Advances in Quantitative Analysis of

Finance and Accounting

Volume 6

Editor

Cheng-Few Lee

World Scientific

Advances in Quantitative Analysis of Finance and Accounting

Volume 6

This page intentionally left blank

Advances in Quantitative Analysis of

Finance and Accounting

Volume 6

Editor

Cheng-Few Lee

Rutgers University, USA

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Advances in Quantitative Analysis of Finance and Accounting – Vol. 6
ADVANCES IN QUANTITATIVE ANALYSIS OF FINANCE AND ACCOUNTING

ISBN-13 978-981-279-168-9
ISBN-10 981-279-168-X

Typeset by Stallion Press
Email: enquiries@stallionpress.com

Printed in Singapore.

Preface to Volume 6

Advances in Quantitative Analysis of Finance and Accounting is an annual publication designed to disseminate developments in the quantitative analysis of finance and accounting. The publication is a forum for statistical and quantitative analyses of issues in finance and accounting as well as applications of quantitative methods to problems in financial management, financial accounting, and business management. The objective is to promote interaction between academic research in finance and accounting and applied research in the financial community and the accounting profession.

The chapters in this volume cover a wide range of topics. In this volume there are 12 chapters, three of them are corporate finance and debt management: 1. *Collateral Constraints, Debt Management, and Investment Incentives,* 2. *Thirty Years of Canadian Evidence on Stock Splits, Reverse Stock Splits, and Stock Dividends*, and 3. *Corporate Capital Structure and Firm Value: A Panel Data Evidence From Australia's Dividend Imputation Tax System*. There are two of the other nine chapters which cover earnings management: 1. *Why is the Value Relevance of Earnings Lower for High-Tech Firms?* and 2. *Earnings Management in Corporate Voting: Evidence from Anti-Takeover Charter Amendments*.

Three of the other seven chapters discuss equity markets: 1. *Evaluating the Robustness of Market Anomaly Evidence,* 2. *Intraday Volume–Volatility Relation of the DOW: A Behavioral Interpretation*, and 3. *Determinants of Winner–Loser Effects in National Stock Markets*. Two of the other four chapters analyze options and futures: 1. *The Pricing of Initial Public Offerings: An Option Apporach* and 2. *The Momentum and Mean Reversion Nikkei Index Futures: A Markov Chain Analysis*.

The remaining two chapters are related to portfolio diversification and quadratic programming: 1. *A Concave Quadratic Programming Marketing Strategy Model with Product Life Cycles* and 2. *Corporate Capital Structure and Firm Value: A Panel Data Evidence from Australia's Dividend Imputation Tax System*. In sum, this annual publication covers corporate finance and debt management, earnings management, options and futures, equity market, and portfolio diversification. Therefore, the material covered in this publication is very useful for both academician and practitioner in the area of finance.

This page intentionally left blank

Contents

List of Contributors

Chapter 1

Elettra Agliardi
Department of Economics
University of Bologna
P.zza Scaravilli 2, 40126 Bologna (Italy)
Tel: +39 0512098668
Email: elettra.agliardi@unibo.it

Rainer Andergassen
Department of Economics
University of Bologna
P.zza Scaravilli 2, 40126 Bologna (Italy)
Tel: +39 0512098666
E-mail: rainer.andergassen@unibo.it

Chapter 2

Paul Y. Kim
College of Business Administration
Clarion University of Pennsylvania
Clarion, PA 16214
Tel: (814) 393-2630
Fax: (814) 393-1910
Email: pkim@clarion.edu

Chin W. Yang
College of Business Administration
Clarion University of Pennsylvania
Clarion, PA 16214
Tel: (814) 393-2609
Fax: (814) 393-1910
Email: yang@clarion.edu

Cindy Hsiao-Ping Peng
Yu Da College of Business, Taiwan
Tel: 011886 34226134
Email: s1444001@yahoo.com.tw

Ken Hung
Department of Finance
National Dong Hua University
Hua-lien, Taiwan
Tel: 360 715 2003
Email: hung_ken@yahoo.com

