - \frac{1}{2}\rho\sigma^2$.

is at risk, the risk averse individual would gladly trade the risky prospect for a guaranteed amount equal to its expected value. Stated differently, the risk averse individual always seeks fair insurance. Risk is noxious.

An equivalent expression uses the idea of a *risk premium*, defined as the difference between the expected value and the certainty equivalent:

$$RP_a = E[w|a] - CE_a \tag{9.4}$$

Thus, an individual is risk neutral if his risk premium is always zero; and he is risk averse if his risk premium is strictly positive whenever choice a has strictly positive probability on at least two different wealth levels.

Now glance back at our running illustration in Tables 9.1 and 9.2. Choice a_2 is riskless, as it offers a guaranteed gain of 240. So its risk premium is precisely zero, regardless of how we specify the utility for wealth. Choice a_1, on the other hand, is risky, as the gain will be either 100 or 400. Using our work in Examples 9.1 through 9.5, where $\alpha = 0.5$ and thus $E[w|a_1] = 250$, we have the certainty equivalents and risk premia displayed in Table 9.5. Notice the risk premium is zero in the linear case, $U(w) = w$, but strictly positive otherwise.

TABLE 9.5: Risk Premia for Choice a_1 with $E[w|a_1] = 250$ and $w_i = 0$

case	utility for wealth w	CE_1	RP_1
1	$U(w) = w$	250	0
2	$U(w) = \sqrt{w}, w \geq 0$	225	25
3	$U(w) = -\exp(-\rho \cdot w), \rho = .001$	238.79	11.21

TABLE 9.6: Derivatives for Table 9.3

case	utility for wealth w	$U'(w)$	$U''(w)$
1	$U(w) = w$	1	0
2	$U(w) = \sqrt{w}, w \geq 0$	$\frac{1}{2\sqrt{w}}$	$\frac{-1}{4}w^{\frac{-3}{2}} < 0$
3	$U(w) = -\exp(-\rho w), \rho > 0$	$\rho\exp(-\rho w)$	$-\rho^2\exp(-\rho w) < 0$

Importantly, now, there is a systematic connection between the attitude toward risk and the utility function. Risk aversion is present if the second derivative is everywhere negative, $U''(w) < 0$, meaning the utility function is strictly concave. Conversely, risk neutrality is present if the second derivative is everywhere equal to zero, $U''(w) = 0$.[11] As verified in Table

[11] Thus, risk neutrality means the slope of $U(w)$ is everywhere constant, while risk aversion means this slope is decreasing in wealth, w. Further note, as we mentioned, that $U(w)$ is unique to positive affine rescaling, and any such admissible rescaling will not alter this conclusion.

9.6, our linear utility case in Table 9.3 is a risk neutrality story while the other two cases are risk aversion stories.

To explore a bit further, in Figure 9.2 we plot the risk premium associated with choice a_1 as a function of initial wealth w_i for three cases: $U(w) = \sqrt{w}$, $U(w) = -\exp(-\rho \cdot w)$ with $\rho = .001$, and $U(w) = -\exp(-\rho \cdot w)$ with $\rho = .002$. For the root utility case the risk premium systematically declines with initial wealth. (This is called a case of decreasing absolute risk aversion.) For the negative exponential cases, however, the risk premium depends on parameter ρ but is independent of initial wealth. (This is called a case of constant absolute risk aversion.)
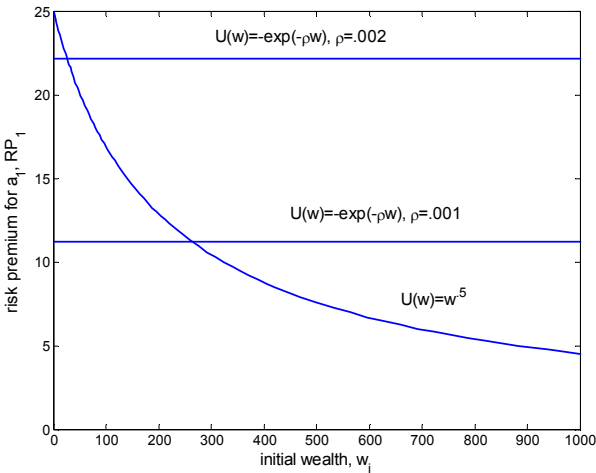
Risk Premium vs. Initial Wealth

This is an important point. Risk aversion may well depend on the status quo. It does not if the individual is risk neutral. In the square root case, risk aversion declines as initial wealth increases. More generally, risk aversion might decline, increase, remain constant, or be some combination thereof as we vary initial wealth. A negative exponential utility function is the only one that displays risk aversion that is everywhere constant.[12]

---

[12] The technical reason is risk aversion is related to the utility function's concavity, and the negative exponential case is the only increasing function with constant concavity. In any event, a little algebra is insightful. Let z denote choice a_1's certainty equivalent under zero initial wealth, so

$$-\exp(-\rho \cdot z) = -.5 \exp(-\rho \cdot 100) - .5 \exp(-\rho \cdot 400)$$

For nontrivial initial wealth a_1's certainty equivalent is the solution to

$$
\begin{aligned}
-\exp(-\rho \cdot CE_1) &= -.5 \exp(-\rho(w_1 + 100)) - .5 \exp(-\rho(w_1 + 400)) \\
&= -\exp(-\rho \cdot w_i) \exp(-\rho \cdot z)
\end{aligned}
$$

The implication is as annoying as it is important. If, and we will at times, think risk and risk aversion are too important to ignore, the compelling place to begin is with constant risk aversion. Otherwise, we are also saying risk, risk aversion and changing risk aversion are all important. This is unlikely to be the case. It behooves us, then, when we explicitly acknowledge risk to do so in a setting of a non-changing attitude toward risk. You will see in later chapters that whenever risk aversion surfaces as something too important to ignore, we treat it with a constant attitude toward risk aversion – the annoying negative exponential case.

Having convinced you this is, indeed, how we should proceed, one additional nuance should be noted. In Figure 9.2 we display the risk premium for choice a_1 for two parametric specifications: $\rho = .002$ and $\rho = .001$. A larger ρ implies a larger risk premium, and thus a lower certainty equivalent. This is also evident in our special case of expression (9.3), where we see, for a given (normal, of course) wealth distribution, that the certainty equivalent is decreasing in ρ. Stated differently, in the negative exponential case parameter ρ is a measure of the individual's risk aversion. For that matter, an arbitrarily small ρ returns us to risk neutrality, while an arbitrarily large ρ signifies paralyzing aversion to risk.

9.5 Information

We now turn to the acquisition and use of information. Expected utility representation offers an additional advantage at this point. It forces the information to be used in a particular way: systematic, consistent revision of the probabilities.

As hinted in our comparison of Tables 9.1 and 9.2, information is equivalently though more intuitively dealt with when probabilities are assigned to states. To append an information option to our running illustration, suppose an information source will reveal one of two possible signals, denoted g (for "good" news) and b (for "bad" news).[13] This leads to Table 9.7's extension of the display in Table 9.2.

---

or

$$\exp(-\rho \cdot CE_1) / \exp(-\rho \cdot w_i) = \exp(-\rho \cdot z)$$

Thus, $CE_1 - w_i = z$. So for nontrivial initial wealth the risk premium turns out to be $RP_1 = E[w|a_1] - CE_1 = w_i + 250 - (z + w_i) = 250 - z$, a constant independent of initial wealth.

More generally, the utility function's concavity is measured by $-U''(w)/U'(w)$, which you will note from Table 9.6 is the constant ρ for the negative exponential case. This concavity measure is called the Arrow-Pratt measure of absolute risk aversion.

[13]This "good" versus "bad" nomenclature reflects the fact observing signal g leads to an upward revision in the probability a_1 delivers the larger outcome. This is pursued in a subsequent note.

TABLE 9.7: Probabilities on States

	state s_1	state s_2	state s_3	state s_4
probability, $\pi(s)$	$.8\alpha$	$.2\alpha$	$1 - \alpha$	0
outcomes				
under a_1	100	100	400	400
under a_2	240	240	240	240
information signal	b	g	g	b

Notice that if we combine states s_1 and s_2 we have the original state s_1 in Table 9.2; and if we combine states s_3 and s_4 we have the original state s_2 in Table 9.2. This applies to the outcomes from each alternative, as well as the probabilities. (E.g., the probability that s_1 or s_2 obtains is $.8\alpha + .2\alpha = \alpha$.) So we continue with precisely the same running illustration. Also notice the way we have modeled the information. Signal g occurs only under states s_2 or s_3, while signal b occurs only under states s_1 or s_4. That is, the information source is modeled as a function from states to signals. This implies the information source forms a partition of the possible states. We emphasize this state-based description is perfectly general, can be awkward at times, and (importantly for us) can be profoundly insightful.

In any event (an awful pun), suppose in our setting of Table 9.7 that we are able to observe the information signal (g or b) before making our choice between a_1 and a_2. This means our choice can depend on what signal we observe. There are, in fact, four distinct patterns of matching the two possible choices with the two possible signals. Let (a_g, a_b) denote the policy where alternative a_g is chosen if signal g is observed and alternative a_b is chosen if signal b is observed. We thus have: (1) select a_1 regardless of the signal, (a_1, a_1); (2) select a_2 regardless of the signal, (a_2, a_2); (3) select a_1 if signal g is observed and a_2 if signal b is observed, (a_1, a_2); and (4) select a_2 if signal g is observed and a_1 if signal b is observed, (a_2, a_1). We have, in other words, four possibilities now, instead of the original two.

Expanding Table 9.7, our four possibilities produce the outcome structure in Table 9.8.[14] Information increases the options in the original decision problem. We can ignore the information, the first two policies in Table

---

[14] An equivalent and perhaps more familiar approach to identifying the optimal policy is to focus directly on the revised probabilities. Suppose we observe signal b (for "bad" news). This means the state is either s_1 or s_4. Since state s_4 has 0 probability, we know for certain that state s_1 is present. Given this, we know choice a_1 will deliver an outcome of 100 while a_2 will deliver an outcome of 240. Advantage a_2! So conditional on observing signal b, the expected utility measure is $E[U|a_2, b] = U(w_i + 240)$.

Conversely, suppose we observe signal g (for "good" news). This means the state is either s_2 (where the risky choice delivers 100) or s_3 (where the risky choice delivers 400). And given signal g, i.e. given the state is either s_2 or s_3, the conditional probability of state s_3 is

9.8 Alternatively, we can exploit the information by varying the underlying choice depending on what the information reveals, the latter two policies in Table 9.8.

TABLE 9.8: Informative Signal g or b

	state s_1	state s_2	state s_3	state s_4
probability, $\pi(s)$	$.8\alpha$	$.2\alpha$	$1 - \alpha$	0
outcomes				
under (a_1, a_1)	100	100	400	400
under (a_2, a_2)	240	240	240	240
under (a_1, a_2)	240	100	400	240
under (a_2, a_1)	100	240	240	400
information signal	b	g	g	b

That said, precisely how the information is best used (or ignored) depends on the outcome structure, the probabilities and the attitude toward risk. Table 9.9 reports the certainty equivalents for each of the four policies for a familiar array of risk attitudes (or utility functions), all for the case

$$\pi(s_3|g) = \pi(s_3)/\pi(g) = (1 - \alpha)/(1 - .8\alpha) > \pi(s_3) = 1 - \alpha$$

Good news, as the odds of obtaining the 400 prize have gone up.

This probability revision is the key to information processing under economic rationality. Let M and N be two events, with respective probabilities $\pi(M) > 0$ and $\pi(N) > 0$. Denote their joint probability $\pi(M \text{ and } N)$. The conditional probability of M given N is given by Bayes' Rule:

$$\pi(M|N) = \pi(M \text{ and } N)/\pi(N) = \pi(N|M)\pi(M)/\pi(N)$$

Now let N be the event information signal g is observed, which is equivalent to learning the state is either s_2 or s_3. From Table 9.7 we see that $\pi(g) = \pi(s_2 \text{ or } s_3) = .2\alpha + 1 - \alpha = 1 - .8\alpha$. Also let M be the event state s_3 is true. But the joint probability of observing g and s_3 being true is simply $\pi(M \text{ and } N) = \pi(s_3 \text{ and } s_2 \text{ or } s_3) = \pi(s_3) = 1 - \alpha$. So we have $\pi(s_3|g) = \pi(s_3)/\pi(g)$, as claimed. Likewise, $\pi(s_2|g) = \pi(s_2)/\pi(g)$.

But which action is best, given the revised probabilities? Well, a_1 now carries an expected utility measure of

$$E[U|a_1, g] = \pi(s_2|g)U(w_i + 100) + \pi(s_3|g)U(w_i + 400)$$

while a_2 carries a measure of

$$E[U|a_2, g] = U(w_i + 240)$$

And it follows that the choice depends on the odds as well as the risk aversion.

Finally, for the sake of argument, suppose the best choice given signal g is a_1. Then the overall expected utility evaluation would be

$$\pi(g)E[U|a_1, g] + \pi(b)E[U|a_2, b]$$

This is precisely the expected utility measure that is transformed to the certainty equivalents in Table 9.9 for the (a_1, a_2) policy.

of zero initial wealth and $\alpha = .5$ (so the respective state probabilities are $.4, .1, .5$ and 0).

TABLE 9.9: CEs for Various Policies, $\alpha = .5, w_i = 0$

utility	$CE_{(a_1,a_1)}$	$CE_{(a_2,a_2)}$	$CE_{(a_1,a_2)}$	$CE_{(a_2,a_1)}$
$U(w) = w$	250	240	306	184
$U(w) = \sqrt{w}$	225	240	295.73	176.76
$U(w) = -\exp(-.001w)$	238.79	240	300.71	181.63
$U(w) = -\exp(-.1w)$	106.93	240	123.03	109.16

Keep in mind that the information can be ignored, the (a_1, a_1) and (a_2, a_2) policies, or it can be used, the (a_1, a_2) and (a_2, a_1) policies. Two patterns emerge. First, since state s_4 is assigned zero probability, policy (a_1, a_2) dominates policy (a_1, a_1), as it produces the same outcome in two of the possible states and a better outcome in a third. Likewise, policy (a_2, a_2) dominates policy (a_2, a_1). This dominance implies that, regardless of risk preference, if the information is to be used, it will be used according to the (a_1, a_2) policy in this case. This is reflected in the certainty equivalents displayed Table 9.9.

Second, even though policy (a_1, a_2) is the best way to use the information, it does not follow that using the information is the best policy. In the first three cases in Table 9.9 it is indeed optimal to use the information. But in the last case, where risk aversion has increased, the outcome prospects even when improved by acting on the information are deemed too risky. The safe project is chosen, in effect ignoring the information.[15]

Information, then, enriches the opportunities. And if one of these new opportunities is preferred, the information is being used and thus improves the quality of our decision. Of course, we should not blithely assume that acquiring information is a good idea. Risk still matters, as we saw in Table 9.9. Moreover, our illustration has purposely assumed the information was costless, or free; and it well might be too costly. This cost might be explicit.[16] For example, we might have to pay for it, as when a consultant is hired. It also may take too much time to produce or decipher the information. This book, for example, contains considerable information (at least in the author's opinion), but cannot be thoroughly studied and

---

[15] Now glance back at Examples 9.2 and 9.3. There we concluded the safe choice was best, for the specified risk attitudes and zero initial wealth. Information, however, leads to the (a_1, a_2) policy for the same risk attitudes (and zero initial wealth). That is, information may lead to a more risky choice!

[16] Here it is important to remember risk aversion may vary with the wealth level, and paying for the information varies the wealth level. Suppose we pay C for the information in the running illustration. Then the gain in wealth will be $100 - C, 240 - C$, or $400 - C$.

deciphered in a few hours. The cost might also be highly implicit. In a strategic setting, it is possible that one player's acquisition of information alters another player's behavior to such an extent the player getting the information is harmed. To illustrate, it is more difficult to sell an automobile if the would-be buyer knows we just had a mechanic thoroughly check the auto.

9.6   An Important Aside

This leads to an important aside that reveals a great deal about the manner in which we are studying accounting. Suppose an integer between 1 and 100 is going to be picked at random. One information source will tell us whether the number is low (50 or below) or high (51 or above). Another information source will tell us whether the number is odd or even. The low/high source tells us nothing about whether the number is odd or even, while the odd/even source tells us nothing about whether the number is low or high.

Suppose we have a chance to bet on the number. If the bet is odd versus even we are in great shape with the odd/even information source; and if the bet is low versus high we are in great shape with the other information source. (Of course, if the other player knows we have this information, there will be no bet.)

Here's the rub. No matter which betting game we face, knowing both odd/even and low/high is as good as knowing just one or the other, and that is as good as knowing nothing. (Again, the other party to the bet is unaware we have access to this information.) There is more information in knowing both than in knowing just one; and there is more information in knowing one than in knowing neither. Unfortunately, comparing odd/even and low/high is problematic. We cannot say one has more or less information than the other.

Suppose we want to identify the best information source without saying too much about the context. No difficulty arises if we know there is more information in one than the other; we would always opt for the one with more information, presuming it is costless.[17] Yet we are not necessarily in the happy case of facing information choices that can be ranked from high to low in terms of the amount of information they offer. Odd/even versus high/low is a case in point. They tell us different things; and cannot say which is better without knowing the context.

Now recall the underlying idea of economic rationality: consistent preferences. We are always able to decide, and we do not cycle. No measure of

---

[17]Recall the earlier pithy comment on variable versus full costing systems, especially Table 6.11, where we noted doing both provides more information.

preference is possible without consistent preferences. Here we cannot decide between odd/even and high/low without knowing the context. This means we cannot make consistent statements about information sources in a context-free manner.

Yet in accounting we often find reference to accounting "principles." Treating accounting as a source of information, we immediately see economic forces preclude an ability to make general statements about which source of information, which accounting method, is best (odd/even versus low/high for example). The term, accounting principles, is a misnomer in the sense it conveys an ability to discern the preferred method of accounting without specifying the context. This is why we always carry context along in our discussion, and why we are reticent in making sweeping statements about the nature of good accounting practice.

Information cannot be studied without specifying the context in which it is to be used. Treating accounting as a source of information implies we cannot study accounting without specifying the context in which it is to be used.

9.7   Summary

Consistent framing of a decision problem under uncertainty is a natural and important extension of our work on decision framing. Here we add to our growing list of assumptions an independence requirement, so that the individual's preference measure can be separated into taste for the outcome (which means risk aversion in our stylized monetary outcome setting) and beliefs concerning the likelihood of that outcome. This results in a setting where preference is measured by the expected value of the utility defined on possible outcomes. The price, so to speak, is we require independence between tastes and beliefs along with consistency and smoothness.

Framing issues, in turn, now enter in two important ways. One is whether we find it more comfortable to assign probabilities to outcomes or to states. The other is whether we find it more comfortable to deal with expected utility per se or with certainty equivalents.

Our machinery, however, has now become quite powerful. It allows us to deal with risk, with aversion to risk, and with the equivalent idea of a risk premium. In addition, it allows us to deal with information in terms of systematically expanding decision alternatives, by allowing a choice to depend on what the information source has to say. This insight is driven by the fact that, under expected utility analysis, information affects beliefs but not tastes.

Some cautionary medicine is also in order. Expected utility analysis carries a price, and we must remember to interpret it as the economic foundation that underlies the art of decision making when uncertainty is

sufficiently important to affect that decision. It is also important to remember that our study of decision making to this point has been a closed system consisting of a single individual (or firm) confronted by choices and consequences. We have yet to introduce any other individual (or firm) into the story.

9.8   Bibliographic Notes

Rationality, especially the expected utility variant, is a controversial subject. Machina [1987] provides a review. Demski [1980] provides an introduction to the connections with measure theory and choice among information sources. Deeper treatments are available in many places; favorites are Krantz and associates [1971] and Kreps [1988]. Howard [1971] offers a compelling case for constant risk aversion. Behavioral finance is based on systematic deviations from expected utility analysis. See Shefrin [2005] and Ross' [2005] counterpoint.

9.9   Problems and Exercises

1. Define and contrast the terms certainty equivalent and risk premium.

2. How does information improve the quality of a decision? What is done in the absence of information? Continuing, a common colloquialism is that of "needed information." For example, accounting policy makers frequently describe their work as providing the information needed by investors. Is the idea of needed information consistent with economic rationality?

3. The text claims the term accounting principles is a misnomer, to the extent that it refers to an ability to design or specify the accounting method without specifying the context. Carefully explain this argument. Why is context so important in using and designing the accounting product?

4. *decision analysis*
Examples 9.1 through 9.3 all assume zero initial wealth. For each utility function determine the maximum value of probability α such that the risky choice is preferred.

5. *decision analysis*
Repeat problem 4 above, assuming initial wealth is $w_i = 400$. Comment on your findings.

6. *certainty equivalents*

   Ralph is anxious to display understanding of the mechanics of risk aversion. For this purpose, five distinct choices are available, with probabilities given in the table below. #4, for example, will result in $2,000$ with a probability of .60 and $3,000$ with a probability of .40. For convenience, Ralph's utility measure is given by $U(w) = \sqrt{w}$, with zero initial wealth. Determine the certainty equivalent for each of the possible choices.

	1,000	2,000	3,000	4,000
#1	.25	.25	.25	.25
#2		.50	.50	
#3		.40	.60	
#4		.60	.40	
#5		.40		.60

7. *information*

   Before making his choice in problem 6 above, Ralph encounters an oracle who will tell him in advance what the outcome of choice #1 will be. The oracle has long admired Ralph and offers this important service free of charge. Now what should Ralph do, and what is the resulting certainty equivalent?

8. *certainty equivalents*

   Ralph is contemplating a lottery. A fair coin will be tossed. If the coin shows "heads," Ralph will be paid 100 dollars. If the coin shows "tails," Ralph will be paid nothing. So the expected value of this lottery is $.5(100) + .5(0) = 50$.

   (a) Initially suppose Ralph's utility for wealth is given by $U(w) = \sqrt{w}$ and that Ralph's initial wealth is zero. Determine Ralph's certainty equivalent and risk premium for this lottery. Why is $CE < 50$? Also, why is $CE > 0$?

   (b) Using the same root utility, now determine Ralph's risk premium for initial wealth $w_i \in \{0, 5, 10, 25, 50, 100, 500, 1,000\}$. Interpret your finding.

   (c) Now let $U(w) = -\exp(-\rho w)$, with $\rho = .01$. Repeat the construction in (b) above. Interpret your finding.

   (d) Again let $U(w) = -\exp(-\rho w)$. Determine Ralph's risk premium for $\rho \in \{.0005, .001, .01, .06, .1, 1\}$. Interpret your finding. What happened to initial wealth in your construction?

9. *information with joint probability frame*

   Return to the risky versus safe choice originally chronicled in Table 9.1. Let initial wealth be zero and $\alpha = .5$. Before deciding Ralph will

observe an information source that will report a signal of g or b. The joint probability of dollar gain (100 or 400) and signal (g or b) under the risky choice is displayed below. For example, the joint probability of 100 and g is $.2\alpha$. The riskless choice, of course, continues to offer a dollar gain of 240. Further assume Ralph's initial wealth is zero.

	100	400
signal g	$.2\alpha$	$1 - \alpha$
signal b	$.8\alpha$	0

(a) Determine Ralph's optimal use of this information for each of the utility specifications in Table 9.9. Also determine the resulting certainty equivalent for each such specification

(b) Write a short paragraph detailing how your analysis differs, or does not differ, from our work in Tables 9.8 and 9.9.

10. *perfect information*
Change the probabilities in Table 9.8 such that the risky choice continues to deliver 400 with probability $1 - \alpha$, but that the information is perfect, meaning it will perfectly reveal the outcome of any risky choice. Do the same for the joint probability display in problem 9 above.

11. *scaling the utility function*
A seemingly awkward part of using the negative exponential utility function is the fact it is negative. Return to Examples 9.3 and 9.5 but consider a utility function of $U(w) = 10 - \exp(.001w)$. Determine the expected utility and certainty equivalent for each of the choices. Explain your findings, relative to the expected utility and certainty equivalents calculated in the original examples.

12. *normal density*
Ralph must select between two lotteries. Either one will net him some cash in the amount w, where w is a normally distributed random variable. The first lottery has a mean of $\mu_1 = 200$ and a variance of $\sigma_1^2 = 50$ while the second has a a mean of $\mu_2 = 210$ and a variance of $\sigma_1^2 = 150$. Ralph displays constant absolute risk aversion, and thus uses a utility function of the form $U(w) = -\exp(-\rho w)$. Determine ρ such that Ralph is indifferent between the two lotteries.

13. *information use*
Return to the setting of Table 9.7, when $\alpha = .5$. This implies respective state probabilities of $.4, .1, .5$ and 0. Now change these respective state probabilities to $.4, .2, .2$ and $.2$.

(a) Using the four utility functions in Table 9.9, and assuming zero initial wealth, calculate the certainty equivalent for each choice when no information is present.

(b) Now assume the information is present. Again using the four utility functions (and zero initial wealth), determine the certainty equivalent for each of the possible policies. Discuss your findings.

(c) What happens in (b) above if initial wealth is very large?

14. *useful and useless information*
Ralph faces a choice problem in which the dollar outcome is uncertain. Ralph thinks of the uncertainty as reflecting natural and economy-wide events. For simplicity, four such events or states are possible, denoted in usual fashion. The four states are equally likely, and Ralph is risk neutral. The outcome structure is displayed below, where you will notice the possible choices are denoted a_1, a_2 and a_3.

	state s_1	state s_2	state s_3	state s_4
a_1	225	225	225	100
a_2	900	900	100	100
a_3	625	625	625	100

(a) Determine Ralph's best choice.

(b) Now suppose Ralph can purchase information before making a choice. The information source will tell whether the actual state is s_1 or s_2 versus s_3 or s_4. You should think of this as telling Ralph whether the state will be "low" or "high" in terms of the indexing system. Put differently, the information will tell Ralph whether the outcome is confined to the two left-hand or the two right-hand columns of the outcome table. How much would Ralph pay for this information (i.e., at what price is he indifferent between buying and not buying the information)?

(c) Consider the case where the information costs 500. Ralph will not pay such a price. The information is not needed. What does Ralph substitute for the lack of information?

(d) Suppose the information structure will reveal whether s_4 is true, that is, whether life is in the left three columns or the right-most column of the outcome table. How much would Ralph pay for this information?

15. *value of information*
Ralph is contemplating four possible choices, cleverly labeled one, two, three and four. The outcome of any choice depends on the state of the economy. For analysis purposes, Ralph models this as four

equally likely states. The net gain to Ralph, as a function of the option chosen and state of the economy, is displayed below.

	s_1	s_2	s_3	s_4
one	100	90	30	20
two	30	20	100	90
three	30	150	30	30
four	30	30	30	150

(a) Calculate the expected utility for Ralph for each of the four choices. Assume here, and only here, that Ralph is risk averse with utility measured by the square root of the gain. What is the certainty equivalent for each of the choices?

(b) Again calculate the expected utility for Ralph for each of the four choices. But assume here and for all remaining parts of the exercise that Ralph is risk neutral.

(c) Suppose before acting Ralph can learn, from an expert fore-caster, whether the state of the economy will be in $\{s_1, s_3\}$ or will be in $\{s_2, s_4\}$. Notice the mnemonic of "odd" or "even." If the forecaster says odd, for example, then Ralph knows the state will be s_1 or s_3, just not which one. What is the maximum amount Ralph would pay for this forecaster's service?

(d) Suppose instead of the "odd" or "even" story a second fore-caster is able to forecast whether the state will be in $\{s_1, s_2\}$ or in $\{s_3, s_4\}$. Notice the mnemonic here of "low" or "high." What is the maximum amount Ralph would pay for this second fore-caster's service? This should be answered on the assumption Ralph does not hire the first forecaster.

(e) Finally, suppose the two forecasters jointly approach Ralph and offer to combine their services. What is the maximum amount Ralph would pay for the joint forecasting service?

(f) You now have three value of information calculations: the first source, the second source, and the two sources together. Notice additivity is not present. The sum of the first two values does not equal the third. Why is additivity not present here?

16. *constant risk aversion and value of information*
Repeat problem 15 above for the case where Ralph's utility is neg-ative exponential, $U(w) = -\exp(-\rho w)$ with $\rho = .001$. (Hint: when deriving the most Ralph would pay for the information it is easiest to convert the no information and information cases to certainty equiva-lents and then take the difference. This short cut depends on constant risk aversion; see the following problem.)

17. *complications with nonconstant risk aversion*
Return to problem 16 above. Now assume Ralph is risk averse with utility function $U(w) = \sqrt{w}$ coupled with zero initial wealth. Determine the maximum amount Ralph would pay for the "low" versus "high" information.

18. *dominance*
Ralph is contemplating various lotteries. The possible prizes are 100, 200, 300 or 400 dollars. Assume in what follows that more is strictly preferred to less dollars. Below are some representative lotteries (For example, lottery c yields 100 with probability .2 and 300 with probability .8.

	100	200	300	400
a	.1			.9
b	.2			.8
c	.2		.8	
d	.25	.25	.25	.25
e	.15	.35	.25	.25
f	.24	.21	.25	.30

(a) Consider lotteries $a, b,$ and c. Which choice from among the three is best? Try an expected utility analysis with several different $U(w)$ functions. Why does lottery a always turn out to be best?

(b) Do the same thing for lotteries d and f.

(c) Now consider lotteries e and f. Exhibit one $U(w)$ function such that e is preferred to f and another such that the opposite holds. What is the explanation?

19. *substituting an expected value for a random variable*
Suppose we want to maximize the expected value of $\theta a - a^2$, over $a \geq 0$, where θ is a random variable. So we want to solve

$$\max_{a \geq 0} E[\theta a - a^2]$$

We now frame this by substituting the random variable's expected value for the random variable. Let $\bar{\theta}$ denote the expected value of θ. Solving

$$\max_{a \geq 0} \bar{\theta} a - a^2$$

will of course locate the solution to the original problem. Discuss the principle of consistent framing that is being employed. What happens to the transparent substitution when risk aversion is present?

20. *expected rate of return*
Ralph is contemplating loaning a cousin 10,000. The loan would be

due in one year, with interest at 18%. Ralph figures the probability the cousin will pay back the loan (plus interest) is .80; with probability .10 only the principal will be paid back; and with probability .10 nothing will be paid by the cousin. Ralph's next best use of the $10,000$ is to invest it at the risk free rate of 4%. Ralph notes that the expected payment in one year is $10,000(1.18)(.8) + 10,000(.1) + 0(.1) = 10,440$. So, he concludes, the funds can be invested at 4% or at 4.4%. The latter is a winner. Carefully discuss how Ralph has used the principles of consistent framing in reducing this to a comparison of expected interest rates. As a starting point, assume Ralph's preferences are measured by the present value of expected wealth, and Ralph has a variety of investments in place.

21. *inconsistent framing attempt*
Ralph manages a two product enterprise. The first (q_1) sells for 400 per unit and the second (q_2) sells for 600 per unit. Estimated unit costs are as follows:

	q_1	q_2
direct labor	80	120
direct material	100	150
overhead	160	240

Overhead is applied to each product on the basis of direct labor dollars (at a rate of 200%), while the overhead LLA is given by

$$OV = 54,000 + .5DL\$$$

where $DL\$$ denotes direct labor dollars. In addition, marketing costs are described by

$$MKT = 30q_1 + 90q_2$$

The firm employs two production departments. The first has a capacity of 400 direct labor hours and the second a capacity of 500 direct labor hours. The first product uses one hour in each department, while the second uses one hour in the first department and two hours in the second, so we must observe

$$q_1 + q_2 \leq 400$$
$$q_1 + 2q_2 \leq 500$$

(a) Determine an optimal output schedule for Ralph.

(b) Ralph is worried about the overhead function in the above problem. To think some more about this, assume actual overhead is one of the following two models, with equal probability: $OV = 63,000 + .25DL\$$ or $OV = 45,000 + .75DL\$$. Absent any additional information, Ralph will implement the schedule determined in part (a) above. How much would Ralph pay for a cost

study that would perfectly reveal which of the two overhead models is in fact correct?

(c) Ralph now tries another exercise. Instead of worrying about the overhead estimate, the estimates of available capacity in the two departments are called into question. Suppose department two's estimate is correct, but the estimate for department one is ambiguous. With equal probability, it will be either 350 or 450 hours. I want to ask you how much Ralph would pay to learn the actual capacity. But this cannot be answered without additional specification. Why can we not answer this question? What is the framing issue?

10
Consistent Framing in a Strategic Setting

We conclude our study of decision framing by introducing the possibility of competitive response, or strategic encounter. Adding a product to our product line may affect one or more competitors and may evoke a response. For example, they may decide to cede the new product's market to us or they may retaliate with vigor. Extensive investment in $R\&D$ may scare off potential competitors, or may lead to even more investment by the competitors. Access to proprietary technology may give us a cost advantage.

The list goes on, and the anecdotes are endless: airline pricing, proprietary versus open architectures in the computer industry, financial aid to students, curriculum development, shortened product development cycles in the auto industry, auction of treasury bills, community based policing, buying or selling a used car, or designing a political campaign.

The common feature in these settings is the consequences of what we do depend in part on what someone else does. One might think the way to proceed is to assign a probability to what this someone else might do and then proceed. Game theory takes a more consistent approach, by simultaneously analyzing the situation from each player's perspective and by combining these perspectives with the notion of equilibrium behavior. In this way, the analysis renders as endogenous the description of what someone else might do.[1]

---

[1] In a related vein, the expected utility view of choice behavior separates an uncertain choice problem into tastes and beliefs. The outcome depends on the choice and on how the uncertainty plays out. There is, however, no notion in which odds are altered by the choice taken. Murphy's Law (the quintessential expression of apprehension) is

We begin with the idea of equilibrium behavior. We then illustrate equilibrium thinking and analysis in a variety of settings: competitors sharing a market, competitors racing to capture a market, competitors bidding for a prize and adversaries haggling over a transaction. From here we broach the subject of internal control, a topic that is central to the accounting library's integrity and, yes, another illustration of equilibrium behavior.

Keep in mind this is all about framing a decision. The tension in such an exercise is always about how much detail to admit into the analysis, be it formal or informal. When the behavior of others is a first order issue, the appropriate frame widens significantly as the behavior of those other players becomes an essential ingredient in the frame.

10.1   Equilibrium Behavior

In the prior two chapters we have concentrated on what is, essentially, a single person decision problem. We viewed the individual as having identified a choice that must be made, and framed that choice in terms of the imperative to select one and only one member from an identified set of alternatives denoted A. We further refined this by invoking sufficient regularity that would allow us to describe this behavior as though it were governed by a criterion function, $\omega(a)$. And this led to choice being modeled as $\max_{a \in A} \omega(a)$.

In turn, the outcome from this choice might well be uncertain, and in a pithy sort of way we refer to the outcome as jointly produced by the individual's choice and by the state that obtains. All of this architecture remains in place, but we now introduce other individuals into the story.

At this juncture just one more decision making individual will do the trick. So now, in an equally pithy sort of way, we refer to the outcome (now one for each individual) as jointly produced by the first individual's choice, by the second individual's choice, and by the state that obtains. Suppose, for example, that you and I bid on a customer's special project. Whether you or I win the competition depends on each of our bids. And how much I make on the project depends on whether I actually won the bid and, if so, whether my cost turned out to be high or low. My outcome thus depends on my choice, on your choice and on the state. Likewise, your outcome depends on my choice, on your choice and on the state.

To smother this setup with notation, think of individual 1 as facing a choice of $a_1 \in A_1$ and individual 2 as facing a choice of $a_2 \in A_2$. The

---

inconsistent with the model, in that it assumes events such as the weather will unfold to cause the most harm given the choice taken. Game theory now enters in the guise of one player's outcome also being influenced by another player's choice.

two individuals simultaneously make their choices. Individual 1's criterion function is denoted $\omega_1(a_1, a_2)$ and individual 2's is denoted $\omega_1(a_1, a_2)$.[2] Note well, individual i's criterion function depends on his choice and on the other individual's choice. In addition, the outcome for either player may well be uncertain, and if so we presume the criterion functions are expected utility measures, thereby accommodating whatever risk attitude is pertinent to the story.

Example 10.1 Suppose our two individuals each face a pair of alternatives. The structure of such an encounter is usually displayed with a simple bimatrix convention (and these are often called bimatrix games). We do this in the display below. It is decoded as follows. Player 1 is called *Row*, and he has two choices: *up* or *down*. Player 2 is called *Column*, and she has two choices: *left* or *right*. In any cell, the numbers are the expected utilities of the two players, listed in the order of *Row* followed by *Column*. For example, if *Row* plays *down* while *Column* plays *right*, *Row*'s criterion measure (expected utility) is 6 while *Column*'s is 1. Notice we have specified the players, their available choices, and their respective criterion measures for each possible combination of choices.[3]

	left	*right*
up	1,0	4,5
down	2,4	6,1

10.1.1   Simultaneous Choice

Two assumptions are now invoked. One is that the individual players know *both* their own and the other individual's possible choices (A_1 and A_2) and their own and the other individual's criterion functions ($\omega_1(a_1, a_2)$ and $\omega_2(a_1, a_2)$). In Example 10.1, this means *Row* and *Column* know the bimatrix. The second assumption is that the choices are made simultaneously and in making their choices the players will engage in equilibrium behavior. Loosely this means *Row* will behave in a manner best for *Row*, given what *Column* is doing; *Column* will behave in a manner best for *Column*, given what *Row* is doing; and the two sets of behavior and expectations will be consistent.

More precisely, the pair of choices (a_1^*, a_2^*) is a (Nash) equilibrium in the noted encounter if

$$a_1^* \quad \in \quad \arg \max_{a_1 \in A_1} \omega_1(a_1, a_2^*) \qquad\qquad (10.1a)$$

$$a_2^* \quad \in \quad \arg \max_{a_2 \in A_2} \omega_2(a_1^*, a_2) \qquad\qquad (10.1b)$$

---

[2] We are describing a 2-person, normal-form game: each player has a specified set of choices or strategies along with a criterion function that depends on each player's choice.

[3] Thus, the entry in each sell is the pair of criterion measures, $\omega_1(a_1, a_2), \omega_2(a_1, a_2)$.

The idea is mutual best response. Consider individual 1. If (a_1^*, a_2^*) is indeed an equilibrium, then his half of the pair, a_1^*, should be the best he can do given individual 2 will choose a_2^*. That is, a_1^* should be a solution to $\max_{a_1 \in A_1} \omega(a_1, a_2^*)$. We state this in the seemingly awkward though very useful arg max notation in expression (10.1a): a_1^* is a feasible value (of the argument a_1) for which $\omega_1(a_1, a_2^*)$ attains its maximum value.[4] Likewise, for individual 2's half of the equilibrium, we require a_2^* be a solution to $\max_{a_2 \in A_2} \omega_2(a_1^*, a_2)$.

Example 10.2 Return to Example 10.1. $(down, left)$ is an equilibrium. If *Row* plays down, *Column*'s best response is to play *left*, as $4 > 1$. Also, if *Column* plays left, *Row*'s best response is to play *down*, as $2 > 1$. *Down* is best for *Row*, given *Column* is playing *left*; and *left* is best for *Column*, given *Row* is playing *down*. Notice how *Row* expects *Column* to play *left* and plays accordingly, and *Column* expects *Row* to play *down* and plays accordingly. Mutual best response is the theme, precisely as depicted in the two halves of expression (10.1).[5]

Games of this sort (two players, each with a finite number of choices, and simultaneous play) may have one equilibrium, as in our illustration. They may have no equilibrium, unless we allow randomized strategies. They also may have multiple equilibria. When invoking equilibrium behavior, we will structure our settings so an equilibrium exists without randomization. Also, in dealing with incentive games we will encounter multiple equilibria, but it will be clear from the context which equilibrium should be emphasized. In general, however, these games are not necessarily as friendly as our example suggests.[6]

10.1.2   Sequential Choice

It is also important to understand the rules of the game matter. To illustrate, suppose we change our simultaneous choice story to one where the first individual makes his choice, the second individual observes that choice and then, with that knowledge, makes her own choice. We will see, in later chapters, that this simple device of sequential choice is the key to understanding the economic forces in performance evaluation.

---

[4]There is no reason to suspect a_1^* is the only (feasible) choice that maximizes $\omega_1(a_1, a_2^*)$, so we require that it be a member of the set of all such maximizers. That is, literally, what (10.1) expresses.

[5]The task of finding an equilibrium here is helped by the fact *Row* has a dominating choice. No matter what *Column* does, *Row* is better off playing *down*: $2 > 1$ and $6 > 4$.

[6]Change *Row*'s expected utility in the lower right cell from 6 to 1. $(down, left)$ remains an equilibrium, but now $(up, right)$ is also an equilibrium. The latter is better for both players, though in general there is reason to expect conflict over which equilibrium is best. Conversely, change *Row*'s expected utility in the upper left cell from 1 to 3. Now we have no equilibrium, absent randomized strategies.

Example 10.3 Return to Example 10.2 but now assume *Row* moves first. What this means is *Row* decides, and that decision is observed by *Column* before *Column*'s decision. Suppose *Row* selects *up*. *Column*'s best response is surely to play *right*. *Row*'s expected utility is 4. Alternatively, if *Row* plays down, *Column*'s best response is to play left. *Row*'s expected utility is $2 < 4$. The equilibrium is now *up* followed by *right*.[7] Here both players prefer sequential play, with *Row* moving first. Alternatively, both prefer *Row* be able to commit to a particular choice. This removes an element of strategy from the encounter and improves both players' prospects.[8]

Notice, as in Example 10.3, how locating an equilibrium is an easier task under sequential choice. For the sake of argument, continue to assume individual 1 makes the initial choice. For any such choice, say $\widehat{a}_1 \in A_1$, individual 2 faces the problem of

$$\max_{a_2 \in A_2} \omega_2(\widehat{a}_1, a_2)$$

Let $R(\widehat{a}_1)$ be a solution. Repeat this for every possible initial choice of $\widehat{a}_1 \in A_1$. This gives us a "reaction function" of $R(a_1)$. And with this reaction function identified, individual 1 now faces the problem of

$$\max_{a_1 \in A_1} \omega_1(a_1, R(a_1))$$

Let a_1^* be the solution to this problem. This gives us an equilibrium of $(a_1^*, R(a_1))$. And it provides an equilibrium outcome of $\omega_i(a_1^*, R(a_1^*))$ for individual i. This is how we located the equilibrium in Example 10.3.[9]

You should recognize this as a strategic version of the third principle of consistent framing: component searches are possible. We basically remove the strategic element by transforming the choice setting to a single person setting, with the second individual's choice represented by the reaction function. Of course, sequential choice is the key that allows us to proceed in this fashion.

10.1.3   Repeated Choice

Another variation on the rules of the game concerns repetition. To give a brief flavor of the possibilities, suppose the simultaneous play game will

---

[7]So much for dominating strategies!

[8]Contrast this with the case where *Column* moves first. The equilibrium is *left* followed by *down*. Simultaneous or sequential play with *Column* moving first is a matter of indifference to the two players.

[9]Here we engage in some serious subtlety. Technically, an equilibrium in the sequential choice case consists of a choice by the first individual and a reaction function for the second, as the second must specify her choice for every possible observed choice by the first individual. This is what accounts for our identification of the equilibrium and the equilibrium outcome.

be repeated twice, and that before entering the second round both players know the choices made in the first round. Let's concentrate on the setting of Examples 10.1 and 10.2, where the equilibrium is $(down, left)$. Notice $(up, right)$ is preferred by both individuals. So suppose individual 2 announces she will play $right$ on the first round and will do so again on the second round, provided individual 1 plays up. Once we enter the second round, whatever happened in the first round is history and the second round equilibrium is nothing other than $(down, left)$. But realizing this, individual 2's attempt at improving both lives is a fantasy. $(down, left)$ is equilibrium play in the first stage as well. Indeed, if the underlying game has a unique equilibrium, then repeating that game a finite (though known) number of times has a unique equilibrium outcome produced by the individuals repeatedly playing the underlying single period or one shot game's equilibrium.

Don't miss the framing implication. In such a repeated setting the strategic encounter can be effectively condensed to a single encounter.[10]

10.2   Sharing a Market

Competing for market share is a natural illustration where the importance of equilibrium analysis comes to the fore. Suppose we have two firms, competing in terms of how much output they place on the market. Think of this as two producers of a perishable commodity, such as fish. Each player's catch is turned over to a wholesaler, who clears the market. Naturally, the market price depends on how many units are offered for sale.

Let q_1 denote the quantity produced by the first firm and q_2 that produced by the second. So total quantity is $q = q_1 + q_2$. With q units placed on the market, the market clearing price is given by the inverse demand function of $\widehat{P}(q)$.[11]   The firms are otherwise identical, with cost curves denoted in the usual fashion, i.e., $C(q_i; P)$, where, recall, P denotes the factor prices. Thus, with firm 1 placing q_1 and firm 2 placing q_2 units on

---

[10] This statement depends on the game being repeated a finite, known number of times. Random abandonment or no abandonment are different stories. You should also be aware that we are entering the world of extensive form games (where we lay out the game "tree" in full detail) and are dealing with subgames (where a subgame is any possible continuation of the game). For that matter, we should also flag the possibilities of some players knowing more than others at different decision points in the game, possibilities that give rise to refinements of the basic equilibrium concept we have articulated.

[11] For the record, a demand function gives quantity as a function of price; and an inverse demand function gives price as a function of quantity. In (the forthcoming) Example 10.4 we assume an inverse demand of $\widehat{P}(q) = 340 - 2q$, which is equivalent to a demand function of $D(\widehat{P}) = .5(340 - \widehat{P})$.

the market, the profit of firm i will be its revenue less its cost, or

$$\Pi_i(q_1, q_2) = \widehat{P}(q_1 + q_2) \cdot q_i - C(q_i; P) \qquad (10.2)$$

Importantly, the market price depends on combined output, and thus each firm's profit depends on its output and, through the market price, on the output of its competitor.

Now suppose the firms view quantity produced and sold as the strategic variable, and simultaneously place their output on the market. The pair of output quantities, (q_1^*, q_2^*), then, constitutes an equilibrium if each is a best response to the other, or

$$q_1^* \in \arg\max_{q_1 \geq 0} \Pi_1(q_1, q_2^*) \qquad (10.3a)$$

$$q_2^* \in \arg\max_{q_2 \geq 0} \Pi_2(q_1^*, q_2) \qquad (10.3b)$$

You should recognize this as a direct application of our equilibrium expression in (10.1).

Example 10.4 Suppose the inverse demand function, or price, is given by $\widehat{P}(q) = 340 - 2(q_1 + q_2) = 340 - 2q$ and that each firm's cost is given by $C(q_i; P) = 100q_i$. This implies, using (10.2), that firm i's profit is given by

$$\Pi_i(q_1, q_2) = (340 - 2(q_1 + q_2))q_i - 100q_i$$

If q_1^* is a best response to q_2^*, it must, as (10.3a) requires, maximize $\Pi_1(q_1, q_2^*)$. This means the derivative vanishes at q_1^*:[12]

$$\frac{\partial \Pi_1(q_1, q_2^*)}{\partial q_1}\Big|_{q_1 = q_1^*} = 240 - 4q_1 - 2q_2^* = 0$$

Likewise, if q_2^* is a best response to q_1^*, it must, as (10.3b) requires, maximize $\Pi_2(q_1^*, q_2)$. This means the derivative vanishes at q_2^*:

$$\frac{\partial \Pi_2(q_1^*, q_2)}{\partial q_2}\Big|_{q_2 = q_2^*} = 240 - 2q_1^* - 4q_2 = 0$$

Solving the two equations in two unknowns provides an equilibrium of $q_1^* = q_2^* = 40$.[13]

---

[12] Let's be clear about this. Given q_2^*, we differentiate firm 1's profit expression:

$$\frac{\partial \Pi_1(q_1, q_2^*)}{\partial q_1} = 340 - 2(q_1 + q_2^*) - 2q_1 - 100 = 240 - 4q_1 - 2q_2^*$$

Now, this must equal zero when $q_1 = q_1^*$. Hence the noted expression.

[13] The solution is symmetric because the cost curves are the same. What would happen if firm 2's marginal cost were 150 instead of 100? Also, to dig a bit deeper with the present case, you can verify the claimed equilibrium by setting, say, $q_2^* = 40$ and performing the maximization indicated by expression (10.3). Try it. You should also find $\Pi_1(q_1^*, q_2^*) = 3,200$.

Several points are in order. First, notice how we attack the problem by simultaneously thinking through the details from our perspective and from that of our rival. Second, we rely on the market to set the price. Competition occurs in terms of quantities that are simultaneously brought to the market.[14] Finally, you should interpret this as a heavily stylized version of two firms competing for market share. For simplicity we presume but two competitors with no further entry, known demand and cost curves, and an understanding that competition takes place using quantity as the strategic variable. But the principle of simultaneously thinking through one's position and that of one's rival holds. This is the essence of strategic framing.

10.3 Racing to Capture a Market

A closely related theme is where competitors race to be the first to market, in a setting where first-mover advantage leads to capture of the market. Pharmaceuticals and, more generally, major patents are illustrations.

To tease out a simple version of such a story, suppose the first to market garners a prize in the amount $\widehat{P}$. Think of $\widehat{P}$ as the value of the patent. Coming in second or worse is of no value whatsoever. Two competitors are present. Eventually one of them will secure the patent (though a guaranteed winner is hardly essential in what follows). The probability of winning the race depends on how much the firm invests in $R\&D$ relative to its competitor. Denote the two investments, respectively, by z_1 and z_2. We assume the probability that firm i wins the race is given by

$$p_i(z_1, z_2) = \frac{1 + z_i}{1 + z_1 + 1 + z_2} = \frac{1 + z_i}{2 + z_1 + z_2}$$

This implies both firms are equally adept at investing in $R\&D$, and that absent a competitor the sole firm could acquire the patent at zero marginal cost, not very realistic but sufficient for our purpose.

Regardless, firm i's expected profit from the race is simply

$$\Pi_i(z_1, z_2) = p_i(z_1, z_2) \cdot \widehat{P} - z_i = \frac{(1 + z_i)\widehat{P}}{2 + z_1 + z_2} - z_i \qquad (10.4)$$

From here we assume the two competitors simultaneously make their $R\&D$ investment decisions. This leads to what, hopefully, is becoming a

---

[14]Historical note: this story, called Cournot competition, originates in 1838, in Cournot's *Recherches sur les Principes Mathematiques de la Theorie des Richesses.* Another story, called Bertrand competition, would have the rivals compete by announcing a price. This gets awkward when the prices differ, because we must then specify what happens following the price announcements (assuming homogeneous goods). Presumably, the entire market goes to the low cost announcer. Conversely, if the prices are the same, we must resolve how total demand is split between the rivals.

familiar equilibrium specification:

$$z_1^* \quad \in \quad \arg\max_{z_1 \geq 0} \Pi_1(z_1, z_2^*) \tag{10.5a}$$

$$z_2^* \quad \in \quad \arg\max_{z_2 \geq 0} \Pi_2(z_1^*, z_2) \tag{10.5b}$$

Example 10.5 Suppose the prize is $\widehat{P} = 1,000$. Each firm's investment must be a best response to the other's, and this implies, once again, that the derivatives of their respective profit functions should vanish at the equilibrium investments. Paralleling what we did in Example 10.4, we have the following pair of conditions

$$\frac{\partial \Pi_1(z_1, z_2^*)}{\partial z_1}|_{z_1=z_1^*} = \frac{(1 + z_2^*)\widehat{P}}{(2 + z_1 + z_2^*)^2} - 1 = 0$$

$$\frac{\partial \Pi_2(z_1^*, z_2)}{\partial z_2}|_{z_2=z_2^*} = \frac{(1 + z_1^*)\widehat{P}}{(2 + z_1^* + z_2)^2} - 1 = 0$$

This provides an equilibrium of $z_1^* = z_2^* = 249$.

10.4   Bidding for a Prize

Bidding is a common trade arrangement: remodeling contractors, professional movers, supply arrangements with auto manufacturers, audit engagements, airframes and engines for airlines, and what have you. And, as you no doubt suspect, equilibrium analysis is central to the story.

To explore this theme, we examine a bidding competition with but two bidders. To set the stage, suppose a potential customer has asked for bids on a special project. Think of this as a construction project, a specialized machine, a custom fabricated product, or a specialized service. It is unique and must be supplied according to the customer's specifications.

The rules are simple. We bid or not. Any bid is prepared without knowledge of any other bid, so bidding takes place simultaneously. The buyer examines the bids. Among those submitting bids, the low bidder wins the contract. The low bidder is paid the winning (low) bid, and supplies the product or service, as specified. In the unlikely event of a tie, one potential winner is randomly selected.[15]

---

[15] Thus, we are dealing with a sealed bid auction. Though we emphasize the sealed bid mechanism, be aware that various types of auctions are used. In an English auction, the auctioneer begins with a low price and solicits successively higher bids. In a Dutch auction, the auctioneer starts out high and lowers the price until someone takes the object. A second price auction is one in which the highest bid wins, but pays the second highest price.

As noted, there are two potential suppliers, firm 1 and firm 2. They are both risk neutral. And each faces an identical cost structure, with incremental cost given by

$$\Delta = \alpha x + \beta y + \gamma z \tag{10.6}$$

α, β and γ are known, positive constants. x, y, and z are independent, identically distributed random variables; with uniform densities between 0 and 1. Such a density has $f(t) = 1$, for $0 \le t \le 1$, and $f(t) = 0$ otherwise.[16] And for later reference, recall that this density has an expected value of

$$E[t] = \int_0^1 tf(t)dt = \int_0^1 tdt = 1/2$$

Thus, each firm's expected incremental cost is simply

$$E[\Delta] = E[\alpha x + \beta y + \gamma z] = \alpha E[x] + \beta E[y] + \gamma E[z] = (\alpha + \beta + \gamma)/2$$

This cost structure probably appears awkward and unintuitive. It is chosen to illustrate a number of points once we understand the bidding behavior. Be patient.

10.4.1 Uninformed Bidders

Now suppose this is all there is to the story: two risk neutral, basically identical firms bidding for the prize. The incremental gain to either firm depends on both bids, as low bidder wins and is paid the winning bid. The first firm's incremental expected gain, then, is given by

$$\Pi_1(b_1, b_2) = \begin{cases} 0 \text{ if } b_1 > b_2 \\ b_1 - \Delta \text{ if } b_1 < b_2 \\ .5(b_1 - \Delta) \text{ if } b_1 = b_2 \end{cases}$$

That is, the first firm's gain is 0 if $b_1 > b_2$, is $b_1 - \Delta$ if $b_1 < b_2$, and is half this amount if $b_1 = b_2$. The second firm's incremental gain follows in parallel fashion.

Further observe that life is often not as well-defined as our story suggests. The specifications may change, at the behest of the supplier, the customer, or both. They may even be negotiated before bids are submitted. The total number of units might be variable, as might the required completion date. Either party might breach; the parties might disagree about quality. The winning bidder might turn around and subcontract with one of the competitors.

[16] This extended illustration is patterned after an auction illustration in Myerson [1991, pp. 132-136], in which two competitors bid for a single object (e.g., an art object) with an unknown but common value.

Now apply our equilibrium definition in (10.1), but remember that the cost is uncertain so we work with the expected value of the gain, $E[\Pi_i(b_1, b_2)]$. In particular, (b_1^*, b_2^*) is a pair of equilibrium bids provided

$$b_1^* \in \arg\max_{b_1} E[\Pi_1(b_1, b_2^*)] \tag{10.7a}$$

$$b_2^* \in \arg\max_{b_2} E[\Pi_2(b_1^*, b_2)] \tag{10.7b}$$

Notice two things. First, bidding the expected cost is equilibrium behavior: $b_1^* = b_2^* = E[\Delta]$. Given one firm is so bidding, the other can do no better than do the same thing.[17] In equilibrium, competition between the two equally informed bidders ensures zero profit. Second, in the equilibrium definition here we have used the expected value of each firm's expected gain, independent of whether the proffered bid is a winning bid. This subtle point will haunt us shortly.

Example 10.6 Suppose the cost expression in (10.6) is given by $\alpha = 10, \beta = 10$ and $\gamma = 40$, thus implying an expected cost of $E[\Delta] = 30$. Now suppose one of the firms will bid 30. The other firm can do no better than bid 30 as well. After all, a bid above 30 is a guaranteed gain of 0 while a bid below 30 is a guaranteed bad deal. Both firms bidding their (identical) expected cost is an equilibrium. The power of competitive sourcing is evident, as the two competitors are perfect substitutes and competition between them leads to trade at the absolute minimum price possible.[18]

10.4.2 Equilibrium Bidding with Private Information

Now change the story so the two firms are not equally (un)informed. Before bidding, firm 1 now observes x and y; and firm 2 observes x and z. Both know x; firm 1 also knows y; and firm 2 also knows z. In addition, firm 1 knows firm 2 is privately observing z, and so on. This is a story in which each bidder knows something about the cost that the other does not know.

For example, one firm might be better at engineering and the other at fabrication. Both know something about the product in question (x), the first also has insight into the engineering part of the story (y), and the second has insight into the fabrication part of the story (z). Notice how our simple model captures this intuitive idea. Firm 1 sees y, but not z and firm 2 sees z, but not y. Both see x.

So what happens when firm 1 observes x and y, while firm 2 observes x and z? Since the three random variables are independent, firm 1 now

---

[17]Think back to the work we have done on product costing. This is a delicate art, suggesting the importance of cost estimation in a bidding exercise. Our illustration, however, leaves no room for product costing debate. We know the incremental cost is $\Delta = \alpha x + \beta y + \gamma z$, period.

[18]What would happen if firm 1's cost were 30 while firm 2's cost were 35?

perceives an expected cost, given x and y, of

$$E[\Delta|x,y] = \alpha x + \beta y + \gamma E[z] = \alpha x + \beta y + \gamma/2$$

Similarly, firm 2 now perceives an expected cost, given observation of x and z, of

$$E[\Delta|x,z] = \alpha x + \beta E[y] + \gamma z = \alpha x + \beta/2 + \gamma z$$

(Notice how the assumption x, y and z are mutually independent simplifies these calculations. Also notice our earlier story in which both firms knew the same thing can be interpreted as one in which $\beta = \gamma = 0$.)

With this preamble, we turn to the question of equilibrium bidding behavior. Two interrelated issues surface. One is that each firm is now bidding based on private information but against a competitor who is also in possession of private information. This means an equilibrium will be described in terms of a pair of bidding functions or strategies, not a specific pair of bids as in the uninformed bidders case.[19] The second issue is the fact that the bids themselves now convey information, because they are based on the firms' private information.

To see this more clearly, suppose firm 1 submits a bid of $b_1 = b$. If $b_2 < b$, it loses the auction, and gains nothing. If $b_2 > b$, it wins the auction outright. What does it gain? It gains the customer's payment less the cost, or $b - \Delta = b - \alpha x - \beta y - \gamma z$. From firm 1's perspective, this is a random variable, as it does not know z. But it now knows something about z. Presumably firm 2 knew z when it bid, and bid in such a way firm 1 won the bidding outright. Does this imply firm 2 perceived a higher cost, suggesting z might be "large?" This will turn out to be the case. For now just express the expected cost, conditional on knowing x, knowing y and knowing firm 2's bid was higher, as follows.

$$E[\Delta|x,y,b_2 > b] = \alpha x + \beta y + \gamma E[z|b_2 > b]$$

At the time of bidding, firm 1 knows x and y; but if its bid actually wins it then knows x and y and the fact it submitted the lowest bid. Intuitively, firm 2 knowing z and bidding above b tells us something about z. Equilibrium analysis will give some meat to this intuition.

Remember we also flip a coin if the bids tie. So we calculate firm 1's expected profit from bidding $b_1 = b$, given x and y, as:

$$E[\Pi_1(b,b_2)|x,y] = 0 \cdot prob\{b_2 < b\} + (b - E[\Delta|x,y,b_2 > b]) \cdot prob\{b_2 > b\}$$

$$+.5(b - E[\Delta|x,y,b_2 = b]) \cdot prob\{b_2 = b\}$$

---

[19] We saw the same issue in the earlier sequential choice game, where the second player's reaction function was essentially a fully worked out strategy, a choice for each possible choice the first player might make.

(This is certainly a mouthful.)

To specify the missing probabilities, we must of course think in equilibrium terms. Firm 1 will have a bidding strategy that specifies its bid for any realization of the information variables it is observing. Denote this strategy or bidding function $b_1(x, y)$. In parallel fashion, denote firm 2's strategy or bidding function $b_2(x, z)$. Equilibrium, now, refers to an equilibrium pair of such functions. After all, firm 1 knows what its information source reported, but is not privy to the private information of the other firm. It is bidding against firm 2's strategy, just as firm 2 does not know firm 1's private information and thus is bidding against its strategy. Paralleling our earlier equilibrium conditions, if strategies $b_1^*(x, y)$ and $b_2^*(x, z)$ are indeed equilibrium bidding strategies, they must be mutual best responses:[20]

$$b_1^*(x, y) \in \arg\max_b E[\Pi_1(b, b_2^*(x, z))|x, y]\forall x, y \in [0, 1] \quad (10.8a)$$

$$b_2^*(x, z) \in \arg\max_b E[\Pi_2(b_1^*(x, y), b)|x, z]\forall x, z \in [0, 1] \quad (10.8b)$$

Notice the point-by-point syntax. For a given x and y, $b_1^*(x, y)$ is a best response to firm 2's strategy, and this is repeated for every conceivable x and y combination. A parallel comment holds for $b_2^*(x, z)$. As you were warned, strategic considerations can greatly expand a decision frame.

It turns out the following strategies form an equilibrium:

$$b_1^*(x, y) = \alpha x + (\beta + \gamma)/2 + (\beta + \gamma)y/2 \quad (10.9a)$$

$$b_2^*(x, z) = \alpha x + (\beta + \gamma)/2 + (\beta + \gamma)z/2 \quad (10.9b)$$

$b_1^*(x, y)$ is a linear function of what firm 1 has observed, just as $b_2^*(x, z)$ is a linear function of what firm 2 has observed. Each bid consists of αx, the commonly known component of cost, plus the expected value of the other two components, $(\beta + \gamma)/2$, plus an "extra" amount depending on the firm's private information: $(\beta + \gamma)y/2$ or $(\beta + \gamma)z/2$.

In the interest of not relying on magic, we will take the time to verify firm 1's half of the equilibrium. To begin, if firm 2 is bidding in this manner, a bid of b by firm 1 will win outright whenever

$$b < \alpha x + (\beta + \gamma)/2 + (\beta + \gamma)z/2$$

Manipulating this expression leads to b winning outright when

$$\frac{b - \alpha x - (\beta + \gamma)/2}{(\beta + \gamma)/2} \equiv g(b) < z$$

From here we can bracket the interesting bids. Given x, the lowest bid firm 2 will submit (which occurs when $z = 0$) is $\alpha x + (\beta + \gamma)/2$; and the

---

[20] This is called a Bayesian equilibrium.

highest bid it will submit is $\alpha x + (\beta + \gamma)/2 + (\beta + \gamma)/2 = \alpha x + (\beta + \gamma)$. There is no reason for firm 1 to bid below the lowest or above the highest conceivable bid of its competitor. But then the interesting bids for firm 1 imply $g(b)$ as defined above ranges between 0 and 1.

Continuing, if $0 \leq z < g(b)$ we have $b > b_2^*(x, z)$. So the bid of b loses outright, and provides firm 1 an incremental profit of 0. Conversely, if $g(b) < z \leq 1$ we have $b < b_2^*(x, z)$ and the bid of b wins outright. Firm 1's incremental profit is $b - \Delta = b - \alpha x - \beta y - \gamma z$. Since a tie actually occurs with probability zero (as z is a continuous random variable), we have the following expression for firm 1's expected incremental profit when it bids b having observed x and y:

$$
\begin{aligned}
E[\Pi_1(b, b_2^*(x, z))|x, y] &= \int_0^{g(b)} [0]dz + \int_{g(b)}^1 [b - \alpha x - \beta y - \gamma z]dz \\
&= \left[(b - \alpha x - \beta y)z - .5\gamma z^2 \right] \Big|_{g(b)}^1 \\
&= (1 - g(b))[b - \alpha x - \beta y - .5\gamma(1 + g(b))]
\end{aligned}
$$

Firm 1, now, wants to maximize its expected incremental profit, conditional on what it knows. So we focus on the point where the derivative of this expression vanishes:[21]

$$
\begin{aligned}
\frac{\partial E[\Pi_1(b, b_2^*(x, z))|x, y]}{\partial b} &= -g'(b)[b - \alpha x - \beta y - .5\gamma(1 + g(b))] \\
&\quad + (1 - g(b))[1 - .5\gamma g'(b)] = 0
\end{aligned}
$$

Two additional steps complete the torture. Substitute $g'(b) = 1/(\beta + \gamma)/2 = 2/(\beta + \gamma)$. Also substitute the earlier expression for $g(b)$. Collecting terms leads to expression (10.9a).

Whew! Now, if we repeat this from the other side we will discover that if firm 1 uses this bidding rule, firm 2's best response is to use the bidding rule that we originally claimed in (10.9b). The two bidding strategies constitute equilibrium behavior. Each is a best response to the other.

10.4.3 Winner's Curse

In Table 10.1, we compile various aspects of this equilibrium bidding story. If, for example, we take the time to substitute the equilibrium bidding strategy into each firm's expected (incremental) profit calculation, we will find the noted expected (incremental) profit expressions. Firm 1, for example, having observed x and y is facing a bidding competition that offers an expected gain of $.5\beta(1 - y)^2$. This is strictly positive as long as $y \neq 1$

---

[21] $E[\Pi_1(b, b_2^*(x, z))|x, y]$ is a concave (and differentiable) function of b, so we know the maximum occurs at the point the derivative vanishes.

(the worst possible cost information) and $\beta > 0$. This stands in stark contrast to the case where the firms have the same information and compete so fiercely they face nil profit prospects. Private information can, indeed, be very useful, even if your competitor knows you are observing something he is not observing. A parallel observation applies to firm 2.[22]

TABLE 10.1: Equilibrium Implications

bidding strategies
$$b_1^*(x,y) = \alpha x + (\beta + \gamma)/2 + (\beta + \gamma)y/2$$
$$b_2^*(x,z) = \alpha x + (\beta + \gamma)/2 + (\beta + \gamma)z/2$$
difference in bids
$$b_1^*(x,y) - b_2^*(x,z) = \frac{\beta+\gamma}{2}(y - z)$$
expected profit given information
$$E[\Pi_1(b_1^*(x,y), b_2^*(x,z))|x,y] = .5\beta(1-y)^2$$
$$E[\Pi_2(b_1^*(x,y), b_2^*(x,z))|x,z] = .5\gamma(1-z)^2$$
expected cost prior to bid
$$E[\Delta|x,y] = \alpha x + \beta y + \gamma/2$$
$$E[\Delta|x,z] = \alpha x + \beta/2 + \gamma z$$
revised expected cost if bid wins
$$E[\Delta|x,y,b_1 < b_2] = \alpha x + \beta y + \gamma(1+y)/2$$
$$E[\Delta|x,z,b_2 < b_1] = \alpha x + \beta(1+z)/2 + \gamma z$$
bias in initial cost estimate
$$E[\Delta|x,y,b_1 < b_2] - E[\Delta|x,y] = \gamma y/2$$
$$E[\Delta|x,z,b_2 < b_1] - E[\Delta|x,z] = \beta z/2$$
bid as expected cost plus markup
$$b_1^*(x,y) = E[\Delta|x,y] + \gamma y/2 + \beta(1-y)/2$$
$$b_2^*(x,z) = E[\Delta|x,z] + \beta z/2 + \gamma(1-z)/2$$

Continuing, let's track firm 1's assessment of its cost as it works through the exercise. Initially, upon observing x and y, we know its expected cost is $E[\Delta|x,y] = \alpha x + \beta y + \gamma/2$, but if its bid is a winner, it also learns something about the z variable that firm 2 has observed. We wind up, as noted in the Table, with a revised cost expectation of

$$E[\Delta|x,y,b_1 < b_2] = \alpha x + \beta y + \gamma(1+y)/2 = E[\Delta|x,y] + \gamma y/2$$

---

[22] If, upon observing x and y, firm 1's equilibrium bid wins, it will be paid the winning bid of $\alpha x + (\beta + \gamma)/2 + (\beta + \gamma)y/2$ and face an expected cost of $\alpha x + \beta y + \gamma(1+y)/2$. This expected cost reflects the fact winning reveals something about z, and will be explained shortly. Revenue less expected cost is $.5\beta(1-y)$. In turn, as the difference in the bids displayed in Table 10.1 reveals, firm 1 wins only when $y < z$, and this has probability $1 - y$. So with probability $1 - y$ we obtain the above conditional expected profit and with probability y we lose the auction and gain nothing. Overall, then, the expected incremental profit is the claimed $.5\beta(1-y)^2$.

Importantly, the fact of winning causes the firm to revise its cost expectation upward. Winning means your original estimate was too low, was downward biased in the amount $\gamma y/2$. This is called the winner's curse.

This expected cost revision is readily verified. Notice, Table 10.1 again, that the two bids differ by a constant multiplied by $(y - z)$. So firm 1's bid is lower only when $y < z$. Prior to winning, firm 1's best estimate of z was $E[z] = .5$, reflecting our convenient uniform distribution assumption. But winning tells us $y < z$, which implies $E[z|y < z] = y + (1 - y)/2 = .5 + .5y$, again thanks to the uniform distribution. This implies we have revised our estimate of z upward from .5 to $.5 + .5y$, so our cost estimate goes up by $\gamma(.5y) = \gamma y/2$. A parallel calculation holds for the second firm, as summarized in Table 10.1.

We belabor this because of its importance. Winning carries information. And in this case you win only when you thought the cost was lower than the competitor thought it was. More precisely, the fact of winning should cause us to raise our perception of the cost of supplying the object in question. This stems from the fact the eventual cost will be the same regardless of producer, and each firm is observing something the other is not observing.

For example, if two competitors bid on repairing a highway, the winning bid should tell us something about what the other competitor thought the cost of performing the repairs might be. Likewise, competitors bidding for oil drilling rights must deal with the fact that if they win the bid their information must have suggested a more valuable oil reserve than that of the competitors. Winning carries information!

Of course rational bidders are aware of the winner's curse phenomenon. In equilibrium each bid is proffered with the understanding that it wins only if the private information on which it is based is more favorable than that of the competitor. The bid is "padded" in recognition of the fact it is a winning bid only if the underlying cost calculation is too low. This is most apparent if we rewrite the equilibrium bidding strategies in terms of "cost plus a markup," as also displayed in Table 10.1. Firm 1, for example, begins with a cost estimate of $E[\Delta|x, y]$ and adds the amount $\gamma y/2 + \beta(1 - y)/2$ to this estimate to form its bid.[23] Notice, with $\beta > 0$ and $\gamma > 0$ that this additional amount is always strictly positive. This "plus" reflects the firm's private information advantage and also protects it against the winner's curse. It is the product of strategic forces.

Example 10.7 For some specific calculations, we return to Example 10.6, where $\alpha = 10, \beta = 10$ and $\gamma = 40$. In that setting the firms were uninformed and, in equilibrium, each bid the uninformed cost assessment of $E[\Delta] = 30$. Now, of course, they are informed. Table 10.2 reports firm 1's equilibrium bid, cost, and profit calculations as a function of y. Study it carefully.

---

[23]Straightforward algebra will confirm this claim that the equilibrium bid can be expressed in this fashion.

The commonly observed information, x, merely revises each firm's assessment of the αx component of the cost, and thus equally impacts their bids, but is of no profit consequence. The important part of the story is the private information. Here you should notice how the naive cost expression of $E[\Delta|x, y] = 10x + 20 + 10y$ (which we normalize to $E[\Delta|x, y] - 10x = 20 + 10y$) increases linearly as we step through various values of y. Adjusting for the winner's curse, though, we have a parallel expression of $E[\Delta|x, y, b_1 < b_2] - 10x = 20 + 30y$. The bias in the naive estimate is $20y$. The sophisticated cost expression increases linearly as we step through values of y, but at three times the rate.