Chapter 3
William D. Brown, Jr.
Department of Business Administration
Stonehill College
Easton, MA 02357
Tel: (508) 565-1256
Fax: (508) 565-1444
Email: wbrown@stonehill.edu

Erin A. Moore
Department of Accounting
College of Business and Economics
Lehigh University
621 Taylor Street
Bethlehem, Pennsylvania 18015-3117
Tel: (610) 758-4962
Fax: (610) 758-6429
Email: erin.moore@lehigh.edu

Ray J. Pfeiffer, Jr.
Isenberg School of Management
Department of Accounting and Information Systems
University of Massachusetts
Amherst, Massachusetts 01003
Email: pfeiffer@acctg.umass.edu
Tel: (413) 545-5653
Fax: (413) 545-3858

Chapter 4

B. Brian Lee
College of Business
P.O. Box 519, MS-2310
Prairie View A&M University
Prairie View, TX 77446
Tel: 936.261.9258
Email: brlee@pvamu.edu

Eric Press
Fox School of Business
335 Speakman Hall
Temple University
Philadelphia, PA 19122
Tel: 215.204.8127
Email: eric.press@temple.edu

B. Ben Choi
91-1712 Newton Street
Victoria, BC V8R 2R2
Tel: 250.598.8717

Chapter 5

Vijay Jog
Professor of Finance
Eric Sprott School of Business
Carleton University
1125 Colonel By Drive
Ottawa, Ontario, Canada, K1S 5B6
Tel: (613) 520-2600 Ext. 2377
Email: vjog@ccs.carleton.ca

PengCheng Zhu
Ph.D. candidate
Eric Sprott School of Business
Carleton University
Email: pzhu@connect.carleton.ca

Chapter 6

Ali F. Darrat
Department of Economics and Finance
Louisiana Tech University
Ruston, LA 71272

Shafiqur Rahman
School of Business Administration
Portland State University
P. O. Box 751
Portland, OR 97207-0751
Tel: (503) 725-3715
Fax: (503) 725-5850
Email: rahmans@pdx.edu

Maosen Zhong
UQ Business School
The University of Queensland
Brisbane, QLD 4072, Australia

Chapter 7

Sheen Liu
Faculty of Finance
Washington State University–Vancouver
Vancouver, WA 98686
Tel: 360 546-9516
Fax: 360 546-9037

Chunchi Wu
Lee Kong Chian School of Business
Singapore Management University
50 Stamford Road
#04-01 Singapore 178899
Email: ccwu@smu.edu.sg

Peter Huaiyu Chen
Department of Accounting and Finance
Youngstown State University
One University Plaza
Youngstown, OH 44555
Tel: 330 941-1883
Email: hchen01@ysu.edu

Chapter 8

Ming-Shiun Pan
Department of Finance and Supply Chain Management
Shippensburg University
Shippensburg, PA 17257
Tel: 717-477-1683
Fax: 717-477-4067
Email: mspan@ship.edu

Chapter 9

Chun-Keung Hoi
Rochester Institute of Technology
106 Lomb Memorial Drive
Rochester, NY 14623-5608
Phone: (585)-475-2718
Fax: (585)-475-6920
Email: ckhoi@saundersrit.edu

Michael Lacina
University of Houston-Clear Lake
2700 Bay Area Boulevard
Houston, TX 77058-1098
Tel: (281) 283-3171
Fax: (281) 283-3951
Email: Lacina@uhcl.edu

Patricia L. Wollan
Rochester Institute of Technology
106 Lomb Memorial Drive
Rochester, NY 14623-5608
Phone: (585)-475-4419 (Phone)
Fax: (585)-475-6920 (Fax)
Email: pwollan@saundersrit.edu

Chapter 10

Herbert E. Phillips
Professor of Finance
Fox School of Business and Management
Temple University
Philadelphia, Pennsylvania 19122
Tel: 215 204 8141
Email: hep@temple.edu