TABLE 10.2: Equilibrium Bids and Profits, $\alpha = 10, \beta = 10, \gamma = 40$

y	$E[\Delta\|x,y]$ $-\alpha x =$ $20 + 10y$	$b_1^*(x,y)$ $-\alpha x =$ $25 + 25y$	$E[\Delta\|x,y,b_1 < b_2]$ $-\alpha x =$ $20 + 30y$	$E[\Pi_1\|x,y] =$ $5(1-y)^2$
0.0	20	25	20	5.00
.1	21	27.5	23	4.05
.2	22	30	26	3.20
.3	23	32.5	29	2.45
.4	24	35	32	1.80
.5	25	37.5	35	1.25
.6	26	40	38	.80
.7	27	42.5	41	.45
.8	28	45	44	.20
.9	29	47.5	47	.05
1.0	30	50	50	.00

Intuitively, the bias increases with y because firm 1's bid increases with y. If firm 1 perceives a high cost (i.e., high y) and still wins the auction, $z > y$ is implied and therefore z is well removed from the original mean of .5. Conversely, if firm 1 perceives a low cost (i.e., low y) and wins the auction, $z > y$ is again implied, but this is now not compelling evidence z is well removed from the original mean.

Next examine the optimal bid of $b_1^*(x, y) = \alpha x + 25 + 25y = E[\Delta|x, y] + 5 + 15y$, where we write it as equal to the (naive) expected cost plus an amount equal to $5 + 15y$. (See Table 10.1.) Notice the firm is not simply adding the winner's curse adjustment (of $20y$) to the cost expression. The bidding strategy is more subtle as the bid affects what the firm will be paid if it wins as well as the probability it wins the competition.

Suppose $y = .1$ is observed. The equilibrium bid is $\alpha x + 27.5$. We know both bids will independently range between $\alpha x + 25$ and $\alpha x + 50$, so the bid of $\alpha x + 27.5$ will win with probability .9. But if this bid wins, we also know $z \geq .1$ and this implies $E[\Delta|x, y = .1, b_1 < b_2] = \alpha x + 23$. This implies the expected gain if the firm's bid of $\alpha x + 27.5$ wins is $\alpha x + 27.5 - \alpha x - 23 = 4.5$.

But, recall, this bid wins with probability .9, and we thus have the expected profit of .9(4.5) = 4.05 reported in Table 10.2

Now check your understanding. Suppose, having observed $y = .1$, that firm 1 decides to be more aggressive and bids $\alpha x + 25$, which basically is a winning bid as firm 2 will bid above $\alpha x + 25$ (unless $z = 0$, a zero probability event). If firm 1 wins regardless of firm 2's bid, its expected cost, conditional on winning (and $y = .1$), must be $\alpha x + 10(.1) + 40(.5) = \alpha x + 21$. So the aggressive bid delivers an expected profit of $\alpha x + 25 - \alpha x - 21 = 4 < 4.05$. The firm wins more often with the lower bid, but gains less in the process.

Private information, then, leads the firms to bid above their estimated cost, a far cry from the equilibrium in Example 10.6. Of course how much to "pad" the cost is an equilibrium calculation. It reflects a delicate balancing of odds of winning against the ex post value of the object if the resulting bid happens to prevail.[24]

Example 10.8 Now step through several versions of the cost specification in (10.6). Details are presented in Table 10.3. Two features are intuitively important. One is the α coefficient reflects the importance of the cost component both parties observe. While this affects their bidding strategies, being common to both it has no effect on their respective strategic assessments or expected profits. So we simply allow it to be some arbitrary amount in what follows. The second important feature is the relative importance of each bidder's private information. The β coefficient reflects the importance of the cost component firm 1 privately observes, and the γ coefficient reflects the importance of the component firm 2 observes. In the first case in Table 10.3 ($\beta = \gamma = 9$), these two components are nontrivial but equal in magnitude. So the firms are on equal footing, and each faces a nontrivial though identical expected profit.

TABLE 10.3: Equilibrium Bids and Expected Profits

β	γ	$b_1^*(x,y)$ $-\alpha x$	$b_2^*(x,z)$ $-\alpha x$	$E[\Pi_1|x,y]$	$E[\Pi_2|x,z]$
9	9	$9(1+y)$	$9(1+z)$	$4.5(1-y)^2$	$4.5(1-z)^2$
ε	ε	$\varepsilon(1+y)$	$\varepsilon(1+z)$	$.5\varepsilon(1-y)^2$	$.5\varepsilon(1-z)^2$
$9-\varepsilon$	ε	$4.5(1+y)$	$4.5(1+z)$	$.5(9-\varepsilon)(1-y)^2$	$.5\varepsilon(1-z)^2$
8	32	$20(1+y)$	$20(1+z)$	$4(1-y)^2$	$16(1-z)^2$

The second case is also one where the firms are again on an equal footing ($\beta = \gamma = \varepsilon$), but what each is learning in private is of minor consequence. To no surprise, their expected profits are of minor consequence and we are getting awfully close to the uninformed case in Example 10.6. Expected

---
[24] The winner's curse would become more pronounced if we had more bidders, each with their own information.

profit here depends on knowing something that is both important and private! In the third case firm 1 is in the driver's seat, as it is observing an important component of cost while firm 2 is not. So it enjoys a distinctly favorable expected profit position. To avoid the appearance of favoritism, we reverse this in the final case.

10.5 Haggling

The bidding story relies on competitive bids, and thus leaves open the question of equilibrium behavior when there are no competitors. This is the subject of haggling. We are all familiar with bargaining, negotiating, or arguing over items of interest. Buying an automobile is a classic example in our society (though internet access is rapidly changing this particular game). Real estate, divorce, labor contracts, and consulting fees provide additional examples.

Each setting evokes particular structural details. We often have a buyer, a seller, and a real estate agent in between in the real estate story. We usually have two lawyers (and perhaps a judge) in between in the divorce story. Individuals particularly adept at negotiation are often center stage in labor negotiations. And then there is the unflattering caricature of the quintessential used car salesperson.

The point is not idle. Haggling becomes interpersonal, and places a premium on the skills of the individuals involved. These skills, in turn, are often buttressed by careful design of the setting. Do we meet in a formal though neutral setting? Does the professor always sit behind the desk when a student complains about a grade? Does the auto salesperson always have to check with the boss, thereby introducing another party into the encounter?

We streamline our earlier setting to highlight some of these issues. Most important, now assume there is one supplier, whom we will term the seller. In this way, the buyer must deal with the seller, and there is no other viable option. Also set $\beta = \gamma = 0$, so the seller's cost is given by $\Delta = \alpha x$.

Let the value of the project to the buyer be V. The net social gain, then, is zero if no deal is struck, and $V - \Delta$ if a deal is struck. Further assume $V > \Delta$, so it makes sense to close a deal. The question then becomes one of how to share the gain of $V - \Delta$ between the two parties.

This depends, to no surprise, on how the individuals play the game. And it may also turn out that socially desirable trade is impeded.

10.5.1 Milquetoast Players

Suppose both parties know V and Δ, and are nonaggressive, cooperative types. A natural solution is to split the difference so that the gain of

$V - \Delta$ is shared equally between the two parties. In fact, this is precisely the solution predicted by the theory of bargaining. The idea, roughly, is that there are gains to cooperation (here, $V - \Delta$), and we also know what happens to each player in the absence of any agreement (here, they each get zero). Treating the parties symmetrically then calls for splitting the gain in this particular setting.[25]

But what happens if our parties are aggressive types? For example, suppose they will simultaneously announce a way to split the gain. If their proposals agree, they have an agreement. If not, they simultaneously announce whether to stand pat or accede to the other's proposal. If both stand pat, no trade occurs. If one holds firm and the other accedes, they have an agreement. If they both accede, they again have no agreement.

Now for the catch. Suppose the buyer proposes to keep $k(V - \Delta), 0 \leq k \leq 1$ of the overall gain. Further suppose the buyer will stand pat if no agreement is reached in the first round. It is routine to verify that seller's best response to such a strategy is to propose keeping at least $(1-k)(V-\Delta)$ in the first round, and to accede in the second. Moreover, buyer's best response to such a strategy is to propose and then stand pat, as noted. We have an equilibrium. Unfortunately, k can be anywhere between 0 and 1, so we have paid a large price (pun) for our attempt to introduce some bargaining details into the exercise. Any solution, any split we had in mind, can be "defended" with an equilibrium argument. We have too much of a good thing. Of course, our friends might naturally focus on the $50-50$ split of $k = .50$.[26]

10.5.2   Private Cost Information

Closer to the earlier bidding story is the case where cost is uncertain and known only to the seller. Suppose the cost, Δ, is either 1 or 2, with equal probability, and $V = 4$. So the gain will be $4 - 1 = 3$ or $4 - 2 = 2$. The

---

[25] Splitting the difference is a familiar and, it turns out, theoretically defensible solution. The theoretical formulation, due to John Nash, envisions the parties as cooperatively searching for a way to split the gain. They do not renege, they do not posture, they do not "game" in any sense. The substance of the axiomatic setup is then the requirement that all gains to trade be pursued (efficiency), that the parties be treated symmetrically, and that irrelevant alternatives not influence the split. With risk neutrality and zero gain nonagreement points, the solution is simply a price P such that the expression $(V - P)(P - \Delta)$ is maximized.

[26] An example of this type is analyzed in Kreps [1990, Chapter 15]. Also notice the indeterminacy would disappear if the $V - \Delta$ term declined the longer it took the parties to reach an agreement. Haggling, after all, takes time, and time carries an opportunity cost in the sense we have left other uses of time outside the formal analysis, not to mention actions of yet other parties. But this take us too deep into the theory of bargaining.

difficulty, now, is how to share the gain when only one party knows the gain.

To put more structure on the story, suppose the buyer makes a take-it-or-leave-it offer to the seller. This offer takes the form of an announced price, P. If the seller says no, the game ends. If the seller says yes, production takes place, the buyer gains $V - P$ and the seller gains $P - \Delta$. That's it; renegotiation is ruled out. The buyer's ability to commit to a take-it-or-leave-it offer gives the buyer substantial bargaining power. Nevertheless, we will find that the seller is somewhat protected by private information about cost.

Now, if the buyer offers $P = 2$, the seller can do no better than always accept. Buyer gains $V - P = 2$, and seller gains $P - \Delta = 1$ if cost is low or 0 otherwise. Alternatively, if the buyer offers $P = 1$, seller can do no better than accept in the low cost event, and reject in the high cost event; either way seller nets precisely 0. Buyer, however nets $V - P = 3$ in the low cost event and 0 in the high cost event.

Remember the odds are $50 - 50$. Would the buyer prefer to gain 2 or $.5(3) + .5(0) = 1.5$. Before jumping to any conclusions, repeat the story for the case where the probability of low cost is .8. Here, the buyer's best choice is to forego the project when $\Delta = 2$, even though $V > 2$.[27]

Recall in the bidding story that privately knowing something about cost gave the bidder an advantage. The same occurs here. We set the rules so the buyer has all the bargaining power, being able to make a take-it-or-leave-it offer. If the buyer knows the seller's cost, Δ, the offer will be $P = \Delta$, no more and no less. By not knowing the seller's cost, though, the buyer is reduced to one of two possibilities. One is to offer a price equal to the high cost. Then trade always occurs, but the seller captures some surplus. The other is to offer a price equal to the low cost. Then trade only occurs under the low cost condition, but none of the surplus is shared.

The buyer cannot capture the full gain, despite the advantageous position. The gain is dissipated in one of two ways: it is shared or trade is cut short in the high cost event.

Also notice how either scheme, namely offer $P = 2$ or offer $P = 1$, can be thought of in terms of a mild negotiation encounter. Buyer says, tell me your cost, and I'll pay $P = 2$, or tell me your cost and we'll deal (at $P = 1$) if the cost is low. We will learn in later chapters that this is a revelation game in which the informed party is induced to reveal what is known. The price, so to speak, is not using that revelation in aggressive fashion. The

---

[27] It is tempting to conclude haggling can be inefficient. If everyone knows everything, we know in this case trade should occur. But it is a mistake to take this "full information" answer and blindly presume we should implement it in the private information setting. We will see this theme repeatedly in subsequent chapters when we explore control problems in depth.

buyer cannot, under the rules of the game, subsequently renege on the take-it-or-leave-it offer.

This suggests, on the surface, renegotiation might be useful. But this is not so. What happens if buyer says, I'll pay $P = 1$, unless you say cost is high, in which case I'll pay $P = 2$? This is just a long-winded version of the initial scheme of paying $P = 2$, whatever the cost.[28]

Example 10.8 To end the story, can you guess what happens when $\Delta = \alpha x$, but x is uniformly distributed as in the bidding story? Let $\alpha = 2$, implying $0 \leq \Delta \leq 2$. Suppose the buyer offers to pay $P \leq 2$. The seller's profit is $P - \Delta = P - 2x$. This is positive for $x \leq P/2$. So the offer will be accepted with probability $P/2$. For $P \leq 2$, the buyer's expected gain is now

$$(1 - P/2)[0] + (P/2)[V - P]$$

Straightforward differentiation implies $P = V/2$ (presuming $V \leq 4$) maximizes this expression. With $V = 4$, the buyer offers the highest possible cost, and trade always takes place. With $V = 3$, the buyer offers $P = 1.5$, and trade is rationed, etc.

10.6   Internal Control

As becomes clear in our trip down the equilibrium highway, equilibrium behavior opens the door to understanding choice in multi-actor settings and "the rules of the game" matter a great deal. Exploiting this latter point is the very essence of internal control. We have stressed the fact the accounting library is a specialized library, one that emphasizes financial records that have a high degree of integrity. It is purposely designed to have a low error rate and to be difficult to manipulate, despite the fact it is influenced by numerous individuals and events.

Nicely said, but just what are these techniques? In broad brush terms they are three in number: parceling out decision rights, designing in re-

---

[28] Though we have been a little casual, you should be able to write down the definition of equilibrium here. The key is one player moves first, just as in Example 10.1. The second player's decision, the seller, depends on the offered price, P, and on the privately observed cost, Δ. Let $R(P, \Delta) \in \{0, 1\}$ denote this decision, with 0 meaning no trade and 1 meaning trade. So

$$R(P, \Delta) \in \arg \max_{R(P,\Delta) \in \{0,1\}} R(P, \Delta)[P - \Delta], \forall P, \Delta$$

And, since the buyer does not know Δ, the equilibrium offer of P^* is defined by

$$P^* \in \arg \max_P E[R^*(P, \Delta) \cdot (V - P)]$$

From here you should be able to convince yourself that the buyer can do no better than pursue one of the noted two schemes of offering $P = 2$ or $P = 1$.

dundancy, and use of incentives. These techniques are used, so to speak, to define the rules of the "accounting library game."[29]

10.6.1   Decision Rights

One control technique restricts the choices an individuals faces, a common ploy that leads to decision rights. Who approves major capital expenditures? Who selects the textbook for the course? Who specifies the depreciation schedule or the revenue recognition rule?

Restricting choice possibilities is a familiar, visible control technique. Fences, limited access highways, turnstiles, and door locks are illustrative. Access codes to a firm's data bank and ID verification to enter the facility are also familiar examples.[30]

The restrictions are often administrative. The parts manager in an auto dealership might be able to purchase parts from a prespecified list of approved vendors. The division manager might be able to authorize capital investments up to $2 million, without turning to central management for authorization.[31]

Similarly, and profoundly, the accounting library will not allow various subjective estimates to enter into its rendering of the firm's financial history. A projected profit bonanza based on a new product is off limits, just as is booking a sale well before standard revenue recognition thresholds have been met. It is important to remember that one way the accounting library's integrity is defended is by restricting not only who has access to the library but what is allowed to be placed in the library.

This is why we are often confronted with the fact the way a well framed decision appears may be far removed from the way in which the accounting library records the consequences of that decision. Similarly, so-called fair

---

[29] Internal control is serious business. A firm could not function without reliable financial records. Regulatory matters are present as well. The Foreign Corrupt Practices Act of 1977 prohibits U.S. companies from bribing foreign political officials. It also requires public companies to maintain adequate (in a cost benefit sense) internal controls. The Sarbanes-Oxley Act of 2002, among other things, dramatically increases regulatory requirements in the internal control arena, including attestation requirements. Civil and criminal penalties, under federal securities laws, are possible actions when internal controls are inadequate.

[30] This hints at the importance of architecture in the design of control systems. Prisons are designed with surveillance in mind; observation structures are built into gambling casinos; and so on. These techniques are not new, either. Moats and limited access were used to combat work force pilferage during the Renaissance.

[31] Two additional applications of this technique should be noted. One is to substitute capital for labor. Currency counting machines and bar code readers come to mind. The other is to rotate the individuals. A records clerk must take a vacation. Auditors must rotate assignments. The dealer in a casino must take breaks. Forced rotation inhibits continuity, a dynamic version of restricting action possibilities. Similarly, managerial promotion has, as a side advantage, a rotation of duties dimension.

value measurements rely on an absence of private information, as evidenced by our exploration of the winner's curse.

10.6.2   Redundancy

A second control technique is to design in redundancy. A lending officer in a large bank oversees a portfolio of loans. Larger loans in this portfolio are examined by a supervisor and perhaps a loan committee. The entire portfolio is also subject to review by the bank's hierarchy. Similarly, a surgeon's work is routinely examined by a surgery committee. We return the students' examinations to provide feedback. A secondary purpose is to provide a check on the grading. And we routinely compare our firm's progress with that of peer firms, so-called benchmarking.

Turning to the accounting library, firms keep detailed inventory records, but also perform an annual physical inventory. Unit costs of manufactured inventory are verified. Cash is reconciled with the bank statement. Likewise, an auditor will replicate a sample of transactions.[32]

10.6.3   Explicit Incentives

A third control technique is use of explicit incentives. The firm's workforce may be compensated with an hourly pay and benefits package, but also with an explicit profit sharing arrangement. The young manager faces an array of financial, nonfinancial, promotion, and peer approval possibilities. A licensed physician might lose the necessary license. The plumber might get a reputation for inadequate service. The inventor might produce a valuable patent. The scientist might win a Nobel Prize. The bank teller who cannot count will be discovered. The parents who neglect their children run the risk of family fracture.

Turning to the accounting library, we want diligent record keeping. Diligence is difficult to observe, so we don't just go to the labor market and purchase so many units of diligence. Instead, we foster diligence. We stress its importance. We seek diligent employees. We randomly check some records (a form of redundancy). We then use the error rate observed in the sample as an indirect measure of the diligence supplied. A low error rate might be accompanied by supervisory approval, financial reward, promotion, self-satisfaction, and so on. A high error rate might be accompanied by supervisory disapproval, lack of self-satisfaction, loss of employment,

---

[32]Internal and external auditors typically perform the audit. Here you should notice how the work of the two sets of auditors creates an elaborate, mutual control mechanism. In addition, the external auditor works within a complex organization, itself subject to public oversight. Thus, in a larger system the milieu even speaks to the question of auditing the auditor.

and so on. This means the record keeper's "compensation" varies with the observed error rate.

There is also a more delicate side to the incentives story. An important internal control issue is making certain the source documents are reliable. Otherwise, the record keeping begins with errors. This is why we see consistency checks, for example, among purchase orders, receiving reports, and vendor invoices. It is important, however, not to put too much pressure on source documents. Otherwise, they may be compromised. The incentive consequences, so to speak, cannot be too severe if we are to have reliable documents.

Suppose division managers are always harassed when their budgets do not show a sizable increase in profitability. This invites, at the margin, distorted budget forecasts.[33] Similarly, suppose a special tools fabrication department has two jobs. One is unusual. The other is similar to several previous jobs. The department must self-report labor hours spent on each job. If heavy cost control effort is directed toward the familiar job (because experience suggests what it should cost), we invite less than reliable source documents. The department is tempted to assign any unusual labor usage to the unfamiliar job.[34]

10.6.4   Equilibrium Behavior

Importantly, these various control techniques reflect equilibrium analysis. In a larger sense, various individuals affect what is recorded in the accounting library. A sophisticated web of controls is overlaid so that care and feeding of the library is achieved, so that the collection of mutual best responses leads to library integrity. This does not mean the library is error free or not vulnerable to opportunism, after all, economic forces (and regulations) lead to a balancing of risks and the underlying costs. But it does mean that it takes considerable bad luck or nefarious behavior to breach its defenses. It is wise, opportunistic design of the "rules of the game" that is the very essence of internal control.

10.7   Summary

Introducing equilibrium behavior at this juncture may seem interstitial. Yet that is hardly the case. The natural pedagogical sequence is to add more factors to the decision setting, and this eventually gets us to strategic

---

[33] The U. S. federal government also faces such an issue. The Gramm Rudman Act requires a balanced budget. Estimates are routinely engineered to balance the budget.

[34] The Allies in World War II consistently over-estimated Axis aircraft losses. The source documents were pilot self-reports, produced in the heat of battle. See Parker [1990].

considerations such as competitive response. In addition, we shall see the same notion of equilibrium behavior will hold the key to our forthcoming study of performance evaluation.

Stepping back from our various vignettes, it is important to remember that strategic considerations are an open-ended topic. The subtlety comes in two waves. One is that some details we worry about are truly chance phenomena (e.g., weather driven) while others are heavily influenced by other actors. Describing what these other actors might be doing calls for a more expansive analysis, one that includes their opportunities and motives as well. This is why we stressed equilibrium behavior as a device for disciplining our intuition.

The second wave of subtlety comes from the fact we must somehow draw the line, deciding what is sufficiently important to be thought of in strategic terms. For example, if we redesign our product and alter the price, what will happen to the product designs and prices charged by competitors? In turn, the demand for our product will, in principle, reflect the qualities of all the offerings and their respective (and presumably adjusted) prices. At what point do we sever the chain of interactions?

Similarly, a firm often has many opportunities to engineer important strategic encounters. Information may or may not be released. The sales force may not know what new products are under development, thereby putting them more on an equal footing with the customers. A reputation for playing hard ball may be carefully nurtured. The sales force may be given high powered incentives so they react aggressively to competitors in a pricing encounter.

10.8   Bibliographic Notes

The world of strategic encounter and equilibrium behavior is dealt with in introductory fashion by Dixit and Nalebuff [1993] and the classic of Luce and Raiffa [1957]. Deeper treatments are available in Gibbons [1992], Myerson [1991], Osborne [2004] and Rasmusen [2004]. Milgrom [1989] provides an introduction to auctions as well as a deeper treatment in Milgrom [2004]. Bajari and Hortacsu [2003] document a fascinating auction saddled with the winner's curse phenomenon, and Gawer and Henderson [2007] document product design races. Nash [1950] is the original source on cooperative bargaining; and a superb introduction is presented in the above noted Luce and Raiffa [1957]. Beyond that, Kennan and Wilson [1993] and Osborne and Rubinstein [1990] provide introductions to more modern themes of private information and alternating offers. Also see the above noted Gibbons [1992], Myerson [1991], Osborne [2004] and Rasmusen [2004]. Strategic planning considerations are explored in a variety of man-

ners. Oster [1999] and Tirole [1988] provide an economics flavor. You might also enjoy Milgrom and Roberts [1987] and Porter [1991].

10.9   Problems and Exercises

1. A central theme of strategic or competitive analysis is equilibrium behavior. What does it mean for strategies to be mutually consistent, in the sense of equilibrium behavior?

2. The text stresses the idea that a wider decision frame is implied by the presence of significant strategic concerns. What does this mean? How does it relate to the concept of equilibrium behavior?

3. *equilibrium analysis*
   Below is a bimatrix game played between protagonists *Row* and *Column*.

	left	*right*
up	60,10	0,12
down	40,-40	2,2

   (a) Locate and interpret an equilibrium when the protagonists make their choices in simultaneous fashion

   (b) What happens if the rules of the game call for sequential choice with *Column* making the initial choice, followed by *Row* making the second choice after having observed *Column*'s choice.

   (c) What happens in the sequential case if *Row* makes the initial choice?

   (d) Return to the simultaneous choice case, but now allow the encounter to be repeated two times. Before making their second choices, each protagonist observes the first round choices. Determine and interpret an equilibrium.

4. *mutual best response*
   Suppose a firm must determine a profit maximizing output quantity. Let q denote this quantity. Selling price is given by $\widehat{P}(q) = 340 - 2q$ (i.e., selling price declines with quantity) and cost is given by $C(q; P) = 100q$. Determine an optimal output and profit for this firm. Now explain the connection between this calculation and Example 10.4.

5. *equilibrium analysis*
   Repeat Example 10.4 for the case where $C(q_i; P) = 200q_i - 18q_i^2 + q_i^3$.[35]

---

[35] You might enjoy Example 2.3 at this juncture.

6. *duopoly and sequential play*
   Return to the duopoly setting in Example 10.4, but now suppose, instead of simultaneous play, that the first firm can announce and commit to a production plan for itself before the second firm decides on a production plan. Find the first firm's best quantity choice, and the second firm's best choice upon hearing the first firm's announcement. How does this change in the "rules of the game" help the first firm? What does it do to the second firm?

7. *mutual best response*
   Return to the setting of Example 10.5, but now assume the prize is $\widehat{P} = 10,000$. Determine the equilibrium investments. Comment on your finding.

8. *fair value*
   Many accounting voyeurs are fond of fair value. Define fair value. Then discuss its application in Examples 10.6 and 10.7

9. *best response bidding*
   Return to the first case in Example 10.8. Assume $\alpha = 0$ and the second firm is bidding according to the noted strategy. Suppose the first firm observes $y = 0.6$. Determine its expected profit if it bids (i) 14, (ii) 14.4 or (iii) 14.8.

10. *cost plus equilibrium bidding*
    Return to Example 10.8. Now suppose we define cost for the first firm as the expected value of its cost given x and y and for the second as the expected value of its cost given x and z. For the third case (Table 10.3), determine the plus that is added to each firm's cost if it bids according to the noted equilibrium. Discuss your finding

11. *winner's curse*
    Suppose our bidding illustration has $\alpha = 0, \beta = \gamma = 10$. Plot firm 1's bid as a function of y. (Glance back at Example 10.7!) Also plot firm 1's expected cost, given it has observed y, on the same graph. Finally, plot firm 1's expected cost, given it has observed y, bid as noted, and won the bidding.

    (a) Carefully explain the relationship among the three graphs.

    (b) We interpret this as an instance of cost plus pricing. Assume cost is defined as in the second graph, i.e., as firm 1's expected cost given it has observed y. What explains the amount that is added to this cost estimate to determine firm 1's bid?

12. *winner's curse*[36]

    Ralph wants to purchase a family heirloom from a neighbor. The heirloom has private value to the neighbor denoted v. Neighbor knows v; Ralph only knows v is uniformly distributed between $v = 0$ and $v = 100$. Neighbor knows this about Ralph. Finally, whatever v is, the value of the heirloom to Ralph is $1.5v$. Neighbor also knows this about Ralph.

    (a) Who should own the heirloom, Ralph or neighbor?

    (b) Suppose the trade encounter between the two individuals proceeds as follows. Ralph offers to purchase the heirloom at price P. If neighbor agrees, Ralph pays P in exchange for the heirloom. If neighbor does not agree, the game ends, and neighbor keeps the heirloom. Now, from Ralph's perspective, what is the expected value (to Ralph) of the heirloom, before any conversation with neighbor? Conversely, suppose Ralph offers price $P > 0$ and neighbor accepts the offer; what now is the expected value (to Ralph) of the heirloom? (As an aside, if Ralph pays P for the heirloom, at what price will Ralph's accountant value the heirloom?)

    (c) What is the equilibrium in this game? Why does no trade take place, despite the fact Ralph is known to value the object higher than the neighbor?

13. *winner's curse*

    Return to problem 12 above, but now assume v is uniformly distributed between $v = 20$ and $v = 120$. Repeat your earlier analysis. Why does trade take place here, for some values of v, but not in the original setting?

14. *rules of the game*

    Our discussion of competitive response focused on several well-defined encounters where the "rules of the game" were well-specified and understood. A larger question addresses the "rules of the game." Return to the haggling illustration in the text where the buyer had a value of $V = 4$, but the seller's cost was privately known to be either 1 or 2. Let θ denote the probability the cost is 1.

    (a) Should trade take place?

    (b) Suppose buyer makes a take-it-or-leave-it offer. Plot buyer's best offer as a function of θ. What is seller's best response?

    ---

    [36] This game, though with a different story, is discussed in Bazerman [1990]. The continuation in problem 13 was contributed by Richard Sansing.

(c) Change the rules so seller makes a take-it-or-leave-it offer. What is seller's best offer? How does it depend on seller's cost and on θ?

(d) Why does trade always occur in the second set of rules but only in some instances in the first? Which set of rules does each player prefer?

15. *sunk cost and bidding*
Return to the bidding story in the text, but now assume $\alpha = 1,000$ and $\beta = \gamma = 10$. We will also now interpret the αx term as a type of design cost that must be incurred before the bidding takes place. So at the time of bidding the αx term is a sunk cost; the firms incurred this cost before submitting their bids. What are the equilibrium bidding strategies? Of course, the firms would not have done this initial design work, and incurred the αx cost, had they looked ahead to the bidding exercise. What might the buying firm do in this instance in order to ensure a supply of bidders?

16. *value of information*
Two competitors are fighting it out as the only merchants on a remote island. Each has two strategies, simultaneous play is the order of the day, and Nature will provide one of two states (with equal probability). If state one obtains, the players' payoffs are given by the following bimatrix game.

	left	*right*
up	10,10	0,12
down	40,-40	2,2

Conversely, if state two obtains, the players' face the following bimatrix game.

	left	*right*
up	4,4	10,0
down	-40,12	10,10

(a) Suppose neither player can gather any additional information. Verify that 7 for *Row* and 7 for *Column* are equilibrium payoffs. (Here they play the game defined by the expected value of the two matrices.)

(b) Suppose *Row* obtains perfect information before acting. *Column* knows this and *Row* knows that *Column* knows, and so on. Verify that an equilibrium has *Row* play *down* no matter what state occurs, and *Column* play *right*, with expected payoffs of *6* for each. (Here *Row* has four strategies: *up* no matter what; *down*

no matter what; *up* in state one and *down* in state two; and vice versa.) How do you explain this equilibrium? Is *Row* better off with the information?

(c) Suppose both *Row* and *Column* acquire perfect information. Determine and interpret the resulting equilibrium behavior.

11
Large versus Small Decisions: Short-Run

We now combine the decision framing and accounting library themes. The accounting library is an important resource in estimating various components of cost and it will also eventually record the financial consequences of the firm's decisions. Moreover, its recording of these consequences will reflect accounting conventions and library choices; and this recording, emphatically, will not fully comport with the decision frames used in the various decisions. The present chapter focuses on short-run decisions, and the succeeding focuses on long-run decisions.

Two concerns should be kept in mind as we proceed. The first is signaled by the chapter's title. "Small" decisions can be treated as variations on the status quo, variations where we have reason to think that our LLAs are reasonably accurate, interactions with other decisions are inconsequential, and strategic issues, interactions with others, are also inconsequential. "Large" decisions contemplate movements sufficiently beyond the status quo that our LLAs come into question, we suspect interactions with other activities may be consequential, or strategic considerations may come into play.

Knowing when to ignore an interaction or when to abandon an LLA is a matter of judgment. The tension of choosing between a readily available or custom made cost construction is ever-present. Short-run decisions are not necessarily small, just as long-run decisions are not necessarily large.

The second concern is the firm's objective. Following earlier treatments, we will usually focus on profit maximization, or expected profit maximization. This has its awkward moments. What do we say about a municipality, a closely held firm, a family firm, or a hospital? The theory of the firm

does not provide much guidance on this score. Fortunately, the principles we explore are robust to whatever the organization's goals happen to be. Unfortunately, we cannot convey these principles with the language of a single goal that all organizations pursue.

With these concerns in mind, we initially study two preliminary topics. First, we consolidate the mechanical aspects of dealing with $LLAs$ to make marginal or small decisions. Second, we review framing considerations with special emphasis on the short-run cost of materials held in inventory. We then explore various prototypical choices: make or buy, product evaluation and customer evaluation. Finally, formal uncertainty is explored, focusing on the implicit benefit of flexibility and the implicit cost of risk followed by interactions with tax considerations.

11.1   Preliminaries

We begin with two preliminary excursions, designed to remind us of the simplifying power of $LLAs$ and the fact what you mean by the cost of something depends on how you have framed the underlying decision.

11.1.1   Break-Even Analysis

Our first preliminary excursion is the ultra simple model of break-even analysis. Consider a mythical single product firm. It expects its output and sales to be somewhere between 250 and 400 units. Let q denote quantity produced and sold. In this region, of $250 \leq q \leq 400$, the firm approximates its total revenue via

$$TR = 600q$$

and approximates its total cost via[1]

$$TC = 150,000 + 100q$$

Combining the two $LLAs$ we approximate the firm's profit via

$$TR - TC = [600 - 100]q - 150,000$$

The quantity in brackets, $600 - 100$, is the product's contribution margin. (In Chapter 6 we constructed the contribution margin by focusing on (normal) variable unit costs and variable period costs.) Given the above $LLAs$ for total cost and total revenue, contribution margin is just the slope of the latter less the slope of the former.

---

[1] Recall, beginning in Chapter 5, we were careful to overlay the $LLAs$ on the underlying economic cost, $C(q; P)$. Keep in mind that the LLA is always an approximation, a local linear approximation.
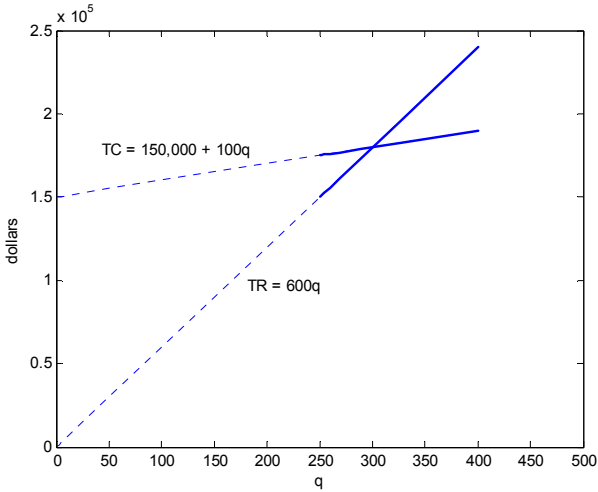
FIGURE 11.1. Break-Even Graph

We graph TR and TC in Figure 11.1. Notice we have interrupted the graphs at $q = 250$ and at $q = 400$, to remind ourselves $250 \leq q \leq 400$ is presumed.

At this level of simplification, two questions might be asked of this portrayal of the firm. One is the effect of a marginal customer. Suppose we are producing (and selling) at some level $250 \leq \widehat{q} < 400$. What is the effect of another customer arriving? We remain, presumably, within the range of the LLA. Revenue would thus increase by 600 and cost by 100. The difference is $[600 - 100]$, the contribution margin. This should come as no surprise. Contribution margin is our estimate of the effect of another unit on revenue less cost.[2]

The other question we might ask is where the two lines intersect.[3] They intersect where $TR = TC$, i.e., where estimated profit is precisely zero. This is called the *break-even* point. The algebra is straightforward. Setting

---

[2]More fundamentally we are asking for the rate at which revenue less cost changes with respect to quantity, or marginal revenue less marginal cost. Given the linearity assumption this is equivalent to asking the profit effect of an additional customer. Naturally, we could extend the calculation to the effect of several more units, always presuming we stay within the relevant range of the LLAs.

[3]To be fair, we also might ask how many units are required to produce an exogenously specified profit, or at what point is the vertical distance between the two lines equal to some specified number.

$TR = TC$, we have

$$600q = 150,000 + 100q; \text{ or}$$
$$[600 - 100]q = 150,000.$$

Hence,

$$q_{BE} = 150,000/[600 - 100] = 300$$

In algebraic terms, the break-even point occurs at the TC intercept divided by the contribution margin. See Figure 14.1. (Of course there is no guarantee q_{BE} occurs within the relevant range, though that is the case for our illustration.)

The economic interpretation is far from straightforward. Of what significance is q_{BE}? The answer depends on what we have left out of this surely very simplified analysis.

Example 11.1 Suppose this really is a single product firm. Further suppose its only short-run options are to shut down for the period, or to produce and sell whatever the market will bear. Let F_0 denote its best estimate of total cost when $q = 0$. Further suppose the firm anticipates demand will fall within the relevant range of $250 \leq q \leq 400$. Also suppose no other effects are associated with temporary shutdown this period.

Now, shutdown provides a loss of F_0. Continuation provides a profit (or loss) of $500q - 150,000$. The worst possibility, since we assume q is in the relevant range, is $500(250) - 150,000 = -25,000$. If $F_0 \geq 25,000$ our choice is clear. Continuation is preferred, since the worst that can happen beats shutdown.

Conversely, suppose shutdown would open the door to short-run leasing of the production facility. The lessee would pay F_0 plus 10,000. This means shutdown offers a guaranteed profit of 10,000, as opposed to the continuation profit of $500q - 150,000$.

Other stories could be told. Notice two things. We had to move beyond the LLAs to think about the shutdown option. Nowhere did q_{BE} enter the analysis. The reason is we were seeking an alternative (continue or shutdown) that leads to the best profit prospects. The break-even point is of no interest in the analysis.

Example 11.2 Now somewhat change the story in Example 11.1. The net dead weight cost of shutdown is F_0. (We have no lease option.) In addition, we do not know what demand will be. We do, however, know it will exceed q_{BE}. Now our options take the form of a guaranteed loss (shutdown) or a guaranteed though uncertain positive profit (continue). Here knowledge that $q > q_{BE}$ considerably simplifies the analysis.

Product market entry vignettes lead to a similar use of the break-even point.

Example 11.3 Suppose our firm produces many products. It is contemplating production of a new product. This new product will only be

produced this period. Its production and sales are totally separate from all other activities. The firm estimates actual demand will be somewhere between 250 and 400 units. Its incremental revenue and incremental cost estimates are given by TR and TC, respectively.

Is this short-run opportunity of any interest? It is if we know $q \geq q_{BE}$. And it is not if we know $q \leq q_{BE}$. Otherwise, our choice will rest on what we think is a good characterization of the probabilistic structure of demand and our attitude toward risk. Beyond that we might ponder whether the absence of strategic concerns at this point reflects analytic malaise or reality.

11.1.2   Framing Subtleties

Our second preliminary excursion is a review of the costing subtleties that may arise when we adopt a decomposed decision frame. For this purpose, consider a firm that manufactures and sells (or simply merchandises) a product in each of two periods. Production and sale are contemporaneous, as the finished product cannot be stored. For simplicity, no interactions with other activities last beyond two periods. So we drop these activities from the story.

Moreover, the only factor of production of interest is direct material. The firm's technology and market opportunities are such that its net cash flow is 100 per unit produced and sold in the first period and 110 per unit produced and sold in the second period. (This way we keep the focus on the direct material in question.) Capacity is limited to 100 units per period. Let q_t denote units produced and sold in period t.

The unusual feature in the story is the direct material market is far from perfect, and as a result the firm opportunistically manages its inventory of this material.[4] At present, the firm has I units of direct material on hand. Let A_t denote units of material acquired at the start of period t, and S_t denote units of material sold at the start of period t.

The following constraints circumscribe the firm's alternatives:

$q_1 \leq 100$ — capacity in period $t = 1$
$q_2 \leq 100$ — capacity in period $t = 2$
$q_1 \leq I + A_1 - S_1$ — supply of material in period $t = 1$
$q_2 \leq I + A_1 - S_1 - q_1 + A_2 - S_2$ — supply of material in period $t = 2$
$q_1, q_2, A_1, A_2, S_1, S_2 \geq 0$ — non-negativity

---

[4] Coase [1968] is the inspiration for this section. Coase remains particularly insightful and eloquent.

Capacity is limited to a maximum of 100 units per period. In addition, production cannot exceed the supply of material in each period. The twist in understanding the material inventory balance is that we allow the firm to use material in production or to sell it in the (previously owned) material market. Therefore, material available for production in the first period is $I + A_1 - S_1$. Similarly, that available for production in the second period is $I + A_1 - S_1 - q_1 + A_2 - S_2$. At present, the firm can purchase material at a cost of 10 per unit and sell it for a net price of 8 per unit. Let P^+ denote the forecast price of material acquired in the second period, and P^- the corresponding selling price in the second period. These latter prices are net of transactions costs in the second period spot markets, and are therefore stated in period $t = 2$ currency.

All transactions occur in cash, at the beginning of the respective period in which they are consummated. The interest rate is 10%. Thus, the present value of any feasible production and sales plan is

$$\Pi = 100q_1 - 10A_1 + 8S_1 + (1.1)^{-1}[110q_2 - P^+A_2 + P^-S_2]$$

(All $t = 1$ transactions take place at the start of the first period and all $t = 2$ transactions take place at the start of the second period.)

Combining the present value expression and the above constraints gives us the following statement of the firm's profit maximization problem.

$$
\Pi^* \equiv \max_{q_1,q_2,A_1,A_2,S_1,S_2 \geq 0} 100q_1 - 10A_1 + 8S_1 \tag{11.1}
$$
$$
+ (1.1)^{-1}[110q_2 - P^+A_2 + P^-S_2]
$$
$$
\text{s.t.} \quad q_1 \leq 100
$$
$$
q_2 \leq 100
$$
$$
q_1 \leq I + A_1 - S_1
$$
$$
q_2 \leq I + A_1 - S_1 - q_1 + A_2 - S_2
$$

TABLE 11.1: Solutions for Two Period Inventory Illustration

case	I	P^+	P^-	Π^*	shadow price	$A_1^*\ (S_1^*)$	$A_2^*\ (S_2^*)$
1	5	10	8.00	18,140.91	10.00	95	100
2	5	15	8.00	18,050.00	10.00	195	0
3	105	10	8.00	19,136.36	9.09	0	95
4	105	15	8.00	19,050.00	10.00	95	0
5	205	10	8.00	20,040.00	8.00	(5)	0
6	205	10	9.95	20,045.23	9.05	0	(5)

Table 11.1 displays the solution for various combinations of beginning inventory and second period material prices. In all cases, we produce at

capacity, with $q_1^* = q_2^* = 100$. The only variations are in how the material inventory is managed. The noted shadow price refers to the marginal value of beginning inventory and is the key to the story.[5] Glancing back at (11.1), notice that beginning inventory arises in the last two constraints; so the marginal value of beginning inventory is the sum of the shadow prices on these two constraints.

Dwell on these six cases. When the beginning inventory balance is $I = 5$, we must acquire 195 units to satisfy production requirements. The prices are such that this is best delayed as much as possible in case 1 and completely done in the first period in case 2. The same pattern emerges in cases 3 and 4, where we have $I = 105$ units. 95 units must be acquired. It is best to do this late in case 3 and early in case 4. Finally, in the last two cases, we have excess inventory. Disposal is best planned early in case 5 and late in case 6.

Now turn to the shadow price on the beginning inventory. In the first two cases, we always buy at least enough inventory in the first period to satisfy first period production requirements. As these acquisitions cost 10 per unit, the shadow price is 10. In cases 3 and 4 we must buy inventory to satisfy the second period's requirements, so timing is an issue. In case 3 we purchase late, implying a shadow price of $P^+/1.1 = 10/1.1 \cong 9.09$. In case 4 we purchase early. Finally, in the last two cases we dispose of extra inventory. In case 5 we dispose immediately, implying a shadow price of 8. In case 6 we dispose late, implying a shadow price of $P^-/1.1 = 9.95/1.1 \cong 9.05$.

Naturally, the beginning inventory would be recorded in the accounting library. Recognition rules are binding, however. We should expect to see it valued in the library at some variation of historical cost (subject of course to lower of cost or market).

With this lengthy setup, we are ready to grapple with our advertised framing subtleties. Suppose it is possible for the firm to manufacture and sell a second product in the first period. Only one unit can be sold, so the choice is between "yes" and "no." Choice of no is neutral. Present and future costs and demands are unaffected. So the no choice leads to a maximum profit of Π^*, as defined in (11.1) and tabulated for various cases in Table 11.1.

The yes choice will require one unit of material, and will net the firm $\widehat{P}$ dollars, exclusive of the material cost. This additional product also will be produced outside the firm's constrained capacity, so $q_1 = 100$ remains feasible. But using an additional unit of material reverberates through the two periods. The yes, choice, thus leads to the following variation on the original formulation, with maximum profit denoted $\widehat{\Pi}^*$.

---

[5] It is the rate at which maximal profit changes with respect to change in beginning inventory: $\frac{\partial \pi^*}{\partial I}$.

$$\widehat{\Pi}^* \quad \equiv \quad \max_{q_1,q_2,A_1,A_2,S_1,S_2 \geq 0} \widehat{P} + 100q_1 - 10A_1 + 8S_1 \quad (11.2)$$

$$+ (1.1)^{-1}[110q_2 - P^+ A_2 + P^- S_2]$$

$$\text{s.t.} \quad q_1 \leq 100$$

$$q_2 \leq 100$$

$$q_1 \leq I + A_1 - S_1 - 1$$

$$q_2 \leq I + A_1 - S_1 - q_1 - 1 + A_2 - S_2$$

Importantly, now, yes is preferred if it increases profit, which amounts to asking whether $\widehat{\Pi}^* > \Pi^*$. Details are summarized in Table 11.2, where $\widehat{A}_1^*$ denotes optimal first period material purchases under the yes choice and A_1^* its counterpart under the no choice, etc.

TABLE 11.2: Solutions for Special Product Extension

case	$\widehat{\Pi}^* - \Pi^*$	$\widehat{A}_1^*$	A_1^*	$\widehat{A}_2^*$	A_2^*	$\widehat{S}_1^*$	S_1^*	$\widehat{S}_2^*$	S_2^*
1	$\widehat{P} - 10$	96	95	100	100	0	0	0	0
2	$\widehat{P} - 10$	196	195	0	0	0	0	0	0
3	$\widehat{P} - 9.09$	0	0	96	95	0	0	0	0
4	$\widehat{P} - 10$	96	95	0	0	0	0	0	0
5	$\widehat{P} - 8$	0	0	0	0	4	5	0	0
6	$\widehat{P} - 9.05$	0	0	0	0	0	0	4	5

To torture this a bit longer, notice the incremental profit associated with this second product is

$$\widehat{\Pi}^* - \Pi^* = \widehat{P} - 10(\widehat{A}_1^* - A_1^*) + 8(\widehat{S}_1^* - S_1^*) + (1.1)^{-1}[-P^+(\widehat{A}_2^* - A_2^*) + P^-(\widehat{S}_2^* - S_2^*)]$$

and this can be thought of as incremental revenue less incremental cost. $\widehat{P}$, of course, is the incremental revenue. The negative of the remaining terms is the incremental cost of the material used for this second product.

Now glance back at the incremental profit calculations in Table 11.2. The second product's incremental cost might be the current acquisition price, the current sale price, the present value of the future acquisition price, or the present value of the future sale price: $10, 8, 9.09$ or 9.05 depending on which of the 6 cases is present. Moreover, in each case this incremental cost is the shadow price on the beginning inventory. This reflects the fact producing the second product is economically equivalent to reducing beginning inventory by one unit.[6] Using a unit of inventoried material for

---

[6] In addition, the linear structure ensures the shadow price holds as we range from zero to one full unit of the second product.

this product displaces that unit of material from its otherwise intended use; and that intended use might be to forestall purchase this period or next or to be sold this period or next period.

The point to all of this should not be missed. One way to frame the decision is to solve the programs in (11.1) and (11.2) and go with the larger profit. In this case you never ask what the second product costs. Another way to frame the decision is to go directly to the incremental profit of $\widehat{P}$-$cost$. Now your decision frame relies on specifying the product's cost.[7] And the appropriate specification could be current acquisition price, current sale price, present value of the future acquisition price, or present value of the future sale price. It all depends. What are the odds the current book value of the inventory provides an adequate answer?

What are the odds valuing the inventory at fair value would provide an adequate answer?[8]

Indeed, the current book value of the inventory can be thought of as a so-called sunk cost. Assume, for the sake of argument, that we paid 7 per unit for the I units of beginning inventory. This cost of 7 per unit has already been expended, and is irrelevant to the decision at hand. A *sunk cost* is a cost that arises from a previous decision that is irrelevant to a present decision. Depreciation is a classic example, along with historical cost of existing inventory.[9]

Regardless, a short-run decision problem does not necessarily reside entirely in the short-run. Similarly, the market price to place on a factor of production in some decomposed decision analysis may be far from obvious. Framing is a subtle art, and the cost expression a particular frame calls for may be close or far removed from what is found in the accounting library.

Get used to it!

---

[7] To complete the tale, return to the three principles of consistent framing. When we focus on the two programs in (11.1) and (11.2) we are engaging in component searches, by maximizing out the other products and material supply choices. Conversely, if we begin with the first program (and Π^*), we have left one set of choices outside the analysis. The opportunity cost of proceeding with the choice implied by the first program is $\widehat{\Pi}^*$, which is equal to $\Pi^* + \widehat{P}$ less the shadow price of I.

[8] Our concern for properly ascertaining material cost in this case disappears if we assume the markets are complete and perfect. We would then be able to buy and sell at the same price; and transactions could be consummated in present or future dollars. The market structure would then fully separate the material management question from the remaining decisions. For that matter, it would be difficult to understand why inventory would be on hand; but if it were, historical and market price would presumably be aligned.

[9] Closely related is the notion of relevant cost, a cost that varies with the decision at hand. Of course, separability issues soon come to the fore when we travel this path.

11.2   Make or Buy

We now examine a "make or buy" or "outsourcing" question. The generic issue is whether to produce some intermediate product or service inside the firm or acquire it from an outside source. Examples are refuse collection in a municipality, the split between internal and external auditing, chips or keyboards in the personal computer industry, permanent or temporary faculty, overtime or temporary labor services in a period of peak demand, and so on.

In turn, the issues involved may be straightforward or involved. The risk of being dependent on an outsider (including loss of proprietary information and opportunistic renegotiation of the terms of trade), concern for quality, managing technology change, and comparative advantage may be important. The outside supplier has similar concerns. For example, considerable investment may be required of the source; and this may place the source at risk. Dual sourcing, in which the item is acquired from two sources (perhaps internal and external) may be advantageous. This provides partial insurance against source failure, and it also may inject discipline into the control problem. It also may be needlessly costly.

A "small," short-run version of this theme is relatively unambiguous. Will the firm's short-run profit be higher with inside or outside sourcing? Questions of risk, long-run effects, quality, source opportunism, and technology change are absent (or are of second order importance). Rather, the intermediate product or service is available in the spot market. Economic forces may have led to excess capacity in this sector, and the firm suddenly finds itself in a position where an unanticipated short-run opportunity may be attractive. A downturn in the construction industry may make outside sourcing of some short-run maintenance attractive.

11.2.1   A Two Product Illustration

Consider a firm that manufactures and sells two consumer appliance products. Numerous parts are purchased from suppliers, including partially assembled components. The manufacturing process entails further assembly of some components, in a subassembly department, and final assembly of the two products, in an assembly department.

Denote the respective quantities that are produced and sold by q_1 and q_2. Capacity is constrained in the two departments:

$$
\begin{aligned}
\text{subassembly:} \quad & q_1 + q_2 \le 6,000 \\
\text{assembly:} \quad & q_1 + 2q_2 \le 10,000
\end{aligned}
$$

These constraints are exogenous, and unalterable. (This is a short-run setting.)

Direct labor (DL_S) and direct material (DM_S) costs in the subassembly department are described by the following $LLAs$:

$$DL_S = 10q_1 + 10q_2$$
$$DM_S = 110q_1 + 200q_2$$

Their counterparts in the assembly department are:

$$DL_A = 40q_1 + 80q_2$$
$$DM_A = 12q_1 + 15q_2$$

Overhead is estimated by

$$OV = 2,000,000 + 3.5(DL_S + DL_A).$$

Here, a plant-wide overhead LLA is used, with direct labor dollars as the synthetic variable. These are the only cost pools in our streamlined story; no other costs (e.g., selling and administrative) are involved.

Let $\widehat{P}_i$ denote the selling price of product i. For any feasible production plan, total revenue is $\widehat{P}_1 q_1 + \widehat{P}_2 q_2$. Total cost is the summation of the above direct labor, direct material, and overhead costs. This gives us a short-run profit expression of

$$
\begin{aligned}
\Pi(q_1, q_2) &= \widehat{P}_1 q_1 + \widehat{P}_2 q_2 - DL_S - DM_S - DL_A - DM_A - OV \\
&= [\widehat{P}_1 - 347]q_1 + [\widehat{P}_2 - 620]q_2 - 2,000,000.
\end{aligned}
$$

You should recognize the expressions in brackets as the respective product contribution margins, not to mention the close connection to normal, variable costing:

	$\widehat{P}_1$	$\widehat{P}_2$
price		
direct labor	50	90
direct material	122	215
variable overhead at 3.5(direct labor)	175	315
contribution margin	$\widehat{P}_1 - 347$	$\widehat{P}_2 - 620$

From here we identify the following program:

$$\Pi^* \equiv \max_{q_1, q_2 \geq 0} [\widehat{P}_1 - 347]q_1 + [\widehat{P}_2 - 620]q_2 - 2,000,000 \quad (11.3)$$
$$\text{s.t.} \quad q_1 + q_2 \leq 6,000$$
$$q_1 + 2q_2 \leq 10,000$$

Suppose $\widehat{P}_1 = 600$ and $\widehat{P}_2 = 1,100$. The solution to (11.3) has $q_1^* = 2,000$ and $q_2^* = 4,000$; with $\Pi^* = 426,000$. In addition, the shadow prices on the constraints are 26 and 227, respectively.

11.2.2   An Unusual Offer

At this juncture, a neighboring manufacturer offers to supply up to 500 units of the components for the second product that are assembled in the subassembly department. The offered price is P per unit. Any component purchased in this manner will not pass through the subassembly department but will enter directly into the assembly phase of the production process.

Let $\widehat{q}_2$ denote units of the second product produced in this fashion. With this added opportunity, de facto additional product, the direct cost $LLAs$ become:

$$DL_S = 10q_1 + 10q_2$$
$$DM_S = 110q_1 + 200q_2$$
$$DL_A = 40q_1 + 80q_2 + 80\widehat{q}_2$$
$$DM_A = 12q_1 + 15q_2 + 15\widehat{q}_2$$

Notice how the cost structure is affected. Since units manufactured with outsourced components skip the subassembly department, DL_S and DM_S are unaffected by $\widehat{q}_2$. Also, since assembly takes place in the same fashion, regardless of component source, $\widehat{q}_2$ affects DL_A and DM_A in the indicated manner. Of course, we also must pay the supplier. (To keep things simple, we further assume no incremental overhead, beyond that implied by the overhead LLA, is associated with the out-sourcing process.) Yet this is also a small decision, so the overall structure of the $LLAs$ remains unaffected.

The contribution margin for this variation on the second product is:

price	$\widehat{P}_2$
direct labor in assembly	80
direct material in assembly	15
component purchase price	P
variable overhead at 3.5(direct labor)	280
contribution margin	$\widehat{P}_2 - 375 - P$

We next expand the profit maximization program in (11.3) to accommodate this outsourcing option:

$$\widehat{\Pi}^* \quad \equiv \quad \max_{q_1, q_2, \widehat{q}_2 \geq 0} [\widehat{P}_1 - 347]q_1 + [\widehat{P}_2 - 620]q_2 \qquad (11.4)$$
$$+ [\widehat{P}_2 - 375 - P]\widehat{q}_2 - 2,000,000$$
$$\text{s.t.} \quad q_1 + q_2 \leq 6,000$$
$$q_1 + 2q_2 + 2\widehat{q}_2 \leq 10,000$$
$$\widehat{q}_2 \leq 500$$

Notice how $\widehat{q}_2$ enters the capacity constraint for assembly operations, but not for subassembly. The idea is to subcontract the first operation to the neighboring firm. Also notice, in light of the offer, we limit this opportunity to a maximum of 500 units. Of course, our framing of the outsourcing option leads to the conclusion the option is desirable if $\widehat{\Pi}^*$ of program (11.4) exceeds Π^* of program (11.3).

Staying with the noted selling prices, now further suppose the component price is $P = 250$. This leads to an optimal solution to (11.4) of $q_1^* = 3,000, q_2^* = 3,000$ and $\widehat{q}_2^* = 500$, along with a profit of $\widehat{\Pi}^* \equiv 436,500$. This represents an increase of 10,500.

Now for some fun. Suppose, instead of $\widehat{P}_1 = 600$ and $\widehat{P}_2 = 1,100$, we have $\widehat{P}_1 = 600$ and $\widehat{P}_2 = 1,150$. This implies the second product is more profitable than in the original case. The increase is sufficient to emphasize the second product in the optimal production plan. In particular, the solution to program (11.3) is now $q_1^* = 0$ and $q_2^* = 5,000$, with $\Pi^* = 650,000$. The shadow prices on the two constraints are 0 and 265, respectively. Only the constraint in the assembly department is binding.

Next, add to this revised price setting (with $\widehat{P}_2 = 1,150$, recall) an outsourcing price of $P = 250$. Expanded program (11.4) provides a solution of $q_1^* = 0, q_2^* = 5,000$ and $\widehat{q}_2^* = 0$, with $\widehat{\Pi}^* = 650,000 = \Pi^*$ and respective shadow prices on the three constraints of $0, 265$ and 0. The outsourcing offer is summarily rejected.

Let's instead frame the outsourcing decision in terms of incremental cost. Using the noted LLAs, and remembering assembly is unaffected by how the subassembly is sourced, we readily construct the incremental analysis in Table 11.3.

TABLE 11.3: Incremental Cost when $\widehat{P}_1 = 600$ and $\widehat{P}_2 = 1,150$

	make	buy	difference
direct labor in subassembly	10	0	-10
direct material in subassembly	200	0	-200
variable overhead at 3.5 direct labor	35	0	-35
outside price	0	250	250
total	245	250	5

This is straightforward, and raises the question of why we did not simply jump to this reduced form, incremental cost frame. After all, outsourcing saves labor, material, and overhead in the subassembly department. And here the savings are less than the offered price of 250. In particular, outsourcing increases the product's cost by 5 per unit.

Well, the reason for our seemingly pedantic approach is to be found in what we mean by the term cost in the reduced frame. To see this, return to the original setting where we assumed $\widehat{P}_1 = 600$ and $\widehat{P}_2 = 1,100$. Is

the outsourcing offer of $P = 250$ attractive here? Only the selling price of the second product has changed; the cost structure has not changed. And we already know outsourcing "costs" 5 per unit more than in-house production of the subassembly for the second product.

Returning to our exhausting, complicated frame based on comparing programs (11.3) and (11.4), we already know the former, where outsourcing is not present, has an optimal solution of $q_1^* = 2,000$ and $q_2^* = 4,000$, with $\Pi^* = 426,000$ and shadow prices on the two constraints of 26 and 227. In contrast, the optimal solution to (11.4) is $q_1^* = 3,000$, $q_2^* = 3,000$ and $\widehat{q}_2^* = 500$, with $\widehat{\Pi}^* = 436,500 > \Pi^* = 426,000$ and respective shadow prices of 26 and 227 and 21 on the three constraints.

So what did we miss with our claim outsourcing was a loser, as it raised the cost of the second product by 5 per unit in the presence of $\widehat{P}_2 = 1,100$? Notice the quantities of both products vary as we introduce the outsourcing alternative. Moreover, the first constraint, the subassembly capacity constraint of $q_1 + q_2 \leq 6,000$, continues to be binding when outsourcing is introduced. Outsourcing saves direct cost and overhead in the assembly department, and it reduces demand on the tight capacity. This capacity effect must be considered when we frame the choice in incremental terms.

In Table 11.4 we repeat our earlier reduced form, incremental cost frame, but with an important addition (pun) designed to introduce this capacity effect. Importantly, we have added a capacity cost of 26, which is the shadow price on the first constraint. This has the effect of changing the incremental cost of outsourcing from 5 to -21. Here it is less costly to outsource the component. Further notice the profit gain of $436,500 - 426,000 = 10,500$ is the cost saving of 21 per unit multiplied by the $\widehat{q}_2^* = 500$ units that are outsourced.[10]

TABLE 11.4: Incremental Cost when $\widehat{P}_1 = 600$ and $\widehat{P}_2 = 1,100$

	make	buy	difference
direct labor in subassembly	10	0	-10
direct material in subassembly	200	0	-200
variable overhead at 3.5 direct labor	35	0	-35
capacity cost	26	0	-26
outside price	0	250	250
total	271	250	-21

The shadow prices are the key to these incremental frames. In the case of Table 11.3, the second product is sufficiently attractive that we

---

[10] With the objective function and constraints in (11.3) being linear, the shadow price in this case remains constant as we move from the original to the improved solution, as the basis has not changed.

maximize its output, and this leads to unused capacity in the subassembly sector, which implies a shadow price of 0 on the corresponding constraint. Outsourcing relieves capacity demands in subassembly, but this relieved demand is of no value in this case. Conversely, in the case of Table 11.4, a mixture of both products is produced and, importantly, all of the subassembly capacity is being used. This is reflected in the shadow price of 26 on the corresponding constraint. Since outsourcing relieves this demand on subassembly capacity, capacity cost now enters the incremental analysis.

Another way to see this is to focus on the expanded setting of the program in (11.4). The shadow price on the third constraint (the one limiting the outsourcing to 500 units) reveals the rate at which profit increases with respect to outsourcing. As noted earlier, in the Table 11.3 case this rate is 0, while in the Table 11.4 case it is 21. In Table 11.4 we note that outsourcing is less costly by precisely 21 per unit. The incremental frame de facto reports the shadow price on this constraint.

The old adage of "no free lunch" is at work here. We can lay everything out in excruciating detail in programs (11.3) and (11.4) or we can work with an incremental frame that seems less demanding. The former approach simultaneously selects all output quantities, while the latter focuses on the outsourced product. Yet the products are interlinked because of the capacity constraints. This interlinkage is explicit in the (11.3) versus (11.4) frame, where the outsourcing price P is the outsourcing cost. But the interlinkage is implicit in the incremental analysis. This is why capacity cost (of 0 or 26) enters the incremental analysis, as another component of the outsourcing cost calculus in the seemingly less demanding incremental analysis. The details are present, whether they are in explicit or implicit form.

11.3   Product Evaluation

A parallel situation arises when we examine the profitability of products in a short-run, small setting. To explore this, we introduce another potential product into the running story. For this purpose we assume the outsourcing option is not available, but otherwise all of the structure in program (11.3) remains, including respective selling prices of 600 and 1,100 per unit. The new twist is short-run market conditions limit each product to a maximum of 2,000 units. So the maximization in (11.3) now becomes

$$\widehat{\Pi}^* \equiv \max_{q_1, q_2 \geq 0} [\widehat{P}_1 - 347]q_1 + [\widehat{P}_2 - 620]q_2 - 2{,}000{,}000 \qquad (11.5)$$

$$\text{s.t.} \qquad q_1 + q_2 \leq 6{,}000$$

$$q_1 + 2q_2 \leq 10{,}000$$

$$q_1 \leq 2{,}000$$

$$q_2 \leq 2{,}000$$

The obvious solution is $q_1^* = 2,000$ and $q_2 = 2,000$, with $\Pi^* = -534,000$. Also, the shadow prices on the four constraints are, respectively, $0, 0, 253$ and 480. Moreover, the firm now has excess capacity: $6,000 - q_1^* - q_2^* = 2,000$ in the subassembly department and $10,000 - q_1^* - 2q_2^* = 4,000$ in the assembly department.

At this point, an unanticipated customer arrives. This customer requests a bid on a special product. Capacity is available. No interactions with present or future products are anticipated. The new customer is "small" in every respect.

Further assume each unit of this potential product will consume one unit of the scarce resource in both the subassembly and assembly areas. Let q_3 denote units of this new product. The capacity constraints now become

$$
\begin{aligned}
\text{subassembly:} \quad & q_1 + q_2 + q_3 \leq 6,000 \\
\text{assembly:} \quad & q_1 + 2q_2 + q_3 \leq 10,000.
\end{aligned}
$$

It is further determined that manufacturing the new product will incur direct labor cost in subassembly of 10 per unit, direct material cost in subassembly of 150 per unit, direct labor cost in assembly of 50 per unit, and direct material cost in assembly of 10 per unit. This implies we have the following LLAs:

$$
\begin{aligned}
DL_S &= 10q_1 + 10q_2 + 10q_3 \\
DM_S &= 110q_1 + 200q_2 + 150q_3 \\
DL_A &= 40q_1 + 80q_2 + 50q_3 \\
DM_A &= 12q_1 + 15q_2 + 10q_3
\end{aligned}
$$

This is assumed to be a "small" decision. Our LLAs extend in ready fashion to accommodate the new alternatives. This is an important assumption.

Now let $\widehat{P}$ denote the selling price per unit. Further recall (back to the make or buy illustration) that the overhead LLA is given by $OV = 2,000,000 + 3.5(DL_S + DL_A)$. We thus have an estimated contribution margin for this potential product of $\widehat{P} - 430$:

price	$\widehat{P}$
direct labor	60
direct material	160
variable overhead at 3.5(direct labor)	210
contribution margin	$\widehat{P} - 430$

It also turns out special tooling will be required if this new product is manufactured. This tooling will cost $15,000$. Thus, if q_3 units are man-

ufactured, and if output of the first two products remains constant, the incremental profit will be

$$[\widehat{P} - 430]q_3 - 15,000$$

We are now prepared to answer various questions. Suppose the price is $\widehat{P} = 1,000$. How many units must be manufactured and sold if this is to be a profitable venture? This is a break-even question. Setting incremental profit equal to zero and solving for q_3 gives

$$q_3 = 15,000/[1,000 - 430] \approx 26.32$$

Similarly, suppose $q_3 = 800$ units. (Notice we have the excess capacity for this many units.) What is the minimum price if this is to be an interesting product? Setting incremental profit equal to zero and solving for $\widehat{P}$ (given $q_3 = 800$) gives

$$\widehat{P} = 430 + 15,000/800 = 448.75$$

More generally, this is a classic short-run optimization exercise. Is this short-run opportunity profitable? We estimate the incremental cost of the first unit to be $15,000 + 430$. Beyond that, we estimate the incremental (and marginal) cost of additional units at 430. The 430 datum is constant, within the relevant range of our LLAs and excess capacity. Moreover, no interactions with other products or competitors are envisioned. Thus, we speak unambiguously of the product's incremental (or marginal) cost. The capacity cost issue in the outsourcing excursion is absent here, as the capacity shadow prices are both 0.

Example 11.4 Stay with the above story, where the incremental cost of q_3 units of this potential product is $15,000 + 430q_3$. Now, however, a competitor is present. His incremental cost is $15,000 + 427q_3$ or $15,000 + 433q_3$ with $50 - 50$ odds. The competitor privately knows his cost. The customer has asked for bids in the form of $15,000 + b \cdot q_3$. She will subsequently announce the required quantity, $400 \leq q_3 \leq 1,400$. The bid, b, is required to be a whole dollar amount and a coin is flipped in the event of a tie. It turns out an equilibrium is for our firm to bid $b = 433$ and the competitor to bid $b = 432$ if she is the low cost type and $b = 434$ if she is the high cost type. You should be able to verify the claim this is an equilibrium. Meantime, notice how this bid decision is small, i.e., completely separated from the other products and within the range of the various LLAs, except for the strategic interaction with the competitor.

11.4   Customer Evaluation

A parallel exercise arises when we think in terms of adding or dropping a particular customer. Again presume a short-run setting, with small effects.

If the customer in question purchases a single specialized product, this is just a repetition of the earlier discussion. If the customer purchases a variety of products, we would repeat the earlier discussion but would focus on the particular set of products purchased.

Interesting issues arise here. For example, the customer may place unusual service demands on us. Alternatively, the customer may provide unusual feedback on product quality or insights into new product proposals.

Continuing along this line, we eventually reach the conclusion this is not really a small problem. Taking on the new customer, or dropping an existing customer, is likely to have effects that interact with other activities, interactions that last well beyond the current period, and that are not well-approximated by the existing set of LLAs. Moreover, the firm's reputation for supporting its customers comes into play, and this is clearly a strategic issue (which returns us to the point of Example 11.4.)[11]

11.5   Uncertainty

The final stop on our overview of small decisions is the question of uncertainty. Surely uncertainty is present. The question is when and how to give it formal standing in a decision analysis. We might ignore it.[12] We might give it implicit standing, for example, by acknowledging our cost estimates are uncertain and then attempting to buttress them with statistical digestion of the accounting library. Or we might formally introduce risk.

The professional manager makes these judgments in an informed manner. Here we explore some dimensions spanned by that judgment.

11.5.1   Option Value of Flexibility

One issue concerns the option value of flexibility. To illustrate, return to the additional product setting of program (11.5), and its original version in program (11.3). Also stay with the original prices of $\widehat{P}_1 = 600$ and $\widehat{P}_2 = 1,100$. Given these prices, we know in the original story of program (11.3) that all of the capacity is used, and that respective shadow prices on the capacity constraints are 26 and 227. We also know that when short-

---

[11] This ubiquitous strain of identifying when a decision is small or large can be further explored in a work force scheduling context. There we face a specific type of make or buy decision: whether to acquire additional labor services from the existing (permanent) work force, from a temporary work force, or from an expanded permanent work force.

[12] We have, in fact, made considerable use of such myopia. In subsequent chapters we will not have this flexibility. Control problems only arise when uncertainty is present. Our conceptual thinking at that time must then recognize uncertainty. It is too important to gloss over in that arena.

run market conditions limit each product to a maximum of $2,000$ units, the firm has idle capacity, and using some of this idle capacity on the proposed third product provides an incremental profit of $[\widehat{P} - 430]q_3 - 15,000$.

Now, however, suppose the short-run market restrictions might well disappear. Let α denote the probability they disappear. This means the new product uses scarce capacity with probability α and idle capacity with probability $1 - \alpha$. Moreover, the scare capacity carries respective shadow prices of 26 and 227.

Further recalling this third product requires one unit of each capacity type, we now conclude the incremental profit is

$$[\widehat{P} - 430]q_3 - 15,000 - (26 + 227)q_3$$

with probability α and

$$[\widehat{P} - 430]q_3 - 15,000$$

with probability $1 - \alpha$.

Presuming risk neutrality, we wind up with the conclusion the expected incremental profit is

$$[\widehat{P} - 430]q_3 - 15,000 - \alpha(26 + 227)q_3$$

This additional term, $\alpha(26+227)q_3$, reflects the option value of the now idle capacity. Using capacity on the third product lessens the firm's flexibility to respond to hopefully improving market conditions. This is what the option value or capacity cost measures. This option value or capacity cost is, of course, central to the incremental decision frame yet has no natural tie to the accounting library. It is simply not recorded in the accounting library.

11.5.2 Cost of Risk

A second issue concerns risk aversion. To lighten the details (I, too, am growing weary), suppose we have identified a short-run opportunity. The profit possibilities are particularly simple. The selling price is $100,000$. The estimated incremental cost is either $70,000$ or $120,000$ with $50 - 50$ odds. This implies the incremental profit will be either $100,000 - 70,000 = 30,000$ or $100,000 - 120,000 = -20,000$, again with $50 - 50$ odds.[13]

On the surface this appears rather obvious. The new opportunity offers an expected gain of $.5(30,000) - .5(20,000) = 5,000$. So absent any significant interactions this is an attractive opportunity.

---

[13] If the selling price were at least $120,000$ we could readily dismiss the uncertainty on grounds of first order stochastic dominance. The worst that might happen is no gain!

However, this is a story with uncertainty, and interactions induced by risk aversion and the risk profile of the firm's existing projects are a possibility. To specify these possibilities, recall (from Chapter 9) we model the firm's risk preference with a utility function defined on wealth (w), denoted $U(w)$ with initial wealth of w_i.[14] In the present case, this initial wealth refers to the firm's existing projects and may well be random.

Continuing, we resort to a state-outcome specification to tie all of this together. For this purpose, four equally likely states will do. Examine the (incremental) cash flow and initial wealth profiles in Table 11.5. The new project is as advertised: a gain of $30,000$ or a loss of $20,000$ with $50-50$ odds. This is combined with existing projects that have an expected value of $45,000$. In case 1, the existing wealth is risk free. In the other cases it is either $70,000$ or $20,000$ with $50-50$ odds. In case 2, the new and existing projects are probabilistically independent. In case 3 they are perfectly negatively correlated (and when combined offer a wealth of $50,000$ regardless of the state), while in case 4 they are perfectly positively correlated (and when combined offer a wealth of either $100,000$ or 0, again with $50-50$ odds).

TABLE 11.5: State-Outcome Specifications

	state s_1	state s_2	state s_3	state s_4
probability, $\pi(s)$	.25	.25	.25	.25
outcomes				
new project	30,000	30,000	-20,000	-20,000
existing projects (w_i)				
case 1	45,000	45,000	45,000	45,000
case 2	70,000	20,000	70,000	20,000
case 3	20,000	20,000	70,000	70,000
case 4	70,000	70,000	20,000	20,000

Now let CE_0 denote the certainty equivalent of the firm's existing projects, and CE_1 the corresponding certainty equivalent when the existing projects are combined with the new project. Notice in Table 11.5 that the new project offers an expected incremental gain of $E[\Delta\Pi] = 5,000$. Now express the two certainty equivalents in tautological though illuminating fashion as follows:

$$CE_1 = CE_0 + E[\Delta\Pi] - RP_1 \qquad (11.6)$$

---

[14]Recall it is somewhat awkward to speak of the firm's utility function or preference measure. The theory of the firm, we have noted, is unsettled in important areas, including what the firm's goals might be in a setting more friction laden than perfect and complete markets. This does not negate the importance of uncertainty. It merely makes it more difficult to understand.

Importantly, the new project is worthwhile only if it increases the certainty equivalent, only if $CE_1 > CE_0$. But this is true only if $E[\Delta\Pi] - RP_1 > 0$. That is, the new customer, the new project, is worthwhile only if its expected incremental gain of $E[\Delta\Pi] = 5,000$ exceeds the risk premium associated with this incremental gain, the RP_1 term in expression (11.6). This risk premium depends on the firm's attitude toward risk and the interaction between the new and existing projects' risks.

To calculate RP_1 we work through (11.6) in the following fashion:

$$RP_1 = CE_0 - CE_1 + E[\Delta\Pi] \qquad (11.7)$$

See Table 11.6 where we do this for the four cases in Table 11.5 and the three utility measures introduced in Table 9.3.

TABLE 11.6: RP_1 for Various Cases and Utility Assignments

utility for wealth w	case 1	case 2	case 3	case 4
$U(w) = w$	0	0	0	0
$U(w) = \sqrt{w}, w \geq 0$	3,349	9,780	-3,792	21,208
$U(w) = -\exp(-\rho \cdot w), \rho = .00001$	3,093	3,093	-3,093	8,918

Naturally, if the firm is risk neutral (i.e., $U(w) = w$), this risk premium is nil, as risk is a matter of indifference. So regardless of how the existing and new risks interact, we find $RP_1 = 0$ and the decision is essentially a small decision. Conversely, if the firm is risk averse but with constant risk aversion (i.e., $U(w) = -\exp(-\rho \cdot w)$), a risk premium of $RP_1 = 3,093$ emerges in case 1, where there is no status quo risk, and in case 2, where the existing and new risks are independent. If the two risks are negatively correlated, the new project actually decreases the risk of the existing projects (as the combined effect of the existing and new projects leads to a guaranteed wealth level of $50,000$), and we then have a risk premium of $RP_1 = -3,093$. But if the interaction between the risks is exacerbating, as in case 4 where double doses of good news or double doses of bad news materialize, the risk premium grows dramatically, $RP_1 = 8,918$. Thus, if the risks do not interact (cases 1 and 2) the decision is essentially small in scope; otherwise the risks interact and the decision is large in scope.

Of course this clear demarcation between small and large reflects constant risk aversion. Turning to the case where risk aversion is present but not constant (i.e., $U(w) = \sqrt{w}$), we find the decision is always large. The reason is the attitude toward risk depends on both the riskiness of the existing projects and their scale. Initial wealth is no longer benign in the assessment of attitude toward risk, a point you were warned about in Chapter 9.

This may appear to be setting a record for convoluted recalculation. But an important point is emerging. Our story began with a comparison

of expected incremental revenue and expected incremental cost. This is a profit oriented calculation, one with natural ties to the accounting library. (If the chapter were not getting too long, we would drive this home by walking through how the accounting library would record the consequences of this decision.)

We then introduced risk aversion and framed the decision in terms of expected incremental profit less a risk premium. The resulting risk premium in expression (11.7) might be nil, significant though the decision is small or significant though the decision is large. And other than the case where it is nil, it is an additional cost component, one that we might label the "cost of risk." This risk premium or cost is central to the decision frame yet has no natural tie to the accounting library. Cost of risk is simply not recorded in the accounting library.[15]

11.6   Interaction with Taxes

Once we start looking for interactions, it seems they appear at every twist and turn. What started out as a nicely contained analysis of small decisions has served to warn us that the boundary between small and not so small is often subtle and idiosyncratic.

Just to turn this message a little deeper, we should acknowledge the importance of taxes. We often think (or claim) income taxes are important in long-run but not short-run decision settings. Suppose the marginal tax rate is some constant τ. For every incremental dollar of profit, the firm then receives $(1 - \tau)$ net of taxes.

Under risk neutrality, maximizing expected profit and maximizing $(1 - \tau)$ times expected profit lead to the same decisions. (This is a simple application of our first principle of consistent framing.) Some modest care is necessary when we move into risk aversion, as we must be careful to distinguish pre- and post-profit utility functions. We should also admit a constant marginal tax rate is not guaranteed. What if a loss occurs? The immediate marginal rate might drop to zero. And it is a short step from here to worrying about the probabilistic nature of loss risks across various projects.[16]

---

[15] This cost of risk phenomenon returns us to the issue of identifying the firm's goal or goals. This will lead us in the following chapter to think in terms of risk and return. There, in the spirit of modern finance, we will use discounting techniques with a market determined discount rate that is appropriate for the risk at hand. Although there is a temptation to invoke this machinery here, doing so would obscure the point that with less than perfect markets the firm itself may have diversification incentives. If so, the boundary between large and small decisions becomes more obscure.

[16] For example, risk neutrality and an increasing marginal tax rate imply risk aversion in pre-tax dollars.

Taxes are not benign.

11.7   Summary

The accounting library is retrospective, while decision making is prospective. The library will record the accounting interpretations of whatever decisions we make. It also may contain important information that helps us analyze a particular decision problem. Nonetheless, the link between our decision analysis and the accounting library may be close or tenuous. It all depends on the decision circumstance, decision frame, and library procedures.

Small decisions are somewhat free of interactions with other activities and strategic concerns and also do not contemplate movement outside the relevant range of the prevailing *LLA*s. Thus, small decisions often exhibit a close link between their portrayal in the accounting library and their analysis. Simple cost-volume-profit exercises, such as estimating the profitability of an additional customer or break-even analysis, are illustrative of this fact. Of course, we should remember the difference between gross and contribution margin here. A short-run profitable venture may appear unprofitable if we record it using full costing procedures and forget to account for the effects on normal, full costing's ubiquitous plug to cost of goods sold. (Glance back at Example 6.1.)

Even so, we should not be naive. The direct material odyssey illustrates the point. There we had an accounting cost, presumably historical cost, of the material in question. Yet the cost of this material in our highlighted decision frame depended in intimate ways on current and future prices and plans. The inherently short-run problem had an inherently long-run connection.

The make or buy, product evaluation, and customer evaluation vignettes all illustrate ways in which small decisions may take on larger dimensions. Introducing uncertainty further clouds the distinction. We stress the distinction because it highlights the important managerial art of knowing how much detail to load into a decision analysis.

On a final note, our study of accounting information in short-run decisions has been silent on who makes these decisions. This depends on the authority structure within the firm. Whether the decision is pondered at a "high" or "low" level in the firm is independent of the importance of framing. Whether the importance of the accounting library varies with location in the firm is another matter. To answer this we would have to specify the availability of other sources of information across the firm. A traveling sales agent knows costs and market conditions. A maintenance supervisor knows repair costs, subcontracting alternatives, and so on. A grocery clerk knows products (e.g., kumquats versus tangerines) and prices.

11.8   Bibliographic Notes

The idea here is to frame a decision with the smallest possible package; a small decision is very accommodating in this respect. The three principles of consistent framing are clearly at work. A more pragmatic view has us use a frame that is further reduced by ignoring modest or small complications and interactions. Admitting this implies the analysis is an approximation (yes, art meets fundamentals once again). Howard [1971] calls this proximal decision analysis. Demski and Feltham [1976] highlight the approximation theme in terms of a simplification and tie it to costing questions. Interactions in the risk domain are a central feature of modern finance (e.g., Ross [2005]). It is also possible interactions among decisions will put serious strain on the expected utility apparatus itself (e.g., Amershi, Demski, and Fellingham [1985]). Interactions between taxes and risk aversion are explored in Fellingham and Wolfson [1985].

11.9   Problems and Exercises

1. What is the distinction between (i) a large and a small decision and between (ii) a short-run and a long-run decision? Give an example where a short-run decision is small, another where a short-run decision is large. Then give two more examples, where a long-run decision is small and where a long-run decision is large.

2. Discuss the relationship between break-even analysis and our earlier study of variable costing.

3. Why are the graphs in Figure 14.1 truncated?

4. *break-even calculations*
   Ralph is planning a visit to the bank to solicit a small business loan. Ralph's business plan, in summary form, is given by the following *LLA*s:

revenue	$TR = 240q$
manufacturing cost	$TMC = 125,000 + 100q$
selling and administrative	$S\&A = 85,000 + 20q$

   (a) The bank, after studying Ralph's numbers, asks what the break-even point is. What is it, and why might the bank be interested in it?

   (b) Ralph then points out that the business plan calls for production of $q = 2,500$ units the first year. Under GAAP style income measurement, using full costing (with a normal volume of $2,500$

units), how many units must Ralph sell in the first year for accounting income to be zero?

(c) Explain the difference between your two break-even calculations.

(d) What would you, as the banker, say to Ralph's comment in (b) above?

5. *large break-even and output calculations*[17]
Ralph's cost curve is piece-wise linear. For output of $0 \leq q \leq 1,000$ units it is given by $C(q; P) = 1,000 + 6q$; for $1,000 \leq q \leq 2,000$ it is given by $C(q; P) = 3,000 + 4q$; and for $2,000 \leq q$ it is given by $C(q; P) = -5,000 + 8q$.

(a) Plot Ralph's cost curve.

(b) Suppose the selling price is $\widehat{P} = 7$ per unit. Plot the implied total revenue curve on your graph in (a); also locate Ralph's break-even point. Repeat for cases where the selling price is $\widehat{P} = 8$ and $\widehat{P} = 9$.

(c) Again assume the selling price is $\widehat{P} = 7$. Locate Ralph's optimal output.

(d) Now approximate Ralph's cost curve with an LLA of $3,000+4q$; notice this approximation is consistent with the optimal output chosen above as well as the original break-even calculation. Suppose the selling price unexpectedly drops to $\widehat{P} = 4.8$. Using the LLA of $3,000 + 4q$, calculate Ralph's best choice of output (somewhere between shutdown and a maximum of $2,000$ units).

(e) What mistake has Ralph made in part (d) above?

6. *cost versus expenditure*
It is often claimed that arranging a long term supplier contract for materials will insulate you from price changes in the materials market. Coase [1968] contends this is erroneous. Carefully analyze the claim. (You may want to reflect on the example surrounding Table 11.1.)

7. *cost versus expenditure*
The material cost illustration in Table 11.1 stresses the difference between an appropriate measure of cost for some purpose and the expenditure on the factor in question. Does a similar comment apply to labor cost? Carefully explain.

8. *cost versus expenditure*
Return to the two period inventory setting of Tables 11.1 and 11.2, where we framed a new product opportunity in terms of incremental revenue less incremental cost.

---

[17] Contributed by Richard Sansing.

(a) Carefully explain the shadow prices on the initial inventory for the six cases in Table 11.1. Also verify these shadow prices using your favorite optimization software, e.g. Excel. Reconcile your approach with that in Chapter 2's Appendix.

(b) Why is this shadow price the appropriate material cost in the incremental cost frame?

(c) Give a variation on the assumed prices and quantities such that the appropriate material cost in the incremental formulation is 8.1. Do not change any of the first period prices.

(d) What happens in the setting of Table 11.1 if we set $P^+ = 15$ and $P^- = 14$? What does this tell you about a formulation that uses these second period prices?

(e) Is the historical cost of the material a sunk cost? Can you provide a frame of the new product decision that makes this most obvious?

(f) Briefly discuss what happens in this setting if the firm faces a constant 40% marginal tax rate on accounting income.

9. *accounting library and decision frame*
Ralph's Library is a two product firm. Quantities of the two products are denoted q_1 and q_2. Ralph has studied the situation, and decided to rely on the following $LLAs$:

direct labor cost	$DL = 90q_1 + 95q_2$
direct material cost	$DM = 50q_1 + 100q_2$
first overhead pool	$OV_1 = 400,000 + 3(DL)$
second overhead pool	$OV_2 = 200,000 + 1(DM);$
selling and adm.	$S\&A = 700,000 + 10q_1 + 20q_2$

In addition, Ralph estimates total revenue via $TR = 860q_1 + 960q_2$. Ralph also faces capacity constraints. Machine hours are limited in each of two departments to a total of 6,000. (Each department has a capacity of 6,000 machine hours.) Machine hour requirements are as follows:

	product 1	product 2
department one	1	2
department two	2	1

So, for example, a unit of the first product requires one machine hour in department one and two machine hours in department two.

(a) Determine Ralph's optimal output and associated maximum profit.

(b) Summarize your recommendation in (a) above with two income statements, one based on full costing and the other based on variable costing.

(c) Ralph's Cousin, a meddling individual, insists product 2 is unacceptable on aesthetic grounds. Ralph decides to examine life on the assumption $q_2 = 0$. What is Ralph's opportunity cost of following the Cousin's suggestion?

(d) Return to part (a) above. Ralph now encounters a new customer. This customer seeks a custom-made item. Ralph estimates this will require direct labor of 150 and direct material of 150. It will also require 2 machine hours in each department. Ralph also is looking for the easy way out. Determine the cost of this modified product such that Ralph should oblige this customer if and only if the offered price $\widehat{P}$ is greater than cost.

(e) Suppose Ralph decides to oblige this new customer in (d) above. Further suppose output and the $LLAs$ turn out just as expected. Everything produced is sold, except the custom-made product for the new customer didn't get shipped until the first day of the next fiscal year; it remains (fully completed) in inventory at year end. What inventory value will the accounting library place on this unit? Assume variable costing is used for this purpose. Carefully explain why this cost recorded in the accounting library differs from that constructed in part (d) above.

(f) Just for fun, return to part (d) above. Formulate an optimization program that uses three products, the first two modeled in part (a) above and the third the one referred to part (e), with a selling price of $\widehat{P}$. Be certain to add the constraint $0 \leq q_3 \leq 1$. Carefully explain why the cost associated with this third product differs from the cost associated with the same product in part (d) above.

10. *accounting library and decision frame*
Ralph produces two products, with respective quantities denoted q_1 and q_2. Product costs are aggregated into direct material (DM), direct labor (DL), and overhead (OV) categories. Relevant $LLAs$ are summarized by $TMC = 160,000 + 420q_1 + 480q_2$ and $S\&A = 10,000 + 80q_1 + 20q_2$. ($TMC$ denotes total manufacturing cost, and $S\&A$ is the total of all period costs.) No inventory is present, as all production is sold. Revenue is given by $TR = 1,000q_1 + 700q_2$. In addition, Ralph's output is limited by the following pair of departmental capacity constraints:

$$2q_1 + q_2 \leq 1,000$$

$$q_1 + 2q_2 \leq 1,000$$

For convenience, no other costs are present.

(a) Determine Ralph's optimal production plan, and the shadow prices for each of the two constraints.

(b) After this original plan is formulated, a customer enquires about production of a single unit of a custom product. Ralph estimates incremental TMC will be 300, and incremental $S\&A$ will be 0. This product will require two "hours" of capacity in each department. Determine the cost of this custom product, assuming Ralph wants to compare the cost of the product to the offered price $(\widehat{P})$ to decide whether to oblige the customer.

(c) Conversely, formulate a three variable optimization model to simultaneously determine the optimal quantities of the two original products as well as the custom product described in (b) above.

(d) Carefully explain why you have two distinct cost figures for the same product, one in part (b) and the other in part (c).

(e) Presuming production of this custom product, what unit cost for this product will be recorded in the accounting library?

11. *statistical analysis*
    Return to the outsourcing illustration in Tables 11.3 and 11.4. Everything remains as specified, and the prices of the products are $\widehat{P}_1 = 600$ and $\widehat{P}_2 = 1,100$. The new feature concerns whether the overhead LLA is well specified. To this end, the following data from the 10 most recent periods are extracted from the accounting library.

t	OV_A	OV_S	DL_A	DL_S	DM_A	DM_S
1	1,428	1,596	256	76	24	612
2	1,811	2,228	446	106	49	985
3	1,775	2,306	428	78	61	986
4	1,005	2,239	205	25	86	1,003
5	1,687	1,701	404	54	113	676
6	1,568	2,502	365	15	130	1,028
7	1,299	2,256	262	82	37	1,066
8	1,625	2,268	366	26	79	1,016
9	1,570	2,405	385	45	122	1,069
10	1,411	1,656	234	98	42	679

Notice we have two distinct overhead pools, one for assembly (OV_A) and one for subassembly (OV_S). No other overhead is present, so total overhead is simply $OV = OV_A + OV_S$. The other cost pool totals refer, of course, to direct labor and direct material in the assembly and subassembly spheres. No other direct costs are present.

(a) Using these data, regress total overhead on total direct labor cost. Is your resulting LLA estimate statistically consistent with that assumed originally?

(b) Using this newly minted overhead LLA, determine the implied contribution margins and optimal solution when the outsourcing option is available (with price $P = 250$ of course).

(c) How much would the firm pay to distinguish between the original and your new statistically estimated LLA?

(d) Further exploration suggests overhead in the subassembly area is best related to direct material (DM_S), while overhead in the assembly area is best related to direct labor (DL_A). Performing the implied regressions, are the data consistent with this conjecture?

(e) Now what happens, using the two overhead LLAs to the contribution margins and optimal solution?

(f) Why does the projected profit vary as you move among the three possible specifications of the overhead structure?

12. *more of the same*
Return to problem 11 above but now assume prices of $\widehat{P}_1 = 600$ and $\widehat{P}_2 = 1,150$. What happens to the optimal solution as you move among the three possible specifications of the overhead structure? Explain your finding.

13. *interactions in customer evaluation*
Return to the product evaluation discussion in the text, where a potential third product, with quantity q_3, was under consideration. Presuming limited market conditions for the other products (of $q_1 \leq 2,000$ and $q_2 \leq 2,000$), and a selling price of $\widehat{P} = 1,000$, we derived a break-even quantity of $q_3 = 15,000/[1,000 - 430] \approx 26.32$. We now assume there are no market constraints on the first two products (thereby dropping the $q_1 \leq 2,000$ and $q_2 \leq 2,000$ constraints). Suppose we acquire the necessary tooling, at a cost of $15,000$, and produce q_3 units of this new product.

(a) Without the noted market constraints, the production of the first two products is affected by production of the third. Suppose $0 \leq q_3 \leq 2,000$. Determine the best choice of q_1 and q_2, given an exogenous q_3 in the noted range. (Recall their respective selling prices are $\widehat{P}_1 = 600$ and $\widehat{P}_2 = 1,100$.)

(b) Now suppose the selling price of the new product is $\widehat{P} = 1,000$ per unit. How many units must be produced and sold if accepting this new product is a good idea?

(c) Carefully explain the difference between the original break-even calculation and that which you performed in (b) above.

(d) Suppose $q_3 = 800$ units. What is the minimum price for this to be an interesting product?

(e) Again presuming $0 \leq q_3 \leq 2,000$, what is the incremental cost of the third product?

14. *possible misspecification of overhead LLA*
Empire Electronics (EE) is an assembler of custom electronic compo- nents. It faces an opportunity to bid on a particular assembly. Direct material is estimated to cost $14,000$ and direct labor $112,000$. Vari- able overhead is allocated at 61.75% of direct labor cost. A special consulting fee of $18,000$ will also be incurred if the assembly project is taken on. (An incremental cost of $213,160$ is thus implied.) This fee reflects the standard retainer arrangement EE has with a human resources consulting firm. Labor is in short supply and EE will hire 2 temporary laborers if it successfully bids on the project. The con- sulting firm does the necessary search, interviewing, applicant testing, and so on, all for a price of $9,000$ per employee supplied. The tempo- rary laborers will leave after the assembly work is completed, with no additional compensation. Moreover, the 61.75% datum stems from a recent cost analysis. The overhead account contains mainly la- bor support activities, fringe benefits, supervision, inventory control, payroll administration and so on. The data listed below were used to produce this estimate.

t	OV (000)	DL$ (000)	no. of hires
1	5,915	11,326	120
2	5,922	11,339	120
3	6,600	12,074	160
4	5,648	11,127	110
5	3,903	8,976	100
6	3,097	6,275	70
7	4,491	8,625	110
8	6,556	11,863	180
9	5,321	9,837	110
10	4,579	9,689	80

At this point a member of the management team questions the cost estimate of $213,160$. Research indicates the consulting fee is always charged to manufacturing overhead. So a question of double counting arises.

(a) Is it correct to combine the regression estimate of variable over- head with the $18,000$ datum?

(b) Provide an estimate of the cost to EE of performing the assembly in question.

15. *inferring competitor's cost*

Ralph is considering entering the custom E-mail device market. His device is customizable and will operate across a variety of systems. Before proceeding, Ralph decides to appraise the competition. The closest is Enterprise Products (or EP). EP markets a similar product, though it lacks the versatility of Ralph's design. A consultant gathers recent data from EP's financials and reports the following regression that relates reported cost of goods sold (cgs) to units sold (q_s) for the EP product: $cgs = 938,248 + 59q_s$. (This is an excellent statistical fit to the underlying data.) Ralph is ecstatic, as the estimated variable cost of 59 per unit is well above Ralph's variable cost of 42 per unit. Alas, the next day Ralph accidentally sees a confidential cost analysis prepared by the accounting group at EP. Their report includes the following regression of total production cost (tpc) on units produced (q_p): $tpc = 2,230,207 + 43q_p$. Carefully explain the difference between the two regressions.

16. *diagnosis of competitive position*[18]

Ralph's Packaging, Inc. (RP) designs and produces specialized packages for a variety of industrial product firms. Most jobs are won on a competitive bid. The major steps in the production process are design, printing, cutting, and assembly. Historically, RP has earned a 14% margin, but lately the margin has been declining and most recently was 7.2%. RP fears its earlier success has led to complacency and its costs are unnecessarily high. The accounting library identifies labor, material, subcontracting, energy, and space costs. Some are broken down by design, printing, cutting and assembly. Since work in process has never been a significant problem, no formal job costing system has been used.

At this point Ralph decides to look a little more closely at the cost conjecture. A recently completed job, job 113, is randomly selected. Ralph searches through the purchase orders, stores requisitions, and subcontracting invoices and locates the following cost items that pertain to job 113:

miscellaneous materials	125
standard packaging materials	1,875
subcontracted printing	425

Working through payroll records, Ralph is also able to identify direct

---

[18] Inspired by an IMEDE case titled Tipografia Stanca S.P.A.

labor time. Three labor groups are present, regular, semi-skilled and skilled. Their respective wage rates are 11, 18 and 22 dollars per hour. The time sheets record the following hours:

	unskilled	semi-skilled	skilled
design	10	11	34
printing	10	18	20
cutting	12		
assembly	10	10	24

Finally, overhead averages 110% of labor cost. (To identify overhead cost, Ralph took the total of all manufacturing cost, subtracted the labor cost that could be identified with specific jobs and the material and subcontracting costs that could be similarly identified.)

(a) What was the unit cost of job 113?

(b) The bid sheet for job 113 shows that, at the time the job was bid, RP estimated the direct cost as follows:

materials	2,100
subcontract work	400
design labor	918
printing labor	799
cutting labor	238
assembly labor	714

for a total direct cost of 5,169. In turn, the job was bid using RP's standard bidding rule of *bid = 180% of estimated direct cost.* For bidding purposes, labor is costed at 17 per hour. What was RP's bid on job 113? Did RP earn a positive profit on this job?

(c) Suppose RP's labor cost does average 17 per hour and that materials and subcontract work is, on average, equal to direct labor cost. Further suppose overhead averages 110% of labor cost. Presuming many bidding successes using the noted bidding rule, what should RP's margin be?

(d) What advice can you give RP? In particular, why do you think their margin is declining?

(e) Do you think they should invest in a sophisticated product costing system?

17. *option value of capacity*
   Ralph's Custom Products (RCP) is a custom manufacturer of material handling equipment. Various just-in-time manufacturing systems

require parts from suppliers that arrive in containers that are specialized to accommodate transportation and handling in the receiving facility. RCP designs and manufactures these containers. A new customer has arrived, seeking a bid on a particular set of containers. The RCP engineer provides the following estimates:

	dept. #1	dept. #2
machine hours required	150	200
direct labor hours required	120	350
cost per hour of direct labor	18	24

RCP uses two manufacturing departments. The direct labor rates include 20% fringe (covering various taxes, vacations, and so on). Overheads in the two departments are budgeted with the following LLAs: $OV_1 = 150,000 + 14MH$ and $OV_2 = 200,000 + 45DLH$, where MH refers to machine hours in department #1 and DLH refers to direct labor hours in department #2. (Respective normal volumes are $MH = 7,500$ and $DLH = 5,000$.) In addition, the engineer estimates total direct material cost will be $12,000$ and shipping costs will total $4,000$.

(a) What are the normal, full cost overhead application rates in the two departments?

(b) What is the minimum price RCP should consider in negotiating with this new customer?

(c) How does the cost datum derived in (b) above relate to the cost that would be reported in the accounting library? More precisely, what product cost would be recorded in the typical accounting library? In turn, by how much would cost of goods sold increase were this product produced and sold?

(d) Now suppose a capacity problem might exist. One of RCP's usual customers might require some modification of containers in use. If so, taking on the new customer will use up slack in department #2's schedule that should be devoted to the existing customer base; and if this happens, RCP will be forced to subcontract 200 direct labor hours, at a rate of 150 per hour. The sales force estimates the existing customer will require this modification with probability α. If RCP is risk neutral, what now is the minimum price it should consider in negotiating with this new customer?

18. *certainty equivalents*
Determine the CE_0 and CE_1 certainty equivalents for each of the cases and utility measures in Tables 11.5 and 11.6.

19. *risk premia*

    Return to the setting of Table 11.5, and the constant risk aversion story of $U(w) = -\exp(-\rho \cdot w), \rho = .00001$. Why is the risk premium $3,093$ when the risks are independent but precisely the negative of this amount when the risks are perfectly negatively correlated? Hint: what is the risk premium when initial wealth is 0 and the project will provide $50,000$ or 0 with $50 - 50$ odds?

20. *taxes and risk aversion*

    Ralph has been offered an interesting gamble. With probability .5, Ralph will gain $500 and with probability .5 Ralph will gain $100. The gain is net of the purchase price; also, the possible outcomes are independent of any other items in Ralph's portfolio.

    (a) Suppose Ralph is risk neutral. Determine the gamble's certainty equivalent.

    (b) Ralph remains risk neutral, but faces a constant marginal tax rate of 40%. Determine the gamble's certainty equivalent. Can Ralph safely ignore taxes in this circumstance?

    (c) Now suppose Ralph is risk averse, with a utility measure defined over wealth w of $-\exp(-\rho w)$ and $\rho = .001$. Repeat parts (a) and (b) above.

12

Large versus Small Decisions: Long-Run

In this chapter we continue our exploration of decision making, but with a focus on decisions with long-run consequences. Again our concerns are with distinguishing large and small decisions and with links to the accounting library.

The large versus small concern is a recurring theme. Have we strained the credibility of our LLAs to the extent they should be modified? Should we worry about interactions with other decisions? Should we worry about strategic dimensions? As we have emphasized, there is no ready-made answer to these concerns. Short-run decisions are not necessarily small, just as long-run decisions are not necessarily large. On the other hand, it seems we should expect many long-run decisions to be large, reflecting the local nature of our LLAs and their natural if not inevitable connection to competitor concerns.

Whether small or large, decisions always have links to the accounting library. Earlier events, recorded in the library, may give important clues to consequences of the contemplated choice. For example, cost experiences with earlier products may be useful in contemplating new products. Similarly, whatever choice is made and whatever consequences follow, some portrayal will eventually be catalogued in the library. Oddly, we usually use one model to analyze a long-run decision and another to reflect its consequences in the accounting library.

Long-run decisions, by definition, have consequences that fall over an extended time frame. This suggests a focus on present value, but also provides a rich setting in which to continue our study of decision framing. It also returns us to the awkwardness of an incompletely specified objective or

criterion measure for the firm. In the fortuitous world of perfect and complete markets, the firm's long-run decisions would be governed by present value maximization, with market specified discount rates. In a less friendly market structure, the firm's objectives are ambiguous, and so is the use of present value analysis. Lacking guidance on this score, we adopt the traditional approach and emphasize present value techniques in the analysis of long-run decisions.[1]

We begin with the present value criterion and a look at yet additional applications of the three principles of consistent framing. We then turn to the question of estimating cash flows. As usual, this is a pragmatic exercise, one that labors the small versus large distinction: interactions with other decisions or projects, LLA adequacy, taxes, and competitive response. Finally, we return to the accounting library and contrast the decision analysis and accounting renderings of the forces that play on the firm in this setting.

12.1   Back to Present Value

In Chapter 8 we introduced a prototypical decision problem: A set of alternatives, denoted A, and a criterion function, denoted $\omega(a)$, are given. One and only one of the alternatives is to be chosen. As depicted in expression (8.1), the best choice is the available one that produces the largest value of the criterion function:

$$\max_{a \in A} \omega(a) \qquad (12.1)$$

In Chapter 9 we stressed the idea that choices lead to consequences and the criterion measure, $\omega(a)$, is reflective of these consequences, e.g., the expected utility of wealth. The world of long-run decisions and present value analysis is simply another application of this portrayal of economic rationality.[2]

Two features come to the fore. First, by definition, the long-run story is one where consequences transpire through time; so the time at which a consequence transpires is an important part of the story. For example, receiving 100 dollars today is rather different from receiving 100 dollars 9 years from today. And given the importance of cash flow timing, we model each alternative, $a \in A$, in terms of the implied sequence of time-dated cash flows.

Second, the criterion function is the present value of the identified cash flow consequences, using an "appropriate" interest or discount rate. In

---

[1] This is not capricious. We have repeatedly encountered ambiguity in specifying the firm's objective. At the same time, we know present value techniques are widely used.

[2] Indeed, we have already encountered long-run, present value analyses in Chapters 3 and 4.

turn, if uncertainty is present, we focus on the expected values of the (uncertain) cash flows, and use a discount rate that is appropriate for the underlying riskiness of the alternative.

Throwing a little more notation at the story, alternative $a \in A$ here amounts to a sequence of cash flows:

$$a = [x_0, x_1, ..., x_T]$$

where x_0 is the cash flow at time $t = 0$, x_1 the cash flow at time $t = 1$, etc. And the criterion function of present value, using interest or discount rate r, is[3]

$$\omega(a) = PV(a) = \sum_{t=0}^{t=T} x_t(1+r)^{-t} \tag{12.2}$$

Many refer to the calculation in (12.2) as net present value (or NPV). The adjective is carried along to remind us we are including the cash inflows and the cash outflows in the calculation. PV seems sufficient.

Example 12.1 Ralph is contemplating expansion. The choice boils down to expand (a_1) or not to expand (a_2). Analysis reveals the proposed expansion will result in the following annual incremental cash flow sequence (000 omitted):

TABLE 12.1: End-of-Period Cash Flows

$t = 0$	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
-504	124	138	123	114	264	7

That is, an immediate investment, or cash outflow, of 504 will be followed by cash inflows of 124 one year hence, 138 two years hence, and so on. Rejection means the opportunity is lost forever; it cannot be deferred.

As the cash flows are incremental relative to not expanding, we have $PV(a_2) = 0$ and, using $r = 12\%$ and (12.2) we have

$$
\begin{aligned}
PV(a_1) &= -504(1.12)^0 + 124(1.12)^{-1} + 138(1.12)^{-2} + 123(1.12)^{-3} \\
&\quad +114(1.12)^{-4} + 264(1.12)^{-5} + 7(1.12)^{-6} \\
&= 30.07
\end{aligned}
$$

Thus, with $PV(a_1) = 30.07 > PV(a_2) = 0$, Ralph prefers expansion to the status quo. Expansion is economically equivalent to collecting a windfall

---

[3]It is time to glance back at our development in Chapter 3, especially expressions (3.1) and (3.2), where we move from prices to an interest or discount rate. In turn, jumping to the uncertain setting in Chapter 9, where we modeled the outcome as being jointly produced by the choice and a random state, the pricing in expression (3.1) would have time and state dependent prices.

gain of 30.07 (000) in current dollars.

Several comments are in order. First, the display format in Example 12.1 should be noted. By convention, we record cash outflow as negative and cash inflow as positive. The cash flows are also depicted as occurring at the end of the respective periods. They are treated as annual amounts, for example. We might want to think in monthly, quarterly, or semi-annual amounts. Practice suggests the annual reckoning is often adequate, as is the assumption cash flows occur at the end of the respective years.

Second, we exploited the opportunity cost framing principle in Example 12.1. Surely there are other uses for Ralph's funds. But these other uses are contained in the formally excluded set of alternatives (A_2 in our formal definition). But here we assume any such funds earn at the rate of $r = 12\%$. In short, $r = 12\%$ is the opportunity cost of capital in the Example. Further notice any investment at $r = 12\%$ has a present value of zero. So we know the best choice in the excluded set has a present value of zero. Therefore, taking the expansion project, with its present value of 30.07, is the best choice overall.[4]

Third, reflect on the way we defined the decision problem in (12.1): one and only one of the alternatives is to be chosen. Thus, if various long-run projects are on the table and some can be taken in various combinations, each such combination or "super project" is listed as a distinct alternative in our formulation. This can be awkward, but is conceptually clarifying (a pithy point for sure).[5]

The present value rule, then, is familiar and straightforward. Implementing the rule, however, begs two important questions: Where did the cash flow estimates come from? Where did the discount rate r come from?

If markets are complete and perfect, we know the cash flows as a function of whatever events beset the economy. We also know the market price for a dollar at every point in time, as a function of whatever events beset the economy. So if markets are complete and perfect, we have our questions answered.[6] Of course, this is by definition. Such a world only exists in textbooks. Real markets are not complete and perfect.

---

[4]More precisely, the opportunity cost of confining the search to the two noted options in Example 12.1 is a present value of 0. It is important to remember our definition of opportunity cost!

[5]Suppose, due to market or organizational imperfections, that available funds are limited. The present value criterion then says pick that combination of projects that maximizes the total present value, subject to not exceeding the available investment funds. This requires, given our presumed frame in (12.1), that A be defined in terms of all project combinations that are feasible in terms of the available funds. The rationalization does get thin. Perfect markets lead to the present value rule, and we now invoke its use when some type of friction is present.

[6]Another way to express this is to remember insurance is always available, for a price, in a regime of complete and perfect markets. In such a regime, any project we might think up is insurable. This implies an equivalent way to evaluate projects is to

Modern finance emphasizes discounting the expected values of the respective cash flows at a discount rate appropriate for the project's risk.[7] Projects with the same risk are in the same risk class. A risk free project, for example, would be discounted at the risk free rate. Similarly, a project in the firm's risk class would be discounted at the firm's weighted average cost of capital, which is tautologically the discount rate appropriate for the risk class. A project in another firm's risk class would be discounted at that firm's cost of capital. But the present value criterion reigns.

In pragmatic terms, the professional manager estimates the cash flows and makes a managerial judgment as to the appropriate discount rate to use in the calculation. The firm itself may have a policy that prescribes the discount rate as a function of the type of investment. For example, new product projects might be discounted at $r = 14\%$, expansion of existing stable projects at $r = 11\%$, and so on.

The firm's approach to identifying, selecting, and managing investment opportunities is likely to have a significant effect on its success. It is serious business. For this reason we typically find considerable involvement, documentation, and monitoring in the investment arena. We also encounter a variety of analysis techniques, techniques that diverge from present value.

---

think of their expected cash flows as occurring for certain, while their necessary expected investment outflow increases by the market demanded insurance premium.

[7]By analogy, we have in earlier chapters distinguished the expected value of a lottery from its certainty equivalent. We might think of this as "discounting" the expected value to restate it in certainty equivalent terms. In a broader sense we are recycling our ever-present theme of explicating the firm's goals or preferences. Whether we are faced with a single or a multi-period exercise, analysis of decision alternatives presupposes some specification of these preferences. This is why, for example, the casually obvious notion of risk is so difficult to define. Consider two lotteries with the same expected value. Call them α and β. Let F be a set of utility functions. Lottery α is less risky than lottery β for "preference class F" if for each utility function in F the expected utility of α is weakly greater than the expected utility of β. The difficulty is that what we mean by risk depends on the utility function. If F only contains well-behaved quadratic functions, then risk is measured by variance. Otherwise, it is not. Therefore, depending on F we can (or cannot) be highly specific about how to measure risk. Varying F alters what we mean by risk; and the broader F is, the more difficulty we have guaranteeing that one lottery is more risky than another. Inherently, then, thinking in risk and return terms presumes we have said something about the firm's preferences. Moreover, with a broad F, we readily find lotteries that are not comparable in terms of risk, just as we can find information sources that are not comparable.

And as if this were not enough, we also must remember the framing possibilities. Suppose mean and variance of all lotteries is what is important. In incremental terms, then, we worry about covariance. To illustrate, let x and y be two random variables. The variance of $x + y$ is the variance of x plus the variance of y plus twice the covariance of x and y. Incrementally, then, adding y to the portfolio increases the variance by the variance of y plus twice the covariance. We confronted this very issue in our analysis of uncertainty in the prior chapter.

12.2   Present Value Pretenders

Two commonly used portrayals that diverge from the present value criterion are internal rate of return and payback. The former focuses on the implied earning rate of an investment, while the latter focuses on the time to recovery of the investment. These are discussed below. A third, the accounting rate of return, will be discussed later in the chapter. Importantly, none of these portrayals is a consistent transformation of the present value exercise.

12.2.1   Internal Rate of Return

Return to our capacity expansion proposal in Table 12.1. Discounting the cash flow sequence at $r = 12\%$ gave us a present value of 30.07. It seems our project would earn more than $r = 12\%$. Its present value, that is, would be zero for some discount rate larger than 12%.

This is where the internal rate of return enters the story. The internal rate of return is that discount rate $r = irr$ such that the present value of a given cash flow sequence is zero. Fix the cash flow sequence. If we think of PV as a function of r, the internal rate of return is the value of r for which PV is zero.

For the case at hand, we solve the following expression to determine the internal rate of return:

$$
\begin{aligned}
PV(a_1) \quad = \quad & -504(1+irr)^0 + 124(1+irr)^{-1} + 138(1+irr)^{-2} \\
& + 123(1+irr)^{-3} + 114(1+irr)^{-4} + 264(1+irr)^{-5} \\
& + 7(1+irr)^{-6} \\
= \quad & 0
\end{aligned}
$$

The solution is $irr \cong .141317 = 14.1317\%$.

Dwell on the intuition. The project calls for us to invest 504 immediately, in exchange for some future cash inflows. Discounting those future cash inflows at $r = 12\%$ gives a positive present value. Discounting them at a rate over 12% lowers their present value. The crossing point, between positive and negative present value, is approximately 14.1317%.

The analogy to break-even calculations should be apparent. See Figure 12.1, where we plot present value of this cash flow sequence as a function of the discount rate, r.

Many find this an intuitive and comfortable portrait. If we take the expansion project, funds will earn at the rate of $irr \cong 14.1317\%$. This is more than their cost of $r = 12\%$. Beyond that, "earning 14.1317%" is a more intuitive statement than "capturing a present value of 30.07."

Unfortunately, there is no guarantee a focus on the internal rate of return is an error free transformation of the present value criterion in (12.2). Two difficulties emerge.
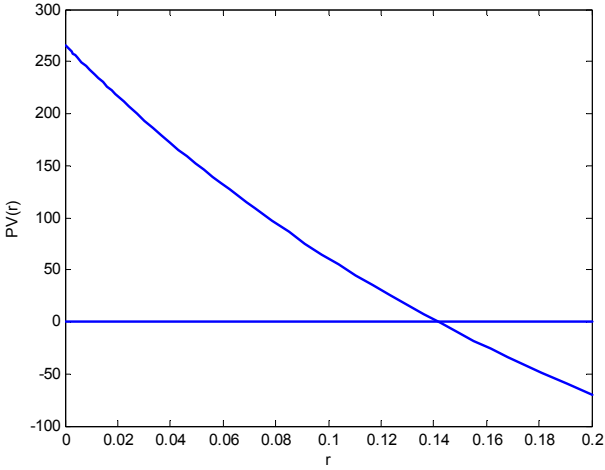
FIGURE 12.1. Present Value as a Function of r

Multiple Internal Rates of Return

One difficulty is the ambiguity caused by multiple internal rates of return. Jump back to the above expression where we solved for irr using the data in Table 12.1. Instead, multiply the expression by $(1+irr)^6$. With $PV = 0$ (by definition of irr), this gives us:

$$\begin{aligned} (1+irr)^6 PV(a_1) \quad &= \quad -504(1+irr)^6 + 124(1+irr)^5 + 138(1+irr)^4 \\ & \qquad +123(1+irr)^3 + 114(1+irr)^2 + 264(1+irr)^1 + 7 \\ &= \quad 0 \end{aligned}$$

This is a polynomial of degree $T = 6$. In solving for irr, we solved a $6th$ degree polynomial.

Recall from algebra that a polynomial of degree T has T roots. The roots might all be the same, in which case they are called repeated roots. They also might be different (or even imaginary). Our example has a single positive root of $irr \cong 14.1317\%$, along with a negative root and two pairs of imaginary roots. With a single positive root, there is no ambiguity as to what the internal rate of return is. In fact, this is the case whenever $a = [x_0, x_1, ..., x_T]$ is such that x_0 is negative and all subsequent cash flows are positive.

More generally, though, multiple roots are problematic.[8] To illustrate, consider a project of $a = [-100, 290, -208]$. Think of this as an environ-

_____

[8] Descartes' Rule of Signs is helpful. Let k be the number of changes of sign in the coefficients of our polynomial. Then the number of positive roots of the polynomial is k or k reduced by an even integer. If x_0 is negative and all other x_t are positive in (12.2),
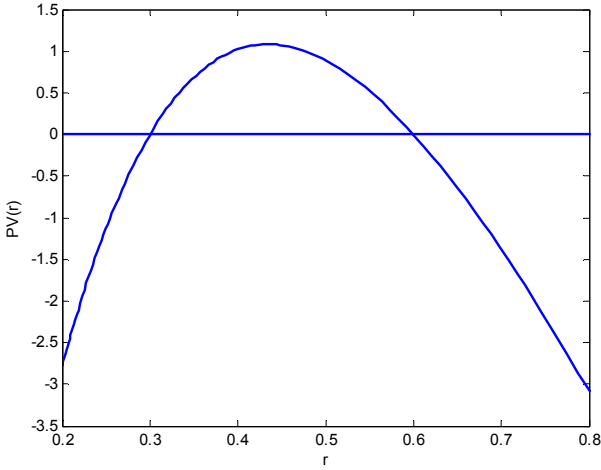
FIGURE 12.2. $PV = -100 + 290(1 + r)^{-1} - 208(1 + r)^{-2}$

mentally sensitive project that calls for significant cleanup or restoration at the end of its useful life.

PV as a function of the discount rate r is plotted in Figure 12.2. PV is zero at $irr = 30\%$ and at $irr = 60\%$. PV is positive between these two values, and negative otherwise. We thus have two values for irr, 30% and 60%! What, then, is the project's irr? There simply is no unambiguous answer.

Even this observation does not exhaust the unusual nature of the illustration. Suppose the appropriate discount rate is $r = 10\%$. We then have a present value of $-100 + 290(1.1)^{-1} - 208(1.1)^{-2} = -8.2645$. If we are using a present value criterion, this project is unacceptable. Yet it has an irr of $30\% > 10\%$ and of $60\% > 10\%$. This further illustrates the fact, we should say tautological observation, that whenever present value and internal rate of return analyses differ, the latter is wrong if the former is correct.[9]

---

we have one change of sign and therefore one positive root. This is the case in Figure 12.1.

[9]Parenthetically, can you identify the source of the inconsistency here? The project has a negative salvage value. If r is sufficiently large, this is not too onerous. For low r it is. (Recall Figure 12.2.) So we want the lower irr below r to make certain the negative salvage value is not too onerous. But then if r is quite large, we again lose interest in the project from a present value perspective. The reason is the intermediate inflows become less and less valuable as we increase r, and eventually are overwhelmed by necessary outflows.

Mutually Exclusive Projects

The second potential inconsistency between present value and internal rate of return frames arises in choice among mutually exclusive projects, which is precisely the format we presumed at the start of this odyssey, expression (12.1). To illustrate, suppose we must select between project 1 and project 2, where for convenience $T = 1$:

TABLE 12.2: Two Projects

	$t = 0$	$t = 1$	irr	PV $(r = 10\%)$
choice a_1	-100	120	20%	9.09
choice a_2	-1	10	900%	8.09

No multiple solution ambiguity is present. But notice a_2 has the higher irr, while a_1 has the larger PV (given a presumed discount rate of $r = 10\%$).

Present value and internal rate of return give conflicting advice here. Maximizing internal rate of return is not consistent with maximizing present value. The difficulty arises because the present value frame assumes the cash flow is reinvested at the presumed rate r, while the internal rate of return frame assumes it is reinvested at irr.

The message is not deep. Present value analysis assumes reinvestment at the exogenously specified r. Internal rate of return analysis assumes reinvestment at the endogenously determined irr. The two may conflict when we face mutually exclusive choices that have unequal investment amounts, unequal investment lives, or even equal investment amounts and lives but at least two periods.

12.2.2   Payback

Another frequently encountered portrayal of investment opportunities is payback. This is simply the minimal length of time for the cumulative net cash flow from the investment opportunity to be positive. In abstract terms, with project $a = [x_0, x_1, ..., x_T]$ the payback period is the minimum time, t_{PB}, such that $x_0 + x_1 + \cdots + x_{t_{PB}} \geq 0$. No discounting is involved. Cash flow beyond t_{PB} is ignored.[10]

_____

[10] A caveat should be noted. If the cash flows beyond t_{PB} are not all positive, we should call the payback period the minimum time beyond which the cumulative cash flow remains positive. With this subtlety, we then worry about cash flows beyond the identified t_{PB} to the extent they might turn the cumulative cash flow negative at a later date. Naturally, $x_0 < 0$ and all other cash flows positive (as in Table 12.1) do not cause any such concern.

Again using the data in Table 12.1 for illustrative purposes, we find $t_{PB} = 6$:

$$\sum_{t=0}^{t=1} x_t = -504 + 124 = -380$$

$$\sum_{t=0}^{t=2} x_t = -504 + 124 + 138 = -242$$

$$\sum_{t=0}^{t=3} x_t = -504 + 124 + 138 + 123 = -119$$

$$\sum_{t=0}^{t=4} x_t = -504 + 124 + 138 + 123 + 114 = -5$$

$$\sum_{t=0}^{t=5} x_t = -504 + 124 + 138 + 123 + 114 + 264 = 259$$

A long payback is often thought of as a risky project. We must wait quite a while before earning positive cumulative cash flow; and this gives us a long time in which to encounter bad luck. This is certainly a casual notion of risk (pun); and it is unlikely to agree with a more sophisticated notion of risk. Payback surely does not provide a reliable frame of a sophisticated investment decision.[11]

Still, we should not be too hasty to condemn the calculation. If the payback is one year, and the project lasts 10 years it sounds attractive. If the payback is 20 years, and the project lasts 21 years, we are immediately suspicious that it is not very interesting. Somewhere in between might be a pragmatic filter. A "short" payback is a signal, other things equal, that risk considerations are not of first order importance. A "long" payback is the opposite signal. For example, does a payback period of 6 years in our running example signal we have a lot of time for things to go wrong?

12.2.3 Framing

We dwell on the possible inconsistencies between present value and alternative analyses because they provide another lesson in the art of framing.

---

[11] You might enjoy the following (thanks to Gordon [1955]). Let x_0 be negative and $x_1 = x_2 = \cdots = x_T = z$ be positive. The payback period is now, roughly, $t_{PB} = |x_0|/z$. Further suppose T is large. The present value, at discount rate r, of the cash inflows is $z[1 - (1+r)^{-T}]/r$. If we now solve for the internal rate of return, we will find that it is approximated by $z/|x_0|$, or the reciprocal of the payback period. This does not imply the payback period is, more broadly, useful. It does, however, remind us to understand the economic environment we encounter and how well various decision frames stand up in that environment.

Our first principle of consistent framing was the irrelevance of strictly increasing transformations. Figure 12.1 tells such a story. A positive PV and irr above r are the same thing in that picture. In fact, holding r constant, a larger PV corresponds to a larger irr. We have an increasing transformation. Figure 12.2 is the absence of an increasing transformation from PV to irr. Similarly, naive analysis of mutually exclusive projects can produce inconsistencies between the present value and internal rate of return portraits.

We suspect these inconsistencies are not of major concern in most practical cases. For example, multiple internal rates of return require multiple sign switches in the cash flow projections. This sounds like replacement problems (where we periodically replace an aging asset such as a large truck) or those with unusual salvage characteristics.

Use of payback presents a somewhat different framing story, even for a pragmatist. If present value is the norm, we might find internal rate of return more intuitive and use it so long as we are not led (too far) astray. Yet, if present value is the norm it is difficult to understand why anyone would bother to compute the payback period. We do know payback is often computed, together with other calculations, such as present value. This suggests ambiguity in the present value frame itself. Perhaps there is concern over well-specifying the cash flow uncertainties or the appropriate discount rate. Payback may then enter as one of several analytic pictures that are taken of the investment opportunities. In this case, we then acknowledge an ambiguous framing exercise coupled with a portfolio of approaches to the framing task.

Present value is an intuitive frame (and criterion). It also has its roots in a world of complete and perfect markets. While tempting to advocate use of a present value frame, we should pause to remember that framing is a managerial art. It is informed by theory and practice, and it is crafted with a heavy dose of managerial judgment. The astute manager knows the frame that is being used, and is tuned to its strengths and weaknesses.

12.3   Cash Flow Estimation

Of course the frames in place rely on judicious estimation of the cash flows. As usual, we stress the importance of managerial judgment. We also emphasize the potential largeness of investment decisions. They can easily cut across boundaries within the firm, call LLAs into question, and raise issues of competitive response in the factor and product markets.

There is, of course, no surefire way to proceed in the estimation game. For that reason we examine a mildly complicated illustration, designed to give a feel for complexity and the role of professional judgment in any such exercise.

12.3.1 An Earlier Story

The cash flow data in Table 12.1 reflect a capacity expansion opportunity in the two product illustration of Chapter 11 (especially section 11.2.1) There the story concerned manufacture and sale of two consumer products. Two production departments were present: subassembly and assembly. Capacity constraints were:

$$\text{subassembly:} \quad q_1 + q_2 \leq 6,000$$
$$\text{assembly:} \quad q_1 + 2q_2 \leq 10,000$$

where q_1 and q_2 denote quantities of the two products.

The short-run cost function was estimated by aggregating direct material, direct labor, and overhead cost components. With an "S" subscript denoting subassembly and an "A" subscript denoting assembly, we assumed the following $LLAs$ for direct labor (DL), direct material (DM) and overhead (OV):

$$DL_S = 10q_1 + 10q_2$$
$$DM_S = 110q_1 + 200q_2$$
$$DL_A = 40q_1 + 80q_2$$
$$DM_A = 12q_1 + 15q_2$$
$$OV = 2,000,000 + 3.5(DL_S + DL_A)$$

We also assumed respective selling prices of $\widehat{P}_1 = 600$ and $\widehat{P}_2 = 1,100$. Any variable selling and administrative is netted out here. We continue this pattern, simply to keep the discussion at a reasonable length.

Pulling these details together, we had contribution margins as follows:

	600	1,100
price	600	1,100
direct labor	50	90
direct material	122	215
variable overhead at 3.5(direct labor)	175	315
estimated marginal cost	347	620
contribution margin	253	480

And maximizing total contribution less the intercepts of the $LLAs$ led to:

$$\Pi^* \equiv \max_{q_1, q_2 \geq 0} 253q_1 + 480q_2 - 2,000,000 \qquad (12.3)$$
$$\text{s.t.} \quad q_1 + q_2 \leq 6,000$$
$$q_1 + 2q_2 \leq 10,000$$

with a solution of $q_1^* = 2,000, q_2^* = 4,000$ and $\Pi^* = 426,000$.

This is our status quo.[12]

12.3.2   The Proposed Project

The proposed project entails increasing the capacity of each department by 1,500 units. This incremental capacity can be purchased for 300,000. It will last 5 years. Estimating the cash flows (besides the purchase price of the additional equipment) requires we foretell how this additional capacity will be used. A good starting point is to redo the maximization in (12.3), but with the added capacity:

$$\widehat{\Pi}^* \quad \equiv \quad \max_{q_1,q_2 \geq 0} 253q_1 + 480q_2 - 2,000,000 \qquad (12.4)$$
$$\text{s.t.} \qquad q_1 + q_2 \leq 6,000 + 1,500$$
$$q_1 + 2q_2 \leq 10,000 + 1,500$$

The solution is $q_1^* = 3,500, q_2^* = 4,000$ and $\widehat{\Pi}^* = 805{,}500$. This implies an annual incremental gain, presumably gain in cash flow, of $\widehat{\Pi}^* - \Pi^* = 379{,}500$.

Investing 300,000 to receive 379,500 per year for five years sounds fairly attractive, if not too good to be true. Our quick and dirty analysis, though, presumes no change whatever in the original cost and revenue structures, not to mention the question of tax effects or strategic nuances. And if this litany of concerns is irrelevant, we indeed have a small, long-run decision.

12.3.3   Is this a Large Decision?

Moving into the large domain, we now document several alterations to our original analysis. This should not be taken as a checklist to be examined in each and every setting, but as a suggestive encounter with the art of estimation.

In working through the alterations, keep in mind we are estimating the cash flows. Think of the original optimization in (12.3) as the production plan that will be in place if no expansion occurs. We then must determine what production plan will be in place if expansion occurs. The difference in cash flows between these two regimes is the cash flow picture we seek. In a broader sense, we do not mean to suggest a well-done estimation exercise rests on a series of short-run optimization exercises. We do mean to suggest it rests on a thorough understanding of what is to be done with the altered set of resources.

---

[12]The original story also entertained a variation on one of the products that used an outsourced component. While we drop this possibility from the current illustration, it is useful to reflect on how the possibility of these types of opportunities further clouds our ability to make precise cash flow estimates.

Our approach consists of three steps. First, we assume the original optimization in (12.3) accurately depicts the status quo. If the investment is not taken, the noted production plan will be in place along with the noted cost and revenue structures. Second, we appropriately alter the cost and revenue structures in this program to determine the production plan that will be in place if the investment proposal is accepted. Third, we then adjust the difference in total contribution margins between the two plans to account for additional cash flow consequences.

Selling Prices

Selling prices depend on market forces. We have also lumped any variable selling and administrative items, for convenience, into the revenue estimates. Expanding output may, or may not, call these estimates into question. The new production schedule, we shall see (and already saw in our tentative analysis where the decision was treated as essentially a small decision) entails considerable expansion of the first product's output. For this reason, we assume the selling price of the first product will drop 1.5% (implying $\widehat{P}_1 = 591$) if capacity is expanded. The second product's selling price remains constant. Total annual revenue is thus estimated to be $591q_1 + 1{,}100q_2$ if expansion takes place. Implicitly, we further assume any effect on selling and administrative costs is negligible.

LLAs

The added capacity will result in altered work flows in the production process. The direct costs are not expected to change, though this is not the case for overhead. Indeed, the change is sufficiently large that separating the overhead into subassembly (OV_S) and assembly (OV_A) pools is called for. The *LLAs* under expansion are estimated to be:

$$OV_S = 1{,}000{,}000 + .4DM_S$$
$$OV_A = 1{,}200{,}000 + 3DL_A$$

Notice the subassembly overhead uses direct material in subassembly as the synthetic variable, while the assembly overhead uses direct labor in assembly as the synthetic variable. Also notice the total of the intercepts is 10% higher than in the status quo. For later reference, no expansion costs or depreciation associated with the proposed expansion are included in these *LLAs*.

The estimated marginal costs and contribution margins are displayed below. The net effect on the firm's marginal costs is to slightly decrease the estimated marginal cost of the first product (from 347 to 336), and to slightly increase that of the second (from 620 to 625).

price	591	1,100
direct labor in subassembly (DL_S)	10	10
direct material in subassembly (DM_S)	110	200
subassembly variable overhead: $.4DM_S$	44	80
direct labor in assembly (DL_A)	40	80
direct material in assembly (DM_A)	12	15
assembly variable overhead: $3DL_A$	120	240
estimated marginal cost	336	625
contribution margin	255	475

The New Plan

At this point we turn to the best use of the expanded resources. This amounts to revisiting our initial attempt in (12.4), but inserting our revised revenue and cost estimates. We have

$$\widehat{\Pi}^* \equiv \max_{q_1, q_2 \geq 0} 255q_1 + 475q_2 - 2,200,000 \qquad (12.5)$$
$$\text{s.t.} \qquad q_1 + q_2 \leq 6,000 + 1,500$$
$$q_1 + 2q_2 \leq 10,000 + 1,500$$

The solution is $q_1^* = 3{,}500$, $q_2^* = 4{,}000$ and $\widehat{\Pi}^* = 592{,}500$. This implies an incremental gain of

$$\widehat{\Pi}^* - \Pi^* = 592,500 - 426,000 = 166,500 \qquad (12.6)$$

This is our estimate of the incremental net cash flow from operations.

Expansion Costs

Next is the question of investment cost. The equipment, we noted earlier, will cost 300,000, and last 5 years. Salvage value is zero. Additional costs, of training new workers, of altering the plant to accommodate the new equipment, and so on will total 90,000. So the immediate cash outflow will be 390,000.

Taxes

Taxes are also important. They are paid periodically throughout the year, though for simplicity we assume they are paid at year's end (as we assume all cash flows occur at year's end). There is also a significant cash flow timing wedge between acquisition and tax consequences of a long-term decision.

Tax law is complex and constantly changing. It is also multifaceted, as the typical story, at least in the U.S., is federal, state and local provisions. A well thought out tax strategy is equally complex. We cannot hope to

introduce all of the specifics at this point, so we will content ourselves with a broad brush treatment.

We assume the combined (e.g., federal, state and local) tax rate is a constant 40% of incremental taxable income. We also estimate incremental taxable income as equal to incremental accounting income, except we use a tax authority mandated depreciation schedule. For tax purposes, we assume the investment will be classified as five-year property. Under the modified accelerated cost recovery system (MACRS) the tax basis of the investment will be depreciated 20% in the year of acquisition, 32% the next year, and 19.2%, 11.52%, 11.52% and 5.76% in the remaining years.

What is the tax basis? Here we assume the acquisition price of 300,000 is the basis, or depreciable amount. We also assume the additional cost of 90,000, associated with worker training and minor plant alteration, will be immediately expensed for tax purposes.

Our mosaic is presented in Table 12.3. Study it carefully. The first thing you will notice is we have rounded the cash flows estimates to the nearest thousand. Given the countless judgments involved it seems, well, silly to pretend we have these judgments nuanced to the very last dollar.

That said, the incremental cash flow from operations is estimated by the difference in our two short-run optimization expressions, (12.6). Recall this is exclusive of any alteration or acquisition costs and is also rounded, so 166,500 is recorded as 167 (000). From there we subtract tax related cash flows and, of course, the $t = 0$ investment expenditures.

TABLE 12.3: Incremental End-of-Period Cash Flows (000)

	$t = 0$	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
cash from operations	0	167	167	167	167	167	0
tax expense depreciation		60	96	58	35	35	17
alteration	90						
taxable income	-90	107	71	109	132	132	-17
tax at 40%	-36	43	28	44	53	53	-7
acquisition	-300						
alteration	-90						
working capital	-150					150	
	-504	124	138	123	114	264	7

Incremental taxes are estimated at 40% of estimated incremental taxable income. For periods $t = 1$ through $t = 5$, incremental taxable income is estimated as incremental cash flow from operations less tax code depreciation on the investment basis of 300,000. Notice the depreciation expense, for tax purposes, continues into period $t = 6$. As we assume the project

lasts 5 years, the only incremental effect in period $t = 6$ is depreciation for tax purposes of 17. This reduces taxable income in period $t = 6$ by 17, and thus reduces taxes payable in that period by $.4(17) \cong 7$.

Also notice what occurs at time $t = 0$. The alteration expenditures of 90 are fully expensed at that time. The incremental effect is to reduce taxable income at that time by 90. This reduces taxes payable in that period by $.4(90) = 36$.

Both calculations, at $t = 0$ and at $t = 6$, presume the status quo taxable income is sufficiently positive that these tax reductions occur at the noted times.

Working Capital

One other item should have caught our attention in Table 12.3: working capital. We assume additional working capital of 150,000 is required. This will be infused at the start of the project and returned at the end of the project's life, at $t = 5$. No tax implications are involved. A more realistic portrait might have the working capital gradually build up at the start of the project, and gradually decline in the final two periods. Nevertheless, this reminds us to search for important cash flows that do not show up in the visible investment and more readily identified periodic amounts.

Imponderables

This gives a deeper picture of how the cash flows we so glibly analyzed in earlier sections were derived. One way or another, managerial judgment enters at every twist and turn. Even with our best insight and patience, there is no guarantee our judgments tally to an accurate picture. We have left many possibilities out of the exercise. Recall, for example, our earlier encounter with this story where subcontracting was a possibility, as was the production of alternate products. We also might worry about new competition, technology changes, price changes (e.g., selling prices, wages, energy, and materials), income tax law changes, employment tax changes, and so on.

Investment decisions are long-run in nature, and usually large decisions. In this case we have altered some of our LLAs. We are also worried about interactions with other decisions. Indeterminate, imponderable facets of the choice are part and parcel of a large decision.

12.4   Rendering in the Accounting Library

The final question in our look at long-run, investment decisions is how these decisions are recorded in the accounting library. Given a short-run, small decision it is highly likely the estimated incremental profit associated with that decision will show up in the firm's accounting library during the period

in question. Where and how it shows up may well differ from the decision frame used, but the profit consequences will be recorded in the library in some fashion.[13] (The major caveat here, recall, is when uncertainty and risk aversion are paramount, yet the accounting library of course does not record risk premia as such.)

Long-run decisions, be they small or large, are a vastly different matter. Here, the consequences unfold through time, and the time at which a present value analysis records events is usually at odds with accounting recognition rules. It is important to remember that present value analysis relies on cash flow estimates and cash flow per se is the antithesis of accrual reporting.

With this in mind, ponder the incremental effect of this expansion project on the firm's periodic income. Presumably cash from operations will be recorded as income in the period in which it is received. The major accrual issues center on the investment's cost, the alteration, and tax recognition. We assume the investment cost itself, 300, is depreciated over its 5 year life on a straight line basis (with no salvage value).

The up-front alteration expenditure of 90 might be expensed immediately, or capitalized and amortized. Following our tax treatment, we adopt the former. Tax recognition, of course, leads to the world of deferred taxes. We simply report incremental tax expense equal to 40% of incremental accounting income. This leads to the display in Table 12.4.

TABLE 12.4: Incremental Accounting Income (000)

	$t=0$	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
cash from operations	0	167	167	167	167	167	0
accruals							
depreciation		60	60	60	60	60	
alteration	90						
pretax income	-90	107	107	107	107	107	0
book tax at 40%	-36	43	43	43	43	43	0
income	-54	64	64	64	64	64	0

As students of accounting, we should understand the calculations in Table 12.4. Notice accrual (i.e., book) income contains no working capital effects. (This is why a statement of cash flows is important.)

---

[13]Now is a good time to remember that, under full costing, the incremental profit consequences will be found with the products themselves combined with whatever effect on the overhead plug transpires.

12.4.1 Accounting Rate of Return

With these calculations before us we are now in a position to introduce the accounting rate of return. The idea is simple. Just as we used the cash flow sequence to think in terms of an internal rate of return on capital invested (where it made sense), we can use the associated accounting income sequence to think in terms of an accounting rate of return on capital invested. Naturally, we take an incremental approach.

Two procedural questions arise. First, since the accounting income varies from period to period, what income number do we want to place in the numerator? The usual answer is the average incremental accounting income over the project's life. Second, what investment base did we have in mind? The usual answer is to take the initial amount that is capitalized. Of course, we might use an average investment amount, we might add in working capital changes, and so on.

Incremental accounting income totals 266 (000). To be consistent with Table 12.3, we average this over $t = 5$ years, or 53.2 per year. With an initial capitalization of 300, this implies an accounting rate of return of $53.2/300 \approx 17.73\%$. Again, we caution, working capital is not included here (though it could be), and we have treated the investment base in a cavalier fashion. Countless variations come to mind.

Calculated in this or any similar fashion, the accounting rate of return is simply a grand average of periodic income divided by investment base. It provides yet another portrayal of the investment opportunity. Equally clear is its departure from the timing considerations that are the central feature of present value analysis.

Though it appears particularly uninteresting, its mere existence should remind us of an important point. The accounting library will record any investment activity. It will record periodic income, asset, and liability manifestations of the investment activity. This means the informed, professional manager is ready and prepared to interpret the library based reports in terms of the investment activity and the accounting conventions at work.

12.4.2 Closing the Gap

More broadly, machinations over the accounting rate of return are symptomatic of the fact there can be a wide disparity between a project's estimated present value and its eventual rendering in the accounting library, even if uncertainty does not play a heavy hand. For example, in our running illustration the firm's manager identifies and adopts a project with a present value of approximately 30 (000), and the accounting library records an immediate incremental loss of 54 (000). Not good!

This leads to devices whose purpose is to close this gap (pun).

Budgets

Budgets are a starting point. Ponder the task of preparing a long term planning budget for this firm. Cash flow projections would include the effects of any planned investment activity. Income and balance sheet projections also would include the effects of any planned investment activity. And there is not a simple, direct relationship between the two projections. Rather, the income and balance sheet projections reflect both the anticipated investment activity and the accounting treatment of that activity.

   Importantly, then, the library rendering *relative to budget* provides an immediate and transparent device for linking long-run decisions and their consequences as measured by the accounting library.[14]

Economic Value Added

Another approach is illustrated by economic value added: integrate the accounting stock and flow calculations.[15] The idea is to reduce each period's accounting income for a capital charge, usually computed as some (appropriate) interest rate multiplied by the beginning of the period's appropriately measured asset base.

TABLE 12.5: Incremental Economic Stocks and Flows (000)

	$t=0$	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
cash flow (CF_t)	-504	124	138	123	114	264	7
continuation							
$\quad PV$ (PV_t)	534	474	393	317	241	6	0
economic income	30	64	57	47	38	29	1
capital charge	_	64	57	47	38	29	1
net	30	0	0	0	0	0	0

To see the idea in its pure form, in Table 12.5 we tell this project's life story using economic income, which of course is based on present value measurement of the project's value through time. This takes us back to Chapter 4 (e.g., Example 4.2). With slight rounding, at time $t=0$ we pay, net of taxes, 504 for a project with a present value of future cash flows of 534 (hence our familiar $PV \cong 30$ datum). So the initial asset value is 534. Next period the asset value is the present value of the then remaining cash

---

[14]This raises the question of why we have not formally introduced budgets. To be sure, budgeting is serious business and essential for any organization's long-run financial health. But once the obvious is acknowledged, there is not much of a conceptual nature to say except for the veracity of the forecasts on which the budget is based. This subject is broached in later chapters.

[15]$EVA^{\circledR}$ is a registered trademark of Stern Stewart & Co.

flows, all at $r = 12\%$ of course. This leads to the asset values (PV_t) and economic income tallies $(I_t = PV_t - PV_{t-1} + CF_t$, just as in Chapter 4).

Notice that once the economic rent of 30 is recorded at time $t = 0$, economic income thereafter is equal to the implied capital charge computed as 12% of the beginning of period present value (i.e., 12% of PV_{t-1}). Using recognition rules based on economic value (as in present value), and netting the implied capital charge places the accounting library and present value analysis on precisely the same footing. Rent of 30 is recognized at inception, and following that the project earns at the required rate of return. End of story.

The pragmatic version of this idea is economic value added. The capital charge is levied, so to speak, but a full-bore economic value approach is not used. (For example, this would require immediate recognition of a 30 incremental profit on project inception!). For the record, we add this nuance to the rendering in Table 12.4. In so doing we also capitalize the alteration cost (of 90). This is done to avoid an incremental loss at inception. Of course, this reverberates through the depreciation (and now amortization) calculations, as well as the tax expense calculations.

TABLE 12.6: Economic Value Added (000)

	$t = 0$	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$
book value						
working capital	150	150	150	150	150	0
alteration	90	72	54	36	18	0
acquisition	300	240	180	120	60	0
	540	462	384	306	228	0
cash from						
operations		167	167	167	167	167
depr. and amort.		78	78	78	78	78
tax at 40%		35	35	35	35	35
income	0	53	53	53	53	53
capital charge	0	47	37	28	19	9
net	0	6	16	25	34	44

Notice we have moved closer to the economic rendering in Table 12.5, but still cling to the bulk of the library's conventions. This is the world of economic value added, a pragmatic approach to moving the accounting library closer to economic fundamentals.

Of course much more could be explored here, but this has become tiresome. Conventional approaches to evaluating long-run decisions place a serious wedge between the decision criterion and frame and the accounting library. To no surprise, this leads to mutations of what one might find

in the library, and to a natural segue into our next topic of performance evaluation.[16]

12.5   Summary

Decision making interacts with the accounting library in two respects: the library contains information that is useful in the decision activity; and the consequences of whatever decision is rendered will be recorded in the library. Much of the managerial art that is brought to bear in long-run decisions concerns recognizing decision opportunities, and subsequently teasing out the large and small components of those opportunities. No sure-fire guidelines can be offered. Skilled management transcends guidelines.

We emphasized present value techniques for framing purposes. This carries considerable weight when markets are well functioning. Otherwise, we encounter ambiguity in delineating cash flows, their riskiness, and the appropriate discount rate. Other, sometimes competing and sometimes complementary, frames surface: internal rate of return, payback, and accounting rate of return are illustrative. Our list is far from exhaustive (e.g., simple urgency or ad hoc sensitivity analysis).

We also emphasized the distinction between small and large decisions for estimation purposes. The accounting library relies on LLAs; and we suspect many long-term investment projects are not that local in nature. Moreover, strategic considerations are likely to be present. Will a competitor follow or be scared off by a major capacity expansion or new product development? Will capacity expansion lead to unstable market prices and the possibility of price wars? Will work force learning from increased production give us a cost advantage? Do some projects also bring options to the table? For example, development of one product might place us in a position to develop another at a later date. The first product project then has two components, the product itself and the option of accessing the second product project. Astute analysis will recognize the value of this option in sorting out whether to pursue the first product. Likewise, the very time to invest is often an open question, another story with serious imbedded options.

Finally, administrative considerations are present. One side to this concerns the assignment of decision rights in the firms. Analysis and choice of minor projects are likely to be decentralized, while analysis and choice of major projects are likely to be at least partially centralized and subject to considerable review (and monitoring). Another side concerns the motiva-

---

[16]It also turns out a much more intimate connection between Tables 12.5 and 12.4 exists. Properly done, the present value at any time is algebraically equal to the accounting book value plus the present value of the future income less capital charge series. This has a natural interpretation in terms of error in book value leads to errors in the future income series, and vice versa. Think about it.

tion of various actors in the firm. Advocating a major project places the manager's reputation on the line. (In this sense, the project produces cash flow and information about those who did and did not advocate it in the first place.) This suggests control considerations are important.

Our study is partially complete.

12.6   Bibliographic Notes

Our treatment of long-run decisions is necessarily brief. Important questions of competitor reaction and investment timing under interest rate uncertainty, for example, have not been discussed. Nor have we broached the subject of financial structure. The starting point for further exploration is a finance text. From there, more sophisticated treatments are in order. Dixit and Pindyck [1994] is compelling and Haley and Schall [1979] remains a personal favorite. Scholes, Wolfson and Erickson [2004] focus on tax effects.

The link between accounting and economic measures surfaces, yet again, here. Horwitz [1984] is a good introduction to the subtlety of using accounting rates of return to make economic inferences, while Salamon [1988] highlights inherent ambiguities. Dillon and Owers [1997] provide an entry to the world of EVA. Sutton [1991] studies industrial structure, bringing together the strategic side of investment and the difficulty in using available data to make economic inference.

Tirole [2006], Arya and Glover [2001] and Arya, Fellingham and Glover [1998] imbed investment decisions in control laden environments and add a timing twist.