Chapter 11

Abu Taher Mollik
(PhD) Applied Finance
School of Commerce
University of South Australia
GPO Box 2471
Adelaide SA 5001, Australia
Tel: +61 8 8302 0526
Fax: +61 8 8302 0992
Email: abumollik@yahoo.com.au or Abu.Mollik@unisa.edu.au

Chapter 12

Ke Peng
School of Management
University of Bradford
Emm Lane, Bradford
BD9 4JL, United Kingdom
Tel: +44 (0)1274 234352
Email: k.peng1@bradford.ac.uk

Shiyun Wang
China Finance Data Center
Southwestern University of Finance and Economics
610074, P. R. China
Tel: +86 (0)28 87099197
Email: swang@swufe.edu.cn

Chapter 1

Collateral Constraints, Debt Management, and Investment Incentives

Elettra Agliardi and Rainer Andergassen
University of Bologna, Italy

This chapter analyses the hedging decisions of an emerging economy which is exposed to market risks and whose debt contract is subject to collateral constraints. Within a sovereign debt model with default risk and endogenous collateral, the optimal choice of hedging instruments are studied when both futures and nonlinear derivatives are available. It is examined in which way the hedging policy is affected by the cost of default and the financial constraints of the economy and some implications are provided in terms of resource allocation.

Keywords: Hedging strategies; financial constraints; default cost; endogenous collateral; emerging markets.

1. Introduction

Emerging markets have been exposed to remarkable market risks and it is by now folk wisdom that, if given a choice, they should be endowed with instruments of hedging against downside risks (see Caballero, 2003; Caballero and Panageas, 2003; Shiller, 2003). Finding out which factors are the fundamental source of volatility for each country — for example, the prices of oil for Mexico, of coffee for Brazil, of semiconductors for Korea, of copper for Chile, and so on — is recognized as a crucial step in order to construct the appropriate hedging instruments, which will be contingent on observable variables (Caballero, 2003). Yet, it remains to be answered the question concerning the proper application of derivative securities that can be used to construct hedging strategies and the optimal hedging policy. The purpose of this chapter is to examine the hedging decisions of an economy which is exposed to market risks and is subject to collateral constraints. The model considered here is a sovereign debt one, with default risk and endogenous collateral.

Collateral is typically used to secure loans. Since the article by Kiyotaki and Moore (1997), it has been pointed out that if collateral is endogenous, then the debt capacity of firms is altered, causing fluctuations in output (Krishnamurthy, 2003). In this chapter, a model is discussed where the use of hedging instruments may affect collateral values and thus, the debt capacity of the debtor.

In most literature relating to the 1980s debt crisis and following the Bulow and Rogoff models (1989, 1991), a given proportion of output or exports are assumed to be available for repayment of outstanding debt. This means that repayment is modeled as an output tax and actual repayment is the minimum of this amount and debt. Alternatively, in other models (Eaton and Gersowitz, 1981; Eichengreen, 2003; Thomas, 2004) a fixed sanction is established in the case of default, which is not a direct claim on the country's current resources and is not received by the creditors, but may represent the future losses due to diminished reputation. In this chapter, a model is developed where the amount of repayment by the debtor country is determined endogenously by an optimizing choice of the debtor and where the two above mentioned aspects of the repayment contract are present. Indeed, the debt contract is a collateralized one, where profits on internationally tradable goods can be used for repayment, constituting the endogenous collateral; additionally, in the case of default, a sanction is imposed which affects nontradable goods, which represents the cost to the debtor of defaulting. Within this framework, hedging may be driven by the desirability to reduce expected default costs. As Smith and Stulz (1985) have shown, by hedging a debtor is able to reduce the likelihood of default by increasing the income it gets in the downside.