12.7   Problems and Exercises

1. What is the relationship between present value analysis of investment proposals and economic income?

2. The project analyzed throughout most of the chapter, Table 12.1, leads to a strictly positive present value. Yet the accounting library is slow to recognize this value, as in Table 12.4. Do the present value and accrual accounting renderings recognize the same cash flow? Why is the accounting treatment less aggressive in reporting the good news of a project with benefits significantly above its costs?

3. *economic versus accounting valuation*
   Below are some projected end of period cash flows. Each potential project requires an initial investment of 10,000. For each project determine the value of x such that the project has a present value

of precisely 0. Do this for interest rates of $r = 9\%, 10\%$ and 11%. Ignore taxes.

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
project 1	5,000	5,000	1,000	2,000	x
project 2	x	5,000	5,000	5,000	5,000
project 3	3,000	5,000	x	4,000	-3,000

(a) Having done this, what would be the economic income in year 2 of each of the projects?

(b) What would be the accounting income in year 2 if straight line depreciation is used.

4. *economic versus accounting valuation*
Ralph is contemplating a capital investment project. No taxes are present, the initial outlay will be 10,000, followed by end of year inflows for the next 3 years, as displayed below.

$t=1$	$t=2$	$t=3$
5,000	x	x

(a) Determine x such that the project has a present value of precisely zero if the interest rate is $r = 9\%$.

(b) Determine the end of period economic value of the project at the end of each period, as well as its economic income each period.

(c) What accounting income will be reported the first year if straight line depreciation is used?

5. *cash flow and accrual estimation*
The data presented in Tables 12.3 and 12.4 are rounded to the nearest thousand. Using the various assumptions noted in their development determine the exact amounts for both of these tables.

6. *more of the same*
Following up on problem 5 above, determine the exact amounts for the economic stock and flow calculations in Table 12.5.

7. *polynomial roots*
Return to the cash flow sequence in Table 12.1. Write out the equation for the present value, as a function of r. Multiply both sides by $(1+r)^6$, as we did in the text. Notice this gives you the equivalent, or future, value at the end of time $t = 6$. Naturally, if the future value is zero the present value is zero.

(a) Determine the future value, assuming $(1+r) = 1.14131679$. Interpret the result.

(b) Determine the future value, assuming $(1 + r) = -0.02681697$. Interpret the result.

8. *polynomial roots*[17]
Consider a two period investment project with cash flow vector denoted $[x_0, x_1, x_2]$. Determine a cash flow vector such that we have two strictly positive internal rates of return.

9. *cash flow estimation*
We now find Ralph managing a two product firm, Ralph's LP, with constrained capacity. The production process consists of fabrication and assembly departments. A service group that supplies maintenance, minor engineering, material handling, etc. to these two departments is also present. Let q_1 and q_2 denote the quantities of the two products. The fabrication department is constrained as follows: $2q_1 + q_2 \leq 300$. Think of this as expressed in hours of direct labor. The assembly department is constrained via $q_1 + 3q_2 \leq 600$. This, too, should be thought of in terms of direct labor hours. (The data are scaled for convenience.) Ralph recognizes nine cost pools. Their nature and *LLA*s, in slightly aggregate form, are detailed below. (Selling and administrative is the only period cost category.) Also, the respective selling prices are 600 and 800 per unit.

selling and administrative	$S\&A = 5{,}000 + 3q_1 + 5q_2$
direct labor in fabrication	$DL^f = 22(2q_1 + q_2)$
direct labor in assembly	$DL^a = 35(q_1 + 3q_2)$
direct material	$DM = 120q_1 + 200q_2$
overhead in fabrication	$OV^f = 5{,}000 + DL^f$
overhead in assembly	$OV^a = 6{,}000 + 3DL^a$
manufacturing service group	$MS = 2{,}000 + DL^f + .2DL^a$

(a) Determine an optimal production plan for Ralph

(b) Now consider expansion of the fabrication department, increasing its capacity from 300 to 450 units (scaled in direct labor hours). The original *LLA*s remain valid under such expansion. Determine the incremental cash flow from operations that would follow from such an expansion. (Assume variable costs are cash expenditures.)

(c) Ralph next concludes this increased activity will reshape the cost structure in the manufacturing service group. Automation will result in an *LLA* of $MS = 12{,}000$, with no variable component whatever. Assuming all of the service group costs (both with and without expansion) are cash expenditures, determine

---

[17] Amusement provided by Rick Young.

the incremental cash flow from operations associated with the proposed expansion.

(d) Ralph anticipates the expansion will be viable for 3 years. For modeling purposes of this sort, Ralph treats all cash flow as occurring at the end of the respective period. Ralph's marginal tax rate is 40%, and positive taxable income from other sources will be present if any of the periods result in a negative tax income. The expansion will cost 30,000 (an immediate outlay). For tax purposes, MACRS will be used (requiring depreciation of 33.33%, 44.45%, 14.81% and 7.41% over a 4-year horizon). No salvage value or costs are anticipated. In addition, minor plant modification will result in an expenditure of 5,000. This will occur at the start of the project (when the investment outlay is made), and will be expensed for tax purposes at the end of the first year. Determine whether this is an attractive expansion proposal. Assume a 9% cost of capital.

(e) Briefly speculate on how risk, learning, competition, and technology change might affect your analysis.

10. *present value versus accounting renderings*
This is a continuation of problem 9 above. Assume, for book purposes, that Ralph uses straight line depreciation. Also assume the 5,000 modification and training expenditure will be expensed in the first year. Prepare a series of proforma statements that detail incremental book income over the next 4 years if this expansion proposal is implemented.

11. *present value versus accounting renderings*
Verify the claim in note 16.

12. *consistent framing*
Return to the illustration in Figure 12.2, where the cash flow sequence is -100, 290 and -208. Assume $r = 10\%$ is the correct discount rate. Suppose we take the initial investment of 100 and invest it at $r = 10\%$. In two periods this will grow to $100(1.10)^2 = 121$. Alternatively, suppose we invest in this project. In one year we receive 290. Take this amount and invest it for one year. Hence, at $t = 2$ we have $290(1.10) = 319$. Of course, we also must pay out an additional 208 at this point. So we have 319 - 208 = 111. In this way, accepting the project is equivalent to investing 100 now and receiving 111 two periods later; rejecting the project is equivalent to investing 100 now and receiving 121 two periods later. How, then, can the project have internal rates of return of 30% and 60%? What principle of consistent framing is violated here?

13. *new product with investment and inventory*

    Ralph is now trying to decide whether to accept a customer's proposal to sign a long-term supplier contract. The customer will require 1,000 or 3,000 units of a specialized assembly in each of the next 3 years. The three periods are independent, and the probability of 1,000 units being required is .5 in each period. (Thus, the expected number of units is 2,000 each period.) The customer is willing to pay an up-front retainer of 90,000, plus 100 per unit ordered and delivered. The customer, though, determines the amount required (resulting in the noted probabilities)

    Ralph's cost analysis reveals direct material will cost 32 per unit and direct labor will cost 18 per unit. Overhead is costed at a full cost rate of 150% of material plus 50% of direct labor. Half of each rate is regarded as variable. Ralph's marginal tax rate is 42% in each period. Ralph is working below capacity and has under-absorbed overhead (our plug in Chapter 6) that is being expensed for tax purposes. This situation is expected to persist for at least 4 more years. (Notice that the retainer of 90,000 will be booked as revenue, both for financial and tax purposes, during the period of production. A reasonable assumption is 1/3 is booked at the end of each of the 3 production periods.) Ralph must also acquire a specialized machine in order to manufacture this assembly. The machine can be acquired for 125,000. It will have zero salvage value at the end of the contract, and will be depreciated for tax purposes on a 3-year MACRS basis (33.33%, 44.45%, 14.81%, 7.41%). In addition, Ralph would be forced to maintain an inventory of 500 units, which would be depleted in the third year. Thus, if the customer orders 1,000 units in the first year, 1,500 will be produced in the first year. If 3,000 units are ordered in the first year, 3,500 will be produced in the first year. Production in the third year will be actual demand less 500 units. (Assume the machine depreciation will be treated as a period cost for tax purposes.)

    For planning purposes, Ralph has decided to ignore estimated tax payments and intraperiod cash flow timing differences. Thus cash flow associated with production occurs at the end of the production year, tax payments occur at the end of the year in question, and so on. This is not accurate, but it is the way Ralph has decided to take an initial cut at the problem. Suppose Ralph is risk neutral and discounts after tax cash flow at a rate of 9%. Should Ralph accept the customer's proposal?

14. *make or buy*

    Ralph's Enterprise (RE) manufactures hydraulic components for the aircraft industry. One element common to a variety of products is a specialized valve that RE manufactures. The estimated cost for this valve reveals the following:

direct material	8.20
direct labor	9.30
variable overhead	3.10

Quality problems have surfaced and RE has decided the existing equipment must be replaced. An automated machine is available, at a cost of 2,500,000. It has a 5 year life, no salvage, and would be depreciated as 5 year equipment (MACRS percentages of 20%, 32%, 19.2%, 11.52%, 11.52% and 5.76%) for tax purposes. Straight line is used for book purposes. RE expects the direct material cost to remain the same, though direct labor will be cut in half if the new equipment is acquired. Also, variable overhead will remain at 1/3 of direct labor cost, although cash outlays presently in the "fixed" overhead (yes, the intercept of the LLA), totaling 450,000 per year, will not be incurred if the existing machine is retired. (It has zero salvage now, as well as a zero tax basis.) The marginal tax rate is a constant 40%. RE anticipates an annual demand of 50,000 valves.

Just before signing the purchase contract for the new equipment, another firm in the industry offers to supply RE all the valves needed, at a guaranteed price of 30 per valve over the next 5 years. The after tax discount rate is 9%.

(a) Which option is best, make the valve with the new equipment or buy the valve from the outside supplier? What qualitative concerns do you see here?

(b) What will happen to the first year's accounting income under each of the alternatives in (a)?

(c) What annual demand for the valve leads to indifference between the two alternatives?

13

Economic Foundations: Performance Evaluation

We now turn to the subject of performance evaluation. Don't miss the signal. Our concern now shifts from the metaphor "What will it cost?" to the metaphor "Did it cost too much?" As we shall see in the ensuing chapters, there is a profound difference between these two questions.

Initially we return to economic foundations. The key, you will see, is arranging for factors, such as managerial services, in a setting of less than perfect markets. Unlike basic commodities, such as gold or oil, managerial services come with a mind of their own, a fact that opens the door to performance evaluation.

We begin with some background remarks. From there we return to Chapter 2, and systematically weaken the market structure so performance evaluation arises in the form of carefully crafted pay-for-performance arrangements.

13.1 Performance Evaluation

In broadest terms, performance evaluation occurs when we make provision (at the time of choice) to evaluate (that choice) at a later date. It is a process of evaluation, of appraisal and assessment. It is also not happenstance. Being able to evaluate presumes we took care to lay in the requisite information in the first place. An important managerial task is planning for subsequent evaluation. For example, use of customer satisfaction measures

in the evaluation process presumes we have found the necessary data to make a reasoned assessment of customer satisfaction.

Performance evaluation is also not without purpose. (Double negatives have a distinguished history in academic writing.) Evaluation is done for a reason. We want to learn and adapt. We also want to appraise the performance of various actors in the firm. In the broadest terms, we envision this as evaluating a choice.

Thus, the professor announces the grading policy at the start of the course, administers various evaluation instruments throughout the course, and eventually assigns grades. The professor learns, and the students' performance is appraised. The facility manager begins the period, say, a quarter, with a budget and various service expectations. At the end of the period, spending is compared to budget; and various service statistics are calculated (e.g., average response time for equipment repairs or office space reconfiguration) and compared with expectations. The manager learns, and others use these data to evaluate the manager's performance.

The fast food manager begins the period with various performance goals, detailing, say, profit, employee training, and customer service expectations. At the end of the period a formal assessment is made, focusing on each stated goal. The legislator must stand for re-election.

In each vignette we retrospectively evaluate a bundle of choices. That is performance evaluation. Well said, but what is the central, organizing idea in this arena?

We address this question by focusing on the evaluation of a manager. Alternatively, we might evaluate a product, product line or manufacturing facility. These tasks were addressed in our earlier study of framing and large and small decisions. Our focus now shifts to the evaluation of a manager.[1]

Imagine a departmental manager in a large firm. Should we evaluate this manager based on cost incurred or profit earned? How should revenue be measured if the department's products are transferred to another department, one managed by a separate manager? Might we use the department's asset base in the evaluation? If so, do we prefer historical or fair value approaches?

Suppose the department consumes maintenance and personnel services produced by other departments. Should costs associated with these services

---

[1] There is an important qualitative difference between evaluating a manager and evaluating a product. We evaluate the product to determine whether it should be continued, modified, or dropped. We evaluate the manager as part of the web of controls used to help insure desirable behavior, as well as to determine whether the manager should be continued, dropped, or continued with modified instructions. The prospect of evaluation and its consequences help specify the environment in which the manager labors. Putting the two together, the firm must worry whether its managerial group is well-motivated to evaluate its product line.

be allocated to the department in question? If so, should they reflect actual or budgeted "prices?" Similarly, suppose a cost overrun occurs and is largely the result of unanticipated direct material price increases. Should this fact be recognized in the manager's evaluation?

What about nonfinancial information? Might we use measures of quality and productivity? Would a supervisor's subjective evaluation be relevant? Should the performance of a peer group be introduced, as in a sales contest or grading on the curve?

Naturally, the manager in question is supplying factors of production, in the form of managerial services. Substitutes abound, consultants come to mind, and complementarities are ever-present, as a well functioning management team increases productivity. The important, distinguishing features of managerial services, once we think of them as factors of production, are two in number. The manager and the employing firm may have conflicting tastes over how best to allocate these inputs; and the services actually supplied rest on unobserved choice behavior by the manager.

The manager may excessively worry over personal career concerns in deciding whether to push a new product proposal, just as the professor may worry more about a current research project than an upcoming class. In analyzing and advocating the new product proposal or in preparing for the class the supplier of services has a variety of options, faces a choice if you will. And we don't really observe that choice. We are thus led to worry about the inputs supplied by the manager or by the professor.

This worry creates a profound juxtaposition with the story in Chapter 2. The perfect market setting portrays the manager as supplying factor inputs and being paid a market determined price per unit. Desired and actual inputs always agree, and no price ambiguity is present. Clouding this picture implies desired and actual inputs need not agree. The price per unit calculation also breaks down, since we do not see the inputs supplied. In this way a market imperfection exposes the economist's world to an interest in managerial performance evaluation.[2]

The solution to this dilemma is to use available information to efficiently infer what the manager did, to efficiently infer the services actually provided. This, as you will see, is the world of pay-for-performance contracting: executive bonus arrangements, up-or-out performance arrangements, sales contests, stock options and, yes, grades.

---

[2]The firm uses inputs, including managerial inputs, to produce outputs. Arranging for the inputs is a trivial task if they are available in perfect markets. If the managerial market is imperfect to the extent we do not necessarily observe inputs supplied, contracting for these inputs must be based on an inference as to what inputs were supplied. This inference is the task of performance evaluation. We make provision at the time of choice, at the time of input supply, retrospectively to evaluate that choice, that input supply.

13.2   A Streamlined Production Setting

To begin the odyssey, return to the short-run setting in Chapter 2, especially expression (2.8) where we examined a single product firm that used two factors of production, labeled z_1 and z_2. A particular short-run cost was defined as follows:

$$C^{SR}(q; P) \equiv \min_{z_1 \geq 0, z_2 \geq 0} P_1 z_1 + P_2 z_2 \qquad (13.1)$$
$$\text{s.t. } q \leq f(z_1, z_2)$$
$$z_1 = \widehat{z}_1$$

The production function, $f(z_1, z_2)$, links output q to the two inputs, and here the first input is fixed at $z_1 = \widehat{z}_1$.

13.2.1   Managerial Service

We now add some additional simplifying assumptions. First, the technology limits the second factor to be but one of two possible amounts: (mnemonically) L ("low") or H ("high") with $L < H$. Second, output is uncertain.[3] It will be either x_1 or x_2, with $x_1 < x_2$. Uncertainty and factor input are linked by the following probability structure.

TABLE 13.1: Probabilities on Output

managerial service	output	
(or action)	x_1	x_2
input H	$1 - \alpha$	α
input L	1	0

The idea is input L guarantees the low output of x_1. Input H, though, will result in the larger output with probability α, and the smaller output with probability $1 - \alpha$.[4] It will also be convenient to interpret this input as "managerial service" or "managerial action."

Finally, the firm is assumed to be risk neutral. It seeks to maximize the expected value of its short-run profit.

To complete the analogy to the setting in expression (13.1), think of output as scaled so the selling price is unity. Output $x \in \{x_1, x_2\}$ then is output expressed in terms of revenue. From here, interpret quantity q as the expected value of the output. So quantity q can be either $q_H =$

---

[3] As you will see, uncertainty is unavoidable here. For simplicity, we treat the output as uncertain, though we might treat the required inputs or even their prices as uncertain.

[4] There is no inherent reason for input L to lead to output x_1 for certain. This is done merely to keep the story as simple as possible.

$(1 - \alpha)x_1 + \alpha x_2$, which requires input H, or $q_L = x_1$, which requires input L.

The firm's expected profit, then, will be $(1 - \alpha)x_1 + \alpha x_2 - C^{SR}(q_H; P)$ if it selects quantity q_H and $x_1 - C^{SR}(q_L; P)$ if it selects quantity q_L.

Specifying the cost function will occupy the remainder of the chapter. Notice, however, that with managerial service or action being the only unspecified input we are free to transform the firm's cost function to the incremental short-run cost of output q. And with such a transformation the cost reduces to the cost of acquiring input H or input L. Framing tricks continue to be a source of simplification!

To recap, we focus on a short-run setting in which all factors of production are fixed, except for managerial service or action. Output is also uncertain. We frame the firm's decision to emphasize revenue less incremental cost, which reduces to revenue less the cost of the remaining factor. We interpret this remaining factor as managerial service or action.

13.2.2   Preferences of the Supplier

The potential supplier of this service or action is an economic actor, with two distinguishing features.

Personal Cost

The first distinguishing feature is the service the manager is asked to supply is not a matter of indifference to the manager. Rhetorically, we assume the supplier incurs a personal cost in supplying managerial services to our firm. The underlying idea is not literally one of personal cost, but rather consumption at work. Bouts of enjoyment, collegial rapport, power, prestige, self-satisfaction, curiosity, drudgery, loss of leisure, pressure to perform, anxiety, and so on, are elements of the typical employment relationship. What we have in mind is something in the employment relationship that is important to the employee but not equally important to the employer. Personality and circumstance will give this more precise meaning. For now we simply acknowledge the general idea. Not all aspects of the employment relationship are valued the same by both parties.

With this in mind, we opt for the simple approach and let c_L denote the personal cost to the manager of supplying input L, and c_H the personal cost of supplying input H. We assume H is more costly, $c_H > c_L$.

In this simple fashion we create a wedge between the firm and the manager. Presuming the manager's services are sufficiently productive, the firm will prefer input H. But the manager is unrelenting in his preference for input L.

Constant Risk Aversion

The second distinguishing feature of the manager is risk aversion. Recall in Chapter 9 how we approached risk aversion by positing a utility function defined on wealth w, and denoted $U(w)$. It will turn out that risk aversion is essential in what follows. Granting this, there is no reason to introduce a changing attitude toward risk, so constant risk aversion is the story. This implies we work with the negative exponential formulation of $U(w) = -\exp(-\rho \cdot w)$, where $\rho > 0$ is the measure of risk aversion.

From here we interpret the noted personal cost in monetary terms. In particular, the manager's wealth will increase by $I - c_a$ if managerial service or input $a \in \{L, H\}$ is supplied and payment I is received from the firm.[5] To avoid clutter we also normalize the manager's initial wealth to zero. (Recall initial wealth has no effect on risk aversion for the constant risk aversion case.)

Throwing a little more notation at the story, we model the manager's preferences with the following utility specification

$$
\begin{aligned}
U(I, a) &= U(I - c_a) = -\exp(-\rho[I - c_a]) \qquad (13.2) \\
&= -\exp(\rho c_a)\exp(-\rho I) \\
&= \exp(\rho c_a) \cdot U(I)
\end{aligned}
$$

where, again, I denotes compensation received from the firm and a denotes managerial service or, generically, action supply. Importantly, the manager cares jointly about compensation received and about managerial action supplied.[6]

The next step is to highlight the transaction with the supplier of this input. In this way we extend our characterization of a firm to include the idea of arranging and managing transactions.[7]

Comparative advantage and transaction technology are now important elements of the larger picture. Performance evaluation, in turn, is a major ingredient in the firm's transaction technology. It is the information glue that supports the trade arrangements.

---

[5] Glancing back at expression (13.1), we are dealing with the second factor here. Interpreting it as managerial service or action, however, we slip in some notation change and denote the factor supply possibilities by $a \in \{L, H\}$.

[6] Another interpretation is the manager is a subcontractor, so the managerial services in question are being outsourced. The personal cost is then readily interpreted as the cost to the subcontractor of performing the desired service.

That said, another straightforward specification of supplier preferences is given by a root utility function combined with separable action cost: $U(I, a) = \sqrt{I} - c_a$. Though easy to manipulate, this has the disadvantage of changing risk aversion.

[7] Viewed in this more expansive manner, the firm and the market are competing institutions for arranging transactions. To illustrate, a firm may internally produce (a largely internal transaction) or externally acquire (a largely market mediated transaction) some subcomponent.

13.3   Transacting with a Perfect Labor Market

Arranging this transaction between the firm and its supplier of managerial action is perfunctory if the trade of service for compensation takes place in a perfect market. As this benchmark provides an important reference point for our excursion into the world of performance evaluation, we begin with such a (textbook) setting. This amounts to specifying the firm's cost function, $C^{SR}(q; P)$, on the assumption the managerial services are acquired in a perfect market, a setting where competition in the labor market ensures a mutually advantageous match of firm and supplier.

The key observation is the manager is not held captive by the firm. Suppose among all alternatives, except working for this firm, the manager's most attractive has a certainty equivalent of $M \geq 0$ (for market). This implies the manager's opportunity cost of working for our firm is $U(M) = -\exp(-\rho M)$.

It is now a simple exercise to specify the cost function. If the firm is to secure input H from this manager, it must offer a payment of I_H such that the manager finds the package attractive. In utility terms, this requires $U(I_H - c_H) \geq U(M)$. Expressing this in certainty equivalent terms, the requirement is

$$I_H - c_H \geq M$$

So the minimum payment to the manager is $I_H = c_H + M$. Similarly, the minimum payment to secure input L is $I_L = c_L + M$.[8]

This happy state of affairs is often called a first-best setting, first-best in the sense there are no contracting frictions.

Example 13.1 To illustrate, let the output uncertainty in Table 13.1 be specified by $x_1 = 10,000, x_2 = 20,000$ and $\alpha = .5$. Further specify the manager via a risk aversion measure of $\rho = .0001$, personal costs of $c_H = 5,000$ and $c_L = 2,000$ and a market opportunity of $M = 3,000$. This implies the cost to the firm of securing input H is $I_H = 5,000 + 3,000 = 8,000$, while the cost of input L is $I_L = 2,000 + 3,000 = 5,000$. The firm's choice is now apparent. Use of input L provides an expected profit to the firm of

$$1(10,000) + 0(20,000) - 5,000 = 5,000$$

while use of input H provides

$$.5(10,000) + .5(20,000) - 8,000 = 7,000$$

Several features of our abstract development should be noted. First, the firm is risk neutral while the manager is risk averse. The best risk sharing arrangement is for the firm to carry all the output risk, and pay

---

[8]In turn, imagine a large number of identical potential suppliers. Competition then ensures the firm's cost of input a will be $c_a + M$.

the manager a wage in exchange for supply of managerial services. This is why we developed the argument assuming the manager would be paid a flat wage. It is also the reason why the manager's risk aversion measure (of $\rho = .0001$) is superfluous in Example 13.1.

Second, the incremental cost of higher output to the firm under this flat wage arrangement is $I_H - I_L = c_H - c_L$. It does not depend on the manager's opportunity cost. This is one reason for using the constant risk aversion specification. It does not allow for an interaction between managerial opportunity cost and incremental cost to the firm. This keeps us focused on essentials.

Third, market discipline together with the manager's personal cost sets the wage. The market guarantees the manager a net of M; so the firm must pay $c_a + M$ for input $a \in \{L, H\}$.

13.4   Transacting in the Face of Market Frictions

We now introduce two frictions that stand in the way of the firm arranging for supply of managerial input. The first is self-interested behavior by the manager. The second is limited information, so the firm has concern for and difficulty in knowing whether the desired services have been supplied.

13.4.1   Self-Interested Behavior

To this point we have implicitly assumed the firm can arrange for supply of any feasible input from the manager. The cost to the firm is determined by the manager's personal cost and market opportunities. The labor market disciplines the trade arrangement.

Subtly tucked away here is the idea any arrangement meeting the market test will be honored. The firm will not renege in paying the manager; and the manager will not renege in supplying the agreed upon input. This is the idea of cooperative behavior. Agreements are honored or enforced with some unmodeled mechanism. The transaction, once agreed upon, will be implemented without a hitch.

We now add one-sided noncooperative behavior to the story. The firm is able to commit to any payment arrangement with the worker, if that arrangement is conditioned on publicly observable events. The firm can commit to pay the manager a flat wage, a bonus dependent on accounting income, a bonus dependent on market share, or whatever. The only catch is the payment can depend only on variables that are publicly observed. Once agreed upon, though, the firm does not renege; the contracted payment

arrangement is costlessly enforceable. The payment arrangement will be honored.[9]

The manager, on the other hand, has no such commitment power. The manager will renege if self-interest so dictates; and self-interest is defined by the manager's expected utility of wealth in our story. To see the power of this assumption, suppose we use the contractual arrangement of the perfect market setting, and seek supply of input H. There the firm offers the manager a flat wage of $I_H = c_H + M$ in exchange for supply of input H.

Having agreed to this arrangement, the manager now faces a decision: should input H or input L be supplied. If H is chosen, the manager will receive the flat wage of $I_H = c_H + M$ and incur a personal cost of c_H for a net (certainty equivalent) of M. Conversely, if input L is chosen, the manager will receive the same flat wage but incur a personal cost of c_L for a net of $I_H - c_L = c_H + M - c_L = (c_H - c_L) + M > M$ as $c_H > c_L$. Choice of input L is compelling. The manager is paid for the more costly input H, but surreptitiously supplies input L and incurs the strictly lower cost of c_L.[10]

Example 13.2 The setting of Example 13.1 is a case in point. With $c_H = 5,000$ and $c_L = 2,000$ and a market opportunity of $M = 3,000$, the perfect market arrangement for securing input H calls for a flat wage of $I_H = 5,000 + 3,000 = 8,000$. Supplying input H nets the manager a certainty equivalent of $8,000 - c_H = 3,000$ ($= M$ of course). But supplying input L provides a net of $8,000 - c_L = 6,000$. Choice of L is compelling.

If we assume the manager can commit to the original terms of the agreement, the manager has no choice to exercise at this point. Input H was agreed upon, and input H will therefore be supplied. If we assume the manager cannot so commit, a choice is on the table. Without the ability to commit, when it comes time to supply the input, the manager must choose between H and L. Opportunistic behavior is invited. When low output (x_1) is observed, the manager can claim H was supplied but bad luck resulted in low output.

The manager's choice is governed by self-interested behavior in this caricature. Input H will be supplied at this juncture only if it is in the manager's self-interest, as defined by the expected value of $U(I, a)$.

---

[9] With the contracted payment depending only upon public observables, a court is in position to confirm and enforce the contractual terms. We should not assume this is always the case. Litigation over employment contracts is not uncommon. We use the assumption of honorable behavior by the firm merely to present a streamlined story in which performance evaluation is substantively important.

[10] This is the previously acknowledged conflict of interest at work. Of course, if the firm wanted input L no conflict would arise. It would pay the manager $I_L = c_L + M$, which is acceptable to the manager and provides no temptation to supply input H. This is why we continue to emphasize the case where the firm seeks supply of input H.

This is not a flattering view of the manager. If we think broadly about the manager's concerns, issues of family, self-satisfaction, intrinsic interest, career development, and so on are all likely to influence what the manager does. A conflict between firm and personal goals seems inevitable. We model this conflict with the assumption of self-interested behavior in the face of personal cost.[11]

While less than flattering, conflict is far from uncommon. Auditing and internal control, for example, would not exist without conflict. Similarly, without conflict we would be hard pressed to explain such phenomena as sizable bonus payments, sales contests, supervision, and piece rates.[12] Indeed the phenomenon is so wide spread it has long been referred to as *moral hazard*.

Recognizing the potential for conflict in this most elementary fashion also, it turns out, reveals a key insight in the art of performance evaluation.

13.4.2   Public Observation of Input

Self-interested behavior, given $c_H > c_L$, implies the manager will supply L, against the firm's wishes. This argument, however, is based on the assumption the firm naively offers a contract paying $I_H = c_H + M$ in exchange for an unenforceable promise to supply H. The firm has other options.

Initially, suppose the manager's supply of input will be publicly observed. This means input supplied can be used in the contracting arrangement between the firm and the manager. Consider the following contract, where the manager's pay depends on the input $a \in \{L, H\}$ supplied.

$$I(a) = \begin{cases} c_H + M \text{ if } a = H \\ 0 \text{ if } a = L \end{cases}$$

---

[11] Technically, we structure the encounter between the firm and the manager as a non-cooperative game. The firm moves first, announcing contract terms of payment arrangement and instruction. This move is then observed before the manager moves, by selecting a feasible input to supply (or by refusing to work for the firm). A best response, or Nash, equilibrium is identified. Yes, the material in Chapter 10 is important.

[12] Without conflict, the manager's pay component that is at risk would be explained by risk sharing. This is an uninteresting explanation, especially in light of a well functioning capital market that exists to orchestrate risk sharing arrangements.

We reiterate that the idea is some returns to employment accrue to the employer while others accrue to the employee; and we posit a conflict stemming from these two return streams. Mark Twain was eloquent on the point of conflict in an employment relationship when he wrote that "...Work consists of whatever a body is obliged to do, and that Play consists of whatever a body is not obliged to do." (*The Adventures of Tom Sawyer*, Chapter 4). Though we tell the story with $c_H > c_L$ we should not interpret this as a model based on an assumption of managerial laziness or aversion to work. It is a model based on differently valued returns to employment, at the margin.

The manager will be paid the perfect market wage if input H is indeed supplied and 0 otherwise.

What might a self-interested manager do at this point? Choice of input H results in a net certainty equivalent of $I(H) - c_H = M$. But choice of input L now nets $I(L) - c_L = -c_L < M$. Choice of H is now compelling.[13]

Example 13.3 Continuing with the setting of Example 13.1, suppose the manager's input is publicly observable. Define the compensation contract by $I(H) = c_H + M = 8,000$ and $I(L) = 0$. Choice of input H nets the manager a certainty equivalent of $8,000 - c_H = 3,000$ while choice of L nets $0 - c_L = -2,000$.

The idea is simple. With input publicly observed, a penalty contract can be used. The manager is paid the same amount as in the perfect market case if the agreed-upon input is supplied; otherwise, a nonperformance penalty is incurred. Opportunistic behavior by the supplier disappears. Self-interest now leads to supply of input H.

With the manager's behavior publicly observed, a simple penalty contract renders a story that mirrors the earlier one in which the manager could commit to supply the promised input. In equilibrium, the manager supplies H and is paid $c_H + M$.

Don't miss the subtle comment, "in equilibrium the manager supplies H." The firm and the manager are here playing a noncooperative game. The firm moves first, by announcing an instruction (supply input H) and a compensation contract ($I(a)$ in this case). Having observed this move by the firm, the supplier accepts the offer and supplies input H. The compensation contract is designed so that it is equilibrium behavior for the manager to comply with the instruction.

Of course the compensation contract must be based on public observables. Otherwise it would not be operational, let alone enforceable.

13.4.3   Limited Public Information

Now suppose the only public observable is the output. The manager's input is not observable, so our penalty contract cannot be used. We cannot specify pay as a function of input, since input (a) is not publicly observed. But we can specify pay as a function of output (x). Abstractly, we envision the following payment schedule:

$$I(x) = \begin{cases} I_1 \text{ if } x = x_1 \\ I_2 \text{ if } x = x_2 \end{cases}$$

---

[13] This assumes a large enough penalty is feasible. If the manager's pay could not fall below 7,000 in the immediately following example, the simple penalty arrangement would not lead to supply of input H based on the perfect market wage.
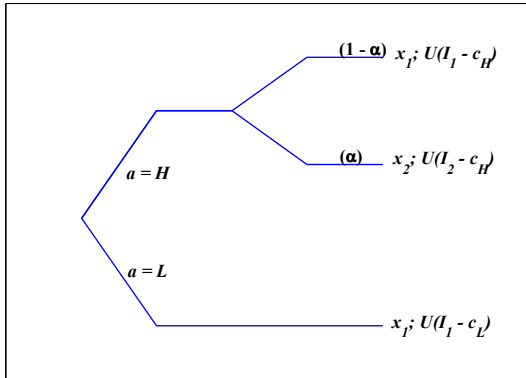
FIGURE 13.1. Manager's Induced Decision Tree at Time of Input Choice

Examine Figure 13.1, where we draw the manager's decision tree at the
point of deciding between input H and input L. For any such payment
function, $I(x)$, let $E[U|a, I]$ denote the manager's expected utility when he
supplies input $a \in \{L, H\}$ and labors under the noted incentive arrange-
ment, I for short. The expected utility calculations are

$$E[U|H, I] = (1 - \alpha)U(I_1 - c_H) + \alpha U(I_2 - c_H) = U(CE_H)$$

and

$$E[U|L, I] = U(I_1 - c_L) = U(CE_L)$$

You should notice in passing the use of a certainty equivalent frame, with
CE_a denoting the manager's certainty equivalent under choice $a \in \{L, H\}$.

If the self-interested manager is to supply H, we must have

$$E[U|H, I] = U(CE_H) \geq E[U|L, I] = U(CE_L) \tag{13.3}$$

This is called an *incentive compatibility* constraint. In designing the com-
pensation arrangement, the firm faces the constraint that the desired be-
havior, supply of H here, be incentive compatible. Goal congruence, the
manager preferring to supply H, is a constraint!

Example 13.4 Naturally, many payment arrangements are incentive com-
patible.[14] Returning to the setup in Example 13.3 and its predecessors,
several incentive compatible payment functions are displayed below.

---

[14]A cautionary note is in order. In more complicated problems it is not a given that
there is a feasible solution to the incentive compatibility problems. We will shun this
particular technicality as we proceed.

case	I_1	I_2	CE_H	CE_L	$E[I|H]$
1	2,000.00	18,000.00	2,092.46	0.00	10,000.00
2	4,000.00	14,000.00	2,798.85	2,000.00	9,000.00
3	2,000.00	25,266.39	3,000.00	0.00	13,633.20
4	5,000.00	12,305.66	3,000.00	3,000.00	8,652.83

To decode this, the first arrangement pays $I_1 = 2,000$ if low output, x_1, is observed and $I_2 = 18,000$ if high output, x_2, is observed. If H is indeed supplied, and recalling $\alpha = .5$ and $c_H = 5,000$, the manager's expected utility calculation is

$$
\begin{aligned}
E[U|H, I] &= -.5\exp(-\rho[2,000 - 5,000]) - .5\exp(-\rho[18,000 - 5,000]) \\
&= -.5\exp(.3) - .5\exp(-1.3) = -\exp(-.0001[2,092.46]) \\
&\approx -.8112
\end{aligned}
$$

The certainty equivalents are $CE_H = 2,092.46 > CE_L = I_1 - c_L = 2,000 - 2,000 = 0$. Further notice that under this arrangement, given the manager supplies H, the firm (being risk neutral) expects to pay the manager

$$E[I|H] = .5(2,000) + .5(18,000) = 10,000$$

The other arrangements follow in parallel fashion. Each is incentive compatible, but they vary in terms of their cost to the firm and attractiveness to the manager. Indeed, the first two would net the manager less than the certainty equivalent of his outside opportunity of $M = 3,000$; and the latter two meet this test but are rather different from the firm's perspective.

As Example 13.4 suggests, there is more to the story here. The payment arrangement $I(x)$, yes, the incentive scheme, must be incentive compatible. But it must also meet the market test of being sufficiently attractive to the manager. This means that the incentive compatible arrangement, the one that motivates supply of H, must also be attractive to the manager in terms of his best alternative, the earlier noted certainty equivalent of M. So the incentive scheme must also satisfy

$$E[U|H, I] = U(CE_H) \geq U(M) \tag{13.4}$$

This is called an *individual rationality* condition. It would, after all, be patently irrational of our presumably rational manager to agree to employment terms that failed the opportunity cost test.

Remember, though, the firm, not the manager, designs the payment arrangement here.[15] And designing this arrangement is akin to arranging

---

[15] Executive compensation is a setting where, arguably, the manager seems to be designing the payment arrangement!

the factors of production so as to minimize the total (expected) expenditure. So the firm, presuming it seeks input H, will search over all payment arrangements that simultaneously satisfy the individual rationality, expression (13.4), and the incentive compatibility, expression (13.3), conditions. Among these, the firm will select the one that minimizes the expected value of $I(x)$, presuming H is indeed supplied.

Paraphrasing the cost function expression in (13.1), then, the firm solves the following program

$$
\begin{aligned}
C(H) &\equiv \min_{I_1 I_2} E[I|H] = (1-\alpha)I_1 + \alpha I_2 & (13.5) \\
\text{s.t.} &\quad E[U|H, I] \geq U(M) \\
&\quad E[U|H, I] \geq E[U|L, I]
\end{aligned}
$$

We seek, that is, the minimum possible (expected) expenditure that will produce output q_H, which in our streamlined setting boils down to finding the minimal possible (expected) expenditure that will guarantee supply of input H. This is nothing other than the incremental short run cost of producing q_H in our streamlined setting. But carrying along the accurate though tedious expression $C^{SR}(q_H; P)$ is simply too much to ask. So we abuse notation even further and call the minimal objective function in (13.5) simply the firm's cost of input H, and denote it $C(H)$. It is your task to remember this is shorthand for our old friend, the firm's economic cost (appropriately framed).

Example 13.5 Return to the illustrative payment functions in Example 13.4. It turns out case 4, with $I_1 = 5,000$ and $I_2 = 12,305.66$ is the optimal incentive arrangement in that setting where the only contractible variable is the output, x. The value of this arrangement to the manager is $CE_H = 3,000$, which equals his opportunity cost of $M = 3,000$. The cost to the firm is $E[I|H] = 8,652.83$. Exclusive of the personal cost, the manager's certainty equivalent of this compensation arrangement is $8,000 = CE_H + c_H$. And 8,652.83 - 8,000 = 652.83 is the manager's risk premium.

This risk premium claim should, of course, be verified. Consider an individual with utility for wealth w given by $U(w) = -\exp(-.0001w)$. Our individual has no initial wealth and faces a lottery of 50−50 odds on 5,000 or 12,305.66. The expected value of this lottery is $.5(5,000) + .5(12,305.66) = 8,652.83$. And if you check, you will see that its certainty equivalent is 8,000, implying a risk premium of 652.83.

As an aside, intuition guides us to the solution to program (13.5). Suppose we have a solution in which $E[U|H, I]$ is strictly greater than $U(M)$. We could then lower each payment a small amount, lowering the firm's cost and not upsetting the other constraint. So anytime we have $E[U|H, I] > U(M)$, we can find a less costly scheme. Therefore, the best scheme must have $E[U|H, I] = U(M)$.

Similarly, suppose we have a scheme in which $E[U|H, I] > E[U|L, I]$. Now the incentive scheme is needlessly strong. Incentives, however, are not a free good. The manager's pay is at risk, and the manager must be compensated for carrying this risk. So, if the incentives are too strong, they can be weakened in a way that lowers the cost to the firm. Hence, the best scheme must also have $E[U|H, I] = E[U|L, I]$.[16]

We therefore have a constraint set of two equations in two unknowns:

$$(1 - \alpha)U(I_1 - c_H) + \alpha U(I_2 - c_H) = U(M)$$
$$(1 - \alpha)U(I_1 - c_H) + \alpha U(I_2 - c_H) = U(I_1 - c_L)$$

Notice this implies $U(I_1 - c_L) = U(M)$, or $I_1 = M + c_L$. (I_1, in fact, is the perfect market solution for securing input L.) Substituting this into the first equality we have

$$(1 - \alpha)U(M + c_L - c_H) + \alpha U(I_2 - c_H) = U(M)$$

And this can be readily solved for I_2.[17]

Regardless, several features of this exercise should be noted. First, we have $I_2 > I_1$. Notice in the above expression that $M + c_L - c_H < M$, as $c_H > c_L$. But this means $I_2 - c_H > M$, or $I_2 > M + c_H > M + c_L = I_1$. This is no accident. We already know $I_1 = I_2$ (a flat wage) won't work. What about $I_1 > I_2$? The manager would then face the prospect of switching to input L, incurring lower cost, and guaranteeing himself the larger prize. What a deal! Simply stated, incentive compatibility, expression (13.3), requires $I_2 > I_1$.

Second, with $I_2 > I_1$ the manager labors under an incentive arrangement. A bonus of $I_2 - I_1$ is paid if high output, x_2, is produced. Of course, this means the manager's wealth is at risk. This is contrary to efficient risk sharing, as the firm is risk neutral. In a sense, then, we trade off efficient risk sharing for incentive compatibility.

Third, with the manager bearing risk, part of his compensation takes the form of a risk premium. We saw this in Example 13.5. To see it more generally, write out the individual rationality condition, (13.4), in a little

---

[16] A more formal argument runs as follows. Delete the incentive compatibility constraint and solve for the best payment scheme. This is our earlier arrangement in which $I_1 = I_2 = 8,000$. We know it is not incentive compatible. The solution must have the constraint imposed and binding, i.e., $E[U|H, I] = E[U|L, I]$. Our intuitive explanation is aided by using the negative exponential utility function and having two possible outcomes and two possible inputs. This is sufficient for our purpose. More generally, solving the design program of minimizing the firm's expected payment subject to individual rationality and incentive compatibility constraints requires additional work. The nonlinear optimization routine in typical spreadsheet software becomes useful at this point.

[17] $U(I_2 - c_H) = \frac{U(M) - (1 - \alpha)U(M + c_L - c_H)}{\alpha} = -\exp(-\rho[I_2 - c_H])$

more detail

$$-(1 - \alpha) \exp(-\rho[I_1 - c_H]) - \alpha \exp(-\rho[I_2 - c_H]) = -\exp(-\rho M)$$

Now multiply by $\exp(-\rho[c_H])$

$$-(1 - \alpha) \exp(-\rho[I_1]) - \alpha \exp(-\rho[I_2]) = -\exp(-\rho[M + c_H])$$

Stare at this for awhile. We have a risky lottery of I_1 or I_2 that has a certainty equivalent of $M + c_H$. This means it has a nontrivial risk premium of its expected value less that certainty equivalent, or

$$RP = E[I|H] - M - c_H > 0 \tag{13.6}$$

This risk premium is a deadweight loss of contracting for managerial action in such a setting.[18]

Fourth, a popular euphemism is that the manager is now paid for results, or "only results count." This masks a subtle and important point. We want the manager to supply input H, but cannot directly observe whether input H is supplied. Output is observed, and we therefore use output to infer input. Casually, high output (i.e., x_2) is consistent with supply of input H, while low output (i.e., x_1) is more ambiguous. This is why the manager is paid more for high output. Output, then, is a source of value to the firm and a source of information in the contracting arrangement.

Fifth, the overall exercise is one of engineering the manager's decision tree, at minimum cost to the firm. Figure 13.1 was designed to convey this insight. At the time of contracting, the manager has three alternatives: reject the firm's offer, accept the firm's offer and supply L (be disobedient), or accept the firm's offer and supply H (be obedient). Individual rationality requires $E[U|H, I] \geq U(M)$, and incentive compatibility requires $E[U|H, I] \geq E[U|L, I]$. The constraints literally ensure the manager's fully formed decision tree rolls back to the conclusion that supply of input H is desirable behavior from the manager's perspective. Indeed, here we further assume that if indifferent the manager will honor the firm's instruction, supply H in this case.[19]

Finally, our story sharply distinguishes the cases of observable and unobservable input. In the former, the cost to the firm of input H is simply the perfect market solution of $I_H = M + c_H$. In the latter, where only output is observed, the cost is this amount plus the above identified risk

---

[18] Notice the firm trades compensation for action. The cost to the firm of the compensation package strictly exceeds its value to the manager. This is the very essence of incentive contracting, where we substitute high powered incentives, i.e., inefficient risk sharing, for lack of better information on which to base the trade.

[19] The alternative is to increase the incentive payment ever so slightly. This creates an annoying complication that offers no practical or intellectual insight. We thus assume when faced with indifference that the manager will follow the firm's instructions.

premium. Unobservable input raises the cost of managerial service. This occurs because output here is an imperfect indicator of input, and thus requires a risky payment to the manager; and we have grounded the model so the cost of the manager's risk bearing is borne by the firm. In this way we readily see that the firm would pay up to this risk premium, expression (13.6), to be able to observe the manager's input.

13.5   The Bad News

You might be growing impatient with this wedding of a simple story and tedious notation. So we pause for reassurance, and in the process will pull out additional insight.

13.5.1   Trivial Managerial Risk Aversion

What happens to the story if the manager is not risk averse? A convenient feature of the negative exponential utility function, recall, is that the parameter ρ measures risk aversion.

In Figure 13.2 we plot $C(H)$ as a function of ρ. The plot uses the data in Example 13.5 and its predecessors, and reflects the optimal incentive scheme when the only public observable is output, program (13.5). Notice how $C(H)$ decreases as ρ decreases and converges to 8,000, the perfect market solution, as ρ goes to zero. When the manager is risk neutral, the firm can just as well contract on output as input. No substantive contracting friction is present when the manager is risk neutral. In such a case the firm would not pay to observe the manager's input, given output is being observed.

The intuition is straightforward. Efficient risk sharing and proper incentives generally are at odds. With efficient risk sharing, the manager receives a flat wage. This creates a free rider problem, as the manager incurs the personal cost of input H but receives none of the benefit. Tilting the payment package allows the manager to share in the benefit of costly input H, but at the implicit cost of inefficient risk sharing.

When the manager is close to risk neutral, tilting the payment package carries a trivial inefficiency. In the limit, the inefficiency disappears.[20] Efficient risk sharing and proper incentives have become one and the same.[21]

---

[20] Conversely, increasing the manager's risk aversion increases $C(H)$. In a more thorough analysis, then, we should allow the choice of input to vary as we indulge in comparative statics; and in the limit we should allow the firm to shut down.

[21] When the manager is risk neutral we have an entire spectrum of equivalent solutions. They all have the property that the expected payment to the manager is 8,000.
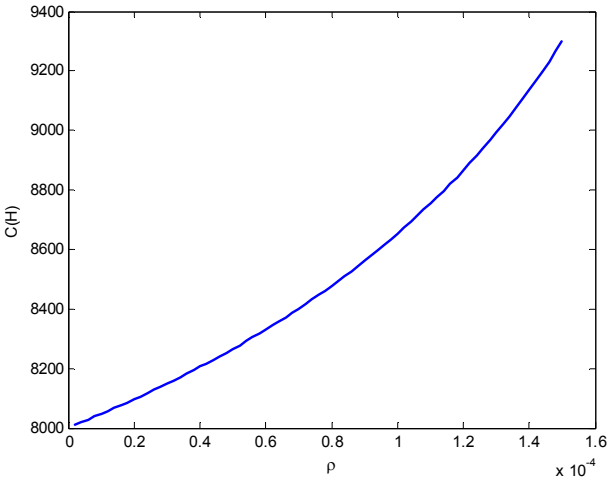
FIGURE 13.2. Cost of Input H as a Function of ρ

13.5.2   Trivial Odds of Low Output under Input H

Now consider what happens as we allow the probability of high output under input H, parameter α, to increase. Our running example uses $\alpha = .5$. In Figure 13.3 we plot $C(H)$ as a function of probability α. Notice how increasing α decreases the firm's cost of input H. Stated differently, the contracting friction is lessened as α increases; and it disappears altogether at the extreme of $\alpha = 1$.[22]

This illustrates the subtlety of the notion that we "pay-for-performance." The firm is arranging for the supply of input H but under difficult circumstances. It cannot see the input that is eventually supplied; and the supplier incurs an unobservable cost in supplying the desired input.

The only indicator of input supply is the output. So the firm uses output to infer input. Output is used as a source of information in the contracting arrangement with the supplier. Now, as α increases, the quality of this information increases. In particular, x_2 becomes more likely given supply of H. With better information, the control problem is more easily solved, the risk sharing inefficiency decreases, and $C(H)$ correspondingly decreases. In the limiting case of $\alpha = 1$, output becomes a perfect indicator of input. If $\alpha = 1$ and we see low output, we know without doubt the manager has not supplied H. This takes us back to the input observable case.

Output is used here to infer input for contracting purposes. Output is both a source of value to the firm and a source of information in dealing

---

[22] You may wonder why the graph begins at $\alpha = .4$. The constraints are infeasible if α is too low. Examine the extreme case of $\alpha = 0$! Now read note 14.
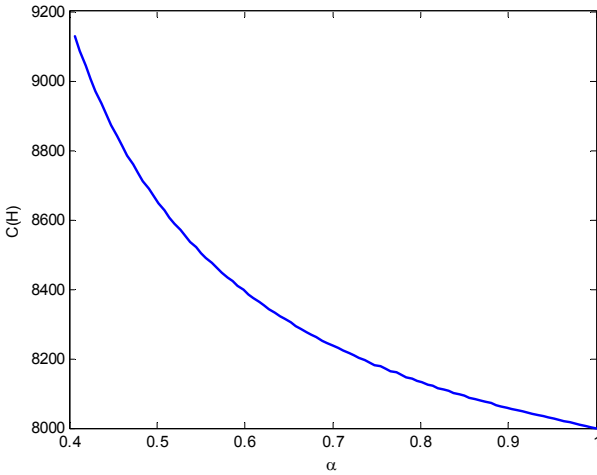
FIGURE 13.3. Cost of Input H as a Function of α

with the contracting frictions. The better output is as an indicator of input, the less costly the contracting friction.

13.5.3   Trivial Incremental Personal Cost

A final drill focuses on the incremental personal cost specification, $c_H - c_L$. The contracting friction is caused by the manager's incremental personal cost and the lack of information. What happens when we hold the information constant but vary the incremental personal cost? In Figure 13.4 we plot $C(H)$ as the manager's personal cost of low input, c_L, increases from 2,000 to 5,000. Holding c_H constant and increasing c_L in this fashion lowers the incremental personal cost of input H. Again using our running example and an optimal incentive function, we see that $C(H)$ declines as c_L increases; and in the limit (where $c_L = 5,000$) we have no substantive contracting friction as there $c_L = c_H$ and the incremental personal cost is nil.

Once the manager agrees to the firm's offer, a choice between H and L must be faced. The incremental personal cost to the manager of supplying input H is $c_H - c_L = 5,000 - c_L$ in this case. As we increase c_L toward 5,000, this incremental cost declines. In this sense, the magnitude of the control problem declines. As this happens, the inefficient risk sharing that is essential to motivate input H (using output to infer input, remember) declines. So $C(H)$ declines. In the limit, the incremental personal cost to the manager is zero, and no control problem is present.
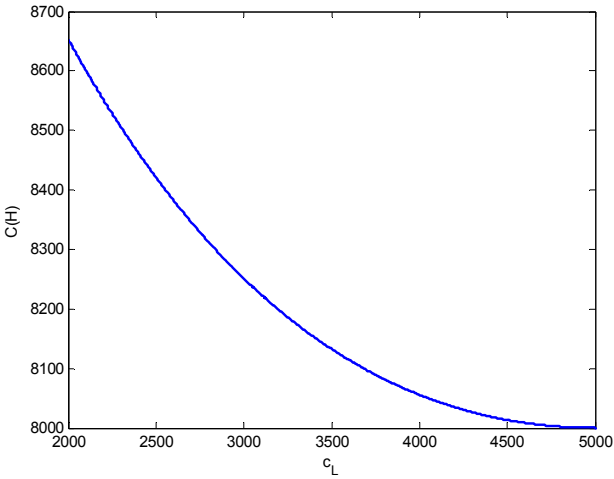
FIGURE 13.4. Cost of Input H as a Function of c_L

13.5.4   The Unavoidable Conclusion

These observations carry an important message for the study of performance evaluation. If the manager is risk neutral (ρ goes to zero), if the output is an unusually powerful source of information (α goes to one), or if the manager's incremental personal cost is trivial (c_L goes to c_H), the firm incurs no additional cost by not being able to observe the manager's input. In these extreme cases there is no demand for additional information to help resolve the control problem. There is no substantive control problem. There is no meaningful conflict of interest between the firm and the manager.

Our study requires a logically consistent story in which performance evaluation is a useful and nontrivial exercise. We therefore must avoid cases where the manager is risk neutral, where other sources of information are definitive in identifying the manager's input, and where the manager's personal cost is not an active friction. There is no reason to evaluate the manager in these cases.

Stated differently, if our stylization of contracting for managerial services is to admit an interest in nontrivial performance evaluation, we are forced to acknowledge several requirements. We must assume the manager is risk averse, $\rho > 0$; otherwise the manager is able to carry the risk of production and will fully internalize the potential conflict. We must assume uncertainty is present in the production process, $\alpha < 1$; otherwise output can be used to infer the manager's input without error and performance evaluation is a trivial exercise. We must assume some inherent conflict, some personal cost of $c_H > c_L$, is present; otherwise there is nothing to control or worry about.

Absent risk aversion, uncertainty, and personal cost, we base our study on a setting where there is no substantive interest in the art of performance evaluation.

The bad news is we must carry along considerable baggage if our story is to admit an economic reason for evaluating the manager's performance. The minimum baggage consists of risk aversion, uncertainty, and an inherent conflict of interest.[23]

Professional management is not an easy task, and neither is the study of professional management.

13.6   A More Expansive View

Common sense and everyday life remind us every organization, every firm, faces conflicts of interest. The response is a well thought out, continuously managed control system. Control systems, in turn, are not cost free. These costs come in explicit form, such as internal auditors and monitors more broadly, and in implicit form, such as organizational arrangements that lend themselves to monitoring.

These costs are often referred to as agency costs, or more narrowly contracting costs. Our streamlined (to be kind) model developed in this chapter exhibits such a cost. If the manager's action is not observed, and the only contracting variable available is output, the firm arranges for equilibrium supply of input H by contracting with a pay-for-performance or incentive scheme. This amounts to using output to infer input, to using output as a measure of performance. Importantly, now, as output is a noisy measure of input, the manager's compensation is uncertain. This results in the risk premium identified in expression (13.6).

This risk premium is the simple model's explication of agency costs. Recall that in the perfect market solution the firm would pay the manager a flat wage of $I_H = c_H + M$ for supply of input H. But, rearranging (13.6), we see in the noted pay-for-performance arrangement that the firm's cost is an expected payment to the manager of

$$E[I|H] = c_H + M + RP = I_H + RP \qquad (13.7)$$

In this manner our little model, as we said, explicates the agency cost theme.

Recalling that our firm is risk neutral, the fact the manager incurs a risk premium signals that we now substitute inefficient risk sharing for lack of

---

[23]It is also possible to create a contracting friction when all parties are risk neutral and uncertainty and inherent conflict are present. The trick is to also assume the manager's pay cannot fall below some lower bound, a type of limited liability requirement. The tack taken here is more compelling, and institutionally vibrant.

performance information. While descriptive, we should be selective in our labeling. The risk sharing is inefficient relative to the case of no contracting friction. But it is the best possible arrangement given the contracting frictions. A more precise description, then, would be the efficient second best arrangement distributes the risk between the firm and the manager in a way that is not efficient in the first best setting. First best refers to the case of no contracting frictions. Second best refers to the case of contracting frictions.[24]

That said, our little model also explicates a powerful, omnipresent theme in the control arena: separation of action choice from control choice. The firm designs the control system by announcing the pay-for-performance arrangement and instruction to supply H. The manager then makes the actual input choice. Separation of duties is a long-standing accounting phenomenon, one that is central to protecting the accounting library's integrity. And, it turns out, this separation phenomenon is central to virtually any control system, including the U.S. Constitution.[25]

Controls are always costly and always present. The firm's web of controls deals with decision rights, control motivated information production, outsourcing (e.g., consultants), elaborate internal labor markets, and physical architecture itself.

It will turn out our little model has a great deal to teach us about this web of controls, but that is getting ahead of the story.

13.7   Summary

This chapter focuses on managerial performance evaluation. The central theme is a firm seeking to acquire managerial inputs in a less than perfect market setting. When input is not observed, directly or indirectly, and when there is a natural conflict between the supplier and the firm, we have an interest in evaluating the performance of the manager. The purpose of the evaluation is to form a basis for inferring the input supplied by the manager.

The solution to this exercise is a pay-for-performance arrangement. Better performance is rewarded. This common euphemism, though, clouds the

---

[24]Another point here concerns the choice of input itself. For simplicity we continue to assume the firm seeks input H, i.e., that the preferred input is the same in the first and second best regimes. In general this will not be the case. Burdening the manager with risky pay increases the cost of managerial services. In addition, different production plans lead to output being more or less informative about the manager's input. Putting the two forces together, it would be unusual to have the choice of input the same in the first and second best settings.

[25]Thanks to Jerry Zimmerman for this important insight.

underlying idea that performance is an indicator of input. Literally, we use performance to infer input.

The model we have sketched will provide considerable insight in subsequent chapters; but we should also acknowledge its heavily streamlined, nearly simplistic nature. We have not addressed such confounding features as taxes, reputation, nonpecuniary rewards, culture and long-term relationships. Taxes, for example, may influence payment arrangements. Some forms of compensation are tax advantaged. Health benefits, work place ambiance, and retirement savings are ready examples. The manager's reputation also may be an important factor in the employment relationship. Defined entry portals may be used so the firm can calibrate the manager's talent and trustworthiness. In this way, some jobs are designed to provide the firm with important information about the nature of its work force. Long-term arrangements, in turn, usually take the form of implicit arrangements. The firm may have a policy of filling managerial vacancies by promotion. It may assign a management team to a particularly troubled division with the understanding that a future assignment will be in a more affable environment.

The list goes on. Our purpose is not to cover the entire spectrum of human resource management. Rather, we want to proceed at this point with the basic insight that we use output, broadly interpreted, to infer input.

Though sufficient for our purpose, we should keep in mind that a broader view of contracting frictions recognizes various types of frictions. So-called moral hazard problems arise when there is the possibility of post-contract opportunism. This was emphasized in our exploration, where the manager but not the firm knew the input supplied. Adverse selection refers to pre-contract opportunism. Buying a used car from its current owner is an example, where the seller but not the buyer knows whether the car is a lemon. And lack of commitment ability can further hinder contracting arrangements. Our manager could not commit to deliver an agreed upon input supply. The firm may not be able to commit to explicit long-term arrangements or to use fairly any information it privately acquires in the evaluation process.

13.8   Bibliographic Notes

The principal-agent model, developed here in highly stylized fashion, has become an important model of trade and resource allocation. Sappington [1991] and Abowd and Kaplan [1999], read together, provide an excellent introduction and survey. Kreps [1990], Chapter 16, provides a comprehensive introduction to the technical details, while more expansive treatments are available in Bolton and Dewatripont [2005] and Laffont and Martimort

[2001]. Our use of the model is patterned after that in Grossman and Hart [1983]. Hart and Holmstrom [1987] provide an overview of the larger picture and Arrow [1974] is particularly eloquent on trade frictions. Bonner and Sprinkle [2002] and Prendegast [1999] offer empirical perspectives, and Bebchuk and Fried [2004] a contrarian perspective.

13.9   Problems and Exercises

1. The central idea in this chapter is that some productive inputs are not acquired in perfect markets, are not necessarily delivered in the quality and quantity intended. In turn, this creates an interest in controls of some sort, controls designed to address frictions inherent in the trade of labor for compensation. The stylized model of personally costly input highlights the use of output in the control apparatus. How, in this model, is output used to facilitate the purchase of input? Why is the supplier paid for performance?

2. Goal congruence is said to exist when members of the management team (or more broadly the work force) share the same goals; and in this perspective goal congruence is seen as an essential objective of organization design. The stylized model presented in the chapter offers a subtly different perspective. Goal congruence is a constraint, manifest in the incentive compatibility restriction requiring the self-interested manager find it personally desirable (or incentive compatible) to behave in the organization's best interest. Carefully discuss this notion of goal congruence as a constraint.

3. The contracting story developed here results in a cost to the firm of input H that we denoted $C(H)$. Without any contracting frictions, the manager would be paid the sum of reservation price plus personal cost, or $M + c_H$. It also turns out the quantity $C(H) - (M + c_H)$ is equal to the manager's risk premium for the compensation risk presuming input H is supplied. Carefully explain why this linkage between incremental cost to the firm and risk premium to the manager arises in the contracting model.

4. *certainty equivalents*
   Verify the certainty equivalent calculations summarized in Example 13.4. Notice, in this case, the certainty equivalents can be calculated in two ways. One method calculates the certainty equivalent of the I_x lottery and then subtracts the c_a cost. The other method focuses directly on the certainty equivalent of the lottery of net gains, $I_x - c_a$. Why are the two methods equivalent here?

5. *manager's opportunity cost*
   Return to Example 13.5, and focus on the noted optimal pay-for-performance arrangement. Recall we assume $M = 3,000$.

   (a) Locate an optimal contract for the following cases: (i) $M = 2,500$; (ii) $M = 2,000$; (iii) $M = 1,000$; and (iv) $M = 0$.

   (b) Carefully explain the emerging pattern.

6. *insurance and incentives*
   The contracting model presented here is called a "hidden action" or "moral hazard" model. The latter term comes from the insurance phenomenon where an insured subject has reduced care incentives. For example, is it likely that the owner-operator of an automobile drives more diligently and less frequently when the auto's insurance has lapsed? Is the implied delicate balancing of risk sharing, or insurance, and proper incentives present in the labor input model? Explain, using the data in Example 13.5.

7. *optimal contract*
   Return to the setting of Example 13.5. Now assume the probability of output x_1 under input L is .9 instead of 1.0. Determine the optimal pay-for-performance arrangement. Carefully explain the difference between this arrangement and that identified in Example 13.5.

8. *optimal contract*
   Ralph owns a production function. Randomness in the environment plus labor input from a manager combine to produce output. The output can be one of two quantities: $x_1 < x_2$. The manager's input can be one of two quantities, $L < H$. Ralph is risk neutral. The probabilities are given below, and you should assume the higher output is sufficiently attractive that Ralph wants supply of input H in all that follows.

	x_1	x_2
input H	.1	.9
input L	.8	.2

   Ralph's manager is risk averse and also incurs an unobservable personal cost in supplying the labor input. We model this in the usual way. The manager's utility for wealth is as given in (13.2), with $c_H = 5,000$, $c_L = 0$, and $\rho = .0001$. Also, the manager's opportunity cost of working for Ralph is a certainty equivalent of $M = 10,000$.

   (a) Suppose the manager is trustworthy and will honor any agreement (or, equivalently, serious penalties are feasible and the manager's input can be observed.) What is the cost to Ralph of acquiring input H?

(b) Suppose the only observable for contracting purposes is the manager's output. Determine the optimal pay-for-performance arrangement. What is the cost to Ralph of acquiring input H? Draw the manager's decision tree and verify the manager can do no better than accept Ralph's terms and then supply input H. What is the manager's certainty equivalent for the payment lottery that is faced?

(c) Why, in your solution to part (b) above, is the manager paid more when the largest feasible output (i.e., x_2) is observed?

9. *shape of optimal incentives*
This is a continuation of problem 8 above. Now assume there are three possible outputs, $x_1 < x_2 < x_3$. The probability structure is listed below, and input H is again desired.

	x_1	x_2	x_3
input H	.1	.8	.1
input L	.7	.2	.1

Determine an optimal pay-for-performance arrangement. Why has Ralph's cost gone up, compared with the setting in the original problem? Also, why is the manager now paid more for intermediate than for the most desirable (x_3) output?

10. *smoothing behavior*
Return to the problem 9 above. Casually, we might interpret the story as one in which the manager receives a bonus when x_2 is produced, but no additional reward if even more is produced. Might the manager now be tempted to inventory or otherwise "hide" output in the short-run once enough output has been produced to qualify for the bonus? What (convenient) assumption in the simple model removes this possibility?

11. *optimal production plan*
Return to Example 13.5. Find specific values for outputs x_1 and x_2 such that in the absence of contracting frictions the firm will contract for input H, but when the noted frictions are in place it will opt for input L. Explain.

12. *risk neutrality*
Return yet again to the setting of Example 13.5, but now assume the manager is risk neutral. Find two distinct pay-for-performance arrangements that will ensure supply of input H and at a cost to the firm of $C(H) = M + c_H = 8,000$. Explain the intuition behind your two solutions.

13. *optimal production plan*
Ralph, who is risk neutral, owns a production process. Production re-

quires input from a manager. This input can be one of three possible quantities: $L < B < H$. Output will be one of two possible quantities: $x_1 < x_2$. The manager is risk averse and incurs a personal cost, precisely as specified in (13.2), with $c_L = 0$, $c_B = 4,000$, $c_H = 10,000$ and $\rho = .0001$. The manager's outside opportunity guarantees a wealth of $M = 40,000$. The output probabilities are as follows.

	x_1	x_2
input H	0	1
input B	.1	.9
input L	.9	.1

(a) Suppose the parties can contract on the output and the input supplied. Determine the best contract from Ralph's perspective that will insure supply of input (i) H, (ii) B, and (iii) L.

(b) Suppose the parties can contract on the output, but not the input. Determine the best contract from Ralph's perspective that will insure supply of input (i) H, (ii) B, and (iii) L.

(c) Let $x_1 = 0$ and $x_2 = 55,000$. Determine Ralph's optimal plan under the contracting conditions in (a) and under the contracting conditions in (b) above.

(d) Let $x_1 = 0$ and $x_2 = 59,000$. Determine Ralph's optimal plan under the contracting conditions in (a) and under the contracting conditions in (b) above. Carefully explain your conclusions.

(e) Let $x_1 = 41,000$ and $x_2 = 46,100$. Determine Ralph's optimal plan under the contracting conditions in (a) and under the contracting conditions in (b) above. Carefully explain your conclusions.

14. *taxes and incentives*

Consider a setting where the manager's input can be L or H and the output can be $x_1 = 10,000$ or $x_2 = 50,000$. The manager's preferences are described in the usual fashion, i.e., (13.2). Let $\rho = .0001$ along with $c_H = c_L = 5,000$ and also set the manager's opportunity cost of working for this firm at $M = 0$. The owner is risk neutral. The output probabilities are noted below and input H is desired throughout the exercise.

	x_1	x_2
input H	.1	.9
input L	.8	.2

(a) Determine and interpret an optimal contract.

(b) Suppose the owner is subject to a 20% income tax (i.e., a tax equal to 20% of the net of $x - I_x$), while the manager faces a zero marginal tax rate. Determine and interpret an optimal contract.

(c) Repeat (b) above for the case where the owner is subject to a 20% income tax on income in excess of 20,000.

(d) Repeat (c) above for the case $c_L = 4,000$.

(e) Repeat (c) above for the case $c_L = 0$.

15. *square root utility*

Ralph owns a production function that uses labor input to produce output. Output will be either $x_1 = 10,000$ or $x_2 = 20,000$. Labor is supplied by an agent. One of three possible supplies will be used: $H > B > L$. The output probabilities are displayed below:

	x_1	x_2
input H	0	1
input B	.5	.5
input L	1	0

Ralph is risk neutral. The agent has a utility function for payment I and labor supply a of $U(I, a) = \sqrt{I} - V(a)$, with $V(H) = 60, V(B) = 30$ and $V(L) = 5$. In addition, the agent's next best offer carries a pay and labor supply package that provides an expected utility of 40. So whatever Ralph dreams up, the expected value of $U(I, a)$ must be at least 40 if the agent is to be attracted.

(a) Ralph must decide which $a \in \{L, B, H\}$ to acquire, and how to compensate the agent for this supply of labor. If the agent can be trusted to supply whatever is agreed upon, it is straightforward to figure out $C(a) = [40 + V(a)]^2$. Determine $C(a)$ for each $a \in \{L, B, H\}$. What is Ralph's best choice?

(b) Now assume output is the only contracting variable. Determine the best compensation package to motivate supply of (i) $a = L$, (ii) $a = B$, and (iii) $a = H$. In turn, what is Ralph's best choice and why does this differ from the best choice in the case where the agent can be trusted?

16. *personal cost*

At this point some reflection is in order. What role does personal cost play in the contracting model developed in this chapter? Does the theory require it be everywhere positive? Is the model based on the idea managers find work repulsive?

14

Economic Foundations: Informative Performance Evaluation

The next issue in our study of performance evaluation is the question of what measures are indeed useful in evaluating a particular individual. This individual, of course, is the manager in our heavily stylized contracting model introduced in Chapter 13. Is the manager best evaluated by focusing on cost incurred relative to output produced, or accounting income or income relative to net assets employed? Should competitor comparisons or customer satisfaction surveys be used? What about his supervisor's personal opinion? The firm's stock price might be, and often is, used, in the guise of options or holding the firm's stock in his personal portfolio.

This seemingly endless list of possibilities condenses to a simple question: is a given measure informative about the manager's action? To develop this important theme we begin by slightly expanding our contracting model, and then transforming the expanded model to focus, laser like, on the information content question. From there we develop the informativeness criterion.

14.1 Slightly Expanded Setting

To fully harvest our growing insight, it will be useful to look at a slightly more general case than that introduced in Table 13.1. We continue with the binary action choice of $a \in \{L, H\}$, with input or action H preferred by the firm, but action L preferred by the manager. This conflict, recall, is modeled by assuming the manager incurs a personal cost of action, denoted

c_a with $c_H > c_L$. In addition, while the firm itself is risk neutral the manager is strictly risk averse and exhibits constant risk aversion. His utility for compensation I and action a, as developed in expression (13.2) in the prior chapter, is given by

$$U(I, a) = -\exp(-\rho[I - c_a]) = \exp(\rho c_a) \cdot U(I) \qquad (14.1)$$

where, again, $\rho > 0$ is the manager's measure of risk aversion.

Into this hopefully familiar stew we now expand the dimensionality of the output. While originally we assumed output could take on but one of two values, we now assume there is some finite though otherwise arbitrary number of possible outputs, say $\{x_1, x_2, ..., x_n\}$.[1] From here let $\pi(x|a)$ denote the probability output x obtains when input a is supplied.

Next assume the only contracting variable is output x, and let I_x denote the manager's compensation when output x is observed. Exploiting the structure in the negative exponential utility function noted above in expression (14.1), the manager's expected utility calculation when he supplies action a and is rewarded according to contract I_x is simply

$$E[U|a, I] = \exp(\rho c_a) \sum_x U(I_x) \pi(x|a)$$

Recalling the firm seeks supply of input H, the optimal payment function is the solution to the following program, which is a simple extension of the program in (13.5)

$$C(H) \quad \equiv \quad \min_{I_x} E[I|H] = \sum_x I_x \pi(x|H) \qquad (14.2)$$

$$\text{s.t.} \quad \exp(\rho c_H) \sum_x U(I_x) \pi(x|H) \geq U(M)$$

$$\exp(\rho c_H) \sum_x U(I_x) \pi(x|H) \geq \exp(\rho c_L) \sum_x U(I_x) \pi(x|L)$$

Again, to refresh our memory, the first constraint is the individual rationality constraint. It requires that accepting the employment terms and supplying action H be at least as attractive as the manager's best outside opportunity, which offers a certainty equivalent of M. The second constraint is the incentive compatibility constraint. It requires that having accepted the terms of employment the manager will indeed find it in his self-interest to actually supply input H.

Example 14.1 To illustrate, specify the manager via a risk aversion measure of $\rho = .0001$, personal costs of $c_H = 5,000$ and $c_L = 2,000$ and a

---

[1] For that matter we might, and will at selected points (a bad pun), assume output can take on any value in some interval. The finite story is less imposing, and will teach us all the essentials.

market opportunity of $M = 3,000$. Output can take on one of four possible values, and the probabilities are as follows:

	x_1	x_2	x_3	x_4
$\pi(x|H)$	.35	.15	.40	.10
$\pi(x|L)$	.20	.80	.00	.00

We find an optimal payment arrangement, which you should verify, of

$$I_x^* = \begin{cases} 8,590.23 \text{ if } x = x_1 \\ 4,272.98 \text{ if } x = x_2 \\ 9,002.35 \text{ if } x = x_3 \text{ or } x_4 \end{cases}$$

This provides the firm a cost of $C(H) = 8,148.70$. And from here the manager's risk premium, which you should also verify, is $RP = 148.70$.

14.2   A Convenient Transformation

As Example 14.1 implies, the optimal trade of compensation for managerial action depends on the manager's preferences, the productivity of his actions, and the information available. It turns out, however, that we are able to transform the contracting problem to an equivalent though notationally less annoying frame. This will be an advantage as the plot thickens, so we pause to put it in place.

Suppose, relative to our original setting portrayed in program (14.2), that we examine a parallel setting in which the manager's outside certainty equivalent and personal costs are normalized via $\widetilde{M} = 0, \widetilde{c}_L = 0$, and $\widetilde{c}_H = c_H - c_L$. Notice we set the normalized certainty equivalent and low action cost to zero, and the high action cost to the difference between the two original personal costs. This implies $U(\widetilde{M}) = U(0) = -\exp(-\rho \cdot 0) = -1$ and $\exp(\rho\widetilde{c}_L) = \exp(0) = 1$. Rather convenient. Let's also stop writing $\exp(\rho\widetilde{c}_H)$, and simply label it $\overline{c} = \exp(\rho\widetilde{c}_H)$. So the design program can now be written in the following economically equivalent but more user friendly format:

$$C(H) \equiv \min_{I_x} E[I|H] = \sum_x I_x \pi(x|H) \qquad (14.3)$$

$$\text{s.t.} \quad \overline{c} \sum_x U(I_x)\pi(x|H) \geq -1$$

$$\overline{c} \sum_x U(I_x)\pi(x|H) \geq \sum_x U(I_x)\pi(x|L)$$

From here, now let $\overline{I}_x^*$ denote the optimal pay-for-performance plan in (14.2) and I_x^* its counterpart in the transformed program (14.3). It turns

out the two solutions are intimately linked (thanks to constant risk aversion):[2]

$$I_x^* = \overline{I}_x^* - M - c_L$$

$$E[I_x^*|H] = E[\overline{I}_x^*|H] - M - c_L$$

$$CE_{I^*} = CE_{\overline{I}^*} - M - c_L$$

$$RP_{I^*} = RP_{\overline{I}^*}$$

That is, the transformed setting's optimal contract is simply a rescaling of the original contract, and the manager's risk premium is unaffected by the transformation.

Example 14.2 To illustrate, return to the setting of Example 14.1, where $M + c_L = 5,000$, but normalize the manager's parameters to $\widetilde{c}_H = 3,000 = c_H - c_L$, $\widetilde{c}_L = 0$ and a market opportunity of $\widetilde{M} = 0$. Everything else remains as in the original example. We find an optimal payment arrangement of

$$I_x^* = \overline{I}_x^* - 5,000 = \begin{cases} 3,590.23 = 8,590.23 - 5,000 \text{ if } x = x_1 \\ -727.02 = 4,272.98 - 5,000 \text{ if } x = x_2 \\ 4,002.35 = 9,002.35 - 5,000 \text{ if } x = x_3 \text{ or } x_4 \end{cases}$$

This provides a cost to the firm of $C(H) = 3,148.70 = 8,148.70 - 5,000$. The payments are scaled by the net normalization, and the risk premium is unaffected: $RP_{I^*} = RP_{\overline{I}^*} = 148.70$. The firm's cost is, again, the manager's perfect market or first best wage of $M + c_H = 0 + 3,000$ plus his risk premium of 148.70.

Importantly, transforming the story in this fashion does not alter the story's economic structure. The transformed pay-for-performance arrangement is simply a rescaling of the original arrangement, and the manager's risk premium remains unaltered. For this reason we proceed, now, with the transformed problem and its reduced assault on our notational senses.[3]

---

[2] To see this, notice that a contract given by $I_x = \overline{I}_x^* - M - c_L$ is feasible in (14.3). Were this not the case, as you can readily verify, $\overline{I}_x^*$ would not satisfy the constraints in (14.2). But if it is feasible it also provides an upper bound for the optimal objective function value in (14.3): $E[I_x^*|H] \leq E[\overline{I}_x^*|H] - M - c_L$. Likewise, $I_x^* + M + c_L$ is feasible in (14.2), which implies $E[I_x^*|H] \geq E[\overline{I}_x^*|H] - M - c_L$. And if you work through the algebra to verify these claims you will see how constant risk aversion, the negative exponential utility function, is heavily relied upon.

[3] With the sanitized notation, the feasibility issue previously hinted at in note 14 in the prior chapter is now readily identified. Incentive compatibility requires

$$\overline{c}\sum_x U(I_x)\pi(x|H) - \sum_x U(I_x)\pi(x|L)$$

$$= \sum_x U(I_x)[\overline{c}\pi(x|H) - \pi(x|L)]$$

$$= -\sum_x \exp(-\rho I_x)[\overline{c}\pi(x|H) - \pi(x|L)] \geq 0$$

That said, reflect on the fact the manager's expected payment is algebraically equal to what that payment would be were the labor market perfect plus the risk premium. (Recall expression (13.7)!) This risk premium is a compensating wage differential, a differential relative to the perfect market wage that compensates the manager for being saddled with undesirable compensation risk. It is our proxy for agency cost. The firm, in turn, would pay up to this risk premium for contracting arrangements that would eliminate the necessity of the manager being saddled with compensation risk.

Of course, you are by now wondering why we don't throw some extra information at the problem, in hopes of reducing this risk premium.

14.3   Informativeness

To look into the question of additional information, we now assume that instead of contracting simply on output x we have a second publicly observed variable that can also be contracted on. This additional variable is observed at the end of the contracting period, along with the output itself. This new information might be some accounting measure of cost, profit or return on investment, a nonfinancial measure, or whatever.[4] It is a form of performance evaluation. We denote it by y. So in principle, the contract can now depend on both x and y. Let I_{xy} denote the manager's compensation when output x and performance measure y are observed.

Next, let $\pi(x, y|a)$ denote the probability that output x and performance measure y are observed given the manager has supplied action or input $a \in \{L, H\}$. The manager's expected utility is simply

$$E[U|a, I] = \sum_{x,y} U(I_{xy} - c_a)\pi(x, y|a) = \exp(\rho c_a) \sum_{x,y} U(I_{xy})\pi(x, y|a) \quad (14.4)$$

From here we just mimic the program in (14.3) to determine the optimal contracting arrangement to acquire input H: minimize the firm's expenditure subject to individual rationality and incentive compatibility constraints.

Now, the probabilities must somewhere differ (otherwise choice between H and L is moot) and at some point we must also have $\overline{c}\pi(x|H) > \pi(x|L)$, otherwise $\overline{c} < 1$ while by definition we must have $\overline{c} > 1$. But then we must also have $[\overline{c}\pi(x|H) - \pi(x|L)]$ switch sign at least once as we move across x. Otherwise the inequality cannot be satisfied. This sign switching requirement is the noted feasibility issue.

[4] If variable y is observed before the manager makes his action choice we must then worry about that choice for each and every possible realization of y. Indeed, such early arriving information, whether public or strictly private to the manager, may improve or worsen the contracting environment. But arriving late the worst possible case is it is useless and therefore ignored.

$$\widehat{C}(H) \equiv \min_{I_{xy}} \sum_{x,y} I_{xy} \pi(x, y|H) \tag{14.5}$$

$$\text{s.t.} \quad \overline{c} \sum_{x,y} U(I_{xy}) \pi(x, y|H) \geq -1$$

$$\overline{c} \sum_{x,y} U(I_{xy}) \pi(x, y|H) \geq \sum_{x,y} U(I_{xy}) \pi(x, y|L)$$

Glance back at the case where we contract only on output, $C(H)$ in program (14.3). One possibility is simply to ignore the new information. That is, one possible solution to program (14.5) is the optimal solution to (14.3), implying I_{xy} would be independent of variable y. So we already know $\widehat{C}(H) \leq C(H)$, with equality only if the new information is useless.[5] This means our new variable, y, is useful if and only if it lowers the contracting friction, if and only if $\widehat{C}(H)$ of program (14.5) is less than $C(H)$ of program (14.3).

Also notice that if the new information is useful, if $\widehat{C}(H) < C(H)$, then the new, improved arrangement must lead to a lower risk premium for the manager. That is, the new information is useful only if it provides a less noisy assessment of the manager's input, a less noisy assessment of his performance.

Example 14.3: To illustrate, assume, as usual, that the manager's input can be H or L, with H preferred by the firm. Output can be either x_1 or x_2 (with $x_1 < x_2$). The manager's preferences are specified by $\rho = .0001$, personal costs of $c_H = 3,000$ and $c_L = 0$ and a market opportunity of $M = 0$. (This is a normalized version of Example 13.5.) The information source will report either $y = g$ or $y = b$. The probabilities are specified in Table 14.1.

Initially suppose the additional information is not available, so output is the only contractible variable. Using the noted probabilities we have $\pi(x_2|H) = \pi(x_2, g|H) + \pi(x_2, b|H) = .40 + .10 = .50$, $\pi(x_1|L) = \pi(x_1, g|L) + \pi(x_1, b|L) = .20 + .80 = 1$. etc. We readily find the optimal pay-for-performance arrangement, the solution to (14.3), is $I_{x_1}^* = 0$ and $I_{x_2}^* = 7,305.66$. The firm's cost is $C(H) = 3,652.83$ and the manager's risk premium is 652.83.

In contrast, when the additional information is available, solving program (14.5) provides the new, improved pay-for-performance arrangement displayed in Table 14.1. Notice $I_{x_2 g}^* = I_{x_2 b}^*$. This reflects the fact output $x = x_2$ can obtain only if input H is supplied, so having observed output x_2 the information cannot possibly tell us anything more about the manager's

---

[5] To verify this, you just insert (14.3)'s solution into the constraints in (14.5), and use the fact $\pi(x|a) = \sum_y \pi(x, y|a)$.

behavior. Conversely, $x = x_1$ is consistent with input H or with input L. Now there is room for the new information to tell us something, and it does as we wind up with $I^*_{x_1 g} > I^*_{x_1 b}$. That is, having observed output x_1, the manager's compensation depends on whether $y = g$ or $y = b$ is observed. Also notice the odds of the manager receiving the high payment, given x_1 obtains, are much higher if he indeed supplied input H as opposed to input L.

TABLE 14.1: Details for Example 14.3

	x_1/g	x_1/b	x_2/g	x_2/b
$\pi(x, y\|H)$	.35	.15	.40	.10
$\pi(x, y\|L)$	.20	.80	0	0
I^*_{xy}	3,590.23	-727.02	4,002.35	4,002.35
I^*_x	0	0	7,305.66	7,305.66
$LR_{xy} = \frac{\pi(x,y\|L)}{\pi(x,y\|H)}$	$\frac{20}{35}$	$\frac{80}{15}$	0	0

$$C(H) = 3,652.83 \ (RP = 652.83)$$
$$\widehat{C}(H) = 3,148.70 \ (RP = 148.70)$$

The new information source is indeed useful here. Otherwise the optimal solution to program (14.5) would not make use of the new signal. The firm's cost (and concomitantly the manager's risk premium) are lower when the information is present. The various details are displayed in Table 14.1. (And you should verify our claimed solution.) Precisely why the information is useful, and how it is best used are linked to the other detail displayed in the Table, the likelihood ratio of $LR_{xy} = \frac{\pi(x,y|L)}{\pi(x,y|H)}$. This will be explored in due course. Be patient.

Before proceeding, suppose we relabel the four x/y combinations in Table 14.1 as, respectively, x_1, x_2, x_3 and x_4. Now return to Example 14.2. We have precisely the same pay-for-performance arrangement! Remember, output is a source of information about the manager's behavior. Appending another information source amounts to contracting on a potentially improved information platform. It is as if the output has become more informative about the manager's behavior.

It is also important to ponder the variations on Example 14.3 presented in Tables 14.2 and 14.3. Both retain the same specification of the manager. The story in Table 14.2 has the same output probabilities, i.e., $\pi(x|H)$ but a different $y = g/b$ information source. It turns out the information is utterly useless. The story in Table 14.3 has a different probability structure, one that regardless of the additional information is not overly informative and thus results in a much larger risk premium for the manager. The unusual feature is that if output x_1 obtains, it is $50 - 50$ odds whether the new information source reports $y = g$ or $y = b$; and the same holds if output x_2

obtains. Yet despite the appearance of being pure noise, the information is useful.

TABLE 14.2: First Variation on Example 14.3

	x_1/g	x_1/b	x_2/g	x_2/b
$\pi(x, y\|H)$	.10	.40	.40	.10
$\pi(x, y\|L)$	.20	.80	0	0
I_{xy}^*	0	0	7,305.66	7,305.66
I_x^*	0	0	7,305.66	7,305.66
$LR_{xy} = \frac{\pi(x,y\|L)}{\pi(x,y\|H)}$	$\frac{20}{10}$	$\frac{80}{40}$	0	0

$$C(H) = 3{,}652.83 \ (RP = 652.83)$$
$$\widehat{C}(H) = 3{,}652.83 \ (RP = 652.83)$$

TABLE 14.3: Second Variation on Example 14.3

	x_1/g	x_1/b	x_2/g	x_2/b
$\pi(x, y\|H)$	.20	.20	.30	.30
$\pi(x, y\|L)$	.20	.40	.10	.30
I_{xy}^*	5,641.26	-5,249.19	9,303.58	5,641.17
I_x^*	-4,176.33	-4,176.33	15,030.32	15,030.32
$LR_{xy} = \frac{\pi(x,y\|L)}{\pi(x,y\|H)}$	$\frac{20}{20}$	$\frac{40}{20}$	$\frac{10}{30}$	$\frac{30}{30}$

$$C(H) = 7{,}347.66 \ (RP = 4{,}347.66)$$
$$\widehat{C}(H) = 4{,}561.84 \ (RP = 1{,}561.84)$$

14.3.1 How the Model Uses the Information

The key to understanding the trio of stories in Tables 14.1, 14.2 and 14.3 is to ask the model how it makes use of the additional information.[6] This returns us to our old friend, the shadow price. Surprise!

Asking the model how it makes use of the additional information amounts to asking for the characterization of the solution to program (14.5). We are trying to minimize the expected value of the payment to the manager, subject to two constraints (the individual rationality and incentive compatibility constraints). Structurally, this is another variation on a problem we have encountered before, and characterized in Chapter 2's Appendix.

---

[6] Just as we began our study by laying out a model of factor choice and asking that model to teach us all it could about cost, we now ask the contracting model to teach us all it can about performance evaluation.

Following that lead, let $\lambda \geq 0$ be the shadow price on the first constraint (the individual rationality constraint) and $\mu \geq 0$ be the shadow price on the second constraint (the incentive compatibility constraint). The Lagrangian, then, is

$$\Psi = \sum_{x,y} I_{xy}\pi(x,y|H) - \lambda[\overline{c}\sum_{x}U(I_{xy})\pi(x,y|H) + 1)]$$
$$-\mu[\overline{c}\sum_{x,y}U(I_{xy})\pi(x,y|H) - \sum_{x,y}U(I_{xy})\pi(x,y|L)]$$

At the optimal solution, of course, the derivatives must vanish so we require I^*_{xy} be such that

$$\frac{\partial\Psi}{\partial I_{xy}}\Big|_{I_{xy}=I^*_{xy}} = \pi(x,y|H) - \lambda[\overline{c}U'(I^*_{xy})\pi(x,y|H)]$$
$$-\mu[\overline{c}U'(I^*_{xy})\pi(x,y|H) - U'(I^*_{xy})\pi(x,y|L)] = 0$$

Dividing by $U'(I^*_{xy})\pi(x,y|H)$ and simplifying slightly our model provides the following answer to how it uses the new information

$$\frac{1}{U'(I^*_{xy})} = \lambda\overline{c} + \mu\overline{c} - \mu\frac{\pi(x,y|L)}{\pi(x,y|H)} \qquad (14.6)$$

Now recall that the shadow price on the incentive compatibility constraint is strictly positive, i.e., $\mu > 0$.[7] But this means the right hand side of (14.6) varies with $\frac{\pi(x,y|L)}{\pi(x,y|H)}$. Therefore the left hand side of (14.6), which is 1 over the manager's marginal utility at the point where compensation in the amount I^*_{xy} is delivered, varies with $\frac{\pi(x,y|L)}{\pi(x,y|H)}$.[8] But this means the optimal pay-for-performance arrangement itself varies with $\frac{\pi(x,y|L)}{\pi(x,y|H)}$. In short, the model uses the information by judiciously varying the manager's compensation, I^*_{xy}, as a function of $\frac{\pi(x,y|L)}{\pi(x,y|H)}$.

As briefly mentioned in passing, this probability ratio is called a likelihood ratio, and is denoted LR_{xy}. It is central to what follows:

$$LR_{xy} \equiv \frac{\pi(x,y|L)}{\pi(x,y|H)} \qquad (14.7)$$

Returning, finally, to the examples in Tables 14.1-14.3, you will see that the optimal incentive contract, I^*_{xy}, varies with LR_{xy}. Indeed, the variation is

---

[7] Were this not the case, the incentive compatibility constraint would not be binding, and the manager would be paid a constant wage regardless of x or y. Oops! And if you think about it, you should also be able to sort out that $\lambda > 0$ as well. But it is $\mu > 0$ that provides the essential insight.

[8] With $U(I) = -\exp(-\rho I), U'(I) = \rho\exp(-\rho I) > 0$.

inverse, with a lower ratio associated with higher compensation. Intuitively, a lower (higher) ratio means the (x, y) combination is relatively less (more) likely if the manager misbehaves, supplies L, than if he behaves, supplies H.[9] Good news, so to speak; and good news is rewarded while bad news is, well, not rewarded.

Moreover, and hopefully to no surprise, precisely the same pattern holds in the case where output, x, is the only contractible variable. In that case, with x the only observable, the controlling probabilities are the $\pi(x|a)$ specifications, and the likelihood ratio is

$$LR_x \equiv \frac{\pi(x|L)}{\pi(x|H)} \tag{14.8}$$

In Table 14.4 we revisit the earlier examples to exhibit the connection between the optimal contract when output is the only contractible variable, I_x^*, and LR_x. Notice that low output (x_1) is bad news. It is more likely if the manager misbehaves, and is not rewarded in the optimal arrangement. We have, as claimed, the same qualitative connection between the optimal contract and underlying likelihood ratio.

TABLE 14.4: Output Only Contracts and LR_x

	x_1	x_2
Tables 14.1 and 14.2 Setting		
$\pi(x|H)$	.5	.5
$\pi(x|L)$	1	0
$LR_x = \frac{\pi(x|L)}{\pi(x|H)}$	2	0
I_x^*	0	7,305.66
Table 14.3 Setting		
$\pi(x|H)$	.40	.60
$\pi(x|L)$	.60	.40
$LR_x = \frac{\pi(x|L)}{\pi(x|H)}$	$\frac{60}{40}$	$\frac{40}{60}$
I_x^*	-4,176.33	15,030.32

With or without the additional information, the model uses the available information in a manner that parallels testing the hypothesis that the manager supplied input H. If the realization of the observables is largely consistent with this hypothesis, a larger payment is delivered. Otherwise, a smaller payment is in order. Information content of the observables is the key.

---

[9]You can sort this out in (14.6). With $\mu > 0$, a higher ratio means the right hand side is lower. This implies 1 over marginal utility is lower, or marginal utility itself is higher; but with risk aversion higher marginal utility implies lower compensation.

Also notice that when additional information is useful, the contracting problem with the manager is resolved with a less risky incentive arrangement. With or without the additional evaluation information, the manager carries the same opportunity cost of employment. But when the additional information is useful, the risk premium necessary to maintain the arrangement's attractiveness declines. This illustrates the insurance side of performance evaluation. Introducing the additional information allows the firm to maintain incentives with less risk placed on the manager. The monitor provides a basis for insuring, to a limited degree, the manager against the noisy relationship between output and input.

14.3.2   The Informativeness Criterion

The net result here is the additional information is useful only if the (x, y) likelihood ratio, LR_{xy} is "more variable" than its output only cousin, LR_x. But to put some insight into this glib observation, we must factor the probabilities. In general, the joint probability of two events, call them α and β, can be written in factored form via

$$\pi(\alpha, \beta) = \pi(\alpha)\pi(\beta|\alpha)$$

That is, the joint probability of events α and β can be expressed as the probability of α multiplied by the probability of β *conditional on* α. So in our setting we write the joint probability of any (x, y) pair given action a as the product of the probability of output x given action a and the probability of signal y *conditional on* output x and given action a

$$\pi(x, y|a) = \pi(x|a)\pi(y|x, a)$$

Now take this factoring and examine the likelihood ratio

$$LR_{xy} = \frac{\pi(x, y|L)}{\pi(x, y|H)} = \frac{\pi(x|L)\pi(y|x, L)}{\pi(x|H)\pi(y|x, H)} = LR_x \frac{\pi(y|x, L)}{\pi(y|x, H)} \qquad (14.9)$$

Bingo! If the new information is useful, the new, improved contract will use the information. This means the optimal solution to program (14.5) will be such that I_{xy}^* is a nontrivial function of signal y. But this is equivalent to saying the (joint) likelihood ratio, LR_{xy} is a nontrivial function of signal y. (This is the point to our earlier sketch of the shadow prices and their role in determining I_{xy}^*.) And saying LR_{xy} is a nontrivial function of signal y is equivalent to saying $\frac{\pi(y|x,L)}{\pi(y|x,H)}$ is a nontrivial function of signal y.

This insight is sufficiently important that we get a bit more formal. First, let's acknowledge $\frac{\pi(y|x,L)}{\pi(y|x,H)}$ is a *conditional* likelihood ratio, conditional on having observed output x, and denote it

$$LR_{y|x} \equiv \frac{\pi(y|x, L)}{\pi(y|x, H)} \qquad (14.10)$$

Second, we say that performance measure y is informative in the presence of output x if the conditional likelihood ratio $LR_{y|x}$ is a nontrivial function of y.

Definition 22 *Performance measure y is informative in the presence of output x if the conditional likelihood ratio $LR_{y|x}$ varies with y for at least one realization of output x.*

Think about this. If the new information is indeed useful, it must have the potential to tell us something important, something useful. We are already observing output x. So whatever it might tell us, we must first control for, condition on, the fact we are already observing x. It must tell us something we are not already learning from x. This means it must be informative, more precisely informative in the presence of output x.

So we wind up with the informativeness criterion: if the additional information is useful then it must be informative in the presence of output x. The conditional likelihood ratio must be a nontrivial function of signal y. It must be telling us something that is important and that we are not already learning from another source.

The model's answer to when an additional performance measure is useful is now in view. If it is useful, it must be informative (in the presence of the original information source).

Note well, the claim is that if the additional information is useful, it must be informative. Can we also say that if it is informative it is useful? In our highly simplified setting this is correct, as it is in a variety of settings. But it is not guaranteed, simply because the additional performance measure might tell us something that is not a concern in the control problem at hand. Running this down takes us into subtle and, if you can stand an awful pun, uninformative territory. So we content ourselves with the statement that if it is useful it must be informative.[10]

With this (finally) in place, we can identify the forces at work in Tables 14.1 through 14.3. The key is informativeness, the conditional likelihood ratio. We calculate it for each of the three cases in Table 14.5. For example, in the Table 14.1 setting, suppose output x_1 obtains. What then is the probability that, say, $y = g$ given output x_1 has been observed and action H has been supplied ? We have

$$\pi(g|x_1, H) = \frac{\pi(g, x_1|H)}{\pi(g, x_1|H) + \pi(b, x_1|H)} = \frac{\pi(g, x_1|H)}{\pi(x_1|H)} = \frac{.35}{.35 + .15} = .70$$

---

[10] To guarantee informativeness implies usefulness we must ensure the control problem is substantive and the new measure speaks to precisely that problem. And, as noted, when contracting on output x alone does not lead to the perfect market solution in our H vs. L setting, informativeness does imply usefulness. Holmstrom [1979] is the classic reference. We explore this issue further in the end of chapter problems and exercises, and in subsequent chapters.

With these conditional probability calculations in place, we then readily display the conditional likelihood ratios in Table 14.5.

In the Table 14.1 setting, notice that when output x_1 obtains, the conditional likelihood ratio is "low" under g and "high" under b. It varies with signal y for $x = x_1$. Conversely, for $x = x_2$ there is no such variation. Overall we know the additional information is useful here, and that means $L_{y|x}$ varies with signal y for at least one output realization.

TABLE 14.5: Conditional Likelihood Ratios

	x_1/g	x_1/b	x_2/g	x_2/b
Table 14.1 Setting				
$\pi(y\|x,H)$	.70	.30	.80	.20
$\pi(y\|x,L)$	.20	.80	0	0
$LR_{y\|x} = \frac{\pi(y\|x,L)}{\pi(y\|x,H)}$	$\frac{20}{70}$	$\frac{80}{30}$	0	0
Table 14.2 Setting				
$\pi(y\|x,H)$	.20	.80	.80	.20
$\pi(y\|x,L)$	.20	.80	0	0
$LR_{y\|x} = \frac{\pi(y\|x,L)}{\pi(y\|x,H)}$	1	1	0	0
Table 14.3 Setting				
$\pi(y\|x,H)$	.50	.50	.50	.50
$\pi(y\|x,L)$	.33	.67	.25	.75
$LR_{y\|x} = \frac{\pi(y\|x,L)}{\pi(y\|x,H)}$	$\frac{33}{50}$	$\frac{67}{50}$	$\frac{25}{75}$	$\frac{75}{50}$

Now turn to the setting of Table 14.2, where we know the additional information is useless. If $x = x_1$, the conditional likelihood ratio does not vary with y ($L_{g|x_1} = L_{b|x_1} = 1$), just as it does not vary with y when $x = x_2$ ($L_{g|x_2} = L_{b|x_2} = 0$). More important here is to notice the imperative of conditioning on the already observed x. To the contrary, suppose we focus on the additional information to the exclusion of output. This suggests an unconditional likelihood ratio of

$$LR_y = \frac{\pi(y|L)}{\pi(y|H)} \tag{14.11}$$

From here we readily see $LR_g = 2 > LR_b = 0$.[11]   The odds on what the signal will be vary depending on the manager's action. The difficulty is we have focused on the new signal itself, to the exclusion of what we are learning from the in-place information. And once we control for this initial supply of information, there is no remaining information content in our

---

[11]Focusing on signal y we have a likelihood ratio of $LR_y = \frac{\pi(y|L)}{\pi(y|H)}$. Using $\pi(x,y|a)$ in Table 14.2 we then calculate the noted likelihood ratios. We revisit this unconditional likelihood ratio in Chapter 16.

additional source. Emphatically, we cannot judge whether an additional performance measure might be useful without first sorting out what we are already learning from existing sources.[12]

Finally, turn to the setting in Table 14.3, where conditional on output, the odds are $50 - 50$ on what the additional information will report, provided H is supplied. It is clear the conditional likelihood ratio nonetheless varies with y, and the information is useful. The trick is to remember what it is we want to know: did the manager behave opportunistically? So the test, so to speak, is the relative odds of what might be reported, given action L versus action H. In this case, in equilibrium (that phrase again), it will appear as if the manager's evaluation (given output) is utterly random. True enough; but it would not be utterly random were he to behave opportunistically. Performance evaluation is all about inferring what transpired, and this means contrasting "on the equilibrium path" with the hypothetical "off the equilibrium path." The threat of what might be reported is an indispensable part of the story.

14.4   Larger Picture

The unrelenting message from the contracting model is the informativeness criterion. An additional performance measure simply cannot be useful in evaluating the manager's behavior unless it brings new information to the table. Bringing new information to the table means we have the potential to learn something new about the manager's behavior. And learning something new about the manager's behavior reduces to the conditional likelihood ratio: conditional on whatever it is we are already observing, can the manager, though his behavior, alter the odds of what might be reported by this additional performance measure? This is what is meant when we say the "threat of what might be reported is an indispensable part of the story."

More broadly, then, information helps mitigate a control problem when it brings additional insight into the manager's behavior in the problematic setting. In our little model the problematic setting is the basic conflict of the firm seeking input H while the manager has a preference for input L.

Somewhat hidden in this odyssey is the fact we know precisely what the control problem is: the firm has a preference for input H while the manager has a preference for input L. So we know precisely where to direct the conditional likelihood ratio test. Rather convenient.

---

[12] If you enjoy notation, the important fact is that $LR_{xy} = LR_x \cdot LR_{y|x} \neq LR_x \cdot LR_y$. To drive this point home, suppose we are observing output x and signal y. Might it be useful to introduce some additional signal or measure, call it z? If you think about it, you will wind up looking at the (doubly) conditional likelihood ratio of $\frac{\pi(z|x,y,L)}{\pi(z|x,y,H)}$.

Life, of course, is not this simple. To illustrate, suppose there are three possible inputs, H, B and L. Again the firm prefers H while the manager prefers L. Is the control "hot spot" H vs. L, H vs. B or both? Suppose it is H vs. B. Then additional information that helps distinguish input H from input L is not useful, as it tells us something that is not of concern in the control problem.[13]

Moreover, managerial tasks are hardly as simple as we portray. Multiple tasks are the norm, such as dealing with short-run and with long-run issues, dealing with this customer base and that customer base, balancing research and teaching, teaching to the test versus to the fundamentals, investing in versus harvesting your human capital. The list goes on. Once this is recognized, balancing tasks becomes another dimension to the control problem.

From here we also admit to multiple managers. Coordination among them can be beneficial, or detrimental. Control problems surely interact. So to no surprise, control problems have multiple parts, and our work to this point focuses on a control problem with but a single part.

But the basic principle holds. Conditional on the other information at hand, we look for performance evaluation variables that help us distinguish desired from opportunistic behavior, balanced from unbalanced task allocations, etc.

14.5   Summary

Performance evaluation addresses the question of whether the manager has behaved in the proper manner, has behaved well. Precisely what this means in our simple contracting model is clear. Given we are already contracting on output, if a second variable, a second performance measure, is useful then it must satisfy the informativeness criterion. The conditional likelihood ratio must be a nontrivial function of the new measure.

This stark criterion imposes two requirements. The new measure must bring something to the table that is not already known; and what it brings to the table must help us distinguish proper or desired from opportunistic behavior by the manager. It is this latter requirement, of testing "on" versus "off" the equilibrium path in the contracting game that distinguishes performance evaluation from other information exercises. Now you know why we sharply distinguish the two metaphorical questions of "What might it cost?" and "Did it cost too much?" The former asks about the equilibrium path while the latter juxtaposes the equilibrium and off equilibrium paths.

---

[13] You saw this in problem 13 in Chapter 13, and will see more of it in Chapter 16.

As reassuring and clarifying as this is, we must also remember the insight is derived in a highly, some would say hyper, simplified setting. But examining these more complicated settings presumes a grasp of the basics, the essence of this chapter.

14.6   Bibliographic Notes

The likelihood ratio connection is developed in Holmstrom [1979, 1982] and Shavell [1979]. Additional threads of the story used here are developed in Stiglitz [1974], Demski and Feltham [1978], and Harris and Raviv [1978]. A general and indispensable reference on the likelihood ratio is Milgrom [1981]. Christensen and Feltham [2005] provide an excellent technical treatment of the information side of contracting, with an accounting flavor. The broader picture is treated in Baron and Kreps [1999], Milgrom and Roberts [1992], Roberts [2004], Lazear and Shaw [2007] and Pfeffer [2007]. We will see in subsequent chapters how this building block is used to examine the performance evaluation theme in settings where multiple tasks, communication, decentralization and coordination are at play.

14.7   Problems and Exercises

1. We have stressed output is itself a source of information in the contracting game. Why is this so? Explain the connection between Examples 14.2 and 14.3.

2. What does it mean in the contracting model for a new performance measure, measure y, to be informative? What is the connection between the measure being informative and whether it is useful in the contracting exercise?

3. What is the connection in the simple contracting model between performance measure y being useful and the manager's risk premium?

4. When a new information variable is introduced into the contracting setting, the parties always have the option of agreeing to a contract that ignores the new information. Why is this so? Now return to Example 14.3 and verify that the optimal contract when contracting on output alone is a feasible solution to program (14.5), where the additional information variable is present. Hint: calculate the manager's certainty equivalent for each of his possible choices, including rejecting the contract.

5. We have emphasized a story where contracting on variable x might be improved by contracting on (x, y). Now reverse the story and

suppose we are initially contracting on variable y. What does it now mean for variable x to be informative?

6. The "shape" of an optimal contract depends on the information, as summarized in the likelihood ratio and displayed in expression (14.6). Now return to problem 9 in Chapter 13. What do you see? Explain.

7. *scaling*

   Ralph faces a recalcitrant manager. While Ralph wants input H, the manager prefers input L. (Yes, we have a prototypical contract story as studied in the chapter.) The manager is risk averse with preferences modeled in the usual fashion of expression (14.1). Assume $\rho = .0001, c_H = 15,000, c_L = 10,000$ and $M = 75,000$. The output probabilities are given below, where you will notice output is the only contracting variable and it can take on one of three possible values.

	x_1	x_2	x_3
$\pi(x\|H)$	.1	.2	.7
$\pi(x\|L)$	.5	.3	.2

   (a) Determine an optimal contract.

   (b) Now scale the setting, as we did in the chapter. What is the optimal contract in the scaled setting? Explain your reasoning.

8. *shadow prices*

   Return to the settings in Tables 14.1, 14.2 and 14.3 and focus on the optimal contract when the additional information variable is present, I^*_{xy}. For each of the three cases determine the shadow prices on the individual rationality and incentive compatibility constraints and verify that the optimal contract satisfies the likelihood ratio condition in expression (14.6).

9. *qualitative shape of optimal incentives*

   Consider a costly input setting in which output (x_i) can take one of four possible values. Input can be either L or H, with H desired by the risk neutral organizer. The input supplier is risk averse, and incurs an unobservable personal cost associated with input supply. We use the usual constant risk aversion specification of his preferences, with $\rho = .0001, c_H = 5,000$ and $c_L = 0$. Let the supplier's next best opportunity offer a wealth of $M = 0$. The output probabilities are given below.

	x_1	x_2	x_3	x_4
$\pi(x\|H)$	.1	.2	.3	.4
$\pi(x\|L)$	.4	.3	.2	.1

   Let I_i denote the payment to the input supplier when output x_i is observed. Without solving for an optimal arrangement, rank the four payments from lowest to highest. Carefully explain your answer.

10. *optimal contract*
    Determine an optimal contract for the setting in problem 9 above.

11. *what will it cost*
    Consider a normalized contracting story patterned after Example 14.3. Everything remains as specified there, except the probabilities are as follows.

	x_1/g	x_1/b	x_2/g	x_2/b
$\pi(x, y\|H)$	.1	.3	.3	.3
$\pi(x, y\|L)$	.2	.6	.1	.1

    (a) Determine an optimal contract. Explain your finding.

    (b) Suppose output here refers to the cost of some project, exclusive of payments to the manager. Let $x_1 > x_2$. Suppose you can observe signal $y = g$ or b, before cost x is observed. Is the signal informative about cost? Explain your reasoning.

    (c) Why is it possible to have an information source that is useless for contracting purposes in the presence of output, yet useful in forecasting what that output will be?

12. *did it cost too much*
    This is a continuation of problem 11 above. Everything remains as before, except the probabilities are given by

	x_1/g	x_1/b	x_2/g	x_2/b
$\pi(x, y\|H)$	.2	.3	.2	.3
$\pi(x, y\|L)$	.7	.1	.1	.1

    (a) Determine an optimal contract. Explain your finding.

    (b) Now is signal $y = g$ or b informative about cost? Explain your reasoning.

    (c) Why is it possible to have an information source that is useful for contracting purposes in the presence of output, yet useless in forecasting what that output will be?

13. *subtle points*
    Suppose we face a contracting problem as studied in this chapter, including but two output levels, two inputs, two possible signals, etc. The manager is specified by $\rho = .0001, c_H = 5,000, c_L = 0$ and $M = 0$. Input H is of course desired. The probabilities are given below, where $0 \leq \alpha \leq 1$.

	x_1/g	x_1/b	x_2/g	x_2/b
$\pi(x, y\|H)$	$.6(1 - \alpha)$	$.4(1 - \alpha)$	$.6\alpha$	$.4\alpha$
$\pi(x, y\|L)$	.25	.25	.25	.25

(a) Does the second information variable, $y = g$ or b, satisfy the informativeness criterion? Explain.

(b) Determine an optimal solution for the case $\alpha = .9$. Is the second information variable useful? Explain.

(c) Repeat (b) for the case $\alpha = 1$. Why can a variable satisfy the informativeness criterion and yet not be useful? Can it be useful without satisfying the informativeness criterion.

(d) What happens here as α becomes small, say approaches .6? (Hint: read note 3.)

14. *good versus bad news*

Ralph is now thinking about evaluating and compensating his manager based on output and on a monitor. The setting is familiar: low (L) versus high (H) input, Ralph desires high input and the manager desires low input. Output will be x_1 or x_2 and the monitor y will report g or b, and this signal will be observed at the time output is observed. The probability structure is given below.

	g/x_1	b/x_1	g/x_2	b/x_2
$\pi(x, y\|H)$	.25	.05	.60	.10
$\pi(x, y\|L)$	.35	.35	.15	.15

(a) Why would an optimal contract pay more for high (x_2) as opposed to low (x_1) output? Why would it pay more for monitor report g than for monitor report b?

(b) Suppose instead the probability structure is as given below. What happens to the good (g) versus bad (g) news interpretation of the monitor's report? Explain

	g/x_1	b/x_1	g/x_2	b/x_2
$\pi(x, y\|H)$	.15	.15	.35	.35
$\pi(x, y\|L)$	.05	.65	.25	.05

15
Allocation Among Tasks

Our study of performance evaluation to date focuses on the case where the manager faces a single task, stylized to choice between inputs H and L. Yet managers face a variety of tasks, and thus the decision of how to allocate their time, energy and talents among those tasks becomes an essential part of the larger picture. For example, the professor devotes time to teaching, research, and administrative duties. The manager of the fast food outlet devotes time to supervision, training, maintenance, communication with central administration, customer contact, and so on. The product line manager devotes time to production and delivery of the current product, to development of the next generation of the product, to workforce and customer bases, and so on.

This suggests our highly stylized control problem has a number of dimensions, within and across time periods. And, it turns out, providing incentives to deal, in an appropriate mix, with a variety of tasks is a profoundly delicate exercise. The reason is allocation among tasks is a multidimensional problem, and we often have information that speaks to only some of the dimensions. For example, the testing of the school teacher's students speaks, with noise, to whether the tested items are being mastered by the students, but is largely uninformative of the teacher's efforts at socializing and helping the students to mature in a safe environment. And, for sure, a strong emphasis on test results will invite effort at the margin that is

skewed to test preparation, simply because performance measures aimed at other equally important activities are difficult to come by.[1]

We begin with a variation on the input H versus input L story that highlights this allocation theme. We then return to the original H versus L story, where the firm wants input allocated to a single task. From there we move to the case where the firm wants a balanced allocation across tasks, to multiple managers, and to the important question of which bundles of tasks should be assigned to which managers. This is a question of assigning decision rights, one that rests on the underlying issue of well motivating a manager to balance his time, energy and talents across a variety of tasks. Following this we conclude, as is our wont, with a look at the larger picture.

15.1   Allocation of Total Input

We continue with the H versus L input story, but with three variations from the original setup. First, whichever input the manager supplies, it must be allocated to or divided up between two tasks. Let $a_i \geq 0$ denote the input allocated to task $i = 1, 2$. The manager then faces the constraint that $a_1 + a_2 \leq H$ if input H is supplied, or $a_1 + a_2 \leq L$ if input L is supplied. Think of this as, for example, dividing his time and talent between short-run and long-run issues, or between two customer bases.[2]

Input H continues to burden the manager with personal cost $c_H > 0$, while for input L the normalized personal cost is $c_L = 0$. No additional personal cost enters the story. So on the surface there is no conflict between the manager and the firm over how to allocate his time and talent, only over how much in total to supply.

Second, the contracting variable is not output per se, but a noisy, weighted tally of the manager's allocation to the two tasks. Specifically, we have a linear measure denoted

$$x = a_1 + \alpha \cdot a_2 + \varepsilon \qquad (15.1)$$

where $0 \leq \alpha < 1$. The idea is the performance measure addresses both tasks, but has a built in bias in favor of the first. For example, $\alpha = 0$ is a case where the performance measure is unaffected by or unable to deal with the second task. $\alpha = 1$, which we will rule out by assumption, is a case where the performance measure treats them as equally weighted. ε, in

---

[1] The movement from Chapter 2's single product firm to Chapter 3's multiproduct setting was a critical step in developing our understanding of the art of product costing. Similarly, the movement from the single task setting of the prior two chapters to a multitask setting is a critical step in developing our understanding of the art of performance evaluation.

[2] Naturally more than two tasks could be imagined, but patience is not a free good. Also notice the short-run versus long-run interpretation, while realistic, is a bit tongue in cheek here as we are, technically, dealing with a single period.

turn, is a noise term. We assume it is normally distributed about a mean of 0 and with a variance of $\sigma^2 > 0$. In short, the contracting variable is a noisy measure of a weighted summation of inputs allocated to each of the two tasks.

Third, while the firm remains risk neutral and the manager risk averse with constant risk aversion, measured as usual by $\rho > 0$, the contract itself is restricted to be a linear (well, affine) function of the contracting variable:[3]

$$I = \omega + \beta \cdot x \tag{15.2}$$

where the intercept ω is a wage or salary component, and the slope β provides an incentive component of $\beta \cdot x$. Think of this as an approximation to an optimal contract, similar to the accounting library's reliance on LLAs.

Now for the trick. In Chapter 9 we introduced a special case of the certainty equivalent calculation where the individual exhibits constant risk aversion and the lottery itself is a normal random variable with mean μ and variance σ^2. Repeating expression (9.3), this specialized setting provides a certainty equivalent of

$$CE = \mu - \frac{1}{2}\rho\sigma^2$$

That is, the certainty equivalent is calculated as the mean less a risk premium of one half the risk aversion measure multiplied by the lottery's risk, as measured by its variance.

With the contracting variable in (15.1) and the contract form in (15.2) the manager's compensation is given by

$$I = \omega + \beta[a_1 + \alpha \cdot a_2 + \varepsilon] = \omega + \beta[a_1 + \alpha a_2] + \beta\varepsilon$$

This is a normally distributed random variable with mean $\omega + \beta[a_1 + \alpha a_2]$ and variance $\beta^2\sigma^2$. So the manager's risk premium is

$$\frac{1}{2}\rho\beta^2\sigma^2$$

and his certainty equivalent is the mean of $\omega + \beta[a_1 + \alpha a_2]$ less the risk premium, or[4]

$$CE = \omega + \beta[a_1 + \alpha a_2] - \frac{1}{2}\rho\beta^2\sigma^2 \tag{15.3}$$

This certainty equivalent for the manager's compensation is the key to what follows.

---

[3] We continue, that is, with the manager's utility measure as in expression (14.1). The personal cost, however, relates to the total input but not its allocation.

[4] This follows from the fact that if ε is a normal random variable with mean 0 and variance σ^2, then for any constans f and g, $f + g\varepsilon$ is a normal random variable with mean f and variance $g^2\sigma^2$.

Don't miss the simplification in (15.3). The risk premium depends on the slope of the contract, β, but not on the manager's input or its allocation. The compensation mean, on the other hand, depends on this input and its allocation. Let's put this trick to work.

15.1.1   An Extreme Case

To begin, suppose the firm wants input H, and all of that input allocated to the first task. This is an admittedly extreme case, but it provides an important benchmark and an opportunity to get more comfortable with the model.

If the manager behaves, meaning supplies input H and sets $a_1 = H$, (and thus sets $a_2 = 0$), he faces risky compensation with a mean of $\omega + \beta H$ and a variance of $\beta^2 \sigma^2$, along with a personal input cost of c_H. Let CE_H denote the corresponding certainty equivalent, net of personal cost. We have

$$CE_H = \omega + \beta H - \frac{1}{2}\rho\beta^2\sigma^2 - c_H \qquad (15.4)$$

This follows directly from (15.3), once we set $a_1 = H$ and $a_2 = 0$ and net the personal cost. Likewise, if the manager misbehaves by supplying input L and setting $a_1 = L$, his certainty equivalent, which we denote CE_L, will be

$$CE_L = \omega + \beta L - \frac{1}{2}\rho\beta^2\sigma^2 \qquad (15.5)$$

where, recall, we have normalized the low input cost to $c_L = 0$.

The firm in this story offers the manager a contract, in the form of (15.2), and an instruction to supply input H and allocate that input to the first task. If this is to be attractive to the manager, it must satisfy the usual individual rationality condition. Using the normalized outside certainty equivalent of $M = 0$, this means the contract must satisfy $CE_H \geq 0$. In addition, if the manager is to be motivated to actually supply H instead of L, the incentive compatibility requirement of $CE_H \geq CE_L$ must also be satisfied. Here, however, the setup conveniently has the risk premium independent of the input choice, so incentive compatibility reduces to $\beta H - c_H \geq \beta L$, or

$$\beta \geq \frac{c_H}{H - L} \qquad (15.6)$$

The risk neutral firm, of course, wants to select the contract parameters, the wage or salary ω and the "piece rate" β, so as to minimize the expected payment to the manager, subject to the contract being acceptable (individual rationality) and motivating the desired behavior (incentive compatibility). And if the manager does indeed behave, the performance measure will be

$$x = H + \varepsilon$$

as the manager is supplying input H and allocating it to the first task. So we have the following program:[5]

$$C(H) \equiv \min_{\omega,\beta} E[I|H] = \omega + \beta E[x|H] = \omega + \beta H \qquad (15.7)$$

$$\text{s.t.} \qquad CE_H \geq 0$$

$$\beta \geq \frac{c_H}{H - L}$$

Example 15.1 Specify the manager by a risk aversion measure of $\rho = .1$ and a personal cost of $c_H = 60$, and define the inputs by $H = 500$ and $L = 200$. Also specify the contracting variable, (15.1), by $\alpha = .7$ and $\sigma^2 = 10,000$. Solving program (15.7), we find an optimal (linear) contract of $\omega^* = -20$ and $\beta = .20$, a claim you should verify.[6]

Intuitively, the manager will supply input H only if incurring the personal cost has a compensating shift in the value of his compensation, i.e., only if $\beta \geq \frac{c_H}{H-L} = \frac{60}{500-200} = .20$. In turn, the cost to the firm of this compensation package is, in total, the manager's personal cost plus his outside opportunity (which we have normalized to $M = 0$) plus his risk premium. This risk premium, which totals $\frac{1}{2}\rho\beta^2\sigma^2$, increases with slope β and is independent of the manager's input. (This is evident in the certainty equivalent expressions (15.4) and (15.5).) Naturally, then, we keep the incentive intensity, the β, as small as possible, consistent with motivating input H. So we have $\beta = .20$.

The intercept, ω, is set to satisfy the individual rationality requirement: $CE_H = \omega + \beta H - \frac{1}{2}\rho\beta^2\sigma^2 - c_H = 0$. We know $\beta = .20$, which provides a risk premium of $\frac{1}{2}\rho\beta^2\sigma^2 = .5(.1)(.04)(10,000) = 20$. With $H = 500$ and $c_H = 60$ we have $CE_H = \omega + .2(500) - 20 - 60 = 0$, or $\omega = -20$. And we wind up with a compensation cost to the firm of $C(H) = \omega + \beta H = -20 + .2(500) = 80$.

Notice, in both program (15.7) and Example 15.1, that we have not concerned ourselves with how the manager allocates his input. To see why, we return to Example 15.1.

---

[5] Both constraints will bind. Otherwise the manager's incentives are overly strong or the salary is overly generous. And with both constraints binding, we have two equations in two unknowns:

$$CE_H = \omega + \beta H - \frac{1}{2}\rho\beta^2\sigma^2 - c_H = 0$$

and

$$\beta = \frac{c_H}{H - L}$$

The solution is $\omega = \frac{1}{2}\rho\beta^2\sigma^2 + c_H - \beta H$ and $\beta = \frac{c_H}{H-L}$.

[6] As with our work in earlier chapters, the contracting examples are scaled to provide an element of numerical convenience. This leads, in this series of examples, to use of a risk aversion measure of $\rho = .1$

Example 15.1 (continued) The manager's performance measure x is given by $x = a_1 + .7a_2 + \varepsilon$ and his compensation is given by $I = -20 + .2x$. If he supplies input $H = 500$, he can allocate his time and talent to the two tasks in any fashion consistent with $a_1, a_2 \geq 0$ and $a_1 + a_2 \leq 500$. Any such feasible allocation provides an expected compensation of $\omega + a_1 + .7a_2$, a risk premium of $\frac{1}{2}\rho\beta^2\sigma^2 = 20$ and a personal cost of $c_H = 60$. This implies a certainty equivalent of

$$CE = \omega + \beta[a_1 + \alpha a_2] - \frac{1}{2}\rho\beta^2\sigma^2 - c_H = -20 + .2[a_1 + .7a_2] - 20 - 60$$

So how do we maximize $.2[a_1 + .7a_2] - 100$ subject to $a_1, a_2 \geq 0$ and $a_1 + a_2 \leq 500$? The answer is simple. Each unit allocated to the first task produces one dollar of certainty equivalent, while each unit allocated to the second produces but 70 cents of certainty equivalent. We have an optimal allocation of $a_1^* = 500$ and $a_2^* = 0$, precisely as desired.

Now you know why we specified the performance measure in (15.1) to put less weight on the second than on the first task (i.e., $0 \leq \alpha < 1$). This ensures the manager's allocation of total input between the two tasks is a trivial exercise. He will allocate everything to the first task because doing so is more personally productive as long as explicit incentives are turned on, as long as $\beta > 0$.[7]

15.1.2 More Information

Arranging for input H that is allocated entirely to the first task has the same flavor as the contracting story we used in the prior two chapters. This extends to additional evaluative information as well. To see this, suppose a second contracting variable is also available, one that will be publicly observed along with the original variable x at the end of the period. This new variable, denoted y, is similar in structure to the original variable. In particular we assume it is given by

$$y = a_1 + \gamma \cdot a_2 + \widehat{\varepsilon} \tag{15.8}$$

---

[7] To dig a bit deeper, suppose the manager has been instructed and motivated to supply input H. This implies the slope of his compensation arrangement is nontrivial, that $\beta > 0$, as implied by (15.6). In certainty equivalent terms, this leads to the following exercise to determine the preferred allocation of his input between the two tasks:

$$\max_{a_1 \geq 0, a_2 \geq 0} \omega + \beta[a_1 + \alpha a_2] - \frac{1}{2}\rho\beta^2\sigma^2 - c_H$$
$$\text{s.t. } a_1 + a_2 \leq H$$

But with $\beta > 0$ and $\alpha < 1$, the solution is $a_1^* = H$ and $a_2^* = 0$. A parallel analysis holds for the off-equilibrium analysis of supplying input L when input H is desired.

where unless otherwise noted we assume $0 \leq \gamma \leq 1$. Also, the $\widehat{\varepsilon}$ noise term is independent of the noise term in (15.1). It is normally distributed about a mean of 0 and with a variance of $\widehat{\sigma}^2 > 0$.

Repeating the earlier setup, but now with two contracting variables, we assume the manager's compensation consists of a salary, ω, and a "piece-rate" applied to each of the performance measures, β_1 and β_2:

$$I = \omega + \beta_1 x + \beta_2 y = \omega + \beta_1[a_1 + \alpha a_2] + \beta_2[a_1 + \gamma a_2] + \beta_1 \varepsilon + \beta_2 \widehat{\varepsilon} \quad (15.9)$$

This compensation lottery, being the sum of two normal random variables, is itself a normal random variable. Its mean is $\omega + \beta_1[a_1 + \alpha a_2] + \beta_2[a_1 + \gamma a_2]$ and its variance, thanks to the two noise terms being independent, is $\beta_1^2 \sigma^2 + \beta_2^2 \widehat{\sigma}^2$. So the manager's certainty equivalent for this compensation package is

$$CE = \omega + \beta_1[a_1 + \alpha a_2] + \beta_2[a_1 + \gamma a_2] - \frac{1}{2}\rho[\beta_1^2 \sigma^2 + \beta_2^2 \widehat{\sigma}^2] \quad (15.10)$$

From here it is a short step to see that motivating input H (which, in turn, leads to $a_1 = H$ and $a_2 = 0$, just as input L leads to $a_1 = L$ and $a_2 = 0$) requires the following slight variation on the original incentive compatibility constraint in (15.6)[8]

$$\beta_1 + \beta_2 \geq \frac{c_H}{H - L} \quad (15.11)$$

It is the sum of the two piece rates that provides the incentive strength in this case.

This leads to the following slight extension of design program (15.7):[9]

$$
\begin{aligned}
\widehat{C}(H) \quad &\equiv \quad \min_{\omega, \beta_1, \beta_2} E[I|H] = \omega + \beta_1 E[x|H] + \beta_2 E[y|H] \quad (15.12) \\
&= \quad \omega + \beta_1 H + \beta_2 H \\
\text{s.t.} \quad &CE_H = \omega + \beta_1 H + \beta_2 H - \frac{1}{2}\rho[\beta_1^2 \sigma^2 + \beta_2^2 \widehat{\sigma}^2] - c_H \geq 0 \\
&\beta_1 + \beta_2 \geq \frac{c_H}{H - L}
\end{aligned}
$$

Example 15.2 To illustrate, stay with the setting in Example 15.1 and assume the new information variable is structured precisely as the original one. So we have $x = a_1 + .7a_2 + \varepsilon$ and $y = a_1 + .7a_2 + \widehat{\varepsilon}$ along with $\sigma^2 = \widehat{\sigma}^2 = 10,000$. Think of this as a random sample of size two. In any

---

[8] Though we skip over the details, we are relying on $\beta_1 + \beta_2 \geq \alpha\beta_1 + \gamma\beta_2$, which is implied by our assumed restrictions on the α and γ weights. Subsequently we will allow, when appropriate, for the second measure to be biased in favor of the second task.

[9] To check your understanding, notice that setting $\beta_2 = 0$ returns us to the initial story summarized in program (15.7).

event, solving program (15.12) provides $\beta_1^* = \beta_2^* = .10$, $\omega^* = -30$ and a cost to the firm of $\widehat{C}(H) = 70$. This cost is less than the original example's cost of $C(H) = 80$.

To provide the intuition, notice that for any contract described by (15.9), the manager's risk premium will be

$$\frac{1}{2}\rho[\beta_1^2\sigma^2 + \beta_2^2\widehat{\sigma}^2] = .5(.1)[\beta_1^2 + \beta_2^2](10,000) = 500[\beta_1^2 + \beta_2^2]$$

We know from our earlier work that the manager will allocate all of his input to the first task. His certainty equivalent if H is supplied is therefore

$$CE_H = \omega + [\beta_1 + \beta_2]H - 500[\beta_1^2 + \beta_2^2] - c_H$$

Similarly, if L is supplied it is

$$CE_L = \omega + [\beta_1 + \beta_2]L - 500[\beta_1^2 + \beta_2^2]$$

Motivating H requires $CE_H \geq CE_L$, or $\beta_1 + \beta_2 \geq \frac{c_H}{H-L} = \frac{60}{500-200} = .20$. This reflects the fact here that the sum of the piece rates is the source of the incentive.

With this observation, it is clear we want to minimize the manager's risk premium while simultaneously ensuring the two piece rates sum to .20. With the additional information, this means we want to minimize $500[\beta_1^2 + \beta_2^2]$ subject to $\beta_1 + \beta_2 \geq .20$. And we thus have $\beta_1 = \beta_2 = .10$. From here, setting $CE_H = 0$ provides $\omega = -30$. But the central observation is the fact the second variable allows us to diversify the noise in the original performance measure, thereby creating a (modest) portfolio of performance measures that, in total, lowers the performance assessment noise and thus the manager's risk premium.

Example 15.2, then, is a case where an additional performance measure is useful. And this returns us to the informativeness criterion. Is variable y in the example informative in the presence of measure x? Definitely. With independence between the two error terms (ε and $\widehat{\varepsilon}$), we readily conclude the conditional likelihood ratio (expression (14.10) to be precise) in this case is

$$L_{y|x} \equiv \frac{\pi(y|x, L)}{\pi(y|x, H)} = \frac{\pi(y|L)}{\pi(y|H)}$$

With independent error terms, knowledge of measure x has no effect on our assessment of variable y's likelihood ratio. The new information is an independent assessment, unaffected by what we have learned from the first measure. And this likelihood ratio surely suggests variable y informs us about the manager's behavior.[10]

---

[10] Remember that y is a normal random variable, with its mean but not its variance affected by the choice between supplying H and L.

Notice, however, that the control problem is well isolated here. It centers on the choice between L and H, because we assume the firm wants the input used exclusively on the first task and the performance measures more heavily weight that first task. Given this, the additional performance measure speaks precisely to the control problem of motivating choice of H over choice of L.

15.2 Balanced Attention to the Two Tasks

The story escalates in subtlety, however, when we depart from the extreme case of allocating all of the input to the first task, in a setting where the performance measures differentially weight the two tasks.. To develop this important theme, we now assume the firm wants input H and half of the input assigned to each task, while the initial performance measure, x, continues to weight the first task more heavily.

Think of this as a case where two tasks are important, but performance on the first task is inherently easier to measure. Production efficiency is easier to assess than the morale of the workforce or the loyalty of the customer base. Likewise, short-run concerns are easier to assess than their long-run counterparts.

15.2.1 Expanded Control Problem

This concern for nontrivial attention to both tasks increases the control problem. The firm must now deal with conflict between itself and the manager over choice between input H and input L, as well as with allocation of that input. On the surface, the allocation issue is not of concern, because the manager has no direct personal interest in how his time and talent are allocated between the two tasks. But the bias in the performance measure, the fact one task is easier to measure than the other, indirectly creates a second conflict.

Most evident is the case where we have a single contracting variable, x as defined in (15.1). We already know the contract's piece rate or incentive component must be nontrivial, that $\beta > 0$. Otherwise the manager has no incentive to supply input H. We also know from Example 15.1 that if performance is measured by (15.1) and if $\beta > 0$, the manager will dedicate whatever input is supplied exclusively to the first task. The reason is his risk premium is unaffected by the allocation, but the mean of his performance is highest when input is so allocated. Given he is paid for performance (i.e., $\beta > 0$), and given his mean performance will be measured as $a_1 + \alpha a_2$, with $\alpha < 1$, it would be personally counterproductive to do anything other than allocate all input to the first task, the task more highly valued by the performance measure.

In fact, the firm faces a stark choice here. It can seek input H with the understanding the second task will then go unattended. Alternatively, it can seek input L, and have that smaller input allocated between the tasks. (Remember, input L requires no explicit pay-for-performance incentive and therefore does not lead to an unbalanced preference for input allocation on the part of the manager.) If neither option is appealing, the firm must find some additional performance measure, or it must find another manager to whom the second task could be assigned, presuming it can be so separated and not polluted by being bundled in some such parallel fashion.

Example 15.3 Return, again, to the setting of Example 15.1, but now assume the firm wants the manager's input equally split between the two tasks. We already know this is impossible if input H is motivated. Input L is another story. If the firm turns off the incentives, sets $\beta = 0$, the manager will receive a flat, guaranteed salary. Naturally he will not incur personal cost c_H to supply input H, but will supply input L. And with the incentives turned off, he will gladly allocate input L as the firm desires. After all, with $\beta = 0$ the evaluation measure does not favor one task over the other.

Turning to specifics, the manager's certainty equivalent, with $\beta = 0$ along with a normalized zero personal cost for input L, is simply $CE_L = \omega$. (This follows from our earlier work in (15.5).) Setting $\omega = 0$, we then have $C(L) = 0$, and $a_1^* = a_2^* = L/2 = 100$. This is the best the firm can do, if it insists on a balanced allocation.

As Example 15.3 reveals, the only way to achieve a balanced allocation of the input is to render the manager's preferences, once passed through the incentive structure, indifferent to the allocation itself.[11] Doing so requires we either turn the incentives off, as in Example 15.3, because they inadvertently distort the manager's preference for allocating his time and talent, or, yes, introduce additional information.

15.2.2   More Information (again)

To round out the story, suppose we (re)introduce the second performance measure, measure y defined in (15.8). Paralleling our earlier work, the manager's compensation contract is given by (15.9), resulting in the certainty equivalent (exclusive of personal input cost) displayed in (15.10).

Presuming we want to motivate input H, the idea, then, is to use the combination of performance measures to remove the inadvertent distortion of the manager's preference for emphasizing the first task. At the mar-

---

[11] Well, if you want to be pithy, we could deviate from the linear structures and dream up a performance measure that varied nonlinearly with the task allocation, and in such a way that the preferred, balanced allocation was the manager's maximizing solution.

gin, the manager must equally value the two tasks, once the incentives sufficiently strong to motivate input H are applied.

To see how to do this, look a little more closely at the manager's certainty equivalent when he supplies input H, but has yet to specify the allocation of that input. For any feasible allocation, and netting out the personal cost of input H, we have, thanks to (15.10), the following expression.

$$CE_{a_1 a_2} = \omega + \beta_1[a_1 + \alpha a_2] + \beta_2[a_1 + \gamma a_2] - \frac{1}{2}\rho[\beta_1^2\sigma^2 + \beta_2^2\widehat{\sigma}^2] - c_H$$

Importantly, the allocation to the two tasks has no effect on the risk premium. Moreover, each unit of input allocated to the first task increases the certainty equivalent by $\beta_1 + \beta_2$. Likewise, each unit allocated to the second task increases the certainty equivalent by $\alpha\beta_1 + \gamma\beta_2$.[12]

If the manager is to be motivated to balance the task assignments, to set $a_1 = a_2 = H/2$, he must equally value the two tasks. This leads to the additional incentive compatibility requirement that

$$\beta_1 + \beta_2 = \alpha\beta_1 + \gamma\beta_2 \tag{15.13}$$

That's it, everything else remains as before.

Pulling all of this together, the best linear incentive function that will simultaneously be attractive to the manager, motivate input H and motivate a balanced allocation of that input comes into view. Just as in program (15.12), we want to minimize the expected payment to the manager (the cost of input H to the firm), subject to $CE_H \geq 0$ (individual rationality) and $\beta_1 + \beta_2 \geq \frac{c_H}{H-L}$ (the original incentive compatibility condition to guarantee input H is motivated). In addition, we have the new incentive compatibility condition in (15.13). In short, we proceed precisely as in program (15.12), except the new, additional incentive compatibility condition is appended.[13]

Example 15.4 Continue with the setting of Example 15.1. Recall that, with $\alpha = .7$ and thus a performance measure of $x = a_1 + .7a_2 + \varepsilon$ (and $\sigma^2 = 10,000$), the cost to the firm of high input was $C(H) = 80$. We also know a balanced allocation of this input is infeasible here, as the single

---

[12]Stated differently, at the margin, the manager values the first task in terms of

$$\frac{\partial CE_{a_1 a_2}}{\partial a_1} = \beta_1 + \beta_2$$

and the second task in terms of

$$\frac{\partial CE_{a_1 a_2}}{\partial a_2} = \alpha\beta_1 + \gamma\beta_2$$

[13]We have skipped over a little work here. In particular, with the manager now indifferent as to how he allocates his input, it follows that his certainty equivalent remains as calculated in the unbalanced case. Suppose H is indeed supplied. With $a_1 = a_2 =$

performance measure more heavily weights the first task. Suppose a second performance measure of $y = a_1 + .6a_2 + \widehat{\varepsilon}$ along with $\widehat{\sigma}^2 = 10,000$ is available. So $\gamma = .6$. If the manager is to equally value the two tasks, the balance requirement in (15.13) must hold, or

$$\beta_1 + \beta_2 = \alpha\beta_1 + \gamma\beta_2 = .7\beta_1 + .6\beta_2$$

Solving program (15.12), but with this balance requirement appended provides an optimal solution of $\beta_1^* = .80, \beta_2^* = -.60, \omega^* = 460$ and a cost to the firm of $\widehat{C}(H) = 560$.[14]

Be certain you verify this solution, and understand how it motivates the desired behavior. First notice the manager has a balanced view of the two tasks, as $\beta_1^* + \beta_2^* = .80 - .60 = .20 = .7\beta_1^* + .6\beta_2^* = .7(.80) - .6(.60)$. Second, the manager is willing to supply high input as doing so has no effect on his risk premium but does increase his expected compensation by his increased personal cost of $c_H - c_L = 60$. In particular, a balanced supply of H provides expected compensation of

$$
\begin{aligned}
&\omega^* + .80[a_1 + \alpha a_2] - .60[a_1 + \gamma a_2] \\
=\quad &\omega^* + .80[H/2 + \alpha H/2] - .60[H/2 + \gamma H/2] \\
=\quad &\omega^* + .2H = \omega^* + .2(500) = \omega^* + 100
\end{aligned}
$$

Likewise, a balanced supply of L provides expected compensation of[15]

$$\omega^* + .2L = \omega^* + .2(200) = \omega^* + 40$$

$H/2$,

$$
\begin{aligned}
CE_{a_1 a_2} &= \omega + \beta_1[H/2 + \alpha H/2] + \beta_2[H/2 + \gamma H/2] - \frac{1}{2}\rho[\beta_1^2\sigma^2 + \beta_2^2\widehat{\sigma}^2] - c_H \\
&= \omega + (\beta_1 + \beta_2)H/2 + (\alpha\beta_1 + \gamma\beta_2)H/2 - \frac{1}{2}\rho[\beta_1^2\sigma^2 + \beta_2^2\widehat{\sigma}^2] - c_H \\
&= \omega + (\beta_1 + \beta_2)H - \frac{1}{2}\rho[\beta_1^2\sigma^2 + \beta_2^2\widehat{\sigma}^2] - c_H
\end{aligned}
$$

A parallel observation holds for the case where L is supplied. From here it follows that the original individual rationality condition suffices for this case, just as the original incentive compatibility condition is required to ensure input H is motivated.

[14] Let's not be too hasty here. Using the various parameters in the setting, we must solve the following program.

$$\widehat{C}(H) \equiv \min_{\omega, \beta_1, \beta_2} \omega + 500\beta_1 + 500\beta_2$$

s.t. $\omega + 500\beta_1 + 500\beta_2 - \frac{1}{2}(.1)[\beta_1^2 + \beta_2^2](10,000) - 60 \geq 0$

$\beta_1 + \beta_2 \geq .20$

$\beta_1 + \beta_2 = .7\beta_1 + .6\beta_2$

Also do not lose sight of the fact the last constraint, the balance constraint, simplifies the other expressions.

[15] These two calculations hold for any allocation of the respective inputs, provided whatever input actually supplied is fully allocated to the two tasks.

Third, accepting and behaving as instructed (and motivated), provides the manager a certainty equivalent of

$$\omega^* + 100 - \frac{1}{2}\rho[\beta_1^2\sigma^2 + \beta_2^2\widehat{\sigma}^2] - c_H$$

$$= \quad 460 + 100 - .05(.80^2 + .60^2)(10,000)$$

$$= \quad 560 - 500 - 60 = 0$$

(Recall the manager's normalized minimal certainty equivalent is 0.)

But why has the firm's cost increased so dramatically relative to the unbalanced case? This is the balance requirement at work. Both evaluation measures underweight the second task, and neutralizing this requires relatively large, counterbalancing incentive weights on the two measures, which leads to a dramatic increase in the manager's compensation risk.

Example 15.5 Contrast Example 15.4 with a setting where everything remains as before except the second performance measure is biased in favor of the second task. $\gamma = 1.2$ will suffice. You should now find an optimal incentive function of $\beta_1^* = .1038, \beta_2^* = .0962, \omega^* = -29.986$ and a cost to the firm of $\widehat{C}(H) = 70.01$. Here it is relatively easy to weight the two measures so as to provide the manager with balanced incentives, as one measure favors the first task and the other favors the second.

15.3 Insight into the Performance Evaluation Game

Examples 15.4 and 15.5 suggest task allocation may be a relatively minor or a relatively major issue. For sure, the control problem has expanded to concern for input supply and for allocation of that input, for allocation of the manager's time, talent and attention, among a variety of tasks. Dealing with a multidimensional control problem leads to unusual delicacy in the performance evaluation game. This leads to a variety of important insights.

15.3.1 Good Information Drives out Bad Information

The intuitive side of this heightened delicacy arises when so-called good information drives out bad information. Return to the single task story for a moment. There the optimal (linear) contract relies on a noisy performance evaluation measure, and nontrivial weight on this measure is essential if input H is to be motivated. But noise in the evaluation measure results in unwanted but unavoidable noise, and thus randomness in the pay-for-performance arrangement. This drives our recurring concern for the man-

ager's risk premium.[16] In a better world, this noise would be removed from the evaluation. Such is the special case where we add the possibility of a second measure, as in expression (15.8), but with no noise whatsoever. (So $\widehat{\sigma}^2 = 0$.)

Example 15.6 To illustrate, return to Examples 15.1 and 15.2. In the first example we have the noted story, where unwanted noise in the evaluation measure results in the manager bearing unwanted compensation risk. In the second example we introduced another noisy evaluation measure. Both measures are used in the manager's evaluation, and this diversification across noisy measures allows us to maintain incentives but with less compensation risk. Not bad. Now change the Example 15.2 story ever so slightly, by assuming the new measure is noiseless, that $\widehat{\sigma}^2 = 0$. Program (15.12) now provides an optimal solution of $\beta_1^* = 0$ and $\beta_2^* = .20$. This maintains the requisite incentives (as $\beta_1^* + \beta_2^* = .20$), and imposes no risk whatsoever on the manager because the performance measure in use is noise free. The new improved information has driven the original information out of the story.

The point here is the second information source is mixed with the first, unless it is of exceptional quality, meaning it measures performance without error. In that case, the original, noisy, information source is driven out by the exceptionally good information source. This is particularly transparent here, as we are not concerned with task allocation. The control problem is one dimensional, and the information speaks to precisely this dimension.

The balanced allocation case, however, is an entirely different matter because the task allocation case there is one in which the control problem is inherently multidimensional. In particular, we face simultaneous, interconnected concern for the total supply of service and for its allocation. Example 15.6, where balance is not an issue, focuses us on noise in identifying the manager's supply, but not allocation, of managerial service. As a result, the label of good versus bad information has precise meaning in that setting.

Contrast this with the same basic setting where balanced allocation is also sought. Now the label of good versus bad information simultaneously rests on noise in identifying total supply as well as on allocation of that supply.

Example 15.7 To see this, stay with the basic setup in the prior examples, but let the second information source offer an unbiased assessment of the two tasks, i.e. have $\gamma = 1$. Initially suppose the firm wants an unbalanced allocation, with the manager devoting himself exclusively to the first task. If both information sources are noisy, e.g., $\sigma^2 = \widehat{\sigma}^2 = 10,000$, it is routine to verify both information sources will be used. (With $\gamma = 1$ we have

---

[16] The same observation holds for the setting in Chapters 13 and 14.

no concern the manager will have a strict preference for the second task.) But, parallel to Example 15.6, if the second information source is noiseless, $\widehat{\sigma}^2 = 0$, it will again completely supplant the first information source.

However, if a balanced allocation of input H is sought, the optimal incentive arrangement disregards the first source regardless of noise in the two sources. That is, the solution to program (15.12), but with the balance requirement in (15.13) appended, provides an optimal solution of $\beta_1^* = 0$ and $\beta_2^* = .20$, regardless of σ^2 and $\widehat{\sigma}^2$. The reason is the first measure is, well, unbalanced while the second is perfectly balanced, and the balance requirement drives out what would otherwise be good information.

To be sure, Examples 15.6 and 15.7 are extreme cases (of no noise whatsoever or of perfectly unbiased assessment). Nonetheless, good information may well, and often does, drive bad information to the point it carries minor, secondary weight in performance assessment. The intriguing and important observation, however, is what we mean by good information is different in the two settings, simply because the underlying control problem has moved from one to two dimensions.[17]

15.3.2   Bad Information Drives out Good Information

Lest we conclude intuition is a highly reliable guide in the performance evaluation game, it is also possible (and observed in practice) that bad information may well drive out otherwise good information. Example 15.3 is a case in point. There the single performance measure is useful in motivating input H when that input is allocated entirely to the first task. However, it is not useful, indeed is caustic, if that input is to be allocated between the two tasks. Otherwise good information is driven out in the balanced setting, because it inadvertently distorts the manager's attention to an unbalanced view of the two tasks.

This phenomenon arises in a variety of settings. The audit firm that heavily stresses not going over budget in performing an audit invites less attention to potentially troublesome areas of the audit engagement. The school system that stresses student performance on standardized tests invites excessive attention to those tests, just as the dean who stresses stu-

---

[17]Recall from (15.13) that a balanced allocation requires piece rates such that $\beta_1 + \beta_2 = \alpha\beta_1 + \gamma\beta_2$. Similarly, an exclusive focus on the first task (when two information sources are present) requires $\beta_1 + \beta_2 \geq \alpha\beta_1 + \gamma\beta_2$. Notice this reduces, in the balance case, to

$$\beta_1 = \frac{\gamma - 1}{1 - \alpha}\beta_2$$

So if the second measure is balanced, if $\gamma = 1$, we necessarily have $\beta_1 = 0$. Similarly, and recalling $0 \leq \alpha < 1$, the balanced case has the two piece rates of opposite signs when $\gamma < 1$, but of the same sign when $\gamma > 1$.

dent evaluations invites excessive attention to matters aimed at student satisfaction, to the potential detriment of student learning.

Information that is good at assessing performance on some tasks may put performance on less well assessed tasks at risk. Another illustration is Example 15.7 where the second measure provides a balanced assessment, as it equally weights the two tasks ($\gamma = 1$). The first measure emphasizes the first task. Suppose it is even noiseless ($\sigma^2 = 0$). As we know, if the firm wants exclusive attention to the first task, this noiseless measure is perfect and drives out the second measure. Conversely, if the firm wants balanced attention to the two tasks, it relies exclusively on the second measure, as any attention paid to the first leads to excessive attention to the first task.

15.3.3   Task Assignment Matters

From here it is a short step to the question of which tasks, or decision rights, the firm should assign to which managers. Naturally, proximity and talent are important here. It makes little sense to assign scheduling in the New York facility to a manager working in the Hong Kong facility. Similarly, it makes little sense to assign the gifted product designer to human resources.

Performance measurement issues are also at work here. Glance back at Examples 15.4 and 15.5. There the firm sought a balanced allocation of the manager's time and talents. In Example 15.4 balance between the tasks was difficult to assess because the two performance measures were close substitutes for one another and both exhibited a bias toward the first task. In Example 15.5, however, balance between the tasks was relatively easy to assess because one measure emphasized the first task and the other emphasized the second task. The measures were, so to speak, natural complements in a setting where a balanced approach to the two tasks is sought.

Now expand this story. Suppose we have two managers and four tasks, with two tasks to be assigned to each manager. One assignment makes it relatively difficult to evaluate performance, as in the Example 15.4 case. The other assignment makes it relatively easy to evaluate performance, as in the Example 15.5 case. Presuming the managers are relatively adept at either of the combinations, it is clear the firm's assignment of tasks to the two managers will be driven by performance evaluation issues.

The firm's control problem appropriately expands to deal with task allocation decisions faced by the manager. And, as the examples suggest, the assignment of tasks to managers is part of the larger issue. A vivid example is separation of duties for internal control purposes, a topic originally broached in Chapter 13. Other examples include whether to subcontract maintenance, to separate initiation from approval of investment projects, or to separate checkout from bagging at the local grocery store. Task assignments are usually thought of in terms of bringing appropriate skills to bear on specific tasks. Underneath is another dimension, that of using task

assignment to put together collections of tasks that put less stress on the firm's control system. The wonderful phrase "organizational architecture" is aptly appropriate here.

15.3.4   Intertemporal Balance

A final insight that flows from the task allocation setting concerns intertemporal balance among tasks, managing the ever-present tension between dealing with today and tomorrow. The key, once again, is the evaluation measures must provide, though time in this case, the basis for instilling a properly balanced view of today and tomorrow.

To see how our (admittedly simple) setting extends in this direction, think back to Example 15.5 where a balanced allocation of tasks was sought in the presence of two performance measures, and the following compensation structure emerged:

$$I = -29.986 + .1038(a_1 + .7a_2 + \varepsilon) + .0962(a_1 + 1.2a_2 + \widehat{\varepsilon})$$

Now reinterpret the story as follows. Input H is up-front investment in human capital by the manager, a_1 is a short-run oriented task and a_2 is a long-run oriented task. The first measure is observed in the first period, while the second measure is observed in the second period. Further suppose the up-front human capital is such that once acquired it limits the manager's ability to effectively re-apply that human capital in an unbalanced fashion. So at the start of the arrangement, the manager perceives a balanced assessment of activities aimed at today and tomorrow, and applies his human capital accordingly. Once tomorrow arrives, he remains motivated to continue supplying $H/2$ to the a_2 task.

This is a bit tongue in cheek, but it illustrates two important points. First, a slight extension of our task story moves us into a multiperiod setting, just as we observed earlier in Chapter 3 (and will see in Chapter 18). Second, properly balancing tasks through time requires the manager equally value the two tasks, and continue to value them appropriately as the exercise unfolds.

Intertemporal balance is the key.

15.4   Stepping Back

While our examples are designed to build intuition and illustrate how control problems interact in a multitask setting, it is important to understand this phenomenon is widespread. For instance, it is routinely claimed the traffic officer does not work under a quota system, emphasizing number of citations issued. To do so would motivate too much attention to citations,

away from other more difficult-to-assess duties. A similar concern for explicit pay-for-performance incentives arises in secondary education. There, recall, the debate over use of bonus payments based on student test scores raises the question of whether this would motivate too much attention to "teaching the test," and away from a variety of other more difficult-to-assess activities.

Even closer to home is the publish or perish game. Suppose the professor's performance is measured by research output. This is, after all, tangible and can be evaluated by peers. This practice also raises the concern of whether it drives out teaching activities. In turn, student evaluations are introduced. This helps address, as we discussed earlier, the control problem of balancing the professor's attention to various tasks. But it also creates other control problems.

These additional control problems come from two directions. First, teaching covers a variety of tasks, including course design, development, and delivery. Today's curriculum must be delivered, and preparations must be laid for tomorrow's curriculum. Introduction of student evaluations raises the question of whether this invites too much attention to the task of delivering the current course, another version of the task allocation idea.

Naturally, student evaluations, course reading lists, examinations, assignments, and personal observation all provide insight into the professor's teaching activities and skills. This leads to the second control problem. The more comprehensive evaluation examines all these sources. Yet the student evaluations are numerically scored and readily tabulated. This invites concern over whether those responsible for preparing the comprehensive evaluation have themselves been comprehensive and thorough. The readily available evidence may drive out the production of other evidence, another example of bad information driving out good information.

These two-sided (or double moral hazard) concerns, in which important control considerations arise on both sides of a relationship, are commonplace. The insurance company worries whether the fact we are insured reduces our diligence; and we worry whether the insurance company is sufficiently frugal in its investment activities so that it can pay should a major claim occur. Is the manufacturer of the consumer durable sufficiently attentive to quality and are we sufficiently attentive to maintenance requirements in the use of the product? Is the manager sufficiently attentive to the variety of assigned tasks? Is the manager's supervisor sufficiently attentive to the task of evaluating the manager's performance?

It is no accident we often find grievance procedures in place. The concerned professor might turn to the university ombudsman. The annoyed new automobile owner might invoke the apparatus surrounding the state's "lemon law." The mistreated arrest victim might turn to the citizen review board. The grieved taxpayer might turn to the IRS problem resolution officer following an abusive, aggressive audit.

Task allocation is a fascinating subject. The full array of managerial art is pressed into play. For which tasks should high-powered incentives be used? Which task combinations properly balance comparative advantage of the individuals and control difficulties? What is the best way to deal with multi-sided control difficulties, as between a manager and supervisor?

15.5   Summary

Performance evaluation is an essential component in the fabric of arranging a trade of compensation for managerial services. This trade is surrounded by opportunistic or moral hazard concerns among the parties, though we have stressed opportunistic behavior by the manager. Moreover, the subtlety and delicacy of the trade arrangement escalate as we move to the multitask setting. Here, some tasks are inherently easier to evaluate than others, but doing so invites unwanted extra attention addressed to the easy-to-evaluate tasks, with a concomitant reduction in attention to the equally important yet difficult-to-evaluate tasks.

This leads to an expanded control problem, one addressed by which tasks are assigned to which managers, by producing additional information and, yes, by producing less information.[18]

15.6   Bibliographic Notes

Holmstrom and Milgrom [1987, 1991] brought the multitask problem into the realm of serious economic scrutiny. This has led to an explosion of papers exploiting the linear contract assumption in a setting of normal random variables. Feltham and Xie [1994] is an early and still wonderfully insightful paper. Christensen and Feltham [2005] and Lambert [2001] provide extensive surveys. Our particular variation on this theme is inspired by Demski, Frimor and Sappington [2006]. Hemmer [2004] provides an important cautionary note on the virtue of imposing contract form.

15.7   Problems and Exercises

1. Separation of duties is a time-honored control technique. Access to the cash register is limited, the inventory clerk does not count the inventory at year's end, and the warden does not grant paroles. Relate

---

[18]Precisely the same control problem expansion arises when we extend the single task setting of Chapters 13 and 14 to a multitask setting. This is explored in the end of Chapter materials, albeit with a slight addition of complexity.

separation of duties to the idea that some combinations of tasks are easier to control than others.

2. If a manager is assigned a single task, we expect high-quality evaluation information to drive out lesser quality evaluation information. For example, a monitor that identified the precise input supplied to the task would render any noisy indicator of input superfluous. Yet when multiple tasks are assigned a single manager, difficulty assessing performance on one of the tasks may overshadow the ability or desire to use high-quality evaluation information on the other task. Carefully explain how this might occur.

3. *source documents*
   Ralph's agent delivers confidential and valuable documents among a number of buildings in a metropolitan area. Rapid delivery is essential, and the agent's average delivery time between locations is an important productivity measure. Ralph is also expected to maintain detailed records: a log of the delivery requests and completions, release and acceptance signatures, and so on. What difficulties do you see with this arrangement, especially the concern for accurate records? What might Ralph do to help ensure accurate records and in such a way that delivery productivity is not compromised?

4. *basics of linear model*
   The contract derived in Example 15.1 has a negative wage of 20. What is the intuition for this (seemingly strange) conclusion? Hint: check out note 5.

5. *basics of linear model*
   Return to Example 15.1, but now assume the manager's outside certainty equivalent is $M = 900$. What is the optimal contract? What does this tell you about our normalization of $M = 0$ in this setting.?

6. *basics of linear model*
   Return yet again to Example 15.1.

   (a) Suppose the manager supplies input H but allocates half of that input to each task. Determine his certainty equivalent

   (b) Now suppose the manager supplies input L. What is his preferred allocation of this input among the two tasks? What is the resulting certainty equivalent?

7. *basics of linear model*
   Return to Example 15.2, but now assume the second variable has a variance of $\widehat{\sigma}^2 = 5,000$. Determine an optimal (linear) contract. Explain the difference between this contract and that in the original setting.

8. *basics of linear model*

   Return again to Example 15.2. Recall that with equal variances and independence this can be interpreted as random sampling the manager's performance. Let n be the size of the sample. Example 15.1 has $n = 1$, while Example 15.2 has $n = 2$. Determine an optimal contract for the $n = 3, 4$ and 5 cases. Write a short paragraph explaining your finding.

9. *balanced allocation*

   Consider a setting similar to that in our string of examples. Let $H = 900$ and $L = 200$. Also, $c_H = 100$ and the manager's risk aversion measure is $\rho = .1$. (Low input cost and the outside certainty equivalent are, as usual, normalized to zero.) The primary evaluation measure is, as usual, biased toward the first task, with $x = a_1 + .5a_2 + \varepsilon$. The noise term, ε, is a normal random variable with zero mean and a variance of $\sigma^2 = 10,000$. A second evaluation measure is also available, as described in (15.8), and the noise term is a zero mean random variable with variance denoted $\widehat{\sigma}^2$.

   (a) Initially suppose the firm wants input H allocated entirely to the first task and that the weighting coefficient on the second performance measure is $\gamma = .9$. Determine an optimal linear contract and cost to the firm for $\widehat{\sigma}^2 \in \{0, 100, 1{,}000, 10{,}000\}$. Interpret your findings, paying special attention to the notion of good versus bad information.

   (b) Next assume the firm seeks a balanced supply of input to each task. Repeat (a) above. Explain your finding.

   (c) Continuing with the balanced case in (b) above, now let $\widehat{\sigma}^2 = 10,000$. Determine an optimal arrangement for balanced supply of input for $\gamma \in \{.7, .9, 1, 1.2\}$. Interpret you findings, again paying special attention to the notion of good versus bad information.

10. *balanced allocation*

    Consider a setting as specified immediately above, except the noise term variances for the two measures are identical: $\sigma^2 = \widehat{\sigma}^2 = 15,000$. Also recall α denotes the weighting on the second task in the first measure (x) and γ denotes its counterpart in the second measure. A balanced supply of input H is sought.

    (a) Determine an (α, γ), such that the two measures are equally weighted in the optimal contract. Provide an intuitive explanation.

    (b) Repeat (a) above, but for the case where the first measure is more heavily weighted.

(c) Repeat (a) above, but for the case where the second measure is more heavily weighted.

(d) Write a short paragraph on the virtues of a balanced approach to such evaluation, one where the measures are, well, equally weighted.

11. *task assignment with an old friend*

Multitask issues are also discernible in a modest extension of our original single task story in Chapters 13 and 14. To see this suppose we have two tasks, each of which will produce output of 10,000 or 20,000, which we will code as "1" or "2" for each task. So the possible joint outcomes are 11, 12, 21, or 22, where 12 denotes outcome 1 from the first and outcome 2 from the second task,etc. As usual, the manager can supply input H or input L, but now to two tasks. The possible input combinations are therefore HH, HL, LH or LL, where, for example, HL refers to input H supplied to the first task and input L supplied to the second. Input H carries a personal cost of 3,000, while the personal cost of input L and the manager's outside certainty equivalent are normalized to 0. The personal cost of HH, then, is 6,000, of HL is 3,000, etc. We also model the manager's preferences in the usual fashion of constant risk aversion, and here assume a risk aversion measure of $\rho = .0001$. The manager will be assigned two tasks, but will not see the outcome of the first task before providing input to the second. Two types of tasks are under consideration. Their respective probabilities are detailed below.

		outcome	
		1	2
task type one			
	$\pi(x\|H)$	.4	.6
	$\pi(x\|L)$	1	0
task type two			
	$\pi(x\|H)$	.4	.6
	$\pi(x\|L)$	.7	.3

(a) Initially suppose this is a single task setting. Determine an optimal contract to provide supply of input H to a type one task. Do the same for a type two task.

(b) Now suppose two of task type one are assigned the manager. Given the possible combinations of inputs supplied to either task, we have the probability structure noted below. Suppose the following pay-for-performance arrangement is offered: -601.30 if "11" is observed, 6,410.97 if "12" or "21" are observed, and 1,0492.43 if outcome "22" is observed. Determine the manager's certainty equivalent for each combination of inputs. Is this con-

tract individually rational and incentive compatible? Explain.
Is the noted contract the optimal contract?

	11	12	21	22
$\pi(x\|HH)$	.16	.24	.24	.36
$\pi(x\|HL)$	.40	0	.60	0
$\pi(x\|LH)$	.40	.60	0	0
$\pi(x\|LL)$	1	0	0	0

(c) Determine an optimal contract for the case where two of task
type two are assigned the manager.

(d) In both cases above, it turns out assigning a pair of type one
or of type two tasks to the manager is more efficient than con-
tracting on each task separately. Verify this claim. What is the
explanation?

(e) Does it remain efficient to assign the pair of tasks rather than
contract for each task separately if the firm wants one of each
type task performed? Explain.

12. *interacting control problems*
Return to the setting immediately above, and concentrate on the first
case where two type one tasks are assigned to a single manager. But
now suppose the manager can delay choice between H and L for the
second task until the output from the first task is observed.

(a) Using the earlier determined pay-for-performance arrangement,
can the manager be counted on to supply input H to the second
task?

(b) Find an optimal pay-for-performance arrangement for this situ-
ation where the manager observes the first task's output before
supplying input to the second task. Explain the difference be-
tween your arrangement and that determined in the original
setting.

13. *task assignment*
Ralph owns a production function. Output can be either x_1 or x_2,
with $x_1 < x_2$. The manager's input can be L or H, with H desired.
Ralph is risk neutral. The probabilities are:

	x_1	x_2
$\pi(x\|H)$	.1	.9
$\pi(x\|L)$	.8	.2

The risk averse manager is modeled in the usual fashion, with personal
cost of high input $c_H = 5,000$ and personal cost of low input (c_L)
and outside opportunity certainty equivalent (M) normalized to zero.

Under constant risk aversion, his risk aversion measure is $\rho = .0001$. The only observable for contracting purposes is the manager's output.

(a) What is the best way to motivate supply of input H by the manager? How much would Ralph pay to be able to observe the manager's input?

(b) Call the above task one. A second task, task two, requires the same personal cost, and so on. The only difference is the probability structure:

$$
\begin{array}{c|cc}
 & x_1 & x_2 \\
\hline
\pi(x|H) & .1 & .9 \\
\pi(x|L) & .7 & .3 \\
\end{array}
$$

Suppose only this task is present. What is the best way to motivate supply of input H? How much would Ralph pay to be able to observe the manager's input?

(c) Now suppose both tasks are present, and Ralph wants supply of input H to both. The output of each task is separately observed. Also, the manager does not see the outcome of the first task before providing input to the second; so the input supply options are H to both, L to both, L and H or H and L. Will the above two incentive schemes motivate supply of input H to both tasks? Verify your claim, and give the intuition.

(d) What is the best way to motivate supply of input H to both tasks?

(e) How much would Ralph pay to observe the input supplied to task one?

14. *aggregation*

Return to problem 13 above. Both tasks are again present, but now only total output is observable. This implies low output from task one and high output from task two cannot be distinguished from high output from one and low output from two. How much would Ralph pay to observe the input to task one? Give an intuitive explanation.

15. *multiple tasks and delayed evaluation*

The manager of a facility that manufactures automobile components is evaluated on the basis of output (relative to an output budget) and cost (relative to a cost budget). Product quality is also important, and it is well recognized short-run performance measures can be favorably influenced by degrading quality. In turn, quality is monitored by inspection, scrap, and rework statistics. Warranty claims that are filed by customers are also important, though they can arise up to four years after the component was manufactured. The firm

tracks warranty claims by component, facility and manager at the time of manufacture. Thus, if the manager is promoted, the warranty statistics will continue to be compiled, thereby stretching out the evaluation period. Comment on this evaluation practice.

Financial reporting requires the firm provide an accrual to estimate the warranty expense and liability at the time of sale. Why does the firm not find this accrual sufficient for the evaluation exercise?

16
Accounting-Based Performance Evaluation

We now turn to the use of accounting measures in the performance evaluation game. Performance evaluation practices are highly varied across firms, ever changing, often contentious, and seriously interactive. Making sense of these practices leads us once again to the theme of an artful rendering of the underlying fundamentals. And just as economic foundations kept us focused on marginal cost in the product costing arena, we now use economic foundations to keep us focused on the information content of some particular evaluation measure.

The typical organization relies heavily but far from exclusively on the accounting library for performance evaluation purposes. The advantages of the accounting library for this purpose are twofold. It stresses financial matters, and financial matters are important. It also stresses integrity, and integrity is important. Performance evaluation can be consequential, as when a promotion is at stake. This places a premium on reliable appraisal. The disadvantage of the accounting library is its limited nature. It is a financial library, and integrity carries an implicit price. The accounting library cannot simultaneously be well protected and capture all we would like to have at our fingertips for evaluation purposes.

The point is simple. We should expect to find a variety of important evaluation insights in the accounting library; and we should expect to look elsewhere for additional, important information. To say "only the bottom line matters" is to reveal a distinctly uninformed and unprofessional conception of the art of performance evaluation. Market share, customer satisfaction, order books, quality, and the subjective opinion of the supervisor are potentially important sources of evaluation insight.

Even so, the accounting library itself offers a bewildering array of evaluation possibilities, simply because of the sheer number and complexity of the transactions it has recorded. Culling the appropriate set of measures from the library is a central issue. For example, should the manager's evaluation reflect or be purged of the effect of unanticipated factor price changes? Should his evaluation be confined to flow measures, such as cost incurred or income earned, or should it also include stock measures, such as assets dedicated to his area of purview? To no surprise (I hope), answers to questions of this sort are guided by the informativeness criterion introduced in Chapter 14.

We begin with the idea of responsibility accounting, which basically catalogues those items in the library that are used in evaluating a manager, those for which the manager is held responsible. This sets the stage for analysis of the major folklore in the performance evaluation arena, that a manager should be held responsible for what he can control, the so-called controllability principle. It also turns out the typical accounting library uses specialized language and algebraic rendering in the evaluation arena, and that is our concluding point. More detailed and nuanced aspects of the evaluation game are explored in subsequent chapters.

16.1   Responsibility Accounting

Responsibility accounting is the generic phrase for the way the accounting products are tailored for purposes of evaluating various managers. The idea is straightforward. A particular manager is held responsible for, is held accountable for, some identified array of accounting measures. In this way the firm assigns responsibility for various accounting outcomes, such as manufacturing cost, product profitability, and division return on investment. Stated differently, responsibility accounting is a blueprint that specifies the accounting measures that are used to evaluate a manager's performance. Responsibility is assigned. Each item in the accounting library is associated with a list of managers who bear responsibility for that item. The pattern is also hierarchical. For example, the measures by which a department manager is evaluated will also be used in evaluating the division's manager.

This is easy enough if we are talking about a small firm, or about the firm's highest level executive. We simply use the entire array of accounting measures in evaluating the manager's performance. Otherwise, we encounter nuances of organization life. After all, a firm is vastly more complicated than a production function that is guided through factor and product market interactions, under the skillful watch of a well-motivated management team. A firm has a life of its own, an ethos. It also enjoys economic

success because it is more efficient than a market at arranging some types of transactions.

16.1.1   Performance Evaluation Vignettes

To dig a bit deeper, consider a supervisor in a manufacturing department. To give the story more content, think of the supervisor in the service department in a large auto dealership. Many customers arrive, requiring a variety of repair services. The supervisor schedules the repair tasks and oversees the work of the mechanics. Standard or budgeted labor times are available for each repair task. A primary evaluation measure is the efficiency with which the various repair tasks are performed, as measured by labor time in relation to the budgeted labor time for each of the repairs. In this fashion, the evaluation process takes the repair tasks performed as a given and asks whether they were done efficiently.

The story does not end here. The dealership asks customers to mail in a service quality questionnaire. The general manager regularly visits the service facility. The service manager is likely to receive a year-end bonus if the dealership as a whole is profitable. We see a mix of accounting data, qualitative assessment by the general manager, nonfinancial data from the customers, and firm-wide profitability used in the evaluation of the service department manager.

Next consider a sales person. This individual contacts and visits many individuals, searching for new customers and managing the implicit relationship between the firm and its customers. The primary evaluation measure is orders received. The sales group is also engaged in a contest. The sales person with the largest total sales for the period receives special recognition, a holiday trip, and a bonus. In this fashion the performance of peer sales personnel is used to evaluate the sales person in question. Performance is evaluated relative to that of a peer group.

This tactic of relative performance evaluation is quite common. Grading students "on a curve" is another illustration. Use of industry comparisons, so-called benchmarking, where an executive is evaluated based on division income relative to the income of competitors, is another illustration. Public schools are often evaluated using spending and student performance measures relative to counterparts at peer schools. Higher education creeps steadily toward this reliance on competitor comparisons as well, with an added flavor of assessments by journalists.

Now envision the manager of a manufacturing facility in an integrated organization. Goods manufactured in this facility are transferred to a marketing group, where warehousing, distribution and so on are handled. Standard manufacturing costs have been established for each product. These set the stage for using actual versus budgeted manufacturing cost, given the list of goods manufactured, as a primary evaluation measure. Other statistics are also used, including summaries of equipment downtime, employee

turnover, on time delivery, and warranty claims that arise from customer use of previously manufactured and sold products.

Primarily the manufacturing manager is evaluated as a cost center, meaning cost incurred relative to budget in light of actual output is the primary accounting-based evaluation measure. This raises the question of whether the manager would be better evaluated as a profit center, meaning revenue and cost are the primary accounting-based evaluation measures. Shipments to the marketing group could be recognized at some agreed-upon internal "price." The manager would then have more of a profit enhancement rather than cost minimization orientation.

Further suppose this manager is asked at times to handle rush orders that are brought to him by the marketing group. This often results in excessive costs, due to general congestion and to use of overtime. Are these excess costs the responsibility of the manufacturing manager? He is, after all, following instructions and responding to the urgency.

Alternatively, consider a retail store manager. Typically he would be evaluated based on the store's income, or revenue and expense. The manager is able to influence, to a degree, the productivity of the store's labor force through work assignments, supervision, and so on. Similarly, activities of the work force indirectly affect the store's attractiveness to customers and thereby influence demand.

Now complicate the story. Suppose central management selects the merchandise that will be stocked. If some of this merchandise does not sell and must be put on sale, the store's revenue will be less than it otherwise would be. Should the manager then receive credit for the sales markdown? For example, suppose the store's revenue totaled 430,000 dollars but would have totaled 500,000 had the same merchandise been sold but with no markdowns. Do we want to evaluate the store manager in terms of 430,000 revenue or in terms of 500,000? More broadly, the question is whether to evaluate the manager based on revenue and expense or based on revenue, expense, and markdowns.

Finally, ponder the plight of the local manager in a fast food chain. Cost control is important, as is revenue growth. Yet the manager has no say over products, prices or, likely, hours of operation. The manager is evaluated as a profit center. A profit goal is negotiated with a regional supervisor, reflecting performance of peer outlets in the chain and local conditions. For example, a nearby construction project may temporarily increase or decrease demand at this outlet. In addition, the manager's performance is rated on a variety of nonfinancial dimensions, relating to employee turnover and training, the outlet's appearance and the quality of the standardized food products offered.

Taken together, we have a variety of evaluation practices. Portions of the accounting library are brought to bear, together with qualitative and quantitative information from a variety of sources. The common theme

is a wide array of evaluation measures, all presumably providing useful information for the evaluation task.

16.1.2   Controllability Principle

Easy to say, but how is the usefulness of some information item to be identified? A common, intuitive evaluation norm is that a manager should be evaluated based on those measures he can control, those he can influence or affect through his actions. For example, the manufacturing manager who is obliged to accept a rush order brought by the marketing group should not be held responsible for overtime costs incurred to get the order out on a timely basis. These overtime costs are not controllable by the manager. Similarly, the manager should be held responsible for the cost of manufacturing the item, once we have removed the overtime costs. The manager can control and is responsible for the ordinary costs of production. Similarly, the manager whose sales are unusually large because of labor strife at a competitor is not to be rewarded for increased sales, just as the manager whose sales have plummeted due to unforeseen and unfavorable exchange rate movements is not to be scolded, or worse, due to a fall off in sales.

   It is time for fundamentals.

16.2   A Closer Look at Controllability

Returning to our stylized contracting model of Chapters 13 through 15, suppose we have two possible variables on which to contract, output and an additional evaluation measure. What does it mean for the manager to control one or both variables? The manager supplies input. So we ask whether the manager's input choice affects the probabilistic description of the variable in question. If the variable's outcome is unaffected by the manager's behavior, the manager does not control the variable. If the variable's outcome is affected by the manager's behavior, the manager does, to some degree, control the variable.

   To be more precise, suppose, in the context of the contracting model, that the manager can supply input L or H. (Allocation of this input among a variety of tasks does not affect the substance of what follows.) Let y denote the evaluation measure in question, and $\pi(y|L)$ and $\pi(y|H)$ its probabilistic description depending on the manager's input supply. Paraphrasing our earlier definition of information content in terms of a likelihood ratio, we again turn to a likelihood ratio, but now focus on the evaluation measure in question. This is the unconditional likelihood ratio, introduced earlier

in Chapter 14's expression (14.8):

$$LR_y \equiv \frac{\pi(y|L)}{\pi(y|H)} \tag{16.1}$$

From here we say performance measure y is controllable if the unconditional likelihood ratio LR_y is a nontrivial function of y.

Definition 23 *Performance measure y is controllable if the unconditional likelihood ratio LR_y varies with y.*

Compared to our earlier work on information content, controllability has an intuitive simplicity of laser-like focus on the measure at hand. If the manager's behavior does not affect the measure's probabilistic description, we have $\pi(y|L) = \pi(y|H)$, and a constant unconditional likelihood ratio. But if his behavior does affect the measure's probabilistic description, we have $\pi(y|L) \neq \pi(y|H)$, which reduces to the fact LR_y, the unconditional likelihood ratio, is a nontrivial function of y.

Yet we know from our earlier work that if an additional measure is to be useful in the single-task contracting model, it must be informative. Its conditional likelihood ratio must somewhere vary with y. Otherwise it carries no new information to the performance evaluation exercise. So what is the connection between informativeness and controllability?

To explore this, we reprise and subsequently extend an earlier illustration, Example 14.3.

Example 16.1: Assume the manager's input can be H or L, with H preferred by the firm. Output can be either x_1 or x_2 (with $x_1 < x_2$). The manager's preferences are specified by the usual constant risk aversion setup with risk measure $\rho = .0001$. In addition, the personal costs are $c_H = 3,000$ and $c_L = 0$ and the (normalized) market opportunity is $M = 0$. Further assume the output probabilities are $\pi(x_1|H) = \pi(x_2|H) = .50$ and $\pi(x_1|L) = 1$.

As a benchmark, if output itself is the only contractible variable it is routine to verify the optimal pay-for-performance arrangement has respective (low and high output) payments of $I_{x_1}^* = 0$ and $I_{x_2}^* = 7,305.66$. The firm's cost is $C(H) = 3,652.83$ and the manager's risk premium is 652.83. (This should be familiar.)

Next (continuing to reprise the earlier example) we introduce an additional performance measure. This measure will report either $y = g$ or $y = b$. The probabilities are specified in Table 16.1, where you should note that absent the additional measure we are back to the benchmark setting of $\pi(x_1|H) = \pi(x_2|H) = .50$ and $\pi(x_1|L) = 1$.

Also notice the new, improved optimal pay-for-performance arrangement displayed in Table 16.1.[1] The additional measure is useful, as we have $I^*_{x_1 g} > I^*_{x_1 b}$. Following our earlier definition in Chapter 14, it is also informative, as the conditional likelihood measure of expression (14.10) is a nontrivial function of measure y (when x_1 obtains). Overall, the firm's cost is reduced to $\widehat{C}(H) = 3{,}148.70$, which implies a risk premium for the manager of 148.70

Now check out the new measure's controllability by the manager. Notice, glancing at the probabilities in Table 16.1, that we have $\pi(g|H) = .75, \pi(b|H) = .25, \pi(g|L) = .20$ and $\pi(b|L) = .80$. The respective unconditional likelihood ratios, using expression (16.1), are $LR_g = 20/75 < LR_b = 80/25$. The measure is controllable by the manager, and likewise useful in evaluating his performance. We are on to something.[2]

TABLE 16.1: Details for Example 16.1

	x_1/g	x_2/g	x_1/b	x_2/b
$\pi(x,y|H)$	.35	.40	.15	.10
$\pi(x,y|L)$	.20	0	.80	0
I^*_{xy}	3,590.23	4,002.35	-727.02	4,002.35
I^*_x	0	7,305.66	0	7,305.66
$LR_{y|x} = \frac{\pi(y|x,L)}{\pi(y|x,H)}$	$\frac{20}{70}$	0	$\frac{80}{30}$	0
$LR_y = \frac{\pi(y|L)}{\pi(y|H)}$		$\frac{20}{75}$		$\frac{80}{25}$

$$C(H) = 3{,}652.83 \ (RP = 652.83)$$
$$\widehat{C}(H) = 3{,}148.70 \ (RP = 148.70)$$

Example 16.2: Now stay with the same setup as in Example 16.1, including the benchmark specification. The only difference is the probabilistic specification of the new measure. Details are summarized in Table 16.2.

---

[1] We have not yet strayed from Example 14.3. Table 16.1 is simply a (strategically) rearranged version of Table 14.1.

[2] Recall we say measure y is informative in the presence of x if the conditional likelihood ratio $LR_{y|x}$ varies with y for at least one realization of x. In turn, Table 14.5 walks us through the calculation of the conditional likelihood ratio for this particular example. So, for instance, we have

$$\pi(g|x_1, H) = \frac{\pi(g, x_1|H)}{\pi(x_1|H)} = \frac{.35}{.35 + .15} = .70$$

and

$$\pi(g|x_1, L) = \frac{\pi(g, x_1|L)}{\pi(x_1|L)} = \frac{.20}{.20 + .80} = .20$$

and thus

$$LR_{g|x_1} = \frac{\pi(g|x_1, L)}{\pi(g|x_1, H)} = \frac{20}{70}$$

Here you will notice the new measure is not useful. The optimal pay-for-performance arrangement ignores the new measure. The new measure fails the informativeness test, as evidenced by the conditional likelihood ratio. It is also not controllable, as evidenced by its unconditional likelihood ratio being unity regardless of what the measure reports.

Again we have agreement between the informativeness test and the controllability test.

TABLE 16.2: Details for Example 16.2

	x_1/g	x_2/g	x_1/b	x_2/b
$\pi(x,y\mid H)$	.10	.10	.40	.40
$\pi(x,y\mid L)$	.20	0	.80	0
I^*_{xy}	0	7,305.66	0	7,305.66
$LR_{y\mid x} = \frac{\pi(y\mid x,L)}{\pi(y\mid x,H)}$	$\frac{20}{20}$	0	$\frac{80}{80}$	0
$LR_y = \frac{\pi(y\mid L)}{\pi(y\mid H)}$		$\frac{20}{20}$		$\frac{80}{80}$

$$\widehat{C}(H) = 3,652.83 \ (RP = 652.83)$$

Example 16.3: Next consider the variation in the probability structure displayed in Table 16.3. Here the new measure is useful and, yes, informative. But, as evidenced by its constant unconditional likelihood ratio, is not controllable by the manager.

The competing tests are now in conflict. The optimal contract uses the additional information to advantage. The new measure is informative, but fails the controllability test. The manager has no influence over what the new measure reports.

TABLE 16.3: Details for Example 16.3

	x_1/g	x_2/g	x_1/b	x_2/b
$\pi(x,y\mid H)$	.15	.05	.35	.45
$\pi(x,y\mid L)$	.20	0	.80	0
I^*_{xy}	3,053.70	6,645.01	-637.14	6,645.01
$LR_{y\mid x} = \frac{\pi(y\mid x,L)}{\pi(y\mid x,H)}$	$\frac{20}{30}$	0	$\frac{80}{70}$	0
$LR_y = \frac{\pi(y\mid L)}{\pi(y\mid H)}$		$\frac{20}{20}$		$\frac{80}{80}$

$$\widehat{C}(H) = 3,557.56 \ (RP = 557.56)$$

Example 16.4: Finally, consider yet another variation, that in Table 16.4. Here the new measure is not useful, is not informative; but is controllable by the manager.

The competing tests are once again in conflict. The optimal contract ignores the additional information. The new measure fails the informativeness test, but is surely controllable by the manager.

TABLE 16.4: Details for Example 16.4

	x_1/g	x_2/g	x_1/b	x_2/b
$\pi(x, y\|H)$	.10	.40	.40	.10
$\pi(x, y\|L)$	.20	0	.80	0
I^*_{xy}	0	7,305.66	0	7,305.66
$LR_{y\|x} = \frac{\pi(y\|x,L)}{\pi(y\|x,H)}$	$\frac{20}{20}$	0	$\frac{80}{80}$	0
$LR_y = \frac{\pi(y\|L)}{\pi(y\|H)}$		$\frac{20}{50}$		$\frac{80}{50}$

$$\widehat{C}(H) = 3,652.83 \; (RP = 652.83)$$

What are we to conclude? Informativeness, as assessed by the conditional likelihood ratio is the gold standard here. And, as evidenced by the four examples, there is no logical connection between controllability and informativeness. In Example 16.1 they are both present, and in Example 16.2 they are both absent. In Example 16.3 we have informativeness and absence of controllability, while in Example 16.4 we have controllability and absence of informativeness. There simply is no logical connection between informativeness and controllability.

The slippage, or error if you want to be less polite, is in how the presence of other information is handled. Controllability asks whether the manager can affect the statistical description of the new measure, regardless of what other information is present. While intuitive and, yes, simple, this ignores what is already being learned from the other information as well as any possible interactions between the existing and the new measures. Informativeness, on the other hand, stresses whether the manager can affect the statistical description of the new measure, conditional on what other information is present.[3]

Both tests boil down to a nonconstant, well-chosen likelihood ratio. The difference is informativeness looks for new information; it is conditional on the existing information. In this sense, informativeness could be interpreted as conditional controllability, controllability that is conditioned on what is already being learned from other information sources. This, however, is

---

[3] Naturally, what we learn from an information source depends critically on what we are learning from other sources. We saw this in the single-person choice setting of Chapter 9. (Problem 9-15 is an excellent refresher on the fact the value of some information source may well vary with what other information sources are present.) The same holds in the performance evaluation game, and this is why there is no logical connection between controllability and informativeness.

just another way of saying the fundamentals come down to whether the conditional likelihood ratio is nonconstant.[4]

16.3   Interpretation of Performance Evaluation Vignettes

Armed with this insight we return to the earlier performance evaluation stories. We begin with the service department supervisor in the auto dealership.

16.3.1   Service Department Manager

The primary measure used in evaluation of the service department manager is direct labor cost. It is used in the format of direct labor cost given the work accomplished, so the focus is on work accomplished (i.e., repair tickets or jobs) and direct labor cost. Actual is compared to budget, where the budget reflects work accomplished. We use the budgeted times allowed for the jobs worked on to raise the information content of the direct labor cost measure. Without knowing which jobs were worked on the direct labor cost would be largely meaningless.

The two underlying variables, then, are direct labor cost and jobs worked on. Yet jobs worked on is largely uncontrollable by the supervisor. In the short-run, it reflects a random arrival of customers. (Of course poor service will eventually affect the supply of jobs!) Together, though, direct labor cost and jobs worked on provide an insightful basis on which to evaluate the service department manager.

---

[4]A slightly more complicated version of the argument arises if we assume the firm is also risk averse. In this case we have two parties, the firm and the manager, who are risk averse. The firm faces a risky choice. It will then be in the interest of both parties to share in the risk. So nontrivial risk sharing will arise. Even without a control problem, then, the manager's compensation would be at risk. Now overlay a control problem. We will then generally see this ideal risk sharing arrangement distorted by a pay-for-performance arrangement that addresses incentive compatibility concerns. And informativeness will again surface as the inherent feature of an evaluation measure that makes it potentially useful in resolving the control problem.

To illustrate, let both parties be risk averse and also assume an evaluation measure that perfectly identifies the manager's input is available. The two parties will then share in the risk of the venture, and the evaluation measure will be used to control the input supplied by the manager. Given we are observing the manager's input, output is not informative. Yet it will be used in the compensation arrangement simply because of risk sharing. We emphasize a risk neutral firm because information content is more readily examined in a setting where ideal risk sharing is trivial. In addition, capital markets exist for sharing risk, and it seems odd to introduce risk sharing as a primary consideration in a labor market transaction.

A heavy focus on cost incurred given the jobs worked on does not, however, reveal the entire story. The supervisor might rush the repairs, cutting quality in the process. Particularly difficult repair tasks may be put off. So customers are invited to mail in a questionnaire; and the general manager periodically visits the service facility. Both activities provide additional information to help infer how well the service manager is performing.

Beyond this the auto dealership relies on its general image to promote sales and service. Some service facility activities spill over into the sales domain. A reputation for good service may help the sales force close a sale, for example. Given the other information, it should come as no surprise that dealership profitability is also used to evaluate the service manager.

In short, the service manager supplies a variety of managerial inputs across a variety of tasks. The evaluation system responds with a variety of measures, including actual cost relative to standard for the work accomplished, firm-wide profit, customer satisfaction, and the general manager's qualitative impressions. The measures are used to infer the manager's behavior. This results in a mix of seemingly controllable and uncontrollable variables. But information content is linked to conditional, not to unconditional, controllability.

16.3.2 Sales Contest

Now turn to the sales person. It seems intuitive that orders booked would be an important, and controllable, evaluation measure. The attendant sales contest, though, introduces the orders booked by another sales person. The peer's sales are not controllable by the sales person in question, just as the exam performance of other students is not controllable by a particular student. Yet important evaluation information is conveyed by this use of relative performance evaluation.

Was the student's performance the result of luck or skill and effort? The exam itself may have been easy or difficult. Another student's score tells us something about whether it was difficult. Similarly, the orders booked by other sales people tell us something about the market and how the product line is faring. Important environmental information is conveyed by using peer performance in the evaluation process.

For example, suppose the output of manager i (where $i = 1, 2$) is equal to that manager's input, a_i, plus noise. Some noise is idiosyncratic, say ε_i, while other noise is common to both environments, say, μ. So manager i's output is $a_i + \varepsilon_i + \mu$. Each manager's output is influenced by the common noise, μ. The difference in their outputs removes this common term. This is the intuitive idea behind relative performance evaluation.[5]

---

[5]Relative performance evaluation requires some commonality in the environments. It also runs the risk of sabotage. Couldn't one sales person encroach on the territory

16.3.3   Profit Center

Now turn to the manufacturing manager in the integrated organization. This manager is evaluated based on cost incurred, given output produced (as is the above service department manager). Additional statistics relating to equipment downtime, employee turnover, timeliness of delivery and warranty claims are also used. These speak, respectively, to issues of maintenance, employee training and morale, scheduling, and product quality. Again we see a mix of measures, designed to aid in the task of inferring what the manager has done.

The novelty in the story is the (oft heard) suggestion we convert the division from a cost to a profit center. This would be done by introducing a measure of revenue into the milieu. We already know the quantity produced and shipped to the marketing division. So to measure revenue of the manufacturing division we must come up with a price.

Consider two extremes. On the one hand, this might be a basic commodity with sales largely driven by market forces and the activities of the marketing division. In this story the product specifications are well established and the manufacturing division simply produces in response to a schedule largely set by the marketing division. It is unlikely actual sales revenue tells us anything substantive about the manufacturing division, given the other evaluation information already in place. This leads us to suspect the best way to measure revenue at the manufacturing division is with a standard or budgeted price per unit.

Measurement is easy, and we seem to get what we pay for here. We already know the units manufactured and shipped. Let q denote this quantity. Suppose the budgeted price is set at 12 per unit. We already know q; $12q$ is hardly going to be useful at this point. Being able to measure profit at the division level simply does not imply the additional measurement of revenue is useful. Here it seems revenue is uncontrollable and conditionally uncontrollable (and therefore useless) in the presence of the other information. A profit center may have more prestige, but prestige and information content are simply not the same.

The other extreme is a specialized product with sales driven by market forces, marketing activities, and the ability of the manufacturing division to help design and eventually produce the product in question. Here it is likely the sales revenue will tell us something about the manufacturing division, despite the other evaluation information in place. One possibility is also to use firm-wide profit to evaluate the manufacturing manager. This brings in

of another or couldn't one student be less than amiable in helping another understand some particular material? Similarly, if the exam is graded on a curve and the students all party the night before, their joint behavior will undermine the information provided by relative performance evaluation. This points to the fact that evaluation is an expansive task.

sales revenue, but commingles the information with randomness associated with various activities in the marketing division. Another possibility is to establish a revenue measure at the manufacturing level. Price at this point, for example, might be negotiated by the two division managers. This runs the risk of being influenced by their relative bargaining skills. It also offers the possibility of a revenue measure that helps infer the activities of the manufacturing manager.[6]

As an aside, recall that a cost center is in place when the evaluation focus is cost relative to output and a profit center is in place when the focus is profit. An investment center, then, is in place when the evaluation focus is profit relative to investment. (You saw this in Chapter 12 where we introduced the accounting rate of return and variations thereon designed to close the gap between accounting and present value based renderings.) Notice the ever increasing array of measures as we move from cost to profit to investment. The choice of which to focus on reflects the fundamentals of informativeness. Beyond that there is not much else to say.

16.3.4 Overtime on Rush Orders

Next, think back to the case of the manufacturing manager who receives a rush order and manufacturing costs are excessive due to overtime. Here the question is whether to evaluate based on total manufacturing cost or total manufacturing cost less the overtime cost. This is a question of whether overtime cost is informative, given that we know total cost. One answer is yes. In this narrative the manager is instructed to run a tight schedule and deal with any rush jobs by using overtime, as necessary. A cost overrun that is due to overtime work on rush orders is then not very interesting. We remove it by tempering the total manufacturing cost with the overtime costs associated with the rush job.

A second answer is in the negative. Here the manager is instructed to keep a relatively tight schedule, but with a modest amount of slack should rush orders appear. A cost overrun that is due to rush orders is now somewhat interesting. We therefore do not remove the overtime cost from the evaluation.

16.3.5 Sales Markdowns

Now return to the department store where central management selects the merchandise to be stocked. Sales markdowns may be used to sell some of this merchandise. If so, should this affect the store manager's evaluation? If not, the primary evaluation measures are revenue and expense. The ques-

---

[6]This transfer pricing arrangement will be explored further in Chapter 18, where we address coordination issues.

tion, then, is whether markdowns provide useful evaluation information given revenue and expense. Suppose market-wide forces heavily influence the price at which merchandise is sold. Markdowns now convey information, as they help remove market based noise from the revenue measure. In the limit, the best evaluation measures might be revenue, markdowns and expense.

Conversely, suppose the manager's sales efforts, display locations, and so on can affect revenue. This argues against using markdowns in the evaluation. Yet markdowns may still be informative. For example, if they are concentrated on a few products the manager was particularly opposed to stocking, this may suggest that sales effort is not being properly allocated. It also may suggest the original stocking decision was not well thought out.

16.3.6   Fast Food Manager

Finally, the fast food manager is evaluated as a profit center, even though he has no control over products or prices. One reason is cost relative to revenue informs about his cost control activities. Another reason is his overall service level and quality can affect demand. Beyond that, the various statistics addressing such things as the outlet's cleanliness are assembled because the emphasis on profit invites short-changing these activities. In turn, these statistics are somewhat subjective in nature; and thus we are likely to find a grievance procedure in place. This allows the manager to contest an evaluation by producing additional information.

The common theme across these vignettes is searching for information content in the presence of whatever else is being used in the evaluation task. That is the central message. We live and work with portfolios of evaluation measures, and each single measure's fundamentals can only be assessed in terms of what it brings to the evaluation task in light of what is being learned from all of the other measures in the portfolio.

16.4   A Caveat

Our euphoria, however, should be restrained by a (previously noted) modest qualification. We have, for pedagogical reasons, explored information content of a possible evaluation measure largely in terms of the single task, binary choice setup, where the manager faces choice between input L and input H. With input H desired this has the advantage of focusing the control concern on whether input L was supplied. We know precisely where the "hot spot" in the control fabric resides. But in the multitask setting, we saw that the hot spot might be total supply of input (i.e., H versus L), allocation of that total input across tasks, or both.

Examples 16.1 through 16.4 walk us through the (lack of) connection between controllability and informativeness in a setting where the control question is singular, input H versus input L. There, the additional measure did or did not speak to this control question, conditional on what was being learned from observing output. The qualification that is missing in this streamlined excursion is the fact the new information must tell us something about the control question, given whatever else we are already observing, but it must also do so at a point in the control fabric that is of concern. To illustrate, we don't quiz the students today on next week's assignments, as we know that at best they are working on current assignments. The quiz would explore whether they are working well in advance of the class schedule, while the pressing control problem is whether they are keeping up with the class schedule.

This is why we stressed in Chapter 14 that informativeness is a necessary but not a sufficient condition for a measure to be useful in the evaluation game. More structured examples follow.

Example 16.5: Return to Example 16.1 and concentrate on the benchmark case where output is the only contracting variable. We know the optimal arrangement is a payment schedule of $I_{x_1}^* = 0$ and $I_{x_2}^* = 7{,}305.66$. Now alter the story so the manager has three input options, L, B or H. L and H are as specified earlier. B is described by $\pi(x_1|B) = .80, \pi(x_2|B) = .20$ and a personal cost of $c_B = 1{,}500$. This enlarged problem, where the firm seeks input H but must worry whether input L or input B was supplied, has precisely the same payment schedule as in Example 16.1. To verify this, solve for the optimal contract on the assumption input B is not present. Given this solution, test it against input B. The manager's certainty equivalent for the risky compensation if he supplies input B is 1,094.50 which is less than the personal cost of input B.

Now consider an additional performance measure. It will report $y = g$ if input B is supplied and $y = b$ otherwise. This is clearly informative, as conditional on observing output what the measure reports depends on the manager's action. However, it is also utterly useless as long as input H is the desired input. The control hot spot is input H versus input L; there is no explicit concern for input B, the concern that is the focus of the new measure.[7]

Example 16.6: Return to Example 15.4 where we are concerned with supply of input H instead of input L, and face the additional complication of wanting this input supply equally split between two tasks. In the solution we relied on a pair of performance measures that would support this input total and input allocation pattern. Now append a third measure. This

---

[7]The shadow price on the incentive compatibility constraint requiring the manager prefer input H to input B is zero. Control hot spots and shadow prices are intimately connected.

measure will report "*yes*" if the manager's input is allocated 25% to the first task and 75% to the other and "*no*" otherwise. This new measure is informative, given the other information. Yet it is useless given the interest in a balanced allocation of the input to the two tasks. As in the prior example, the additional measure speaks to a corner of the control problem that is of no explicit concern.

The advertised qualification should now be clear. If a measure is to be useful in the performance evaluation game it must be simultaneously informative and informative about an aspect of the control problem that is of explicit concern. It must inform about a control hot spot. Informativeness alone, remember, is a necessary, but not a sufficient condition to append an additional measure to the portfolio of evaluation measures.

16.5   The Language of Expectations

A concluding stop on our introduction to the art of performance evaluation concerns language. While the economic fundamentals invite compact abstraction in various forms, such as "the measure will report $y = g$ or $y = b$," the accounting library archives a nearly overwhelming array of data. In response, specialized language and algebraic procedures have evolved and become commonplace.

The idea is straightforward. Consider a cost pool whose recorded cost will be used in the evaluation of a manager. We already know something about this pool. It is described by an LLA of $C = a + bx$, reflecting synthetic variable x. This LLA is central to the language. It also is likely to be based upon a budget, a plan for acquisition and use of resources, of factors, reflected in the cost pool.[8] The precise source of this LLA or budget varies across firms. The deeper issue is assembling information that is privately held by the various managers in order to construct an appropriate budget. This issue is taken up in the following chapters, so for now we simply assume the LLA, the budget, is given.

To add a bit of context, suppose we are dealing with an overhead cost pool, and that the synthetic variable is direct labor cost. From here suppose the actual overhead cost totaled $\widehat{C}$. This, presumably, speaks to the manager's performance but leaves open countless questions.

So we also introduce the synthetic variable into the stew. One possibility is to key on the actual level of the synthetic variable. Denote it $\widehat{x}$; this would be actual direct labor cost in our specific context. At this point,

---

[8]The budget, recall, reflects expected or anticipated consumption of resources in the cost pool, as a function of synthetic variable x. We also find authorization considerations in the world of budgeting. The budget, for example, may authorize expenditures up to the limit defined by the LLA, a common technique in government organizations.

then, we know the LLA, the cost incurred $(\widehat{C})$, and the actual value of the synthetic variable $(\widehat{x})$.

Now comes the language. Typically, this array of data is collapsed into what is called an accounting variance:

$$var \equiv \widehat{C} - a - b\widehat{x}$$

The total cost incurred, relative to the LLA's projected cost in light of the actual value of the synthetic variable, $\widehat{x}$, is showcased by this device.

Three points are in order. First, no new information is being produced by this juxtaposition of the actual and predicted cost. This is why we refer to accounting variances as a specialized performance evaluation language. Second, this usage of the term variance is commonplace and should be distinguished from statistical variance. The accountant and the statistician both use the term "variance," in precise yet distinct fashion. Third, we are keying on cost incurred relative to the LLA's prediction given the actual value of the synthetic variable, hence the phrase "language of expectations."

Adding yet additional information to the story leads to an algebraic decomposition of the accounting variance into components. To illustrate, suppose q units were actually produced. (Think of q as a vector.) Also let x^q denote the projected level of the synthetic variable had it been used efficiently in the production of q. In our specific context, then, we now know the overhead pool's LLA, the overhead cost incurred $(\widehat{C})$, the total labor cost or actual value of the synthetic variable $(\widehat{x})$, and the projected total labor cost had it been used efficiently in production of q (x^q). Keying the overall variance off the projected rather than actual level of the synthetic variable leads to the following decomposition.

$$\widehat{C} - a - bx^q = [\widehat{C} - a - b\widehat{x}] + [a + b\widehat{x} - a - bx^q]$$

The first bracketed term, $[\widehat{C} - a - b\widehat{x}]$, is our original variance and reflects total cost in the pool relative to the LLA's projection based on the actual level of the synthetic variable. The second bracketed term, $[a + b\widehat{x} - a - bx^q] = [b\widehat{x} - bx^q]$, reflects the estimated impact on the cost pool of synthetic variable usage relative to projected usage given actual output.

Variances and their decomposition, as we have emphasized, are commonplace features of the accounting library. Fundamentally, they are information conveyance devices. For example, in the above decomposition we are simply reporting actual cost in the pool, the actual value of the synthetic variable, and output. We then massage these observations with the cost pool's LLA and with the projected value of the synthetic variable had it been used efficiently in the production of q.[9]

---

[9] Getting slightly more abstract, think of the firm's total cost as a cost pool. Entering the period we projected factor prices of P while the factor prices eventually turned out

16.6   Summary

Performance evaluation is a well-practiced art. Multiple measures is the norm, ranging from financial to nonfinancial, from quantitative to qualitative, from periodic to occasional, and so on. Practice is varied and ever changing. For this reason, we emphasize an artistic rendering of fundamentals, based on an explicit control problem (of inherent conflict over supply of managerial input) and the use of information to resolve that control problem (by inferring the input supplied). Moreover, there is no guarantee that information useful in making a decision is useful in evaluating that decision. This is why we have stressed, beginning in Chapter 13, the metaphorical switch from "What will it cost?" to "Did it cost too much?"

Even so, the fundamentals are uncompromising. The imperative to evaluate based on controllable performance is intuitive, appealing, and unfortunately incomplete. An additional measure can be useful in the evaluation task only if it conveys new information. The central feature is informativeness (conditional controllability), not (unconditional) controllability. The professional manager's task here is to sort out which potential measures carry additional information (in a cost effective fashion) into the evaluation arena. This task is vastly more delicate than identifying a list of controllable performance indicators.

We should also flag the importance of the firm honoring its advertised evaluation practices. Evaluation places a burden on the firm, a burden that it may find tempting to diminish. And it is here that the comparative advantage of the accounting library comes to the fore. It is well defended and thus more difficult for either party to manipulate in the potentially high stakes game of performance evaluation.

16.7   Bibliographic Notes

Gordon [1964] explores responsibility accounting in terms of designing internal prices to which the managers should respond. Baiman and Demski [1980] link responsibility accounting to the information content of the measures for which the manager is held responsible, while Antle and Demski [1988] explore the disconnect between information content and the accountant's use of controllability. Laux [2006] stresses the control hot spots theme. The associated hierarchical structure of responsibility accounting is examined in Demski and Sappington [1989]. Arya, Glover and Radhakr-

---

to be $\overline{P}$. With an actual cost of $\widehat{C}$, we have the following decomposition:

$$\widehat{C} - C(q; P) = [\widehat{C} - C(q; \overline{P})] + [C(q; \overline{P}) - C(q; P)]$$

ishnan [2007] emphasize interactions among control problems, both within and across managers, in understanding conditional controllability. Gjesdal [1981] explores the subtle differences between evaluating a product and evaluating a manager, as do Feltham and Xie [1994]. Budde [2007] links the concept of a balanced scorecard to a multitask setting. Merchant [1989] and Bouwens and van Lent [2007] provide important institutional insight. Solomons [1965] is a classic reference on performance measurement at the divisional level, use of investment centers, and so on and Ijiri [1975] is a classic reference on library integrity, with an emphasis on a measure's "hardness." The language and calculation of accounting variances are standard fare (pun) in most managerial accounting textbooks. Structuring such a system on a planning model is developed in Demski [1967].

16.8   Problems and Exercises

1. The idea of responsibility accounting is straightforward: we hold a manager responsible for those accounting measures that tell us something about that manager's performance. Carefully discuss this idea. What does it mean to hold the manager responsible for an accounting measure? What does it mean that an accounting measure tells us something about a manager's performance?

2. Responsibility accounting focuses on the use of accounting measures in evaluating a manager. Might a manager be held responsible for nonaccounting measures? Give an example. How does the use of nonaccounting measures relate to the notion of responsibility accounting?

3. Discuss the difference between an evaluation measure being controllable versus conditionally controllable by a manager.

4. The informativeness criterion can be interpreted as saying that the information content of a potential evaluation measure must be controllable by the manager in question; otherwise, the particular measure cannot possibly be of any use in evaluating the manager. Carefully discuss this idea and relate it to the notion of conditional controllability.

5. *informativeness and off-equilibrium randomness*
Consider a variation on Examples 16.1 through 16.4, where everything remains as specified except the probabilities. The probabilities are partially specified by $\pi(x_2|H) = .9$, and the $\pi(x,y|L)$ noted below. You will notice the (x,y) combinations are utterly random under input L. Specify the remaining probabilities in such a manner that the additional measure (y) is both informative and controllable.

Verify your claim by calculating the unconditional and conditional likelihood ratios. You should also provide and interpret the optimal pay-for-performance arrangement in your specified setting.

	x_1/g	x_2/g	x_1/b	x_2/b
$\pi(x,y|H)$				
$\pi(x,y|L)$	.25	.25	.25	.25

6. *informativeness and on-equilibrium randomness*
   Consider another variation on Examples 16.1 through 16.4, where everything remains as specified except the probabilities. The probabilities are partially specified by $\pi(x_1|L) = .9$, and the $\pi(x,y|H)$ noted below. You will notice the (x,y) combinations are utterly random under input H. Specify the remaining probabilities in such a manner that the additional measure (y) is both informative and controllable. Verify your claim by calculating the unconditional and conditional likelihood ratios. You should also provide and interpret the optimal pay-for-performance arrangement in your specified setting.

	x_1/g	x_2/g	x_1/b	x_2/b
$\pi(x,y|H)$	.25	.25	.25	.25
$\pi(x,y|L)$				

7. *designer informativeness*
   Consider yet another variation on Examples 16.1 through 16.4, where everything remains as specified except the probabilities. The probabilities are specified below. Determine a value for $\pi(y = g|H) = \alpha$ such that (1) the monitor is useless and not controllable, (2) useful and allows zero risk premium, (3) useful and implies observing $y = g$ is "good" news and (4) useful and implies observing $y = g$ is "bad" news. Verify each of your claims by solving for the associated optimal contract.

	x_1/g	x_2/g	x_1/b	x_2/b
$\pi(x,y|H)$	$.2\alpha$	$.8\alpha$	$.2(1-\alpha)$	$.8(1-\alpha)$
$\pi(x,y|L)$	.45	.05	.45	.05

8. *cost allocation*
   We know cost allocation is commonplace, and a distinctly accounting phenomenon. We saw its use in decision framing, but what about performance evaluation? Give two institutional illustrations, one where cost allocation would be informative in evaluating a manager and another where it would not. As a hint, you might consider some overhead pool in the first case and the CEO's use of the corporate jet in the other.

9. *multitask setting*
   A popular metaphor is that of a balanced scorecard, that a variety of measures should be used in evaluating a manager and that they

should be treated in balanced fashion. Taking this literally, return to the multitask setting in Chapter 15, and especially the general setup in Examples 15.4 and 15.5. The firm seeks a balanced supply of input H to two underlying tasks. Two performance measures, specified in slightly different fashion than in Chapter 15, are available: $x = a_1 + \alpha a_2 + \varepsilon$ and $y = \gamma a_1 + a_2 + \widehat{\varepsilon}$. As before, ε and $\widehat{\varepsilon}$ are independent, normal random variables, and each has a variance of 10,000. Remaining details can be found in Examples 15.4 and 15.5.

(a) Provide a specification of α and γ such that the optimal (linear) contract equally weights the two measures.

(b) Provide a second specification such that the first measure is weighted more heavily than the second in the optimal (linear) contract.

(c) Are these measures controllable? Are they informative?

(d) Contrast this exercise with problem 15-10.

10. *informativeness and usefulness*
Verify the claim in Example 16.5.

11. *informativeness and usefulness*
Ralph, who is risk neutral, owns a production process. One of three feasible labor inputs, $L < B < H$, must be selected and H is desired. The labor supplier's preferences are modeled in the usual constant risk aversion fashion, with risk parameter $\rho = .0001$. The supplier's outside opportunity offers a normalized certainty equivalent of $M = 0$; and his personal costs are given by $c_H = 4,000, c_B = 1,000$ and $c_L = 0$. The contracting arrangement is limited to payment based on output, which can take on one of two possible values, x_1 or x_2. Output probabilities are displayed below.

	x_1	x_2
$\pi(x\|H)$	.10	.90
$\pi(x\|B)$	.70	.30
$\pi(x\|L)$	1	0

(a) Determine an optimal pay-for-performance arrangement.

(b) Suppose it is possible to install an additional measure, or monitor. This monitor will report bad news if input L is supplied and good news otherwise. Is this monitor useful?

(c) Is the monitor in (b) controllable? Is it informative (conditionally controllable)? Carefully explain this case of a serious control problem, a monitor that is both controllable and conditionally controllable, and yet is not useful.

12. *randomized monitoring*

This is a continuation of problem 8 in Chapter 13. Everything remains as before, except Ralph now has an information source. For a cost of 4,000 the source will report, without error, whether the manager supplied input H or input L. If input H is reported, the manager will be paid $I_H = 15,000$. If input L is reported, the manager will be fired, with a payment of $I_L = 0$. (No negative payments are allowed.) This is certainly effective, but far too costly.

(a) Suppose Ralph can commit to buying the information only when output x_1 is observed. Ralph will then pay 0 whenever the information is purchased and reveals input L was supplied and 15,000 otherwise. Will this motivate supply of input H? Will the manager have any compensation at risk, in equilibrium? Is this a good idea?

(b) Now suppose Ralph can commit to buying the information only when output x_1 is observed and only then with probability β. (Think of this as random monitoring.) Ralph will again pay 0 whenever the information is purchased and reveals input L was supplied and 15,000 otherwise. Find the lowest β that will motivate supply of input H. Is this a good idea?

(c) Next, consider a more elaborate plan. Set $\beta = .20738$. Pay 0 if the information is purchased and input L is reported. Pay 15,130.97 if x_1 is observed, the information is purchased and reports input H; pay 13,909.36 if output x_1 is observed and the information is not purchased; and pay 15,098.98 if output x_2 is observed. Is this a better idea? What is the explanation?

(d) Do you perceive any incentive problems on the part of Ralph at this point?

(e) Finally, notice all of the above presumes the lowest possible payment to the manager is 0. What happens if there is no such constraint, if the manager can be penalized with impunity?

13. *risk taking and insurance*

A major retailer, at one time, moved toward more centralized buying of merchandise that would be inventoried by its many locations. Each such location was evaluated in terms of profit earned at that location. To account for the overhead costs of centralized buying, the retailer booked the centrally purchased merchandise at each store according to the formula of invoice plus $t\%$. With t at 10, for example, an item costing 1,000 would be "sold" by center to the retail outlet for 1,100. The retail managers were not totally pleased with some of center's merchandising decisions, and complained that, at times,

they were stuck with merchandise that could not be sold. The re-
tailer dealt with this by allowing, upon approval, a markdown. To
illustrate, suppose the above noted 1,000 item was initially listed at
1,600 retail, but marked down 400. Suppose it sells for 1,600 - 400
= 1,200. The retail manager is now credited with 1,600 in revenue
and the shortfall of 400, the markdown, is debited to the above noted
centralized buying overhead account. What tensions are created by
the move toward more centralized purchasing in this case? How are
some of these tensions ameliorated by the markdown arrangement?
What are the likely consequences? What do you suspect will happen
to the percentage t as time goes on?

14. *flexible budget*

Consider a manager who produces goods or services according to cus-
tomer demand. The accounting library uses an estimate of total cost
based on an LLA of $TC = F + vq$, where q is some aggregate measure
of output. This is, of course, a flexible budget. Is the "flex" in the
flexible budget useful in evaluating the manager? If you know total
cost, is it likely learning output will bring additional, useful infor-
mation to the evaluation task? Carefully explain. Can the manager
control output?

15. *nonfinancial measures*

Performance evaluation has a long history. For example, Bokenkot-
ter [1979, page 153] reports the following practice in the Medieval
Church. "The tasks of the bishop were many and varied: administra-
tive, judicial, and spiritual. One of his chief duties was to conduct
visitations of the religious institutions in his Diocese. He usually held
the visitation in the local church and would summon the clergy of
the area and several laymen to attend. After verifying the credentials
of the clergy, the bishop would interrogate the laymen about the be-
havior of the clergy – whether they performed their duties properly,
whether they wore the clerical dress, whether they frequented taverns
or played dice. And the laity too had to answer for their conduct. Fi-
nally, the bishop would inspect the physical state of the church and
the condition of its appurtenances." Carefully discuss this practice.

16. *evaluation practices*

Suppose you, as manager, have just been moved to a new location.
One of your initial tasks is a quick study of how the individuals whom
you will now supervise have been evaluated. You are particularly in-
terested in how the items in the accounting library are used for this
purpose. The notion of controllability implies this quick study is a
relatively easy task: select a particular individual and ask which of
the many accounting measures might that individual control. The
notion of informativeness (conditional controllability) is not so ac-

commodating. How does it imply that your quick study should be organized?

17. *local and firm-wide bonus determinants*
A common practice is to define an overall bonus pool in terms of how well the firm has performed. For example, the pool might be a percentage of accounting income. A division manager's share in this pool, in turn, is heavily influenced by how well that manager has performed, for example, in terms of profitability of the division managed. Implicitly, then, the manager is evaluated in terms of local and global measures. Discuss this practice.

18. *change of pace*
Ralph is studying principal-agent problems. The principal is risk neutral, while the agent is risk averse and also incurs a personal cost. The agent can supply input L or input H. The agent's preferences are given by $\sqrt{payment} - c(input)$ where the agent's input cost is 15 if input H is supplied and 0 if input L is supplied, i.e., $c(H) = 15$ and $c(L) = 0$. The principal wants the agent to supply input H. Output will be either x_1 or x_2. Probabilities are defined by $\pi(x_1|L) = .5$ and $\pi(x_1|H) = .2$. In addition, the agent must face an expected utility of at least 50 "units" if this is to be an attractive option. Further suppose an additional measure or monitor is available for contracting purposes. The monitor will report $y = g$ or $y = b$.

(a) Determine an optimal pay-for-performance arrangement for each of the following four cases:

		x_1/g	x_1/b	x_2/g	x_2/b
case 1:	$\pi(x, y|H)$	.18	.02	.72	.08
	$\pi(x, y|L)$	.05	.45	.05	.45

		x_1/g	x_1/b	x_2/g	x_2/b
case 2:	$\pi(x, y|H)$	.10	.10	.40	.40
	$\pi(x, y|L)$	.25	.25	.25	.25

		x_1/g	x_1/b	x_2/g	x_2/b
case 3:	$\pi(x, y|H)$	.19	.01	.01	.79
	$\pi(x, y|L)$	.01	.49	.19	.31

		g/x_1	x_1/b	x_2/g	x_2/b
case 4:	$\pi(x, y|H)$	.16	.04	.16	.64
	$\pi(x, y|L)$	.40	.10	.10	.40

(b) Determine the conditional and unconditional likelihood ratios for the new measure in each of the four cases.

(c) In which of the four cases is the monitor useful? In which of the four cases is the monitor controllable by the agent? Carefully explain your findings.

19. *cost versus profit center and information content*
    Ralph owns a production function and seeks the services of a man-
    ager. The manager's input can be L or H; Ralph desires input H.
    The manager will oversee a process that will incur a cost that is
    low or high. Denote the scaled cost possibilities as $cost = 1$ and
    $cost = 2$. Associated revenue, in scaled format, will be $revenue = 4$
    or $revenue = 5$. The $cost/revenue$ probabilities are displayed below.
    In addition, while Ralph is risk neutral, the manager is modeled in
    the usual constant risk aversion fashion with risk aversion parameter
    $\rho = .0001$, personal costs of $c_H = 4,000$ and $c_L = 0$ and a normalized
    outside opportunity measure of $M = 0$. Contracting is confined to
    the observable cost and revenue.

	cost/revenue			
	1/4	1/5	2/4	2/5
$\pi(cost/revenue|H)$	.49	.41	.01	.09
$\pi(cost/revenue|H)$	.15	.15	.35	.35

    (a) Suppose Ralph treats the manager as a cost center. Determine an
        optimal pay-for-performance arrangement. Why is the manager
        paid more when a low cost is observed?

    (b) Suppose Ralph treats the manager as a profit center. Deter-
        mine an optimal pay-for-performance arrangement that uses the
        cost and revenue outcomes. Explain the structure of the optimal
        arrangement. Why is revenue a useful contracting variable when
        used in conjunction with the cost observation?

    (c) What general principle for choosing between a cost center and a
        profit center evaluation is illustrated here? How would this apply
        to choice between a profit center and an investment center?

    (d) Suppose instead of contracting on cost and revenue, the par-
        ties contract on profit itself, or revenue less cost. Determine an
        optimal-pay-for-performance arrangement; contrast it with the
        one determined earlier. What does this imply about heavy re-
        liance on a summary measure of performance, such as profit in
        a profit center or return on investment in an investment center?

20. *return on investment*
    Ralph manages a regional home products store. A variety of hard-
    ware, lumber, and small appliance items are stocked and sold to the
    general public. A smaller portion of the business deals with com-
    mercial customers. The store is one of many such outlets owned and
    operated by a large firm. Center, or central management, locates the
    various stores, makes merchandising and supply decisions, provides
    advertising, and so on. The store managers deal with the day-to-day
    operations of their store. A regional manager assists the store man-
    ager on such items as merchandising, the need for special promotions

when selected products don't sell, and so on. The major financial evaluation that Ralph labors under is the store's return on investment. The major assets are land, building, display fixtures, and inventory. The debt is centrally held. Cash is centrally managed. Ralph's particular store has been unusually successful in recent years, and center is contemplating expansion. Ralph, in a dark moment, has begun to worry that the expansion will be disruptive in the short-run, but will also bring additional land and buildings, not to mention inventory, onto the store's balance sheet. Running the numbers suggests the accounting rate of return will drop from 18% to around 11% if the expansion goes forward and sales keep pace at their current level of dollars per square foot of display space. Carefully appraise the firm's evaluation practices.

21. *current cost*

Ralph manages a consumer products outlet. Inventory is important. Customers will not return if the outlet is out of stock; and inventory carrying costs are far from trivial. Changing prices are also an issue. At present the historical cost of the outlet's inventory is 2,036,000 (based on LIFO), while the current cost of the inventory is 2,345,000. A compensation consultant has suggested Ralph be evaluated on the basis of current cost performance, where inventory would be valued at current cost and holding gains would be recognized as income. Ralph is suspicious, since this will add to the income measure's volatility. Carefully discuss the use of current cost measures in the evaluation context. Do you see any connection with the way GAAP handles foreign currency translation or fair value more broadly?

17
Communication

Gathering and communicating information are important managerial tasks. The production manager has superior information and insight in the production sphere, just as the product line manager is in the best position to forecast demand. Product development teams often combine engineering, manufacturing, industrial design, and marketing experts. Various managers are likely to have insights into competitor strengths and weaknesses. The manager faced with a budget shortfall is likely to know better than most the major events that led to the shortfall. The same manager probably contributed important information when the budget was originally set.

These communication activities expand the list of tasks assigned to the manager, and put additional stress on the relationship between the firm and its management team. Consider a familiar example: our auto fails to start and we have it towed to a garage. The mechanic quickly examines the problem. At this point the mechanic has an information advantage. We ask for a quote. The mechanic knows a somewhat padded quote will be advantageous. Eventually we agree to terms, and return when the repair is completed. The mechanic again has an information advantage. Did the repair take as long as noted on the bill? Was it necessary to replace the noted parts? Are the replaced parts fairly priced?

Of course, various institutional features come into play. The mechanic is required to offer us the replaced parts, so we can personally inspect them. Our permission is required if the repair bill is to exceed the original estimate. The mechanic has a reputation to uphold. For that matter, we might internet search our particular set of troubles, or perhaps seek a second opinion from a competing mechanic.

Parallel concerns and institutional arrangements surface inside the firm. The periodic planning process will solicit opinions from various managers. Budget analysts and consultants serve to diminish the information advantage of the managers whose opinions are being solicited. After the fact measurement of the managers' performance will be reconciled against their forecasts. Longer term reputation considerations are important in the managerial "game."

Our exploration begins with the overly familiar managerial input setting, but now expanded to endow the manager with private information and to task him with self-reporting that private information. This allows us to study how various facets of a control problem interact and, in particular, how we extend the web of controls to address communication incentives. From here we enrich the exploration by introducing various timing and interaction wrinkles. Finally, we examine a repeated setting and the age-old issue of earnings management or "income smoothing." The goal is to understand how the firm might go about providing incentives to communicate and participate productively in the management exercise.[1]

17.1   Self-Reporting Incentives in the Managerial Input Model

Initially we append private information and a communication task to our managerial input model. For this purpose we continue with the basic theme in Chapter 16 of using an additional measure to assist in evaluating the manager's performance. (Constant risk aversion and our normalized frame, with $c_L = M = 0$, remain in play.) The twist is this hopefully useful additional measure is now privately observed by the manager, and the only way to use it in the evaluation process is to rely on the manager's communication of what was (privately) observed.

We begin with a benchmark illustration.

Example 17.1: As usual, assume the manager's input can be H or L, with H preferred by the firm. Output can be either x_1 or x_2 (with $x_1 < x_2$). The manager's preferences reflect the usual constant risk aversion setup with risk measure $\rho = .0001$. In addition, the personal costs are $c_H = 3,000$ and $c_L = 0$ and the (normalized) market opportunity is $M = 0$. The output probabilities are given by $\pi(x_1|H) = .25$ and $\pi(x_1|L) = 70$. An additional performance measure is also available. This measure will report either $y = g$ or $y = b$. The probabilities are specified in Table 17.1, along with the associated optimal pay-for-performance arrangement. You will note

---

[1] Naturally, a specialized version of the story is where we provide incentives to record entries in the accounting library in a timely and accurate fashion.

the additional measure is informative and useful, and lowers the manager's
risk premium from 472.48 to 112.49.

TABLE 17.1: Details for Example 17.1

	x_1/g	x_1/b	x_2/g	x_2/b
$\pi(x,y\|H)$	.05	.20	.65	.10
$\pi(x,y\|L)$	.45	.25	.05	.25
I_{xy}^*	-2,594.45	3,055.79	3,688.67	2,334.20
I_x^*	-1,593.84	-1,593.84	5,161.26	5,161.26

$$C(H) = 3,472.48 \ (RP = 472.48)$$
$$\widehat{C}(H) = 3,112.49 \ (RP = 112.49)$$

We next alter this familiar story in order to exhibit a communication
task. First, we assume the manager privately observes the $y = g$ or $y = b$
signal after supplying input (H or L), but before output (x_1 or x_2) is
observed. Second, the manager is instructed to report this observation,
by sending a message that $\widehat{y} = \widehat{g}$ or $\widehat{y} = \widehat{b}$ was observed. (Notice we
notationally distinguish what was observed from what was claimed to have
been observed.) Third, the manager's compensation will now depend on
the self-reported measure and the subsequently observed output. (So we
denote it $I_{x\widehat{y}}$.)

Example 17.2: To see how this might work, return to Example 17.1,
but now assume the additional measure is privately observed by the man-
ager and self-reported during the noted time frame. Further suppose the
public information contract in Example 17.1 is carried over to the private
information setting. So the manager's contract is given by

	$x_1/\widehat{g}$	$x_1/\widehat{b}$	$x_2/\widehat{g}$	$x_2/\widehat{b}$
$I_{x\widehat{y}}$	-2,594.45	3,055.79	3,688.67	2,334.20

What will our manager do? Well, supplying input H and self-reporting
precisely what was observed privately will replicate the public information
case and provide the manager a net certainty equivalent of $0 \ (= M)$. This is
encouraging, but unfortunately temptations abound. Suppose the manager
surreptitiously supplies input L and also reports $\widehat{y} = \widehat{b}$ regardless of what
was observed privately. With $\pi(x_1|L) = .70$ and $c_L = 0$ this provides a
(net) certainty equivalent of 2,833.79 as

$$.70U(3,055.79) + .30U(2,334.20) = U(2,833.79)$$

We have a problem.

17.1.1 Expanded Options

The difficulty showcased in Example 17.2 is that asking for a self-report of a private observation invites the manager to anticipate how that self-report will be used. And this anticipation calculus is unlikely to be very friendly toward the use of the underlying information were it public. In the public case the information can be used without concern for its integrity; but in the private information case, how the information will be used can affect its integrity.

Giving precise meaning to this observation requires some care. When assigned this communication task, the manager's options multiply in number. For any choice of input, H or L, the manager also has four communication strategies: report what was observed privately, report $\widehat{y} = \widehat{g}$ regardless of what was observed, report $\widehat{y} = \widehat{b}$ regardless of what was observed, or always report the opposite of what was observed. This is sufficiently important that we shroud it in notation, by denoting, respectively, the four communication strategies $(\widehat{g}, \widehat{b}), (\widehat{g}, \widehat{g}), (\widehat{b}, \widehat{b})$ and $(\widehat{b}, \widehat{g})$.[2] Think of the four strategies as truth, stuck on $\widehat{g}$, stuck on $\widehat{b}$ or perverse.

All together, then, the manager has 8 distinct strategies when confronted with choice of input and choice of communication. The resulting probabilities, based on the setting of Example 17.1, are displayed in Table 17.2.

TABLE 17.2: Output, Report Probabilities for Example 17.1

	$x_1/\widehat{g}$	$x_1/\widehat{b}$	$x_2/\widehat{g}$	$x_2/\widehat{b}$
$\pi(x, \widehat{y}\|H, (\widehat{g}, \widehat{b}))$	.05	.20	.65	.10
$\pi(x, \widehat{y}\|H, (\widehat{g}, \widehat{g}))$	.25	0	.75	0
$\pi(x, \widehat{y}\|H, (\widehat{b}, \widehat{b}))$	0	.25	0	.75
$\pi(x, \widehat{y}\|H, (\widehat{b}, \widehat{g}))$	.20	.05	.10	.65
$\pi(x, \widehat{y}\|L, (\widehat{g}, \widehat{b}))$	.45	.25	.05	.25
$\pi(x, \widehat{y}\|L, (\widehat{g}, \widehat{g}))$	.70	0	.30	0
$\pi(x, \widehat{y}\|L, (\widehat{b}, \widehat{b}))$	0	.70	0	.30
$\pi(x, \widehat{y}\|L, (\widehat{b}, \widehat{g}))$	.25	.45	.25	.05

$\pi(x, \widehat{y}|H, (\widehat{g}, \widehat{b}))$ is the probability of observing output x and self-report $\widehat{y}$, given the manager supplies input H and pursues communication strategy $(\widehat{g}, \widehat{b})$. As this specific communication strategy is truthful, it replicates the public observation probabilities under input H, $\pi(x, y|H)$, displayed in Table 17.1. Similarly, if the manager supplies input L but pursues truthful communication, $\pi(x, \widehat{y}|L, (\widehat{g}, \widehat{b}))$ replicates $\pi(x, y|L)$ of Table 17.1. Com-

---

[2]Private observation possibilities (g, b) are mapped into $(\widehat{g}, \widehat{b}), (\widehat{g}, \widehat{g}), (\widehat{b}, \widehat{b})$ or $(\widehat{b}, \widehat{g})$. It turns out that this is a slightly overbearing way in which to exhibit the control problem, but it is also the most straightforward.

bining input H with always reporting $\widehat{g}$, stuck on $\widehat{g}$, leads to

$$\pi(x_1, \widehat{g}|H, (\widehat{g}, \widehat{g})) = \pi(x_1, g|H) + \pi(x_1, b|H) = .05 + .20 = .25$$

and

$$\pi(x_2, \widehat{g}|H, (\widehat{g}, \widehat{g})) = \pi(x_2, g|H) + \pi(x_2, b|H) = .65 + .10 = .75$$

along with $\pi(x_1, \widehat{b}|H, (\widehat{g}, \widehat{g})) = \pi(x_2, \widehat{b}|H, (\widehat{g}, \widehat{g})) = 0$.

Be certain you verify the other probability expressions in Table 17.2. They are the key to understanding the manager's behavior.

To see this, suppose the firm announces a compensation arrangement (depending on public observation of output x and the manager's self-report of $\widehat{y}$), denoted $I_{x\widehat{y}}$. Further suppose the manager then supplies input $a \in \{L, H\}$ and pursues communication strategy $m \in \{(\widehat{g}, \widehat{b}), (\widehat{g}, \widehat{g}), (\widehat{b}, \widehat{b}), (\widehat{b}, \widehat{g})\}$. His expected utility measure is then given by

$$E[U|a, m, I] = \sum_{x,\widehat{y}} U(I_{x\widehat{y}} - c_a)\pi(x, \widehat{y}|a, m) = \exp(\rho c_a) \sum_{x,\widehat{y}} U(I_{x\widehat{y}})\pi(x, \widehat{y}|a, m)$$

$$(17.1)$$

17.1.2   Incentive Compatible Resolution

Now suppose the firm seeks input H coupled with truthful self-reporting. This means the contractual arrangement must invite the desired input and the desired self-reporting behavior. To reduce this to an optimal contracting problem, we mimic the program in (14.5) to minimize the firm's expected expenditure subject to individual rationality, but now subject to a host of incentive compatibility constraints. This provides the following design program.[3]

$$\widehat{C}(H, (\widehat{g}, \widehat{b})) \quad \equiv \quad \min_{I_{x\widehat{y}}} \sum_{x,\widehat{y}} I_{x\widehat{y}}\pi(x, \widehat{y}|H, (\widehat{g}, \widehat{b})) \qquad (17.2)$$

$$\text{s.t.} \qquad E[U|H, (\widehat{g}, \widehat{b}), I] \geq -1$$
$$E[U|H, (\widehat{g}, \widehat{b}), I] \geq E[U|H, m] \text{ for all } (a, m) \neq (H, (\widehat{g}, \widehat{b}))$$

The first constraint is our usual individual rationality condition, requiring desired behavior meets the manager's market test, here normalized to an

---

[3] You may be wondering why we insist on complete, honest revelation. It turns out that in these types of optimal contract games, where the contract designer can commit to how a communication will be used (and where there is no substantive restriction on the communication technology itself), any equilibrium can be recast into an equivalent one in which honest, full communication is motivated. This is intuitive and makes the modeling easier. It also illustrates, once again, the importance of framing.

outside certainty equivalent of $M = 0$. The second constraint is a family of constraints, requiring that the desired behavior be weakly preferred to any other combination of input and reporting strategy. This onslaught of incentive compatibility requirements reflects the ongoing theme that appending more tasks, and thus options, to the manager's slate increases the dimensionality of the control problem.

Example 17.3: Let's try this with the setting in Examples 17.1 and 17.2. The optimal solution to program (17.2) is displayed in Table 17.3. For comparison purposes, the Table also displays the optimal contract when the additional measure is publicly observed (though you have to remember it is based on variable y, as opposed to variable $\widehat{y}$ in our notational setup).

TABLE 17.3: Optimal Contract for Example 17.3

	$x_1/\widehat{g}$	$x_1/\widehat{b}$	$x_2/\widehat{g}$	$x_2/\widehat{b}$
$I^*_{x\widehat{y}}$	-2,443.24	-984.64	5,356.71	2,763.35
I^*_{xy}	-2,594.45	3,055.79	3,688.67	2,334.20
I^*_x	-1,593.84	-1,593.84	5,161.26	5,161.26

$$\widehat{C}(H, (\widehat{g}, \widehat{b})) = 3,439.11 \ (RP = 439.11)$$
$$\widehat{C}(H) = 3,112.49 \ (RP = 112.49)$$
$$C(H) = 3,472.48 \ (RP = 472.48)$$

Relative to the public case, the information is used less aggressively. This is reflected in the fact the largest payment occurs in the public case, as does the lowest payment. The information is used in a mooted fashion in order to maintain candor in the communication. This also shows up in the noted risk premium. Here we are better off relative to having no additional information, but not as well off were this additional measure public.

Table 17.4 reports the manager's (net) certainty equivalent for each of the 8 strategies displayed in Table 17.2. The last two (supply L and report via stuck on $\widehat{b}$ or perversely) are, in fact, the binding constraints in the design program. These are the control hot spots. We continue to have concern for the manager's choice of input, but also now have concern for his candor.[4]

Some benchmarking is in order. We are contrasting two cases: one where all performance evaluation information is public and another where some of it is privately observed and self-reported by the manager. As noted in laying out the self-reporting design program in (17.2), the self-reporting program consists of the public information design program coupled with a number of additional (incentive compatibility) constraints. This carries two important

---

[4]This suggests additional public information that speaks either to the manager's input or candor would be valuable in the contracting arrangement.

implications. First, self-reporting cannot improve on the public information case. In general, preserving the integrity of the self-report forces us to use the self-reported information less aggressively than its public counterpart. This is key to maintaining its integrity, just as restricted recognition rules are key to maintaining the integrity of the accounting library.

TABLE 17.4: Certainty Equivalents for Example 17.1

manager's policy	CE under $I^*_{x\widehat{y}}$	CE under I^*_x
H and $(\widehat{g}, \widehat{b})$	0.00	0.00
H and $(\widehat{g}, \widehat{g})$	-231.21	0.00
H and $(\widehat{b}, \widehat{b})$	-1,313.30	0.00
H and $(\widehat{b}, \widehat{g})$	-1,516.35	0.00
L and $(\widehat{g}, \widehat{b})$	-670.16	0.00
L and $(\widehat{g}, \widehat{g})$	-670.16	0.00
L and $(\widehat{b}, \widehat{b})$	0.00	0.00
L and $(\widehat{b}, \widehat{g})$	0.00	0.00

Second, the firm always has the option of not listening to the self-report (given it can so commit). To illustrate, Table 17.4 also displays the manager's (net) certainty equivalent if the contracting arrangement ignores his communication, and thus reverts to contracting based on output alone. The I^*_x solution is always feasible in the private information case, given the information arrives after the manager supplies his input. This implies listening is never deleterious in this setting.[5]

The deeper question, then, is when does listening to the manager actually improve the contracting arrangement, when, so to speak, does self-appraisal improve the arrangement? Example 17.3 suggests optimism, while the following example is less encouraging.

Example 17.4: Return to Example 16.1, as summarized in Table 16.1. Presuming the additional measure is privately observed and self-reported (again, after the manager acts but before output is observed), we find no possible use for the self-report. Details are summarized in Table 17.5. The difficulty here is a lack of information. In the self-reporting case, output is used to infer the manager's input and to infer his veracity. But in this example, high output (x_2) only occurs under high input; and output

---

[5]Remember, this is information that is confined to our managerial input model. If we expand the story so that, for example, any public information or even revelation of private information is observed by a competitor, we may be harmed and thus strictly prefer the information be private and not revealed.

thus turns out to not be sufficiently informative to maintain veracity and nontrivial use of the self-report.[6]

TABLE 17.5: Optimal Contract for Example 17.4

	$x_1/\widehat{g}$	$x_1/\widehat{b}$	$x_2/\widehat{g}$	$x_2/\widehat{b}$
$I^*_{x\widehat{g}}$	0	0	7,305.66	7,305.66
I^*_{xy}	3,590.23	-727.02	4,002.35	4,002.35
I^*_x	0	0	7,305.66	7,305.66

$$\widehat{C}(H,(\widehat{g},\widehat{b})) = 3,652.83 \ (RP = 652.83)$$
$$\widehat{C}(H) = \widehat{C}(H) = 3,148.70 \ (RP = 148.70)$$
$$C(H) = 3,652.83 \ (RP = 652.83)$$

Naturally, the manager possessing potentially useful information invites a search for other devices, such as another injection of related public information, into the stew. The larger picture, however, should not be missed: self-reporting calls for careful attention to self-reporting motives.   This should not be interpreted as suggesting such a low opinion of human behavior that honesty in communication must be motivated at each and every turn. Rather, the idea is that putting too much pressure on communication is likely to cause the communication's quality to decline. How do we capture this in our stylized managerial input model? The easiest way is to stay with preferences defined over wealth, and address directly the question of motivating communication. This keeps the clutter to a minimum, and allows us to address the basic point.

An analogy with budgeting is insightful. Suppose center and a division manager are negotiating a budget. The division manager has superior information. If times are likely to be good and the manager knows this, admitting it is tantamount to receiving a budget with increased performance requirements. Center must be more accommodating if it wants to encourage the manager to reveal the information. The implicit cost of motivating the manager to reveal private information is a commitment to less than aggressive use of that information.[7]

---

[6]A similar result obtains in the linear contract, multitask setting of Chapter 15. With the linear contract we remove any ability to substantively discipline the manager's self-reporting by juxtaposing that self-report with a public observable.

[7]For example, if the manager is always treated in an abusive fashion when bad news is conveyed, bad news, when present, will not be communicated in a timely fashion. If center always raises the quota every time it is met, the stage is set for underachieving the quota. The cost center manager who negotiates productivity goals with the division manager and then finds that manager more and more insistent on continued improvements will have a natural reluctance to agree to significant improvement goals. If the governor offers amnesty to tax deadbeats, while the attorney general announces a policy of aggressive prosecution of all who come forward, the amnesty program will have few

It should also be noted that, more broadly, acquisition of this information in the first place may be a plus or a minus. Accurate weather forecasting is useful across society. On the other hand, it is important that the bank manager but not the public know the combination to the vault. Similarly, research and development often lead to advantages in the product market. Equally obvious is the fact life insurance cannot be purchased retroactively.[8]

17.2   Variations on a Theme

This theme of designing communication incentives extends well beyond our setting of a manager who acquires private information after supplying input, but prior to public observation of output. The manager's private information might be in place before the parties contract. For example, the consultant arrives with considerable industry expertise. The information might arrive after output is observed, or it might arrive after the contract is struck but before input is supplied. For that matter, the firm rather than the manager may be the one with private information.

17.2.1   Late Arrival of Private Information

Suppose the manager's private information arrives late in the game, or he is simply unable to communicate this observation until after the output itself is observed. This delay is unimportant if the information eventually becomes public, as it presumably would be publicly observed in (the nick of) time to guide the payment of pay-for-performance incentives. In the private information case, though, this reporting delay is fatal. The information cannot be used at that point.

To see this, suppose $y = g$ is privately observed by the manager and, before declaring this fact, the manager sees high output (x_2). All that is at stake at this point is the manager's pay; the personal cost is sunk. So we must have $I_{x_2 \widehat{g}} \geq I_{x_2 \widehat{b}}$; otherwise $\widehat{b}$ will be declared. Conversely, suppose our friend observed $y = b$ along with high output. Motivating accurate

---

takers. If the manager who offers a new product idea is reminded constantly that future promotion depends on the success of the product, the firm will find a shrinking supply of new product ideas. If the partner in charge of the audit engagement downgrades the audit team manager's performance whenever the audit is over budget, the audit team is encouraged to underreport overtime or to lower the quality of its audit efforts.

[8]For that matter, the CPA exam is proctored. Periodic financial reports are audited. Tax filings are randomly audited. Hospitals must submit quality control records to an accreditation agency. Fraudulent reporting stories have surfaced in all of these arenas. And on the other side of the fence, the Fifth Amendment to the U.S. Constitution states "...nor shall any person...be compelled in any criminal case to be a witness against himself...."

reporting now requires $I_{x_2\widehat{b}} \geq I_{x_2\widehat{g}}$. Together, the inequalities imply we must have $I_{x_2\widehat{g}} = I_{x_2\widehat{b}}$. A parallel argument implies $I_{x_1\widehat{g}} = I_{x_1\widehat{b}}$. The self-report cannot be used. Our slight change in timing reversed the sequence of self-report followed by public observation of output. This is the heart of the problem.

Remember that output plays two roles here, as a source of value and as a source of information about the manager's behavior. In the latter capacity it is potentially informative about the manager's input supply and about his candor. Reversing the sequence destroys its ability to address the self-reporting control problem. The output is no longer available to discipline the reporting behavior. This points out the importance of timing in these types of encounters. More significant is it reminds us self-reporting incentives must be designed into the exercise.

17.2.2   Two-Sided Opportunistic Behavior

Of course, private information is not simply the province of the manager; it may well reside with the other party to the contract. The manager's supervisor, for example, might privately gather impressions of the manager's skill and dedication. Can we trust the supervisor to be fair and thorough in developing and reporting this performance appraisal? It is no accident we find formal grievance procedures in place in many such circumstances. Similarly, we are in possession of considerable private information when confronted by the tax auditor; and the tax auditor knows a great deal more than we do about reporting patterns that have surfaced in compliance audits.

To reinforce this observation, consider what happens in our setting when the firm, instead of the manager, privately observes information variable y and subsequently proffers a claim that $\widehat{y}$ was observed. This transpires, again, after the manager has acted but before the output is observed. Now we must worry about the incentives of both parties to the trade arrangement; and output's information sphere extends from potentially informing about the manager's action choice to the firm's reporting choice as well.[9]

This leads to a nested control environment and more complicated equilibrium structure in the evaluation game. The payment arrangement will now be designed so the manager will find it optimal to supply input H in

---

[9] The firm now labors under a pair of incentive compatibility conditions. Having observed $y = g$, factual self-reporting by the firm requires

$$\pi(x_1|g,H)I_{x_1\widehat{g}} + \pi(x_2|g,H)I_{x_2\widehat{g}} \leq \pi(x_1|g,H)I_{x_1\widehat{b}} + \pi(x_2|g,H)I_{x_2\widehat{b}}$$

Likewise, factual self-reporting of $y = b$ requires

$$\pi(x_1|b,H)I_{x_1\widehat{b}} + \pi(x_2|b,H)I_{x_2\widehat{b}} \leq \pi(x_1|b,H)I_{x_1\widehat{g}} + \pi(x_2|b,H)I_{x_2\widehat{g}}$$

Notice how the subsequently arriving output observation is used to discipline the firm's self-reporting.

anticipation the firm will factually report its private observation.. And the firm will find it optimal to factually report its observation in anticipation the manager will supply input H.

These increased control concerns may, or may not, have a major effect on the parties ability to use the firm's self-report to advantage. Putting the private information in Example 17.3 in the hands of the firm is paralyzing, in that the only way to maintain the firm's candor is to structure the contract to not use its self-report. In the forthcoming example, however, it turns out just the opposite is true. The optimal contractual arrangement when the information is public is also optimal when the firm privately observes and self-reports that information.

17.2.3 Early Arrival

Another possibility is for the private information to arrive before the manager acts. This further increases the control problem, because we must now also worry about act selection for each possible private information event. This expands the incentive compatibility concerns, for sure. But this information acquisition is also potentially useful because the manager's act can now be fine-tuned to the emerging environment. Our running example is illustrative.

Example 17.5: Consider the setting in Table 17.6. Set $c_H = 3,000$ and the manager's risk aversion measure at $\rho = .0001$ (along with the usual normalization of $c_L = M = 0$). The information arrives just before the manager acts. Notice we have $\pi(g) = \pi(b) = .50$. Moreover, under $y = b$ inputs L and H are equally productive, while input H is more productive under $y = g$. This suggests an input policy of H in the good (g) environment and L in the bad (b). Absent any contracting frictions, this would lead to an expected cost to the firm of $.50c_H + .50c_L = 1,500$.

Now suppose this critical information is public. The optimal contract, denoted I_{xy}^* in Table 17.6, leads to a risk premium of 28.42 (and overall expected cost to the firm of 1,528.42).[10] Notice incentives are applied only in the good environment, where input H is desired. Contrast this with the case where the information is available (to guide the input choice), but is not used in the payment arrangement. This is contract I_x^* which, due to its information diet, needlessly maintains strong incentives in the bad environment.

Conversely, suppose the manager privately observes the information. Not asking him to self-report results in the noted I_x^* setting. Contract $I_{x\hat{y}}^*$ (where we interpret the public information as the self-reported variable)

---

[10]This contract is located by minimizing the expected payment, presuming H under g and L under b, subject to overall individual rationality, H beats L under g and the reverse under b from the manager's perspective. Try it!

implements the desired input and candid self-reporting by the manager. Relative to the public information story we again see less aggressive use of the information. We even see incentives are active in the bad environment, as the privately informed manager must be deterred from opportunistic reporting.[11]

Finally, and in sharp contrast, placing the private information in the hands of the firm returns us to the public information setting. You can readily verify that if the firm self-reports its private observation, in time for the manager to act thereon, it will gladly self-report with candor when the I^*_{xy} contract is in place.

TABLE 17.6: Details for Example 17.5

	x_1/g	x_1/b	x_2/g	x_2/b
$\pi(x,y\|H)$	.05	.25	.45	.25
$\pi(x,y\|L)$	.50	.25	0	.25
$I^*_x\ (RP=130.36)$	-747.18	-747.18	2649.31	2649.31
$I^*_{xy}\ (RP=28.42)$	0	0	3,396.49	0
$I^*_{x\widehat{y}}\ (RP=90.64)$	-1,869.55	-366.02	3,286.91	1,186.05

17.2.4   Counterproductive Information

This set of illustrations also offers an opportunity to explore our earlier, cryptic observation that information can be counterproductive.[12]   Table 17.7 is such a case. Suppose input H is desired regardless of whether $y=g$ or $y=b$ obtains. If variable y is publicly observed after the manager acts, surreptitious supply of L is readily detected (in the $y=b$ environment). Designing the usual penalty contract will then provide, in equilibrium, supply of input H in exchange for the first best wage. However, if this information becomes available before the manager acts, either privately to the manager or publicly, it allows the manager to know beforehand whether the strong control system is working, and this sabotages our clever penalty contract. Not good.

TABLE 17.7: Counterproductive Information

	x_1/g	x_1/b	x_2/g	x_2/b
$\pi(x,y\|H)$	.05	.0	.45	.50
$\pi(x,y\|L)$	.25	.25	.25	.25

---

[11] We should be careful here. The only distortion identified is the amount of risk placed on the manager as a consequence of the best pay-for-performance arrangement. In a richer setting, we would expect the production plan to be altered as well.

[12] The haggling story in Chapter 10 where the seller has private information is a case in point.

17.3   Intertemporal Considerations

An important, wide-spread version of the private information theme occurs when we have multiple periods and face the issue of recognizing particular events or transactions in the proper time frame. Often, it seems, revelation of bad news is delayed in the hope a little good luck will follow and allow the self-reporter to bundle the revelation with some good news. So-called income smoothing or earnings management is the quintessential example. The key ingredients for such a story, based on equilibrium behavior, are private information and multiple reporting opportunities.

To bring this under our self-reporting umbrella, suppose the firm's output can be $x = 1$, 2 or 3 units. (Any number, providing we have at least 3, will do the trick, so we opt for the minimalist version.) As usual, the manager's personally costly input can be L or H, with the latter desired by the firm. The new feature is this gets repeated. A contract is agreed upon, the manager acts, first period output is observed, first period payment is made, the manager acts again, second period output is observed, and second period compensation is made. Let's assume the interest rate is zero, to avoid even more clutter. Also assume output in each period depends only on that period's input, and is governed by the same probability mass, $\pi(x|a)$. So if x_t denotes output in period t and a_t the corresponding input in period t, we are assuming

$$\pi(x_1, x_2|a_1, a_2) = \pi(x_1|a_1)\pi(x_2|a_2)$$

This is an important assumption in what follows. Thanks to constant risk aversion, it turns out the two-period contract is simply the corresponding one-period contract, repeated twice. In other words, a stationary incentive contract provides the best arrangement for arranging for the manager's services over the two period horizon.

Example 17.6: Assume the probability structure displayed in Table 17.8. Set the risk aversion measure at $\rho = .0001$, along with $c_H = 5{,}000$ and (normalized) $c_L = M = 0$. The optimal pay-for-performance arrangement is also displayed in the Table. Importantly, this payment arrangement is now in place for each of the two periods. In addition, we have significant decreasing returns to good news, as compensation increases much more as we move from $x = 1$ to $x = 2$ units, than from $x = 2$ to $x = 3$ units.

TABLE 17.8: One-Period Version of Example 17.6

	$x = 1$	$x = 2$	$x = 3$
$\pi(x|H)$	.10	.20	.70
$\pi(x|L)$	.70	.20	.10
I_x^*	-1,729.79	5,576.94	6,291.24

To introduce concern for intertemporal self-reporting issues, now suppose the manager is able to smooth the output (and thus earnings) series. In particular, assume a sequence of 1 unit in the first period followed by 3 in the second ($x_1 = 1$ and $x_2 = 3$) can be reported as 2 units each period ($x_1 = x_2 = 2$). The same holds for 3 units in the first period followed by 1 in the second. Presumably, the manager has sufficiently intimate detail of various events that he is able to restructure their appearance in this manner.[13]

Unfortunately, this restructuring is seriously tempting as 1 unit in the first period followed by 3 in the second will net total compensation of $-1,729.79 + 6,291.24 = 4,561.45$ which is considerably less than the total compensation of 2(5,576.94) if 2 units are reported each period. Our nice, stationary incentive structure, it seems, is vulnerable to earnings management temptations.

As a passing aside, we now face the choice of tolerating earnings management or of altering the incentive structure to remove such temptation (or producing information sufficient to detect such behavior). However, all of these options lead to more costly transactions relative to the case where no such intertemporal opportunism is possible.

That said, also notice that tolerating earnings management keeps the expected output constant each period, but reduces its (statistical) variance.[14] This is the key to the next observation.

Suppose, instead, it is possible for the manager to engage in undetected earnings management only if he supplies input H each period. The idea is careful attention to the tasks at hand brings with it, as a by-product, the ability to surreptitiously manage earnings in the noted manner. If so, a "smooth" earnings series is now good news, as it is consistent with high input in each period. In this way, a smooth series serves as a signal of good behavior. This allows us to construct a more efficient contract than the one in Table 17.8.[15]

Intertemporal considerations, then, are yet another dimension of communication exercises. To the extent the firm's management team can affect the time at which, say, transactions are recorded, we are dealing with a form of communication. How this particular timing option arises has a great

---

[13] This is a heavy-handed assumption, but it allows us to sidestep introducing an elaborate information structure that allows the manager to gauge existing and forthcoming output or demand.

[14] With the noted probability structure and input H each period, the expected output is 2.6 units period period. Absent earnings management, the per period variance is .44, while with earnings management it is .30.

[15] In such a case, 1 followed by 3 units, or vice versa, is a sure sign of opportunistic behavior. The only way to avoid this in the first period is to supply input H. If first period output is 1 or 3 units, the manager again can avoid this sign only by supplying input H in the second period. The only event in which risky compensation is introduced is when first period output is 2 units.

deal to say about whether it is friction reducing or friction increasing. A "smoothed" earnings series may, that is, be a signal of unusual talent and effort in managing the firm, or it may be a signal of unwanted opportunism.

17.4   The Larger Picture

Regardless, the general theme remains. Self-reporting raises the question of motivating that self-reporting. Introducing self-reporting options into the relationship calls for an expansion of the web of controls to address incentives that are lurking in any such encounter. As we have warned repeatedly, increasing the manager's tasks is unlikely to lessen the control problem!

Naturally, increased control concerns may lead to the introduction of other sources of information. A consultant might provide a second opinion. Other managers might be solicited. For example, the audit committee of the board of directors may communicate directly with the internal audit staff, as well as top management. This suggests a type of reporting tournament between lower and upper management. The manager's reporting history is also likely to be important. For example, the manager whose forecasts are always confirmed, to the penny, by subsequent accounting reports will be suspected of gaming, as will the manager whose forecasts are always well below actual results.

These incentives might also be influenced by the way the firm uses the information communicated. As a rule, we expect less aggressive use of communicated as opposed to publicly observed information. Firm reputation also plays a role. The firm that has a long history and culture of encouraging participation has invested in stable self-reporting incentives.

Moving further beyond our stylized model, we also expect contracts to be incompletely specified. Details will be filled in later, or renegotiated, as circumstances warrant. Contracting is a costly exercise itself; and unforeseen circumstances are a possibility. The proverbial "whistle blower" is a case in point. Bringing bad news forward leaves the whistle blower vulnerable to retribution.

Finally, these communication incentives may be influenced by nonpecuniary factors. Communication may well be a forum for recognition, for exhibiting skills (or lack thereof), or for building group cohesiveness. A well-maintained participation ethos may lead to commitment to and personal identification with the firm's goals, a commitment and identification that are indispensable in firm success.

From the micro details of our stylized manager reporting a good or bad environment to the sweeping picture of management style, the underlying message is consistent. Communication begets concern over incentives to play the communication game.

17.5   Summary

Communication is a natural extension of the managerial input story. If communication is to be engaged, the firm's control problem expands to accommodate the managers' incentives to play the enlarged game. Concerns of this nature are all around us. The judicial system relies on its reputation to convince the state's witness that its immunity offer will be honored. We don't ask the students to grade and self-report their final examinations. We do ask the cost center manager for a cost forecast, and then evaluate the manager based on actual cost and that forecast.

Communication incentives are provided by the firm's culture and by the way the firm manages the communication encounters. Aggressive use is not conducive to full, timely, and accurate revelation. This, of course, surfaces in our managerial input model, where we saw self-reported information used in a less aggressive fashion than would be the case were that information publicly observed. It is also important to understand that this is a two-way street. The firm itself may be prone to opportunistic behavior, as when it faces the temptation of selectively honoring commitments or manipulating information flows to its work force.

These concerns also extend to the accounting library. (This is evident in our discussion of earnings management.) The firm's internal control system uses a variety of devices, including separation of duties, to maintain the integrity of the financial records. Separation of duties introduces a reporting tournament, a type of relative performance evaluation. Similarly, auditing introduces an independent check. We are also careful not to put too much stress on the accounting numbers.

The consistent theme is that controls and managerial activity are coextensive. The firm's controls must be as expansive as the managerial activity that is contemplated. The firm provides the environment in which the work force labors. Call it the web of controls, the environment, or what have you. The theme of controls that extend to the entire array of managerial activity is the central point.[16]

17.6   Bibliographic Notes

A modeling trick, a frame, used in our study of communication is to motivate honest revelation. Under fairly broad conditions (See note 3.) this is

---

[16] Our theme of providing requisite incentives for communication has been focused by our use of the managerial input model. Stepping back, we might think of the various managers as experts that are called upon to offer a prediction. Here we confront human cognition and the fact simple linear models often outperform expert predictions in cases where the outcomes (e.g., bankruptcy or pathology) can be confirmed. This adds another layer to the communication story.

done without loss of generality. Any other equilibrium behavior in these cases can be converted to equilibrium behavior in which candid, full communication is motivated. This motivation, in turn, is ensured by a commitment to "underutilize" the communication. Myerson [1979] is an important reference for this "revelation principle." In this respect, the focus on candid, full communication is yet another illustration of framing. Christensen [1982] studies participation incentives and returns to information in the contracting model. Of course, with the trick of motivating full communication it is always possible to guarantee the information is communicated simply by committing to ignore it. This raises the question of when it makes sense to listen to the informed party in the first place. Dye [1983], among others, examines this question. Demski and Sappington [1991] study double moral hazard versions of the communication story. Nonpecuniary returns to participation are examined in a variety of places, including Becker and Green [1962]. Hofstede [1967], Hopwood [1972], Merchant [1989], and Swieringa and Moncur [1975] provide field study evidence on, among other things, communication and budget participation. More or less voluntary disclosure is another form of communication; Dye [2001] and Verrecchia [2001] provide extensive reviews. Demski and Sappington [1987] and Lambert [1986] focus on motivating an agent to acquire the private information in the first place. Fellingham, Newman and Suh [1985] is the reference for the stationary contract, while Demski [1998] is the inspiration for the earnings management story. Arya, Glover and Sunder [1998] connect earnings management and the above noted revelation principle; and Ronen and Yaari [2008] provide a comprehensive look at the subject.

17.7   Problems and Exercises

1. Our study of communication stresses the theme that a control system must be coextensive with the control problem it is designed to address. If a manager is called upon to supply input and to communicate, for example, the control system must deal with both input supply and communication incentives. Carefully explain this theme.

2. The stylized contracting model accommodates the idea that a well-informed player might be induced to communicate what is privately known. This requires we pay attention to incentives. After all, it would be naive to expect the player to freely give away any private-information-based advantage. In turn, revelation incentives take the form of a commitment to "underutilize" the communicated information. Explain this general principle of nurturing communication incentives by less than aggressive use of what is communicated.

3. *control hot spots*
   Return to the self-reporting setting of Example 17.3. What are the
   shadow prices on the constraints in program (17.2) in this setting?
   How do the shadow prices relate to the control hot spots in this
   setting?

4. *information arrives after manager acts*
   Ralph is contracting with the usual (overly familiar) manager, in a
   setting where the manager's input can be high (H) or low (L), with H
   desired. While Ralph is risk neutral, the manager exhibits constant
   risk aversion ($\rho = .0001$), a high input cost of $c_H = 5,000$ and the
   usual normalizations of $c_L = M = 0$. The only contracting variables
   are output (x_1 or x_2) and a monitor (g or b). The monitor report is
   observed after the manager acts, but before the output is observed.
   The probabilities follow.

	x_1/g	x_1/b	x_2/g	x_2/b
$\pi(x,y|H)$	.05	.25	.45	.25
$\pi(x,y|L)$	.45	.25	.05	.25

   (a) Find an optimal pay-for-performance arrangement that will im-
       plement the H input when only output can be used for contract-
       ing purposes.

   (b) Repeat (a) for the case where the monitor is publicly observed
       and both the monitor and output can be used for contracting
       purposes.

   (c) Repeat (b) for the case where the manager privately observes
       the monitor and communicates this observation; so output and
       the agent's claim as to what the monitor is reporting can be
       used for contracting purposes.

   (d) Carefully contrast your three solutions above.

   (e) What happens in part (c) above if the manager's communication
       is delayed until after the output is observed?

5. *information arrives before manager acts*[17]
   Repeat your analyses in problem 4 above, but now under the assump-
   tion the manager privately observes the good (g) or bad (b) news after
   contracting but before acting. Assume Ralph desires input H in the
   good news case and input L in the bad news case.

6. *two-sided opportunism*
   Return to the setting of Example 17.3. Now assume the firm, rather

---

[17] Suggested by Richard Sansing.

than the manager, privately observes $y \in \{g, b\}$ after the manager acts but before output is observed. Determine an optimal contract that will simultaneously motivate the manager to supply input H and the firm to self-report its observation. Also write a short paragraph describing the equilibrium specification in the various incentive compatibility constraints.

7. *two-sided opportunism*
   What happens in Example 17.4 if the firm rather than the manager possesses the private information?

8. *communication and input supply incentives*
   Return to Example 17.5. Draw the manager's decision tree for the first two cases and verify he can do no better than accept the offered terms, supply H in the good environment, and supply L in the bad environment. Also determine his certainty equivalent in each case, at the point he has observed the information and is about to make his input choice. Why is the manager's compensation independent of output in the bad environment in the case where public information is used?

   Now draw the manager's decision tree for the third case. Verify he can do no better than accept the offered terms, supply H and reveal the good environment if that environment is observed, and supply L and reveal the bad environment if that environment is observed. Determine his certainty equivalent at the point he has observed the information and is about to make his input and reporting choice. Why is the manager's compensation at risk in the bad environment?

9. *valuable private information*
   Return to Example 17.5. Suppose no information is available, and the firm desires supply of input H. Determine an optimal pay-for-performance arrangement, and contrast it with the case where the information is privately obtained by the manager but not communicated. How much would the firm pay for the manager to observe the environment before acting? Does this amount depend on whether communication is feasible? Why?

10. *private information with negative value*
    Return to Table 17.7, and assume the manager is as specified in problem 4 above. Input H is desired, regardless of any information. Initially suppose no information is available, either publicly or privately. Determine and interpret an optimal pay-for-performance arrangement. (Assume the manager can post a large performance bond.) Next, suppose the g or b environment is privately revealed to the manager before acting; this revelation cannot be communicated to the firm. Determine and interpret an optimal pay-for-performance

arrangement. How much would the firm pay to keep the manager from observing this information? Does the manager benefit from having the private information? Why?

11. *information management*
   In problems 9 and 10 above you have encountered numerical examples where private information in the hands of a manager is or is not in the best interests of the firm. Sketch two corresponding institutional settings, one where the firm wants the manager to be informed and one where it does not. Is it possible a firm might want to exclude some information from the accounting library for strategic or control purposes?

12. *hidden reserves*
   A familiar contention when a new management team takes over is the suspicion that various expenses associated with the outgoing team have been aggressively identified, thereby creating some hidden "reserves" for the new team. How does this relate to our general theme of motivating communication?

13. *communication from subcontractor*
   Ralph is trying to finish a rush job for a favored customer. The schedule is tight and Ralph can save 8,000 in overtime cost if part of the job is turned over to a local subcontractor. The subcontractor's cost is either 4,000 or 6,000. The subcontractor knows its cost, but Ralph is uninformed. Let α be Ralph's probability the subcontractor's cost is low (i.e., 4,000). Ralph is risk neutral; time is critical and Ralph must make a take-it-or-leave-it offer to the subcontractor.

   (a) Suppose $\alpha = 0$, so the subcontractor is a high-cost type and Ralph knows it; what should Ralph do?

   (b) Suppose $\alpha = 1$, so the subcontractor is a low-cost type and Ralph knows it; what should Ralph do?

   (c) Determine Ralph's optimal strategy for all possible values of α. Why does Ralph forego trade with the subcontractor on occasion, even though it is common knowledge such trade would be mutually beneficial?

   (d) The strategy you determined in (c) above can be interpreted as one where Ralph designs a contract in which trade will take place at known terms, depending on what the subcontractor claims the cost is; and the subcontractor is motivated to candidly reveal that cost. Provide such an interpretation. Why does Ralph commit to "underutilize" the subcontractor's revelation?

14. *evaluation dynamics*
   An apparel manufacturer centrally plans production schedules and

treats each manufacturing facility as a cost center. Well-engineered standards are in place for each facility, and the major evaluation measure is cost incurred relative to budgeted cost given the output achieved. Output quotas are closely monitored. Depending on market conditions, the production plan will be revised on a monthly basis. It also turns out that a facility that has exceeded its output quota can expect a more ambitious quota whenever the schedule is revised. An internal review of operations has discovered the production managers routinely hold back some output whenever they exceed their quota. The resulting secret safety stock is then used to cushion the inevitable shortfall when the quota is not met. In response, the review team has recommended the manufacturing facilities be upgraded to profit centers. This would, they argue, elevate the prestige of the production managers, make them more conscious of the larger goal of profitability, and better align their local interests with those of center. Evaluate the review team's suggestion.

15. *accounting library*

The accounting library, as we have stressed, is well defended. Revenue recognition is an important policy instrument in this regard. We delay recognition of revenue, and hence income, until the earnings cycle is largely complete. That said, suppose the manager has private information about the firm's customer base. The accounting library would not admit this information on a timely basis, preferring instead to honor the revenue recognition rule. Explain this, especially given the theme in the chapter of underutilizing private communication as the implicit price for ensuring its integrity.

16. *communication in root utility case*

Ralph's manager acquires information after acting, but before output is realized. As usual, Ralph, is risk neutral. The manager has preferences for cash income z and labor input a given by $\sqrt{z} - V(a)$. Two labor inputs are possible, H or L. Ralph seeks supply of H. Conflict is present, as $V(H) = 20 > V(L) = 0$. Also, the manager demands an expected utility of 40 to sign on with Ralph. It turns out that weather plays an important role in the production process. Suppose the weather can be *dry*, *regular*, or *wet* with equal probability. The output possibilities (interpreted as cash before any payments to the manager) are as follows:

	dry	regular	wet
input H	11,000	11,000	5,000
input L	11,000	5,000	5,000

(a) Suppose the manager acts in a self-interested manner and that only the output can be contracted on. Determine an optimal pay-for-performance arrangement.

(b) Now suppose a monitor is available. This monitor will report good news if the weather is not *wet* and bad news if the weather is *wet*. The monitor's report will be publicly observed at the end of the game. Determine an optimal pay-for-performance arrangement.

(c) Next, suppose the monitor will be privately observed by the manager, after the manager acts but before the output is observed. The manager can now tell Ralph what was observed, and the contract can depend on the claimed observation as well as the publicly observed output. Determine an optimal pay-for-performance arrangement.

(d) What roles are played by output in part (c) above?

18

Coordination

The final stop in our growing list of tasks assigned to various managers is the seemingly obvious task of coordination. On one level this is the task of making certain all the details of the firm's mission come together in harmonious fashion. The cross-country flight relies on a well-maintained aircraft, proper fuel and flight plan, and the services of many air traffic controllers. Arrival at the destination airport presumes a waiting gate and prepared ground personnel. By the same token, it does little good to release new product advertising when the distribution channels are empty. The assembly line would exhibit gridlock without well-executed arrivals of component parts and skilled labor. Shopping for the dinner party is made much easier by knowing the recipes for the dishes that will be prepared and served. Traffic lights serve a useful function. Coordination is a well-practiced, vital art.

On another level, coordination concerns the incentives to provide the variety of pieces that come together in harmonious combination. It does little good to design and advertise a product of exceptional quality, and then saddle the manufacturing arm with stringent production quotas. It is also counterproductive to stress a long-run view while emphasizing short-run incentives.

Our exploration begins with a brief look at aggregate budgeting. This provides an opportunity to remind ourselves of the importance of financial coordination, complete with attention to the underlying communication incentives. Next we reprise our earlier theme of balancing short-run and long-run tasks, a type of intra-manager coordination. From there we turn to a divisionalized setting with trade between the divisions, a type of inter-

manager coordination. And just to remind ourselves it is possible to have too much of a good thing, we conclude with a look at coordinated sabotage of a control system. The tension between coordination that serves the firm's interests with coordination that serves the individuals' interests is ever present.

18.1   Master Budgets

We have casually and intuitively used the term "budget" throughout our study. In most general terms a budget is a projected set of consequences.[1] Given the environment and given our plan of action, we project sales will total 14,500 units during the coming (fiscal) year; we project the pipeline will be 75% complete by the end of the quarter; we project our personal finances will be under control in four months. A master budget is an all-inclusive budget that brings all the firm's activities into a single picture. It is the firm's most inclusive projection of the consequences of its activities.

18.1.1   Aggregation into a Global View

We often think of the master budget as a projection of what the firm's financial statements will look like at the end of the period in question. These are called *pro forma* financial statements. What will the ending balance sheet look like, given the various production and sales activities we anticipate, given the capital investment and financial transactions we anticipate, and so on? What will the income statement look like? What about the cash flow statement?

This forces a global look at the firm and provides a reference point for interpreting the financial statements at the end of the period. It also provides the foundation for working capital management.

18.1.2   Disaggregation into a Sea of Coordinated Details

Another side of the master budget is the underlying details that have been aggregated into the pro forma statements. These details are important. They speak to the coordination that is essential for the firm to move forward with minimal friction.

Details at this point are overwhelming, as they should be. Imagine a sizeable firm with sales in the millions, covering a variety of products.

---

[1]We speak of a projection here as though it were a single number or specific event. This is common practice. Do not assume, however, budgets are never prepared in probabilistic format. What are the odds our revenue projection will be exceeded? What product warranty statistics do we anticipate and how much risk do we face in this regard? What are the odds our competitor's diversification strategy will fail?

These projections are broken down by product type and subperiod, say, by quarter. They are meshed with tentative production schedules that reflect existing inventory, production capacity, and desired ending inventory. In turn, the production schedules are further broken down into schedules for the various factors of production. Work force schedules and adjustments are recognized. For example, hiring and training plans may be called for, acquisition of various materials must be arranged and scheduled. New material handling devices may be called for, requiring design and testing before bids are solicited. The myriad factors we combine into an overhead pool must be thought through and coordinated with the tentative production plan. This eventually provides the overhead *LLA*s. A parallel pattern emerges in the marketing and administrative areas. These details combine to provide an operations budget for the firm.

Investment activities are also part of the stew. Here tentative plans for various investments, say, equipment replacement and expansion (or divestiture) are detailed. These details combine to provide the capital budget.

Finally, we have the cash budget. These various activities call for an enormous number of transactions between the firm and external entities. Payrolls must be met, deposits covering withholding must be made with the appropriate agencies, suppliers must be paid, customer payments must be monitored, and so on. This does not happen by accident. Short-term (or long-term) financing may be necessary. Short-term investment opportunities may be available. Detailed, micro management of the firm's working capital is an essential financial service.

Coordination relies on an enormous array of carefully meshed details.

18.1.3   Authorization and Communication

The master budget enterprise is also an authorization and communication vehicle. Just as the annual teaching schedule is an important communication to the faculty, product development and promotion plans are an important communication to the consumer product company's sales force. The discipline of the master budget has the virtue of bringing these plans into common view. Likewise, assembling the countless details relies on (candid) communication from various parties, as the underlying information is widely dispersed in the firm.[2]

Authorization is also part of the story. Suppose a 10% increase in the work force is contemplated. The human resources group requires instruction or authorization to proceed with the search and hiring. The master budget exercise is such a vehicle. Similarly, a research and development group may

---

[2]This should not be interpreted as an endorsement for complete communication of all plans. Communication is not cost free; and strategic concerns are present. For example, you may not want your competitors to know of your product development plans.

operate in largely decentralized fashion, with little explicit direction. Here control is channeled through the authorization process, in effect authorizing an expenditure ceiling.[3]

A governmental entity heavily relies on the master budget as an authorization vehicle. The typical municipal budget, for example, includes an overall spending total coupled with a detailed breakdown into line items. These line item breakdowns serve as spending authorizations. For example, the budget might include a line item totaling $2 million for equipment. In effect, spending up to $2 million in this category has been authorized.[4]

18.1.4   Ties to Responsibility Accounting

The master budget enterprise also has close ties to responsibility accounting. As explored in Chapter 16, we expect to see some of the data in the accounting library used in the evaluation of a particular manager. These accounting summarizations, in turn, are likely to be compared with a budget. The budget has its roots in the earlier master budget exercise.

The master budget enterprise is clearly serious business. Formalized, rhythmic planning is essential for coordination. It is also a costly activity. Time is devoted to the task, not to mention a variety of staff. Budget planning models (and more formalized scheduling models) are often used. Product costing models may also be used, especially in a setting where product development and redesign are frequent activities occurring throughout the budget cycle.

Of equal importance is a well managed communication infrastructure. The underlying budget will be flawed, perhaps tragically, if it has been misinformed or informationally starved.

18.2   Short-Run versus Long-Run Coordination

We now turn to a highly specific coordination theme, that of properly attending to short-run and to long-run tasks. Is the professor maintaining his human capital while, simultaneously, dealing with current students?

---

[3]Many refer to costs of this sort as discretionary fixed costs. They are discretionary in the sense central management decides on the overall level of activity, and hence cost, that will be incurred. This is the primary control point. Also, they are generally constant across contemplated output variation and thus viewed as fixed costs.

[4]At this point encumbrance accounting comes into play. Tracking expenditures in such a category on a cash basis is not very timely. It's likely to be too late when the bills arrive, since commitments in excess of the authorized spending may already be in place. Encumbrance accounting takes an extremely aggressive approach to recognition. If the supplies are ordered, the overall total of $2 million is immediately written down, or "encumbered" to reflect this commitment.

Is the CEO properly balancing short-run accounting performance while continuing to position the firm for extended growth into the future? Is the audit manager responding well to the pressures to bring the audit in under budget while simultaneously assigning tasks to the team in a way that builds their experience base? Balancing short-run and long-run tasks is an ever present coordination issue.

18.2.1   Task Balance, Again

To dig deeper into this coordination issue, we return to the multitask setting of Chapter 15 and expand on our earlier cryptic introduction of this particular tension. To set the stage, recall the manager continues to face a high versus low, H versus L, input story, but with the added feature any such input is allocated among tasks.

To give this a short-run versus long-run flavor, further assume the story now lasts for two periods. First period input can be allocated to short-run, $a_{SR} \geq 0$, and to long-run, $a_{LR} \geq 0$, tasks. Conversely, and for simplicity, second period input is allocated to a single second period task, here denoted $a_2 \geq 0$. So, if high input is supplied each period, we require $a_{SR} + a_{LR} \leq H$ and $a_2 \leq H$.

The first period performance measure reflects only first period, short-run activity and is given by

$$x_1 = a_{SR} + \varepsilon_1 \tag{18.1}$$

where ε_1 is a zero mean normal random variable with variance σ_1^2. The second period performance measure reflects the first period's long-run activity as well as the second period's activity. It is given by

$$x_2 = \theta a_{LR} + a_2 + \varepsilon_2 \tag{18.2}$$

where θ is some known parameter and ε_2 is a zero mean normal random variable with variance σ_2^2 and independent of ε_1.

θ plays the role of measuring the manager's long-run activities, albeit with delay. $\theta = 1$ indicates perfect capture of these effects (in expectation), while $\theta < 1$ indicates a downward biased assessment, typical of the accounting library where we slow down recognition to help protect the library's integrity.

The manager's compensation is "linear" in the performance variables, and across the two periods is given by

$$I = \omega + \beta_1 x_1 + \beta_2 x_2 \tag{18.3}$$

ω, then, is the two period wage, β_1 is the piece rate on first period performance and β_2 is the piece rate on second period performance.

While the firm is, again, risk neutral, the manager's preferences depend on total compensation less total personal cost. The personal costs, as in

the earlier setup, relate to input supplied but not its allocation. To keep things as uncluttered as possible, the manager's personal cost of high input in either period is $c_H > 0$, while the low input costs and outside opportunity certainty equivalent are normalized at 0. If the manager now supplies high input in each period and receives compensation totaling I, we assume his preference measure is given by

$$U(I, H, H) = -\exp(-\rho(I - 2c_H))$$

The manager, then, exhibits constant risk aversion with respect to total compensation less total personal cost. For convenience, no discounting is involved (and the same applies to the firm). Importantly, there is no inherent concern for one period versus the other, or for immediate as opposed to long term issues. Any such tension will be induced inadvertently by dealing with the input supply incentives. This inadvertent tension is at the center of our short-run versus long-run coordination issue.

To fill out the story, once a contract is agreed upon the manager makes his first period input supply choice and its subsequent allocation to short-run and long-run tasks. First period performance, x_1 in expression (18.1), is then observed. Following this, the manager makes his second period input supply choice, second period performance is observed, x_2 in expression (18.2), and total compensation is delivered. (Given no discounting, the time at which compensation is delivered is a matter of indifference, and focusing on the total in this manner keeps us focused on essentials.) The firm seeks supply of high input (H) in each period, and a 50-50 split of first period input between short-run and long-run tasks.

From here we invoke the certainty equivalent machinery, exploiting constant risk aversion and normal random variables, developed in Chapter 15. At the time of contracting, the manager's certainty equivalent, presuming supply of input H in each period and the noted 50-50 allocation, is[5]

$$\omega + \beta_1[.5H] + \beta_2[.5\theta H + H] - \frac{1}{2}\rho[\beta_1^2\sigma_1^2 + \beta_2^2\sigma_2^2] - 2c_H \qquad (18.4)$$

Likewise, at the start of the second period, having supplied (and appropriately allocated) high input in the first period and having observed x_1, we have the following certainty equivalent if high input is supplied in the second period

$$\omega + \beta_1[x_1] + \beta_2[.5\theta H + H] - \frac{1}{2}\rho[\beta_2^2\sigma_2^2] - 2c_H \qquad (18.5)$$

---

[5] Recall that with constant risk aversion and a lottery that is a normal random variable with mean μ and variance σ^2, the individual's certainty equivalent is the mean less one half the risk aversion measure multiplied by the variance, or $CE = \mu - \frac{1}{2}\rho\sigma^2$. Translating to the current setting, we have a mean, under the noted supply and allocation, of $\omega + \beta_1[.5H] + \beta_2[.5\theta H + H]$ and a variance of $[\beta_1^2\sigma_1^2 + \beta^2\sigma_2^2]$. Netting the personal costs provides the noted expression.

as the only remaining uncertainty at this time is the second period performance measure.

From here it is easy to see that motivating input H in the second period requires

$$\beta_2 \geq \frac{c_H}{H - L} \tag{18.6}$$

a hopefully familiar requirement.[6] Similarly, if the manager is to allocate first period input to both short-run and to long-run tasks, he must value them equally. Glancing back at the initial certainty equivalent expression in (18.4), it is clear this boils down to another familiar requirement of

$$\beta_1 = \theta\beta_2 \tag{18.7}$$

And, with this balance guaranteed, motivating input H in the initial period requires

$$\beta_1 \geq \frac{c_H}{H - L} \tag{18.8}$$

Together, conditions (18.6), (18.7) and (18.8) provide the incentive compatibility requirements if the firm is to motivate the desired input and its allocation. Locating the optimal piece rates is now a familiar exercise.[7] We minimize the expected payment to the manager, subject to these incentive compatibility requirements and to the usual individual rationality requirement. This amounts to minimizing the overall risk premium under which the manager labors, subject to the noted constraints.[8]

---

[6] At the start of the second period, with knowledge of x_1 and having behaved as noted in the first period, the manager's certainty equivalent with L supplied in the second period would be

$$\omega + \beta_1[x_1] + \beta_2[.5\theta H + L] - \frac{1}{2}\rho[\beta_2^2\sigma_2^2] - c_H$$

Requiring the certainty equivalent in (18.5) to be larger than this expression leads to the noted condition.

[7] Notice (18.6) ensures the manager sets $a_2 = H$, just as the other two incentive compatibility conditions ensure $a_{SR} = a_{LR} = .5H$.

[8] Given that both short-run and long-run tasks are being undertaken, the balance requirement in (18.7) implies expected compensation will be $E[I|H, H] = \omega + \beta_1 H + \beta_2 H$, and the manager's corresponding certainty equivalent reduces to $CE_{HH} = E[I|H, H] - \frac{1}{2}\rho[\beta_1^2\sigma_1^2 + \beta_2^2\sigma_2^2] - 2c_H$. With an outside certainty equivalent of $M = 0$, then, the optimal piece rates and two period wage are the solution to

$$\min_{\omega, \beta_1 \beta_2} \quad E[I|H, H]$$

$$\text{s.t.} \quad CE_{HH} \geq 0$$

$$\beta_2 \geq \frac{c_H}{H - L}$$

$$\beta_1 = \theta\beta_2$$

$$\beta_1 \geq \frac{c_H}{H - L}$$

To get back to the short-run versus long-run theme, notice that the coordination tension surfaces only in expression (18.7), the second of the incentive compatibility requirements. If this issue is not present, the other two incentive requirements would lead to piece rates of $\beta_1 = \beta_2 = \frac{c_H}{H-L}$ (as anything larger would needlessly burn risk premium and anything lower would not produce the desired inputs). By construction, our little story is one where creating incentives for input in each period may, and likely will, inadvertently create a short-run versus long-run tension. This is explored in the following examples.

Example 18.1 Assume a two period setting patterned after Example 15.1. The manager is described by a risk aversion measure of $\rho = .1$ and a personal cost of $c_H = 60$. The inputs are $H = 500$ and $L = 200$; and noise in the performance measures is specified by $\sigma_1^2 = \sigma_2^2 = 10,000$. Also let parameter $\theta = 1$ in the second period performance measure. You should verify that the optimal piece rates here are $\beta_1^* = \beta_2^* = .20$ and that the manager's risk premium totals 40.

There is no inherent short-run versus long-run conflict here. Dealing with the input supply incentives leads, naturally, to a balanced view of the first period's short-run and long-run tasks. This, recall, is the meaning of $\theta = 1$; it implies the performance scores will capture short-run and long-run activities in unbiased fashion. Literally, input incentives lead to $\beta_1^* = \beta_2^* = .20$, and the balance requirement in (18.7) is redundant. Balance is not a control issue here.

Example 18.2 Continuing with the same setting, now let parameter $\theta = .80$. This means the evaluation measures provide a downward biased estimate of the first period's long-run activities. We know from Example 18.1 that setting $\beta_1 = \beta_2 = .20$ is the least costly way to motivate high input in each period. Unfortunately, these piece rates, coupled with $\theta < 1$, inadvertently create a bias in favor of short-term activities in the first period. The only way to correct this, absent additional information, is to increase the second period's piece rate from $\beta_2 = .20$ to $\beta_2^* = .25$. We are forced to more heavily weight second period performance in order to compensate for this bias. Coupled with $\beta_1^* = .20$, we thus restore the necessary balance of $\beta_1^* = \theta\beta_2^* = .80(.25) = .20$. The solution is not without cost, as increasing the second period's piece rate increases the manager's risk premium from 40 to 51.25. The control hot spots, which correspond to the positive shadow prices in the design program, are now input supply in the first period, constraint (18.8), and the short-run versus long-run balance constraint (18.7).

Example 18.3 By way of contrast, change Example 18.1 so $\theta = 1.25$. This means the evaluation measures provide an upward biased estimate of the first period's long-run activities or, more realistically, a downward biased estimate of short-run activities. In any event, the incentives that ensure high input in each period ($\beta_1 = \beta_2 = .20$) now inadvertently create

a bias in favor of the long-run task. Maintaining a balanced view (a good pun) now requires we raise the first period's piece rate, resulting in $\beta_1^* = .25$ along with $\beta_2^* = .20$. The risk premium is again 51.25.

All together, these examples illustrate the delicate interplay among performance evaluation, input incentives and allocation of those inputs among tasks, here short-run versus long-run coordination.

18.2.2   Additional Frictions

As comforting as this multitask story appears, we have judiciously ignored a variety of frictions. Commitment powers are lessened when the time horizon expands. For example, we may have an unusually harmonious working relationship with our supervisor, and then find our supervisor has switched jobs or been promoted. We, too, may switch jobs or be promoted. In addition, the sheer complexity of designing a lasting, long-term contract is overwhelming. We therefore expect incomplete contracts and renegotiation to occur. We also do not condone absolute hands-tying labor supply commitments. The manager can quit. Performance bonds and non-competition clauses are possible, but the fact remains that the manager can reenter the labor market through time.

Reliance on less than complete contracts brings up the possibility of implicit contracts. It is "understood" the manager's pay is keyed to labor market conditions. It is "understood," whenever possible, promotion will be from within. It is "understood" the accounting library and its array of responsibility accounting subtleties will not be changed with any frequency, resulting in the proverbial moving target.[9]

Career concerns also come into play. The manager's human capital and reputation can be affected by current period activities. Working in the new product arena may, as a by-product, put the manager in a position to learn the ins and outs of emerging technology in some area. Similarly, working with an established product may diminish the manager's possibilities of keeping current with this emerging technology. Each set of results provides additional evidence as to the manager's skill and talents.

For example, the understanding may be the manager is compensated at a level comparable to that of comparable managers in roughly comparable firms. Compensation consultants are used periodically to calibrate this arrangement. Further suppose our manager has been highly successful and is generally regarded as a top performer. Will this induce unusual risk aversion as the manager seeks to protect this reputation? Conversely, suppose our manager has been floundering and is generally regarded as a middling performer. Will this induce unusual risk seeking, as the manager seeks the big hit that will raise this reputation?

---

[9] These issues, endemic in a dynamic setting, are discussed further in Chapter 19.

Control concerns are also not one-sided. The firm may be less than attentive to its promise to evaluate performance. Promised rotation through a variety of assignments may not be forthcoming. Good performance may be met by ever increasing demands for better performance.

Short-run versus long-run balancing is important for the firm, and it is important for the individual. The picture is one of intra-manager coordination that takes place through time. The tensions and pitfalls are enormous.

18.2.3   Balancing Devices

This should not suggest that balancing short-run and long-run considerations is insoluble.[10] Rather, it is a dimension to firm life that requires nurturing with professional skill. One avenue is additional information (surprise). Places where concern for short-run versus long-run tensions is particularly strong invite additional monitoring. If we don't want maintenance cut back in difficult times, we may want to monitor maintenance activity. If we are worried the present management team is not devoting sufficient resources to new product development or manufacturing improvement, we may want to engage an external consultant to perform a strategic audit of their activities.

Performance statistics can also be pointed toward a longer horizon. For example, change in net worth over a 10-year period places an emphasis on growth and downplays the importance of short-run variations in income. The firm's equity price, presuming common shares are traded in an organized market, is a significant source of information. It is surely based on a variety of information sources and forward looking, though from a valuation as opposed to evaluation perspective. In this way we interpret managerial stock ownership or stock options as an evaluation-compensation arrangement that uses the security price as a performance statistic.

Another avenue is attitudes and arrangements within the firm. The firm can nurture a particular view of short-run versus long-run tensions. More direct orchestration is also used. To illustrate, we often find a committee is used to pass judgment on major investment proposals. One reason is to assemble a variety of experts to explore the desirability of major proposals. A second reason is to ensure communication among the managers. A third reason is to ensure, given managerial mobility within and between firms, that someone is still in the firm when the fruits of this investment decision take shape.

---

[10] For that matter, accrual accounting is designed to reflect short-run and long-run forces, suggesting a natural role in helping manage this tension. Equally clear, though, is the fact the accounting library's integrity must be protected, so we should expect the accrual process to be incomplete in its rendering of anticipated future effects. Glance back at Example 18.2.

We also should not forget the manager's reputation. The trick is to recognize when the manager's reputation concerns work for the firm. For example, a manager nearing retirement has few remaining options to influence his reputation, while a younger manager may well cultivate a good reputation, hopefully to the firm's benefit as well.

Short-run versus long-run coordination tensions are yet another multitask story. The unusual features are its importance to the firm and management team, the possibilities for inadvertent creation of a short-run bias and its natural place in the firm's dynamics.

18.3   Inter-Manager Coordination

Of course managers do not work in a vacuum, and our coordination saga naturally leads to inter-manager coordination issues. One form of this is the earlier discussed master budget foray. Another is when divisions within a divisionalized firm trade with one another. Examples are numerous. The branch bank writes loans funded by center. Goods manufactured in a foreign subsidiary are marketed by the domestic division. *R&D* in one division leads to a patented pharmaceutical that is manufactured in a second division and marketed by a third division. Component parts are manufactured in one division and assembled in another. The large audit firm loans audit personnel to another office. Coal is mined in one division and used to generate electricity in another. The political science professor teaches a course in the business school.

The common theme is trade between divisions. This raises the issue of motivating desirable trades or, if you will, of coordinating the divisions' activities. The unusual features are these trades occur under the umbrella of the parent firm, and the trade itself takes the form of exchanging goods and services for accounting (as opposed to real) currency. *Transfer pricing* is the name given to the accounting library's procedures for recording trades of this nature.[11]

Institutionally, trades of this sort are commonplace, though typically comprise but a portion of the firm's activities. They are also typically treated in decentralized fashion, meaning the trades are consummated by the divisions themselves, subject of course to how the firm management has chosen to regulate such trade.

---

[11]In this way the respective divisions are credited or charged as a function of the quantity transferred. A transfer pricing arrangement is an accounting procedure that orchestrates credits and charges in this fashion on the basis of the quantity transferred. Cost allocation attempts the same thing (e.g., when we allocate the cost of some central service to the divisions) but generally must rely on synthetic as opposed to actual output measures.

18.3.1 *Inter-Division Trade in the Face of Control Frictions*

To bring these issues to life in a setting where control frictions are present, we again resort to our multitask model. Assume we have two managers, now interpreted as division managers. Manager $i = 1, 2$ can supply high or low input (what else), denoted $a_i \in \{L, H\}$. High input is desired from each. In addition, the performance measure for manager i takes on the familiar form of $x_i = a_i + \varepsilon_i$, where ε_i is a zero mean normal random variable with variance denoted σ_i^2. The two noise terms are independent. We should think of these performance measures as division profit measures.

To limit notation, the two managers are (nearly) identical, with the overly familiar normalized preference setup of constant risk aversion with risk aversion measure ρ, along with low input personal cost of c_L and outside opportunity M normalized to zero. The single difference between the managers is their respective personal costs of high input, denoted $c_{H_1} > 0$ and $c_{H_2} > 0$.

Each manager is compensated with a linear arrangement, denoted $I_i = \omega_i + \beta_i x_i$. Notice we assume each manager is evaluated solely on the basis of his respective division's numbers. (This keeps us focused on essentials.)

Were this the totality of the story, it would be rather boring. However, in addition to these normal activities and tasks, the managers may have an opportunity to jointly and profitably produce and sell an additional product or service. For simplicity, any such opportunity arises after the managers have supplied their inputs. If the managers coordinate, the first division will incur additional cost in the amount c while the second will receive additional net revenue in the amount R. Therefore, if the managers harvest such an opportunity, the first manager's performance measure will decline by c while the second manager's will increase by R.

To resolve this imbalance, we inject some accounting currency, in the form of a transfer price. If the managers coordinate, if they trade, the second division "pays" the first division T for its goods or services.

Let $q \in \{0, 1\}$ denote the coordinated activity. Putting all of this together, the division manager's performance measures are given by

$$x_1 = a_1 + (T - c)q + \varepsilon_1 \tag{18.9}$$

and

$$x_2 = a_2 + (R - T)q + \varepsilon_2 \tag{18.10}$$

Trade, then, results in incremental gain of $R - c$ to the firm as a whole; of this amount, $T - c$ winds up in the first division's accounts or performance measure and $R - T$ in the second division's accounts or performance measure. Notice that the assumed evaluation setup ensures the managers will find any such trade mutually beneficial only when $T - c \geq 0$ and $R - T \geq 0$.

Next suppose the first division's cost will be "high" or "low," with $\overline{c} > \underline{c}$ while the second division's net revenue will also be "high" or "low," with

$\overline{R} > \underline{R}$. Further suppose $\overline{c} > \overline{R} > \underline{c} > \underline{R}$. This implies the trade is profitable to the firm only when the first division's cost is "low" and the second division's revenue is "high." The first manager privately learns his division's cost, $c \in \{\underline{c}, \overline{c}\}$, and the second manager privately learns his division's revenue, $R \in \{\underline{R}, \overline{R}\}$, (after having supplied their respective inputs).

The probability that profitable trade obtains, that $R = \overline{R}$ and $c = \underline{c}$, is denoted α. The question is how to efficiently motivate profitable trade, profitable coordination, given the underlying control problems.

18.3.2 Regulation of Inter-Division Trade

Numerous approaches to dealing with this coordination issue arise. The firm itself might play a heavy hand by tasking the manager's with communicating their private information to center and thereby centralizing the choice of whether to pursue coordinated production. At the other extreme, the firm might task the division managers with resolving the issue between themselves, effectively decentralizing the choice. This, in turn, opens a variety of approaches to how the transfer price might be determined and how aggregate the evaluation measures might be. Our purposes are best served by digging deeper into a single approach.

Assume the coordination choice is decentralized to the two managers and any trade effects are simply aggregated into their respective performance measures, as implied by expressions (18.9) and (18.10). The firm itself sets the policy on the transfer price T, as opposed to relying on negotiation between the managers or perhaps available market prices for similar goods or services. (More about this later.) In our stylized setting, the firm simply announces a T and instructs the managers to engage in any mutually acceptable trade. Presuming $\underline{c} \leq T \leq \overline{R}$, this will ensure profitable trade takes place.

To see this, we return to the certainty equivalent machinery. Suppose the first division's manager supplies input $a_1 \in \{L, H\}$, and no coordination with the second division transpires. It is, by now, old hat to see his certainty equivalent is

$$CE_{a_1} = \omega_1 + \beta_1 a_1 - \frac{1}{2}\rho\beta_1^2\sigma_1^2 - c_{a_1}$$

Conversely, if trade with the second division were to transpire, his certainty equivalent would be

$$CE'_{a_1} = CE_{a_1} + \beta_1(T - c)$$

for cost realization $c \in \{\underline{c}, \overline{c}\}$.

From here two important facts emerge. First, if the manager is to supply high input we again face the requirement that[12]

$$\beta_1 \geq \frac{c_{H_1}}{H - L}$$

Parallel comments apply to the second manager, leading to

$$\beta_2 \geq \frac{c_{H_2}}{H - L}$$

Second, with $\beta_i > 0$ (as $c_{H_i} > 0$, recall), it now follows the managers will favor any trade for which $T \geq c$ and $R \geq T$. Therefore, with any transfer price of $\underline{c} \leq T \leq \overline{R}$, both managers will seek only those trades that are profitable to the firm, as it is only those trades that increase their evaluation and hence compensation.

Setting the best transfer price, however, turns out to be surprisingly subtle. The reason is the trade opportunity is injected into a setting where other control issues are present. The underlying control problems necessitate strictly positive piece rates, $\beta_i > 0$, while the evaluation measures themselves also depend on whether trade occurs, $q \in \{0, 1\}$. Profitable trade improves the evaluation scores, but adds to the compensation risk as trade obtains only with probability α.

It is the interaction of this additional risk effect with the underlying control problems that guides the choice of transfer price T. This is explored in a series of examples.[13]

---

[12] Though the algebra is a bit dense, recall that profitable trade occurs with probability α. (We will momentarily verify jumping on such a trade opportunity is incentive compatible.) This means the manager faces certainty equivalent CE_{a_1} with probability $1 - \alpha$ and certainty equivalent CE'_{a_1} with probability α, given input choice a_1. So, motivating choice of H requires

$$(1 - \alpha)U(CE_H) + \alpha U(CE'_H) \geq (1 - \alpha)U(CE_L) + \alpha U(CE'_L)$$

From here, a little algebra, using the fact $U(CE'_{a_1}) = U(CE_{a_1}) \cdot \exp(-\rho\beta_1(T-c))$, leads to the noted expression.

[13] The design program should be clear, so we forego elaboration, other than to note its basic structure. Let $E[U_i|T, a_i]$ denote manager i's expected utility measure. For the first manger, for example, and remembering $U_i(\cdot) = -\exp(-\rho(\cdot)) = U(\cdot)$, we have

$$E[U_1|T, H_1] = (1 - \alpha)U(CE_{H_1}) + \alpha U(CE'_{H_1})$$

From here, the firm's program is

$$\min_{\omega_i, \beta_i, T} \quad \omega_1 + \omega_2 + (\beta_1 + \beta_2)H + \alpha(\beta_1(T - \underline{c}) + \beta_2(\overline{R} - T))$$

$$\text{s.t.} \quad E[U_i|T, H_i] \geq U(M) = -1, \ i = 1, 2$$

$$\beta_i \geq \frac{c_{H_i}}{H - L}, \ i = 1, 2$$

$$\underline{c} \leq T \leq \overline{R}$$

Example 18.4 Assume a setting patterned after Example 15.1. Each manager is now described by a risk aversion measure of $\rho = .1$ and a personal cost of $c_{H_i} = 60$. The inputs are $H = 500$ and $L = 200$; and noise in the performance measures is specified by $\sigma_1^2 = \sigma_2^2 = 10,000$. With the usual normalization, low input costs and the outside certainty equivalents are set to 0. For benchmark purposes, initially suppose there is no trade opportunity. We find, to no surprise, that $\beta_1^* = \beta_2^* = .20$, each manager incurs a risk premium of 20 and the firm's cost is $160 = 2(20) + 2(60)$.

Now introduce profitable trade by assuming $\overline{R} = 200, \underline{c} = 100$ and $\alpha = .5$. The piece rates are unaltered, as the control problem surrounding input supply has not changed. The optimal transfer price is $T^* = 150$, which evenly splits the gains to trade between the two divisions. Each manager's risk premium increases from 20 to 21.20, and the firm's overall compensation cost increases to 162.40. The risk premia increase because, thanks to the possibility of profitable trade governed by $\alpha = .5$, each manager's compensation is now more risky. And with identical control problems in the two divisions, we equally allocate the profitable trade noise by having the divisions equally share the gains to trade.

In equilibrium, the managers are neither harmed by the additional risk nor explicit beneficiaries of profitable trade. The firm holds rational expectations, anticipates the probabilistic increment to their performance scores, and reflects this in the compensation function.[14] The transfer price of 150, in turn, is designed to simultaneously motivate trade and to keep the managers' respective evaluation measures, expressions (18.9) and (18.10), as informative as possible. Emphatically, the transfer price is designed to motivate communication and minimize noise. The managers learn each other's private information from that communication, not from the price.[15]

Example 18.5 Now lower the second manager's personal cost of high input from 60 to $c_{H_2} = 15$. This implies the control problem is worse in the first division, and is reflected in the now optimal piece rates of $\beta_1^* = .20 > \beta_2^* = .05$ and transfer price of $T^* = 105.79$. With the control problem much worse in the first division, that division's evaluation measure

---

The firm selects base wages and piece rates for each manager along with the transfer price to minimize expected compensation, subject to the familiar individual rationality and incentive compatibility conditions (for high input). A little thought should convince you that $\beta_1 = \frac{c_{H_1}}{H-L}$ and $\beta_2 = \frac{c_{H_2}}{H-L}$ (as higher piece rates burn risk premia and lower ones are infeasible). Profitable trade incentive compatibility is achieved by requiring $\underline{c} \leq T \leq \overline{R}$; and risk issues drive precisely where in this $[\underline{c}, \overline{R}]$ interval the transfer price is set.

[14] Remember, the manager's normalized certainty equivalents are, in equilibrium, precisely zero.

[15] Clearly, a more refined evaluation system that separately identified any gains to trade would be useful here, as it would allow removal of evaluation noise associated with adding the trade effects to the evaluation measures. Institutionally, however, the aggregation approach is commonplace.

is relatively more important, and the transfer price is set to move most of the noise associated with profitable trade to the second division, where it does less damage. (Conversely, if the lowered personal cost were in the first division, the transfer price would be $T^* = 194.21$.)[16]

Though well dressed with many convenient details, these examples teach us a great deal about transfer pricing. Essential information, whether a profitable transfer opportunity exists, is available at the division level. The managers are motivated to use their private information to coordinate their activities. The accounting measures are designed to be useful in motivating and evaluating these activities. The transfer pricing arrangement is used to encourage profitable trade in light of the surrounding control problems and to make the division profit measures more informative in light of the tasks assigned to the managers. The transfer price is not used to carry information from one manager to another. The managers directly speak to each other. Transfer pricing is engaged to provide a superior aggregate set of performance statistics.[17]

18.3.3   Variations on a Theme

This trade regulation theme arises in a variety of settings, and is approached in a variety of manners. There might, for example, be a close substitute for the first division's product or service available in the market. The firm would then face a make or buy issue. There is also a dynamic side to the story, as trade opportunities are likely to arise time and again.

This leads to search for a convenient, hopefully robust policy. As hinted at earlier, the firm might rely on the managers themselves to set the transfer price, in effect a negotiated arrangement. It may decree that transfers be priced at standard or actual variable cost, or at standard or actual full cost. It may decree that transfers be priced at full cost plus a percentage markup. If an active market is present, it might decree use of the market price or use of the market price less a discount.[18] Glancing back at Examples 18.4 and 18.5 where no such active market is available, you will notice the former has the flavor of a transfer price based on full cost plus a markup, while the latter has the flavor of a transfer price close to variable cost.

---

[16] It should be clear that using both division measures to evaluate each manager will lead to a slight improvement, as, at the margin, this helps control for the noise of profitable trade. We pursued the more myopic formulation because it keeps the essential tensions in clear view. Institutionally, that word again, a common practice is to use, among other things, division income, firm-wide income and the firm's equity price in evaluating its division managers.

[17] This basic insight can also be teased out of the Chapter 13 model, though in not nearly so clear a fashion.

[18] The firm may also establish a grievance procedure, so that it becomes actively involved in regulating trade at the discretion of the division managers.

The variety of policies noted above, all of which are used in practice, may appear unwieldy or disconcerting. The underlying idea, though, is to provide a division profit measurement procedure that helps implement a decentralized structure. With various sources of information and a variety of control frictions, we should expect a variety of ways to bring additional information into the division profit measurement apparatus.[19]

A similar comment applies to the world of taxation. For tax purposes, the firm prefers to park as much profit as possible in the division with more favorable tax circumstances. It is not surprising, then, that state, national, and foreign tax authorities have a great deal to say about transfer pricing practice in the measurement of taxable income, as opposed to dealing with the firm's control problems.

18.4   Coordinated Sabotage

The final stop in our look at coordination raises the question of whether there might be too much coordination. Certainly the "over-centralized" firm is too controlled, too coordinated from the top. Likewise, the officious, bureaucratic procedure suggests too much coordination. A deeper side to this question also exists. Large-scale fraud and bribery require coordination across individuals. Here the coordination is done with the intent of bypassing the firm's internal controls.[20]

More subtle forms of dysfunctional coordination also occur. Suppose the students in one class have a midterm in another class. Study time for the first class will be diminished while the students study for the midterm in the other class. No explicit coordination has occurred. Self-interest leads to the seemingly coordinated behavior in which no one is prepared for the first class. A work slowdown occurs when the labor force complies with each facet and nuance of the labor contract.

The classroom and slowdown illustrations arise in the context of relative performance evaluation. The idea, recall, is to use the performance of one individual as a gauge for the other, presuming they labor in related environments. Grading on a curve, recall, implies a relative as opposed to absolute standard. Using an absolute standard exposes the students to the risk of an unusually difficult examination instrument. Grading on a curve removes most of this risk.

The link to coordination is easily spotted when we recast this in the setting of our managerial input model. To see this with minimal additional

---

[19] This, of course, flies in the face of identifying a general rule for pricing inter-division transfers. Are you surprised?

[20] The Foreign Corrupt Practices Act explicitly prohibits a variety of corrupt practices; it also requires that adequate accounting records and internal controls be maintained. The Sarbanes-Oxley Act, in turn, dramatically increases these latter requirements.

overhead, a good pun, return to the two managers in the divisionalized setting. Everything remains as before, with three exceptions: First, the noise in the respective evaluation measures is all common noise. Glancing back at expressions (18.9) and (18.10), this means $\varepsilon_1 = \varepsilon_2$. Second there is no trade opportunity, so $\alpha = 0$. Third, the personal costs of input H are identical (and denoted c_H).

The common noise term invites use of relative performance evaluation or, more broadly, benchmarking with peers. Consider paying each agent their first-best wage if their measures agree and a large penalty otherwise. Given our normalization and presuming the two managers are identical, this leads to the following payment structure

$$I_1(x_1, x_2) = I_2(x_1, x_2) = \begin{cases} c_H \text{ if } x_1 = x_2 \\ z \ll c_H \text{ otherwise} \end{cases}$$

This creates an evaluation game between the two agents. Clearly, an equilibrium in the game is for each manager to supply input H. After all, if one is so doing, not following suit guarantees a serious penalty for both. (This presumes such penalties are feasible, of course.) And behaving in this manner results in payment of c_H less personal cost of c_H, thereby matching the outside normalized certainty equivalent of $M = 0$.

Unfortunately our little device invites coordinated sabotage. What if one player supplies input L? The other player's best response is to supply input L as well. We have a second equilibrium in the game! And the managers are better off with this second equilibrium. Each faces a certainty equivalent of $c_H - c_H = 0 = M$ in the first equilibrium, compared to $c_H - c_L = c_H > 0$ in the second. Not good.

Example 18.6 To illustrate, let's use the specification in Example 18.4. Given the normalization of $M = 0$, both players supplying input H leads to identical performance scores, and payment of $c_H = 60$. Netting the personal cost, we have a certainty equivalent of 0 for each. If one supplies H and the other L, their measures disagree, and both receive the penalty compensation of z. Netting the personal costs leads to certainty equivalents of $z - c_H = z - 60$ for the high input manager and $z - c_L = z$ for the low input manager. Finally, if both supply L, their performance scores are again identical, resulting in payments of 60 to each. Netting the normalized personal cost of $c_L = 0$ leads to a certainty equivalent of 60 for each. In effect, we have an evaluation tournament summarized in the following bimatrix game display, where both supplying input H or both supplying input L are evident equilibria.

	H_2	L_2
H_1	0, 0	$z - 60, z$
L_1	$z, z - 60$	60, 60

Excessive coordination is excessively tempting here. Notice, however, that when the two performance measures disagree, the manager with the

higher score has surely supplied input H, and just as surely the other has supplied input L. This suggests a way to destroy the unwanted equilibrium. Consider leaving the second manager's compensation as previously specified, but offer the first a prize in the amount $\widehat{z}$ if the performance scores differ but his is higher than that of the second manager. Given this, suppose the second manager supplies input L. The first manager's certainty equivalent is c_H if he too supplies input L, but it is $\widehat{z} - c_H$ if he supplies H, and this exceeds c_H presuming $\widehat{z}$ is sufficiently large. This destroys the unwanted equilibrium in the performance tournament.[21]

Example 18.7 Return to the setting of Example 18.6, but set $\widehat{z} = 200$. This produces the following bimatrix game display, where it is evident jointly supplying input H is the unique equilibrium.

	H_2	L_2
H_1	0, 0	140, z
L_1	z, $z - 60$	60, 60

If the managers approach their relative performance evaluation game as a noncooperative exercise, the equilibrium calculus is compelling. Each now supplies input H. What if, on the other hand, they decide to cooperate? They might agree to have the first manager supply H and the second L, and then split the overall net gain 50-50. This beats what they gain by playing noncooperatively. Equally clear, they might agree simply to play the joint supply of L combination. Excessive coordination is not a happy thought.

Our little yarn is acquiring a life of its own. We began with a setting where relative performance evaluation is called for. Coordination temptations then enter, as the orchestrated competition between the managers can be turned off by playing a second and more advantageous equilibrium. The retort is to drive a wedge between the managers, offering an unusually high prize for stellar performance. This removes the earlier coordination temptation, but at the cost of introducing another. Now the managers have even more of a reason to abjectly collude.

We don't design control systems to make every collusion or circumvention possibility unrewarding. A balance is struck, defending against some and taking our chances against others. If the maitre d', waiters, and bartender all conspire, the restaurant owner is surely at risk. If the real estate developer is a crook, the silent partners are surely at risk. If the division management team decides to take an enormously risky strategy and under-inform central management, the firm is at risk.

The trick is to understand the limits of coordination. There can be too much, and there can be too little. The well-run firm knows when and where

---

[21] What surfaces is a type of whistle blower arrangement.

to take advantage of cooperative tendencies, and when to worry about them.

18.5   Summary

Coordination activities are a center piece of firm life. The firm exists because it is better able to manage various types of transactions. This leads to the study of coordination. Here we encounter the seemingly mundane issue of making certain the details fit together, presuming well conceived incentives foster the coordination process. Intra-manager coordination concerns, especially well balancing, or coordinating, short-run and long-run activities, are also part of the picture. To no surprise, these coordination issues place additional burdens on the performance evaluation exercise.

Inter-manager coordination issues are particularly stark in a divisionalized firm where divisions encounter opportunities to trade goods and services between or among themselves. This leads to the subject of transfer pricing, where we emphasized recording inter-division trades for profit measurement purposes. One way or another, the underlying issues are motivating desirable trade and engineering informative division profit measures. Subtleties abound, as the resolution of the transfer price issue depends on the control fabric into which the trade option has been inserted. Our recurring theme surfaces yet again: more tasks generally expand the inherent control problem.

Finally, we recognize that the firm can have too much of a good thing. Coordination is not cost free. We expect less than complete coordination to be the rule. Moreover, coordination can subtly shift from being advantageous to being dysfunctional. Carefully coordinated behavior can sabotage the firm's control system, just as surely as it can pave the way for efficiency gains.

18.6   Bibliographic Notes

Coordination has been studied extensively. Anthony [1965; 1988] stresses a managerial perspective. Marschak and Radner [1972] focus on information differences at different locations in a firm, and coordination of the local decision behavior, in the absence of control problems. Balancing an agent's allocation across periods in a contracting model is examined in numerous settings. A good introduction is provided by Antle and Fellingham [1990], Gibbons and Murphy [1992], Holmstrom and Ricart I Costa [1986] and Lewis and Sappington [1989]. Coordination among agents, excessive to the point it creates difficulties for the control system, has led to elaborate whistle blower games as in Ma, Moore, and Turnbull [1988] and Rajan

[1992]. In the limit, the agents may collude. This is studied by Tirole [1986] and Suh [1987].

Divisionalized management and transfer pricing have also been the subject of considerable study. To provide some entry to this literature, Hirshleifer [1956] is a classic transfer pricing reference. Solomons [1965] emphasizes accounting subtleties in division performance measurement. Tang [2002] provides important institutional details. Ronen and McKinney [1970] and Groves and Loeb [1979] highlight the strategic side. Harris, Kriebel, and Raviv [1982] introduce control considerations explicitly tied to input supply. Antle and Eppen [1985] link the attendant control problems to capital rationing. Dye [1988] emphasizes information content of the division performance measure. Baiman, et al. [2007] emphasize an auction mechanism for controlling inter-division trade, while Vaysman [1998] and Baldenius, Reichelstein and Sahay [1999] emphasize negotiation and Baldenius and Reichelstein[2006] external market guides. Swieringa and Waterhouse [1982] stress behavioral connections. Holmstrom and Tirole [1991] study the interaction between transfer pricing and firm form. Eccles [1985] provides a connection to the firm's strategy. Interactive control problems with second sourcing are highlighted in Anton and Yao [1987] and Demski, Sappington, and Spiller [1987]. Comparative advantage at organizing trade is stressed by Williamson [1985]. Our particular exposition is based on Christensen and Demski [1998].

18.7   Problems and Exercises

1. Our study of coordination sweeps across master budget, short-run versus long run, inter-division trade and sabotage issues. What is the common theme?

2. Transfer pricing uses prices and quantities to record trade between divisions. In general terms this is often thought of as using a price mechanism to guide such trade. To what extent is this analogy correct? Discuss the similarities and differences when trade passes (i) between two divisions in the same firm or (ii) between two independent entities in an organized market.

3. We used the managerial input model to highlight the importance of allocating the gains to inter-division trade between two divisions. In that setting, how do the managers learn of possible gains to trade and what role is played by the allocation of any gains to trade?

4. In laying out a transfer pricing exercise, we were careful to append it to an existing control problem and to wrap the benefits to trade in uncertainty. What purpose is served by this elaborate staging?

5. *biased evaluation*
   Compare the piece rates in Examples 18.1 and 18.2. Provide an intu-
   itive explanation for their equality in the first setting and inequality
   in the second. What would likely happen here if, in Example 18.2,
   the firm also had access to another measure of long-run activities
   (e.g., the firm's security price)?

6. *shadow prices*
   We have repeatedly stressed the connection between shadow prices
   in an incentive design program and what we call control "hot spots."
   Determine the shadow prices in Examples 18.1 and 18.2. What are
   the control hot spots in these two settings?

7. *short-run versus long-run incentives*
   Return to Examples 18.1 and 18.2, but change the specification of the
   second period's evaluation noise from $\sigma_2^2 = 10,000$ to $\sigma_2^2 = 15,000$.
   Determine the optimal (linear) contracts for both the $\theta = 1$ and
   $\theta = .80$ cases. Carefully explain your findings, especially with respect
   to the optimal piece rates and shadow prices.

8. *trade of output for accounting currency*
   United Management has a divisionalized structure. Division B has
   encountered an opportunity to provide specialized manufacturing for
   an established customer. The customer will pay 100. The catch is divi-
   sions A and B will both have to contribute manufacturing resources.
   A will do the preliminary work and then transfer the semifinished
   product to B, and B will then complete the manufacturing and de-
   liver the item to the customer. The cost will total 60, with 50 incurred
   in division A and 10 in division B. The transfer price, from B to A,
   is set at the amount T.

   (a) Assume the opportunity is taken. Determine the incremental
       profit (i) to the firm; (ii) to division A; and (iii) to division B.
   (b) Provide journal entries on division A's books to record all activ-
       ity associated with this opportunity. Include entries for work in
       process, cost of goods sold, revenue, and so on. (Do not close any
       temporary accounts.) For convenience, assume all cost incurred
       by A is associated with cash expenditures.
   (c) Do the same for division B's books.
   (d) When Ralph, an employee of United Management, prepares con-
       solidated financial statements, will the consolidation process,
       working from your above entries, result in a firm-wide incre-
       mental profit that agrees with your answer in (a)? Explain.

9. *trade of output for fungible currency*
   Return to problem 8 above. Now assume division A is unable to

accommodate division B, and B must, as a result, go to an outside source. This source is paid the amount P. Everything else remains as before.

(a) Repeat the earlier question.

(b) Write a brief paragraph explaining the differences between your two sets of answers.

10. *classical analysis*

Ralph's firm consists of divisions A and B. All of the output of A is transferred to B, where it is processed further and then sold. No costs are incurred at center. The outputs are coordinated, implying $q_A = q_B$. The market price for the finished product is presently 450 per unit; and the division's short-run cost structures are as follows:

$$
\begin{aligned}
C_A^{SR}(q_A; P) &= 200 + 450q_A - 10q_A^2 + (1/6)q_A^3 \\
C_B^{SR}(q_B; P) &= 300 + 250q_B - 10q_B^2 + (1/6)q_B^3
\end{aligned}
$$

(a) Determine the firm's optimal output and corresponding profit.

(b) Suppose B can order any quantity from A, and will be charged a transfer price of T per unit. A is obliged to produce as instructed. Find a T such that maximizing its division income will lead B to prefer the output quantity you determined in (a) above.

(c) Suppose A can manufacture any quantity it desires and will be credited with an internal revenue of T for each unit. Find a T such that maximizing its division income will lead A to prefer the output quantity you determined in (b) above.

(d) As a serious lesson in the art of coordination, we appear to be making a mistake here. What is our mistake?

11. *noisy gains to trade*

Return to Example 18.4, where each manager's compensation is determined by $I_i = \omega_i + \beta_i x_i$. Determine each manager's wage, ω_i, for three cases: $\alpha \in \{0, .5, 1\}$. Explain your finding.

12. *noisy gains to trade*

Return to Example 18.4 but now assume $H = 900, L = 300$ and the high input personal costs are $c_{H_1} = c_{H_2} = 120$.

(a) Determine the optimal contracts and transfer price.

(b) Repeat for the case $c_{H_1} = 60$. Provide an intuitive explanation for your findings.

(c) For both cases, determine how much the firm would pay to separately observe trade between the divisions. Explain.

13. *sourcing dispute*[22]

Ralph's Firm is a large, decentralized firm. Each major product group is manufactured and marketed by a separate division. The divisions are free to trade among themselves as opportunities arise. Each division is treated as an investment center. The managers' compensation depends on the performance of their divisions, relative to expectations, and the performance of the firm as a whole. Division A has developed a new consumer product and is lining up final production plans. A critical subcomponent can be manufactured by division B or acquired from an outside supplier. Division A asked for formal bids. Three were received: division B bid 1,350 per hundred, Western Industries bid 950 per hundred, and Calzig bid 957 per hundred. Western is a well-known, reliable subcontractor. Calzig is a competitor of division B.

The division A manager is ready to accept the Western bid but decided to check with division B one final time. The B manager insisted the bid of 1,350 was solid and would not be lowered. Business is picking up, the B manager explains, and the announced policy of pricing all products at full cost plus the usual 11% markup would be followed. B's variable cost appears to be about 850 per hundred. The B manager also pointed out that they helped in the product engineering work and "understood" that they would be the favored supplier if the product ever went into production. It was also pointed out that A's projected profit margin was 420 per hundred, and this was based on an estimated price of 1,400 per hundred for the subcomponent in question. Before A has time to contact Western, an urgent message from central management arrives. Division B has complained to center that A is about to source with an outside supplier. The firm is forced to respond and has called a teleconference for the following morning. What should center do?

14. *insurance arrangements*

A large bank evaluates commercial lending officers in terms of the profitability and quality of their loan portfolios. When a loan is consummated, the loan officer "borrows" the principal from center at an internally posted rate. The internal rate depends on the maturity of loan. If the loan is a fixed rate loan, the loan officer is charged the posted rate on the outstanding balance each period. The rate used is fixed at the internal rate at the time the loan was booked. In this way the lending officer is insured against interest rate movements, but not against default risk (to the extent default is not related to interest

---

[22]Inspired by Harvard Business School case 158-001, titled "Birch Paper Company."

rate movements). Carefully analyze this transfer pricing arrangement. How would a variable rate loan be treated?

15. *internal cost of funds and rationing*[23]
Ralph manages a decentralized firm where division managers have significant authority to make production and investment decisions. All capital expenditures must be approved by center. Divisions must submit detailed capital budgets prior to approval by center. This is one area where Ralph is somewhat disappointed with decentralization, as divisions show a marked tendency to pad their budgets. Slack in the budgets makes life more pleasant at the divisions. If a division's budget is successfully padded, division personnel have an easier time meeting the budget. Think of this as the division personnel consuming the slack in the budget. That said, Ralph is aware one of the divisions has a capital project that will yield a cash flow of 100 at the end of one year. Ralph believes this project will cost 75 or 65, with equal probability. The division, though, knows with certainty what the cost will be. This leads to a concern the division might pad its budget on this project. To prevent this, a project auditing program, which will discover the actual project cost and report directly to center, is being considered. How much would Ralph be willing to pay for the project auditing of this capital project? Ralph's opportunity cost of capital is 20%, and Ralph is risk neutral. For simplicity, assume the budget is submitted and funds are provided to division at the beginning of the year. The benefits of the project (100) will be available to Ralph at the end of the year.

16. *relative performance evaluation*
Ralph is at it again. Output from the production process owned by Ralph can be x_1 or x_2. The manager's input can be L or H. Ralph is risk neutral. The manager is the usual normalized, constant risk aversion type. His risk aversion measure is $\rho = .0001$; and $c_H = 4,000$ along with $c_L = M = 0$. The probabilities are given by $\pi(x_1|H) = 0.1$ and $\pi(x_1|L) = 1$. Ralph wants supply of input H. The only observable for contracting purposes is the manager's output. Pretty standard stuff so far.

(a) Determine an optimal pay-for-performance arrangement.

(b) Suppose Ralph owns two such production processes and employs an identical manager on each. Further suppose the two environments are perfectly correlated. So if both managers supply input H, their outputs will always agree (both x_1 or both x_2). Suppose Ralph offers to pay each 4,000 if their outputs agree and -10,000

---

[23] Contributed by John Fellingham.

otherwise. Verify that if one manager supplies input H the best the other can do is supply input H.

(c) What happens in the above arrangement if one manager supplies input L? Is the other's best response also to supply input L? How do you think the managers will play the game?

(d) Amend Ralph's scheme so that, in the game played between the two managers, both supplying input H is a unique equilibrium. Give an intuitive explanation for why your modification leads to a unique equilibrium. What difficulty is associated with your scheme?

17. *encouraging profitable investment*[24]

Ralph's firm is always looking for new, innovative products. A manager in Ralph's firm every now and then discovers a new product. Any such discovery is privately known, and it is up to the manager to reveal to Ralph the new product idea. Any new product will eventually result in success (S) or failure (F). The odds of success are higher if the manager is of higher talent. This is because higher talented managers are better at identifying high-quality projects and are also better at implementing them. People inside and outside the firm observe whether a new product proposal is brought forward, and whether it succeeds or fails. In this way, the labor market learns when a particular manager brings forward a new product and whether that product turns out to be successful. (Gossip can be quite powerful.) The manager's reputation, in other words, improves if a product proposal is brought forward and if the product turns out to be successful. A failed product lowers the manager's reputation. No product proposal is a somewhat intermediate story, because we have to worry about whether the reputation is influenced by a lack of proposal. Let's forget about this latter possibility.

Any new product is risky to Ralph's firm; it is also risky for the proposing manager as any such investment proposal places that manager's reputation at risk. What does this do to product development incentives in Ralph's firm, and what might Ralph do to address the situation?

---

[24]Inspired by Holmstrom and Ricart I Costa [1986].

19

End Game

Our study of managerial uses of accounting information has focused on two seemingly innocuous questions: *what might it cost*, and *did it cost too much*? The former is code for a valuation exercise and the latter is code for an evaluation exercise. We have learned that the answers to the two questions differ, as they are fundamentally different. We have also learned that the answers are hardly lacking in complexity or subtlety, but are invariably to be found in well practiced art that is informed by fundamentals.

There is more to learn, both in terms of fundamentals and their artful application. But just as with coordination in the prior chapter, we can have too much of a good thing. So it is time to conclude our odyssey. We begin with a more dynamic perspective, one that emphasizes concurrent use of the accounting library in addressing the two questions. This leads to a richer, more vibrant view of the firm, and takes us inexorably into the reality of incomplete arrangements and contracts, unanticipated events and governance. From here, we conclude with the mantra of professional responsibility. There simply is no substitute for responsible behavior.

19.1 Concurrency

For pedagogical purposes we approached the two metaphorical questions in stages, moving from product costing to decision making to performance

evaluation.[1] This creates, if not implies, a sequential picture of designing the accounting library, making important decisions, and then recording and evaluating the results. Organizational life is, of course, much richer.

Consider a division manager who has responsibility for manufacturing and marketing a line of consumer products. The manager deals with a wide array of tasks. Included are evaluating the performance of the division's management team and focusing the product line strategy in light of changing consumer tastes, technology and competitor behavior. The manager also deals with a wide array of information sources: product market statistics, product development trends, trade association publications, consultants, subordinates, output, and sales and productivity measures, to name a few. The image is one of being overwhelmed by information sources.

Accounting enters at this point. It is a well-protected source of financial summarization. Division income is disaggregated into sales and, to the extent possible, expenses by product group. A similar categorization is likely for important assets, such as inventories. Responsibility accounting refocuses these data on individual members of the management team. These summarizations are drawn from the library, and rest on the recognition rules, LLAs, aggregation and cost allocation policies in place. Special studies will also supplement these periodic summarizations, as appropriate.

At this point the manager straddles the past and the future. Looking backward, one task is to sort out how well the management team has performed. Another is to sort out how well the division's strategy has performed. In this sense, there is a recurring theme of "learning by doing." Experience is accumulating and being interpreted.

The management team may be performing admirably, though product market woes have depressed financial performance; or the team may be riding the crest of an unusually healthy product market. The quality improvement program may be paying dividends. The product design team may be breathing new vigor into the product line, or providing a forum for revisiting old and, it was hoped, long buried political frictions. The new manufacturing technology may be turning out as planned. The technology adopted by a competitor may be providing them a troublesome edge, or a helpful annoyance. These are specific renditions of our metaphorical "did it cost too much" question.

Looking forward, the manager faces the task of identifying the next steps in the division's unfolding history. These steps are informed by a variety of information sources, including the many nuances that can be discerned from recent events and what strategy the division was following as these events

---

[1]In Chapters 2 through 7 we studied product costing and the accounting library. In Chapters 8 through 12 we studied managerial decision making with an emphasis on the "what might it cost" theme. In Chapters 13 through 18 we studied managerial performance evaluation, with an emphasis on the "did it cost too much" theme.

unfolded. They are also informed by the emerging assessment of the firm's capacity and abilities, including those of the management team. These are specific renditions of our metaphorical "what will it cost" question.

Change and surprise are present. Some events are more endogenous than others. Transition considerations are also at work. For example, adoption of a new manufacturing policy may set off an extended adjustment of inventories throughout the manufacturing and marketing network (as excess inventories are depleted and others repositioned). This, too, will cloud the immediate picture as the management team continues to struggle with the best balance of short-run and long-run considerations, both in its direction of the division's activities and its interpretation of recent results.

Product costing, decision making, and performance evaluation are concurrent activities. For example, one of the division's products may be lagging. What does it cost to manufacture and distribute this product? What is the best estimate of the competitors' costs? Could out-sourcing some components lower the product cost? Is managerial replacement suggested? Have incentives been inappropriately weighted, perhaps stressing short-run performance to the extent investment in product updating has lagged?

The concurrency theme is one of working in the middle of the firm's history. The future holds sufficient promise to worry about careful decision making. The past holds sufficient relevance to be able to inform the next round of decision making. It also provides the informational foundation for the firm's performance evaluation activities. This raises the question of how plans, policies, strategies and even the accounting system might be changed to reflect the firm's changing circumstance.

19.2   Governance

Once we acknowledge this richer, more complex setting, we encounter the ever present prospect of change. Products and services ebb and flow, technology advances, consulting fads evolve, political winds drift, the firm's alliances and basic structure evolve. Some change is endogenous, most appears to be exogenous. Some is anticipated, while significant portions appear to be unanticipated.

19.2.1   Incomplete Contracts

This suggests another managerial task, not to mention fact of organizational life, is anticipating the inevitability of being confronted with the unanticipated. Contractual arrangements are typically incomplete; and contracts can always be changed by mutual consent (marriage being an exception, where the state has a say). The remodeling contractor may encounter unanticipated structural problems in the building. The professor may be

asked to teach outside an area of expertise as student demand ebbs and
flows.

Altering contractual arrangements, in light of altered events or even in
response to a renegotiation opportunity, is part of the picture. Renego-
tiation, on the one hand, can provide a vehicle for efficiently adapting to
uncontractible events. In our stylized managerial input model, for example,
if the firm observes but cannot contract on the manager's input, it has the
option of renegotiating the contract presuming the manager has behaved.
This would allow removal of the underlying evaluation risk. On the other
hand, renegotiation options can be inefficient. For example, the ability to
renegotiate the basic contract in the managerial input model, absent any
new information, leads to unraveling, because the parties will not be able
to forego redressing the inefficient risk sharing in the middle of the game.
Anticipating this renegotiation, the power of the well designed pay-for-
performance arrangement collapses. More broadly, an inability to commit
to not renegotiate is often associated with a reduced supply of contractible
information, simply because informationally starving the renegotiation en-
counter removes inefficient opportunism possibilities.

Formally altering contractual arrangements is a small part of the dy-
namic picture. Incomplete arrangements beget implicit arrangements, as
discussed in Chapter 18. It is understood that particular actions will be
honored. Accepting a particularly onerous or risky managerial assignment
will be remembered as the manager's career unfolds. A delayed evaluation
will be forthcoming. Historical precedents will be honored.

Here we encounter so-called implicit contracts, basically non-contractual
agreements that are sustained by repeated play of a trading encounter. For
example, the firm honors evaluation commitments and announced promo-
tion policies and the management team works effectively because the parties
expect to continue the arrangement despite lack of an explicit contract.

Even well designed implicit arrangements have their limits. And it is
here that we encounter various institutional arrangements.

Carefully designed property rights is one such arrangement. Suppose
the firm seeks a delivery service in which a driver uses an automobile to
transport various items to various destinations. In a perfect market setting
it would not matter who owned the auto. Capital markets would supply
the requisite capital, and the driver's use of the auto would be independent
of ownership. With imperfect markets, the story is quite different. The
reason is residual ownership is no longer a matter of indifference.

On the labor market side, the parties will have considerable difficulty
foreseeing and contracting around all the things the driver might do with
the auto (or with critical factors of production more generally). What they
do know is the auto is likely to be heavily used (some would say abused) if
the firm owns or leases the auto and simply supplies it to the driver. After
all, the driver's use of the auto cannot be monitored, and the driver may
not be around later to be confronted with the real depreciation. Therefore,

it matters who owns the auto. Property rights are important. The driver who owns the auto will internalize the effects of real depreciation, while the driver who does not will be in a free rider (pun) position. This suggests the driver should own the auto. (Conversely, if specialized maintenance is important to ensuring the driver's safety, we would worry about the firm's incentives to supply proper maintenance if it rather than the driver owned the auto.)

The capital market likely cuts in the other direction. The firm may be more financially sound, and thus better able to acquire the auto. It is also likely to be better at carrying the risks of ownership. In addition, the firm may be in a better position to capitalize on tax benefits associated with equipment ownership. Some taxi drivers own their cabs while others do not. Some employees use a company auto while others do not. These are not accidental arrangements.

The underlying story is one of trading frictions, or transaction cost. Property rights for capital equipment that lasts beyond the trading encounter confer a type of residual claim on the owner of the equipment. This matters when contracts are incomplete or, for that matter, informationally starved. A well-crafted trading arrangement exploits this aspect of asset ownership.

Another institutional arrangement designed to cope with the limits of contractual arrangements is governance mechanisms. The idea is straightforward: decision rights for dealing with the unanticipated are vested in some fashion. Governance bodies themselves are familiar institutional arrangements for dealing with events as they unfold. Major sports leagues have their oversight arrangements. The typical union contract contains a well-specified grievance structure. The family firm has the still active first generation family member. The U.S. Constitution carefully specifies executive, legislative, and judicial powers. The university has an ombudsman.

These bodies deal with unforeseen events. In a sense, they complete trade arrangements or alter the trading environment as circumstances dictate. The role of the judiciary when an unforeseen product liability is encountered is illustrative. Other examples are dealing with the impact that television has on the structure and conduct of major league baseball, resolution of inter-division conflict in a global banking institution, and curriculum design at your favorite university.

Going a bit further, we should expect the firm's activities and its governance abilities to be well matched. For example, it may be more efficient to house a high risk product development venture in a separate firm. A larger, more stable firm likely has a variety of activities. The key players in the new product venture will be worried about their future if the product flops. If this is a stand-alone firm, the worry takes the form of what the labor market sees and might offer. If this is part of a much larger firm, the worry takes the form of what the internal labor market sees and might offer. Do we want the key players worrying about their future in the one arena or the other?

Ideally their efforts would be focused on the new product. At the margin, career interests will creep into the setting. Are these best controlled by the firm or the labor market? If the large firm's governance arrangement is adept at this game, it might be best to house the new product venture inside the firm. Otherwise, a stand alone arrangement is preferable. For example, the large firm may be able to provide reasonably adequate career insurance for the key players; after all, everyone understands the venture is high risk. On the other hand, the large firm may provide too many opportunities for the key players to worry about ingratiating themselves with other key players, as a precaution against failure of the venture.

19.2.2   Accounting Governance

This governance function naturally extends to the accounting library. We don't see elaborate plans to alter the firm's accounting policies in response to technology and market changes. Instead we see restrained (some would say glacial) behavior, often tied (some would say too closely) to financial reporting requirements and the opinions and advice of the external auditor. GAAP itself is defined by an elaborate (and currently changing) governance arrangement.[2]

In the larger picture this has merit. Accounting provides just one among many sources of information. It is designed to be comprehensive, yet well defended. One version of being well defended is being difficult to change. Imagine a divisionalized firm in which the division managers could routinely alter the split between expensing and capitalizing various expenditures, could routinely vary revenue recognition policies, and could routinely switch among various product costing models. The periodic accounting rendering of divisional events would likely become a game of "catch me if you can." The U. S. federal government appears to be particularly adept at this game.

This is why we see such things as attention to GAAP in a debt covenant (don't change the rules in the middle of the game) and frequent use of consultants when a major change in accounting policy is contemplated (have a third party play an important governance role).

Accounting governance is part of the larger, dynamic picture of organization life. Once we recognize multiple sources of information and the demands for library integrity, we recognize accounting governance is likely to be slow moving. This is not because accountants are wedded to stable procedures, but because the accounting library offers a well-defended, consistent approach to summarizing the firm's financial history. Don't assume accounting policies are frozen in time. Do recognize that a steady, cau-

---

[2]Recall, from Chapter 7, that a firm switching from an impressionism to a modernism approach to product costing is almost surely experiencing a traumatic, albeit endogenous, event.

tious approach to change is one of the prices of library integrity. The firm's activities change faster than its accounting policies, for a reason.

19.2.3   Governance Failures

In painting this brief picture of the dynamic side of firm life, it is essential we return to the "too much of a good thing" theme. Governance failures are not uncommon. In fact, they appear to occur in episodic fashion.

Innovations, it seems, take on the nature of fads: new devices for structuring risky transactions to keep them out of the accounting library, new transactions designed with earnings management in mind, aggressive sale of subprime mortgage instruments (coupled with hedge funds speculating on related derivative instruments), or curricula driven by student evaluations are illustrative. Once one of these phenomena acquires a life of its own, mimicry ensues and we have a form of industry or economy wide contagion. Eventually, the bubble bursts or the activity comes under widespread scrutiny. In the end we have yet another episodic governance failure.[3]

19.3   Responsibility

This returns us to the theme in Chapter 1 of a professional (quality) manager. The professional manager is *well prepared.* Artful rendering of fundamentals is an essential skill, just as it is essential the subtleties of the particular economic climate be understood. Informed professional judgment and action are daily tasks. This is why we have carefully avoided rules, recipes, and guidelines for the use of accounting information. This is also why we have stressed an expanded, nearly boundless view of the managerial task. The professional manager is a well-prepared artisan (yes, informed by fundamentals).

The professional manager is also *responsible.* Fiduciary responsibilities are present, but this only scratches the surface. Ethical and moral responsibilities are also present. Trade arrangements, indeed most modern era economic interactions, become impossible without the rudiments of trust and honor. This, too, only scratches the surface. The professional manager has a responsibility that runs deeper than efficiently administering trade arrangements. The professional manager is both well prepared and responsible. These are constant, ever present traits. They are not to be invoked opportunistically.

---

[3]Precisely how or why these episodes occur is unclear. We model them in terms of herding behavior, where followers imitate leaders, followed by an information cascade that leads to their demise.

19.4   Summary

Product costing, decision making, and performance evaluation themes are simultaneously engaged in the managerial task. Concurrency of the "what will it cost" and "did it cost too much" metaphors is the nature of the game. Placing these activities in their natural context raises issues of dealing with complexity, change and the unanticipated. Though our sketch of this deeper, vibrant fabric is necessarily brief, the emergence of governance in all its dimensions should be clear. In the end, governance is essential. Successful governance rests on responsibility. There is no substitute.

19.5   Bibliographic Notes

This dynamic theme merges into work on organizations. Perrow [1986] offers an expansive critique. Arrow [1974], Holmstrom and Tirole [1989], Milgrom and Roberts [1992], and Williamson [1985] emphasize economic foundations and trading frictions. Sappington and Stiglitz [1987] stress information based trading frictions in identifying whether production of a particular good or service is best located in the public or private sector. Mookherjee [2006] stresses delegation, Drymiotes [2007] interdependent monitoring agents and MacLeod [2007] broader based enforcement institutions. The ability of renegotiation to incorporate uncontractible information into trading arrangements is explored in Demski and Sappington [1991] and Hermalin and Katz [1991], while the inefficiency side is identified by Fudenberg and Tirole [1990]. Information flow under renegotiation conditions is highlighted by Demski and Frimor [1999]. Feltham, Indjejikian and Nanda [2006] continue this theme, but imbed it in a concurrency setting. Breach, dissolution, and ownership changes are explored, respectively, in Stole [1992], Cramton, Gibbons and Klemperer [1987] and Meyer, Milgrom, and Roberts [1992]. Demski, Frimor and Sappington [2004] address accounting change in response to a manager learning manipulation skills. Sunder [1996] tightly connects accounting and organization life, while Ijiri [1975] emphasizes library integrity. Demski [2003] and Lev [2003] highlight the governance failures in the infamous Enron saga and surrounding events.

19.6   Problems and Exercises

1. Why are the metaphorical questions "what might it cost" and "what did it cost" fundamentally different questions?

2. The dynamic theme of decision making and performance evaluation emphasizes use of the accounting library simultaneously for decision making and performance evaluation purposes. Here the basic library

building blocks of aggregation, *LLA*s, and allocation enter. This suggests a managed tension, a tension between using these building blocks to better serve decision making and to better serve performance evaluation interests. Is this a correct view of the accounting library?

3. Accounting governance is visible (and contentious) in the world of financial reporting, as evidenced, by FASB, IASB and GASB activities. Yet accounting governance is important inside the firm and hardly independent of the attendant external reporting environment. Carefully discuss this theme.

4. *an old friend*
   Return to one of your favorite illustrations, Example 13.5, where the following probability structure was assumed:

	x_1	x_2
$\pi(x|H)$	.5	.5
$\pi(x|L)$	1	0

   Further recall the optimal pay-for-performance arrangement used I_1 = 5,000 and I_2 = 12,305.66

   (a) Now suppose both parties observe the manager's input. The firm acquires this information before the output is observed. The catch is the parties cannot contract on their joint observation of the manager's input. This might be due to "contracting costs" (though hardly believable in this simple story) or the impossibility of a third party ever verifying the manager's input. Consider the following arrangement: initially set the above pay-for-performance arrangement in place; then, if the firm sees input H, offer to exchange the manager's risky compensation for its certainty equivalent of 8,000. Is this scheme incentive compatible for both the manager and the firm? Does it, in equilibrium, allow for use of the input observation by the parties? Would Ralph be pleased? What is the explanation?

   (b) Now suppose the firm does not observe the manager's input. But it knows, under equilibrium behavior, that the original control system motivates supply of input H. So after the input has been supplied, why not simply offer to renegotiate the manager's contract and exchange the risky pay for its certainty equivalent? Will this scheme work? Explain.

5. *labor market conditions*
   We usually have difficulty writing long term contracts. Suppose a manager is known to the labor market and has a reputation (good or bad). The employer cannot write an iron-clad long-term contract,

and instead the parties periodically renegotiate, with knowledge of the manager's then current market value. How do the market forces help and how do they impede the structuring of a well-functioning employment relationship?

6. *factors of production*
Ralph's firm is expanding. It tentatively plans to add a sales force. The sales force will require the usual trappings of an automobile, personal computer, state-of-the-art communication equipment, and so on. Since no precedents have been set, the firm has an open mind on the question of who should own this equipment. Discuss the ownership issue. Would ownership matter in the world of Chapter 2?

7. *time not on our side*
Consider a Chapter 15 style multitask setting, especially Example 15.1 where two tasks are present and the firm wants all of the input assigned to the first task. As in that case, the performance evaluation measure is given by

$$x = a_1 + \alpha \cdot a_2 + \varepsilon$$

where a_1 is the desired task and ε is our ever present error term. a_2, however, is a task that is of no value to the firm, but surely influences the performance measure.

(a) Suppose $\alpha < 1$. Does the firm need to worry about the manager engaging in manipulative behavior?

(b) Suppose the story is repeated a number of periods. Each cycle improves the manager's understanding of how to exploit the system by applying his time and talent to the manipulation task. This means that parameter α increases with each repetition. Does this tell you anything about job rotation or accounting change in general?

8. *accounting governance*
Accounting provides a financial library. Its comparative advantage is integrity. The accounting library is consciously designed to be difficult to manipulate. This means it will have less than aggressive recognition rules, and be slow to change its recognition rules. Discuss the reason for slowness to change recognition rules. Is it a surprise financial reporting is subject to an elaborate, external governance structure? Critics often contend this external governance structure unduly influences the firm's internal accounting. Carefully analyze this contention.

References

[1] Abowd, J., and D. Kaplan, "Executive Compensation: Six Questions that Need Answering," *Journal of Economic Perspectives* (Fall, 1999).

[2] Amershi, A., J. Demski and J. Fellingham, "Sequential Bayesian Analysis in Accounting," *Contemporary Accounting Research* (Spring, 1985).

[3] Anderson, S., W. Hesford and S. Young, "Factors Influencing the Performance of Activity Based Costing Teams: A Field Study of ABC Model Development Time in the Automobile Industry," *Accounting Organizations and Society* (2002).

[4] Anderson, S., and S. Young, "The Impact of Contextual and Process Factors on the Evaluation of Activity-Based Costing Systems," *Accounting Organizations and Society* (1999).

[5] Anthony, R., *Planning and Control Systems: A Framework for Analysis* (Division of Research, Graduate School of Business Administration, Harvard University, 1965).

[6] Anthony, R., *The Management Control Function* (Harvard Business School Press, 1988).

[7] Antle, R., and J. Demski, "The Controllability Principle in Responsibility Accounting," *Accounting Review* (October, 1988).

[8] Antle, R., and G. Eppen, "Capital Rationing and Organizational Slack in Capital Budgeting," *Management Science* (February, 1985).

[9] Antle, R., and J. Fellingham, "Resource Rationing and Organizational Slack in a Two-Period Model," *Journal of Accounting Research* (Spring, 1990).

[10] Anton, J., and D. Yao, "Second Sourcing and the Experience Curve: Price Competition in Defense Procurement," *Rand Journal of Economics* (Spring, 1987).

[11] Arrow, K., *The Limits of Organization* (Norton, 1974).

[12] Arya, A., J. Fellingham and J. Glover, "Capital Budgeting: Some Exceptions to the Net Present Value Rule," *Issues in Accounting Education* (August, 1998).

[13] Arya, A., and J. Glover, "Option Value to Waiting Created by a Control Problem," *Journal of Accounting Research* (December, 2001).

[14] Arya, A., J. Glover and S. Radhakrishnan, "The Controllability Principle in Responsibility Accounting: Another Look," in *Essays in Accounting Theory in Honour of Joel S. Demski,* R. Antle, F. Gjesdal and P. Liang, eds. (Springer, 2007).

[15] Arya, A., J. Glover and S. Sunder, "Earnings Management and the Revelation Principle," *Review of Accounting Studies* (1998).

[16] Baiman, S., and J. Demski, "Economically Optimal Performance Evaluation and Control Systems," *Journal of Accounting Research Supplement* (1980).

[17] Baiman, S., P. Fischer, M. Rajan and R. Saouma, "Resource Allocation Auctions within Firms," *Journal of Accounting Research* (December, 2007).

[18] Bajari, P., and A. Hortacsu, "The Winner's Curse, Reserve Prices, and Endogenous Entry: Empirical Insights from eBay Auctions," *Rand Journal of Economics* (Summer, 2003).

[19] Baker, K., and R. Taylor, "A Linear Programming Framework for Cost Allocation and External Acquisition when Reciprocal Services Exist," *Accounting Review* (October, 1979).

[20] Balakrishnan, R., and K. Sivaramakrishnan, "A Critical Overview of the Use of Full-Cost Data for Planning and Pricing," *Journal of Management Accounting Research* (2002).

[21] Baldenius, T., and S. Reichelstein, "External and Internal Pricing in Multidivisional Firms," *Journal of Accounting Research* (March, 2006).

[22] Baldenius, T., S. Reichelstein and S. Sahay, "Negotiated versus Cost-Based Transfer Pricing," *Review of Accounting Studies* (June, 1999).

[23] Banker, R., S. Datar and S. Kekre, "Relevant Costs, Congestion and Stochasticity in Production Environments," *Journal of Accounting & Economics* (July, 1988).

[24] Banker, R., and S. Hansen, "The Adequacy of Full-Cost Based Pricing Heuristics," *Journal of Management Accounting Research* (2002).

[25] Baron, J., and D. Kreps, *Strategic Human Resources: Frameworks for General Managers* (Wiley, 1999).

[26] Basu, S., and G. Waymire, "Recordkeeping and Human Evolution," *Accounting Horizons* (September, 2006).

[27] Bazerman, M., *Judgment in Managerial Decision Making* (Wiley, 1990).

[28] Beaver, W., *Financial Reporting: An Accounting Revolution* (Prentice Hall, 1998).

[29] Bebchuk, L., and J. Fried, *Pay without Performance: The Unfulfilled Promise of Executive Compensation* (Harvard University Press, 2004).

[30] Becker, S., and D. Green Jr., "Budgeting and Employee Behavior," *Journal of Business* (October, 1962).

[31] Bokenkotter, T., *A Concise History of the Catholic Church* (Image Books, 1979).

[32] Bolton, P., and M. Dewatripont, *Contract Theory* (MIT Press, 2005).

[33] Bonner, S., "Judgment and Decision-Making Research in Accounting," *Accounting Horizons* (December, 1999).

[34] Bonner, S., and G. Sprinkle, "The Effects of Monetary Incentives on Effort and Task Performance: Theories, Evidence, and a Framework for Research," *Accounting Organizations and Society* (2002).

[35] Bouwens, J., and L. van Lent, "Assessing the Performance of Business Unit Managers," *Journal of Accounting Research* (September, 2007).

[36] Buchanan, J., *Cost and Choice: An Inquiry in Economic Theory* (Markham, 1969).

[37] Budde, J., "Performance Measure Congruity and the Balanced Score-card," *Journal of Accounting Research* (June, 2007).

[38] Chambers, R., *Applied Production Analysis* (Cambridge University Press, 1988).

[39] Chapman, C., A. Hopwood and M. Shields, *Handbook of Management Accounting Research, Volumes I and II.* (Elsevier, 2007).

[40] Christensen, J., "The Determination of Performance Standards and Participation," *Journal of Accounting Research* (Autumn, 1982).

[41] Christensen, J., and J. Demski, "The Classical Foundations of "Modern" Costing," *Management Accounting Research* (1995).

[42] Christensen, J., and J. Demski, "Product Costing in the Presence of Endogenous Subcost Functions," *Review of Accounting Studies* (1997).

[43] Christensen, J., and J. Demski, "Profit Allocation under Ancillary Trade," *Journal of Accounting Research* (Spring 1998).

[44] Christensen, J., and J. Demski, *Accounting Theory: An Information Content Perspective* (McGraw-Hill/Irwin, 2003).

[45] Christensen, J., and J. Demski, "Factor Choice Distortion under Cost-Based Reimbursement," *Journal of Management Accounting Research* (2003a).

[46] Christensen, P., and G. Feltham, *Economics of Accounting: Performance Evaluation*, Vol. 2 (Springer, 2005).

[47] Clark, J., *Studies in the Economics of Overhead Costs* (University of Chicago Press, 1923).

[48] Coase, R., "The Nature of Costs," in *Studies in Cost Analysis*, D. Solomons, ed. (Irwin, 1968).

[49] Cooper, R., and R. Kaplan, *The Design of Cost Management Systems* (Prentice Hall, 1991).

[50] Cramton, P., R. Gibbons, and P. Klemperer, "Dissolving a Partnership Efficiently," *Econometrica* (May, 1987).

[51] Dawes, R., *Rational Choice in an Uncertain World* (Harcourt Brace Jovanovich, 1988).

[52] Debreu, G., *Theory of Value* (Cowles Foundation, Yale University, 1959, renewed 1987).

[53] Demski, J., "An Accounting System Structured on a Linear Programming Model," *Accounting Review* (October, 1967).

[54] Demski, J., *Information Analysis* (Addison-Wesley, 1980).

[55] Demski, J., "Cost Allocation Games," in *Joint Cost Allocations*, S. Moriarity, ed. (University of Oklahoma Center for Economic and Management Research, 1981).

[56] Demski, J., "Performance Measure Manipulation," *Contemporary Accounting Research* (Fall, 1998).

[57] Demski, J., "Corporate Conflicts of Interest," *Journal of Economic Perspectives* (Spring, 2003).

[58] Demski, J., and G. Feltham, *Cost Determination: A Conceptual Approach* (Iowa State University Press, 1976).

[59] Demski, J., and G. Feltham, "Economic Incentives in Budgetary Control Systems," *Accounting Review* (April, 1978).

[60] Demski, J., and H. Frimor, "Performance Measure Garbling under Renegotiation in Multi-Period Agencies," *Journal of Accounting Research* (1999 Supplement).

[61] Demski, J., H. Frimor and D. Sappington, "Efficient Manipulation in a Repeated Setting," *Journal of Accounting Research* (March, 2004).

[62] Demski, J., H. Frimor and D. Sappington, "Audit Error," *Journal of Engineering and Technology Management* (2006).

[63] Demski, J., and D. Sappington, "Delegated Expertise," *Journal of Accounting Research* (Spring, 1987).

[64] Demski, J., and D. Sappington, "Hierarchical Structure and Responsibility Accounting," *Journal of Accounting Research* (Spring, 1989).

[65] Demski, J., and D. Sappington, "Resolving Double Moral Hazard Problems with Buyout Agreements," *Rand Journal of Economics* (Summer, 1991).

[66] Demski, J., D. Sappington and P. Spiller, "Managing Supplier Switching," *Rand Journal of Economics* (Spring, 1987).

[67] Dillon, R., and J. Owers, "EVA as a Financial Metric: Attributes, Utilization, and Relationship to NPV," *Financial Practice and Education* (Spring/Summer, 1997).

[68] Dixit, A., and B. Nalebuff, *Thinking Strategically: The Competitive Edge in Business, Politics and Everyday Life* (Norton & Company, 1993).

[69] Dixit, A., and R. Pindyck, *Investment under Uncertainty* (Princeton University Press, 1994).

[70] Dye, R., "Communication and Post-Decision Information," *Journal of Accounting Research* (Autumn, 1983).

[71] Dye, R., "Intrafirm Resource Allocation and Discretionary Actions," in *Economic Analysis of Information and Contracts*, G. Feltham, A. Amershi and W. Ziemba, eds. (Kluwer, 1988).

[72] Dye, R., "An Evaluation of 'Essays On Disclosure' and the Disclosure Literature in Accounting," *Journal of Accounting & Economics* (December 2001).

[73] Drymiotes, G., "The Monitoring Role of Insiders," *Journal of Accounting & Economics* (December, 2007).

[74] Eccles, R., *The Transfer Pricing Problem: A Theory for Practice* (Lexington Books, 1985).

[75] Fare, R., and D. Primont, *Multi-Output Production and Duality: Theory and Applications* (Kluwer, 1995).

[76] Fellingham, J., P. Newman and Y. Suh, "Contracts without Memory in Multiperiod Agency Models," *Journal of Economic Theory* (1985).

[77] Fellingham, J., and M. Wolfson, "Taxes and Risk Sharing," *Accounting Review* (January, 1985).

[78] Feltham, G., and J. Xie, "Performance Measure Congruity and Diversity in Multi-Task Principal-Agent Relations," *Accounting Review* (July, 1994).

[79] Feltham, G., R. Indjejikian and D. Nanda, "Dynamic Incentives and Dual-Purpose Accounting," *Journal of Accounting & Economics* (December, 2006).

[80] Fishburn, P., *Utility Theory for Decision Making* (Wiley, 1970).

[81] Fremgren, J., "The Direct Costing Controversy – An Identification of Issues," *Accounting Review* (January, 1964).

[82] Fudenberg, D., and J. Tirole, "Moral Hazard and Renegotiation in Agency Contracts," *Econometrica* (1990).

[83] Gawer, A., and R. Henerson, "Platform Owner Entry and Innovation in Complementary Markets: Evidence from Intel," *Journal of Economics & Management Strategy* (2007).

[84] Gibbons, R., *Game Theory for Applied Economists* (Princeton University Press, 1992).

[85] Gibbons, R., and K. Murphy, "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence," *Journal of Political Economy* (1992).

[86] Gjesdal, F., "Accounting for Stewardship," *Journal of Accounting Research* (Spring, 1981).

[87] Gordon, M., "The Payoff Period and the Rate of Profit," *Journal of Business* (October, 1955).

[88] Gordon, M., "The Use of Administered Price Systems to Control Large Organizations," in *Management Controls: New Directions in Basic Research*, C. Bonini, R. Jaedicke and H. Wagner, eds. (McGraw-Hill, 1964).

[89] Green, D., Jr., "A Moral to the Direct Costing Controversy?" *Journal of Business* (July, 1960).

[90] Grossman, S., and O. Hart, "An Analysis of the Principal-Agent Problem," *Econometrica* (January, 1983).

[91] Groves, T., and M. Loeb, "Incentives in a Divisionalized Firm," *Management Science* (March, 1979).

[92] Gupta, M., "Heterogeneity Issues in Aggregated Costing Systems," *Journal of Management Accounting Research* (1993).

[93] Haley, C., and L. Schall, *The Theory of Financial Decisions* (McGraw-Hill, 1979).

[94] Harris, M., C. Kriebel and A. Raviv, "Asymmetric Information, Incentives and Intrafirm Resource Allocation," *Management Science* (June, 1982).

[95] Harris, M., and A. Raviv, "Some Results on Incentive Contracts with Applications to Education and Employment, Health Insurance, and Law Enforcement," *American Economic Review* (March, 1978).

[96] Hart, O., and B. Holmstrom, "The Theory of Contracts," in *Advances in Economic Theory, Fifth World Congress*, T. Bewley, ed. (Cambridge University Press, 1987).

[97] Hemmer, T., "Lessons Lost in Linearity: A Critical Assessment of the General Usefulness of LEN Models in Compensation Research," *Journal of Management Accounting Research* (2004).

[98] Hermalin, B., and M. Katz, "Moral Hazard and Verifiability: The Effects of Renegotiation in Agency," *Econometrica* (November, 1991).

[99] Hirshleifer, J., "On the Economics of Transfer Pricing," *Journal of Business* (July, 1956).

[100] Hirshleifer, J., *Investment, Interest, and Capital* (Prentice-Hall, 1970).

[101] Hofstede, G., *The Game of Budget Control* (Van Nostrand Reinhold, 1967).

[102] Holmstrom, B., "Moral Hazard and Observability," *Bell Journal of Economics* (Spring, 1979).

[103] Holmstrom, B., "Moral Hazard in Teams," *Bell Journal of Economics* (Autumn, 1982).

[104] Holmstrom, B., and P. Milgrom, "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica* (March, 1987).

[105] Holmstrom, B., and P. Milgrom, "Regulating Trade Among Agents," *Journal of Institutional and Theoretical Economics* (March, 1990).

[106] Holmstrom, B., and P. Milgrom, "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, & Organization* (Special Issue, 1991).

[107] Holmstrom, B., and J. Ricart I Costa, "Managerial Incentives and Capital Management," *Quarterly Journal of Economics* (November, 1986).

[108] Holmstrom, B., and J. Tirole, "The Theory of the Firm," in *Handbook of Industrial Organization*, R. Schmalensee and R. Willig, eds. (Elsevier, 1989).

[109] Holmstrom, B., and J. Tirole, "Transfer Pricing and Organizational Form," *Journal of Law, Economics, & Organization* (Fall, 1991).

[110] Hopwood, A., "An Empirical Study of the Role of Accounting Data in Performance Evaluation," *Journal of Accounting Research* (Supplement, 1972).

[111] Horwitz, I., "Misuse of Accounting Rates of Return: Comment," *American Economic Review* (June, 1984).

[112] Howard, R., "Proximal Decision Analysis," *Management Science* (May, 1971).

[113] Hwang, Y., J. Evans and V. Hegde, "Product Cost Bias and Selection of an Allocation Base," *Journal of Management Accounting Research* (1993).

[114] Ijiri, Y., *Theory of Accounting Measurement* (American Accounting Association, 1975).

[115] Johnson, H., and R. Kaplan, *Relevance Lost: The Rise and Fall of Management Accounting* (Harvard Business School Press, 1987).

[116] Kaplan, R., "Variable and Self-Service Costs in Reciprocal Allocation Models," *Accounting Review* (October, 1973).

[117] Kaplan, R., and S. Anderson, *Time-Driven Activity-Based Costing* (Harvard Business School Publishing, 2007).

[118] Kennan, J., and R. Wilson, "Bargaining with Private Information," *Journal of Economic Literature* (March, 1993).

[119] Krantz, D., R. Luce, P. Suppes and A. Tversky, *Foundations of Measurement* (Academic Press, 1971).

[120] Kreps, D., *Notes on the Theory of Choice* (Westview Press, 1988).

[121] Kreps, D., *A Course in Microeconomic Theory* (Princeton University Press, 1990).

[122] Kydland, F., and E. Prescott, "Rules Rather than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy* (1977).

[123] Laffont, J., and D. Martimort, *The Theory of Incentives: The Principal-Agent Model* (Princeton University Press, 2001).

[124] Lambert, R., "Executive Effort and Selection of Risky Projects," *Rand Journal of Economics* (Spring, 1986).

[125] Lambert, R., "Contracting Theory and Accounting," *Journal of Accounting & Economics* (December, 2001).

[126] Lazear, E., and K. Shaw, "Personnel Economics: The Economist's View of Human Resources," *Journal of Economic Perspectives* (Fall, 2007).

[127] Laux, V., "The Ignored Performance Measure," *Journal of Economics & Management Strategy*, (Fall, 2006).

[128] Lev, B., "Corporate Earnings: Facts and Fiction," *Journal of Economic Perspectives* (Spring, 2003).

[129] Lewis, T., and D. Sappington, "Inflexible Rules in Incentive Problems," *American Economic Review* (March, 1989).

[130] Lipsey, R., and K. Lancaster, "The General Theory of Second Best," *Review of Economic Studies* (1956).

[131] Luce, R., and H. Raiffa, *Games and Decisions* (Wiley, 1957).

[132] Luenberger, D., *Introduction to Linear and Nonlinear Programming* (Addison-Wesley, 1973).

[133] Ma., C., J. Moore and S. Turnbull, "Stopping Agents from Cheating," *Journal of Economic Theory* (December, 1988).

[134] MacLeod, W., "Reputations, Relationships, and Contract Enforcement," *Journal of Economic Literature* (September, 2007).

[135] Machina, M., "Choice under Uncertainty: Problems Solved and Unsolved," *Journal of Economic Perspectives* (Summer, 1987).

[136] Marschak, J., and R. Radner, *Economic Theory of Teams* (Yale University Press, 1972).

[137] Merchant, K., *Rewarding Results: Motivating Profit Center Managers* (Harvard Business School Press, 1989).

[138] Meyer, M., P. Milgrom and J. Roberts, "Organizational Prospects, Influence Costs, and Ownership Changes," *Journal of Economics & Management Strategy* (Spring, 1992).

[139] Milgrom, P., "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics* (Autumn, 1981).

[140] Milgrom, P., "Auctions and Bidding: A Primer," *Journal of Economic Perspectives* (Summer, 1989).

[141] Milgrom, P., *Putting Auction Theory to Work* (Cambridge University Press, 2004).

[142] Milgrom, P., and J. Roberts, "Informational Asymmetries, Strategic Behavior, and Industrial Organization," *American Economic Review* (May, 1987).

[143] Milgrom, P., and J. Roberts, *Economics, Organization & Management* (Prentice Hall, 1992).

[144] Miller, B., and G. Buckman, "Cost Allocation and Opportunity Costs," *Management Science* (May, 1987).

[145] Mookherhee, D., "Decentralization, Hierarchies, and Incentives: A Mechanism Design Perspective," *Journal of Economic Literature* (June, 2006).

[146] Myerson, R., "Incentive Compatibility and the Bargaining Problem," *Econometrica* (January, 1979).

[147] Myerson, R., *Game Theory: Analysis of Conflict* (Harvard University Press, 1991).

[148] Nadiri, N., "Producers Theory," in *Handbook of Mathematical Economics*, Volume II, K. Arrow and M. Intriligator, eds. (North-Holland, 1987).

[149] Nash, J., "The Bargaining Problem," *Econometrica* (January, 1950).

[150] Nisbett, R., and L. Ross, *Human Inference: Strategies and Shortcomings of Social Judgment* (Prentice Hall, 1990).

[151] Noreen, E., "Conditions under which Activity-Based Cost Systems Provide Relevant Costs," *Journal of Management Accounting Research* (1991).

[152] Noreen, E., and N. Soderstrom, "Are Overhead Costs Strictly Proportional to Activity? Evidence From Hospital Service Departments," *Journal of Accounting & Economics* (July, 1994).

[153] Noreen, E., and N. Soderstrom, "The Accuracy of Proportional Cost Models: Evidence from Hospital Service Departments," *Review of Accounting Studies* (1997).

[154] Osborne, M., *An Introduction to Game Theory* (Oxford University Press, 2004).

[155] Osborne, M., and A. Rubinstein, *Bargaining and Markets* (Academic Press, 1990).

[156] Oster, S., *Modern Competitive Analysis* (Oxford University Press, 1999).

[157] Parker, R., *Struggle for Survival: The History of the Second World War* (Oxford University Press, 1990).

[158] Parker, R., G. Harcourt and G. Whittington, *Readings in the Concept and Measurement of Income* (Philip Allan, 1986).

[159] Paton, W., *Accounting Theory* (1922; reissued in 1962 by Accounting Studies Press).

[160] Perrow, C., *Complex Organizations: A Critical Essay* (Random House, 1986).

[161] Pfeffer, J., "Human Resources from an Organizational Behavior Perspective: Some Paradoxes Explained," *Journal of Economic Perspectives* (Fall, 2007).

[162] Porter, R., "A Review Essay on Handbook of Industrial Organization," *Journal of Economic Literature* (June, 1991).

[163] Prendergast, C., "The Provision of Incentives in Firms," *Journal of Economic Literature* (March, 1999).

[164] Rajan, M., "Cost Allocation in Multiagent Settings," *Accounting Review* (July, 1992).

[165] Rasmusen, E., *Games and Information: An Introduction to Game Theory* (Blackwell Publishing, 2007).

[166] Roberts, J., *The Modern Firm: Organizational Design for Performance and Growth* (Oxford University Press, 2004).

[167] Rogerson, W., "Overhead Allocation and Incentives for Cost Minimization in Defense Procurement," *Accounting Review* (October, 1992).

[168] Ronen, J., and G. McKinney, III, "Transfer Pricing for Divisional Autonomy," *Journal of Accounting Research* (Spring, 1970).

[169] Ronen, J., and V. Yaari, *Earnings Management: Emerging Insights in Theory, Practice, and Research* (Springer, 2008).

[170] Ross, S., *Neoclassical Finance* (Princeton University Press, 2005).

[171] Ross, S., et al., *Corporate Finance: Core Principles and Applications* (McGraw-Hill/Irwin, 2006).

[172] Salamon, G., "On the Validity of Accounting Rate of Return in Cross-Sectional Analysis: Theory, Evidence, and Implications," *Journal of Accounting and Public Policy* (1988).

[173] Sappington, D., "Incentives in Principal-Agent Relationships," *Journal of Economic Perspectives* (Spring, 1991).

[174] Sappington, D., and J. Stiglitz, "Privatization, Information and Incentives," *Journal of Policy Analysis and Management* (Summer, 1987).

[175] Sargent, T., *Bounded Rationality in Macroeconomics* (Oxford University Press, 1993).

[176] Scholes, M., M. Wolfson and M. Erickson, *Taxes and Business Strategy: A Planning Approach* (Prentice Hall, 2004).

[177] Shavell, S., "Risk Sharing and Incentives in the Principal and Agent Relationship," *Bell Journal of Economics* (Spring, 1979).

[178] Shefrin, H., *Behavioral Corporate Finance* (McGraw-Hill/Irwin, 2005).

[179] Solomons, D., *Divisional Performance: Measurement and Control* (Financial Executives Research Foundation, 1965; also published by Irwin in 1968).

[180] Solomons, D., "The Historical Development of Costing," in *Studies in Cost Analysis*, D. Solomons, ed. (Irwin, 1968).

[181] Sorter, G., and C. Horngren, "Asset Recognition and Economic Attributes – A Relevant Costing Approach," *Accounting Review* (July, 1962).

[182] Spulbur, D., *Regulation and Markets* (MIT Press, 1989).

[183] Stigler, G., *The Theory of Price* (Macmillan, 1987).

[184] Stiglitz, J., "Risk Sharing and Incentives in Sharecropping," *Review of Economic Studies* (April, 1974).

[185] Stole, L., "The Economics of Liquidated Damage Clauses in Contractual Environments with Private Information," *Journal of Law, Economics, & Organization* (October, 1992).

[186] Suh, Y., "Collusion and Noncontrollable Cost Allocation," *Journal of Accounting Research* (1987 Supplement).

[187] Sunder, S., "Simpson's Reversal Paradox and Cost Allocation," *Journal of Accounting Research* (Spring, 1983).

[188] Sunder, S., *Theory of Accounting and Control* (South-Western, 1996).

[189] Sutton, J., *Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration* (MIT Press, 1991).

[190] Swieringa, R., and R. Moncur, *Some Effects of Participative Budgeting on Managerial Behavior* (National Association of Accountants, 1975).

[191] Swieringa, R., and J. Waterhouse, "Organizational Views of Transfer Pricing," *Accounting, Organizations and Society* (1982).

[192] Tang, R., *Current Trends and Corporate Cases in Transfer Pricing* (Greenwood Publishing, 2002).

[193] Tirole, J., "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," *Journal of Law, Economics, & Organization* (Fall, 1986).

[194] Tirole, J., *The Theory of Industrial Organization* (MIT Press, 1988).

[195] Tirole, J., *The Theory of Corporate Finance* (Princeton University Press, 2006).

[196] Vancil, R., *Decentralization: Managerial Ambiguity by Design* (Dow Jones-Irwin, 1979).

[197] Vaysman, I., "A Model of Negotiated Transfer Pricing," *Journal of Accounting & Economics* (June, 1998).

[198] Verrecchia, R., "An Analysis of Two Cost Allocation Cases," *Accounting Review* (July, 1982).

[199] Verrecchia, R., "Essays on Disclosure," *Journal of Accounting & Economics* (December, 2001).

[200] Weil, R., "Allocating Joint Costs," *American Economic Review* (December, 1968).

[201] Whittington, G., *The Elements of Accounting: An Introduction* (Cambridge University Press, 1992).

[202] Williamson, O., *The Economic Institutions of Capitalism* (Free Press, 1985).

[203] Zimmerman, J., "The Costs and Benefits of Cost Allocation," *Accounting Review* (July, 1979).

Index