The present chapter is most related to the literature on risk management. Recently, a few articles have studied the optimal choice of hedging instruments of a firm when either futures or options are available. It has been shown that in the model of competitive firms with output price uncertainty, where all input decisions are made simultaneously prior to resolution of uncertainty, hedging with futures does provide a perfect hedge and there is no scope for nonlinear instruments such as options as pure hedging instruments. Albuquerque (2003) characterizes optimal currency hedging in three cases, namely in the presence of bankruptcy costs, with a convex tax schedule, and in the case of a loss-averse manager. In all these cases, he shows that futures dominate options as hedging instruments against downside risk. Batterman *et al.* (2000) study the optimal choice of hedging instruments of an exporting firm exposed to exchange rate risk, when both currency futures and standard options are available. They show that the hedge effectiveness of futures is larger than that of options.

Wong (2003) studies the optimal hedging decision of an exporting firm which faces hedgeable exchange rate risk and nonhedgeable price risk, when price and exchange rate risk have a multiplicative nature. This source of nonlinearity creates a hedging demand for nonlinear payoff currency options distinct from that for linear payoff currency futures. Moschini and Lapan (1992) analyze the problem of hedging price risk under production flexibility, yielding nonlinearity of profits in output price, and show that there is a role for options even when the use of futures is allowed. In Froot *et al.* (1993) it is shown that firms may decide not to hedge fully, if there is correlation between investment opportunities and the availability of funds; moreover, options may be needed in addition to futures to implement the optimal hedge when there are state-dependent financing opportunities.

In this chapter optimal investment and hedging decisions are characterized. It is shown that the decision to use nonlinear hedging strategies in addition to futures contracts can be optimal in relation to market conditions and financial constraint of the economy. In particular, it is shown in which way the optimal hedging decision is affected by the cost of default. In addition to a short position in futures, either concave or convex hedging with options is optimal, depending on the size of default costs. In particular, it is found that if default costs are sufficiently large, options are used for financing purposes, that is, to increase financial resources when these are needed for investment purposes. If default costs are sufficiently low, options are employed for speculative motives, i.e., financial resources are reduced when they are needed for investment purposes. The present results are thus closely related to those of Adam (2002, 2004) who shows how firms employ nonlinear hedging strategies to match financial resources against financial needs at different time periods.

The remainder of the chapter is organized as follows. Section 2 describes the model and the hedging problem of the economy. Section 3 contains the optimal hedging choices of a futures and straddles. Section 4 concludes. All proofs are in the Appendix.

2. The Model

The model is a two-period model of sovereign debt with default risk.[1] Consider an economy having access to a technology producing an internationally tradable and a nontradable good, denoted by y_{T} and y_{NT}, respectively. In the

[1] For a survey of the literature about sovereign debt, see Eaton and Fernandez (1995), in *Handbook of International Economics*, Grossman and Rogoff (eds.). Amsterdam: Elsevier.

production quasifixed inputs (e.g., capital goods) and variable inputs (e.g., labor) are used. The economy has no initial endowments. Thus, in order to produce, firms have to borrow capital from abroad. Borrowing is done with collateralized one-period-ahead debt contract in order to purchase and use in the production functions $k + z$ units of capital, where k and z are the units of capital employed in the production of y_{NT} and y_{T}, respectively. Only the internationally tradable good can be used as a collateral.

At time 1 the price of the internationally tradable good p is not known with certainty and the economy must commit to production plans by choosing the level of investment z and k in capital goods. The price of the nontradable good is known, constant over time.

In what follows, it is assumed that at time 1 producers can take positions in the futures market and in the option market to hedge their exposure. At time 2 uncertainty is resolved and the economy chooses the level y_{T} (y_{NT}) conditional on z (k) and on the open futures and options positions determined at time 1. The risk free interest rate is normalized to 0.

2.1. *Time 2*

At time 2, when price uncertainty is resolved, the usual profit maximization yields:

$$g(z, p) = \max_{y_{\mathrm{T}}}\{py_{\mathrm{T}} - c_1(y_{\mathrm{T}}, z)\}$$

where $c_1(y_{\mathrm{T}}, z)$ is the variable cost function which is conditional on the level of z. In what follows, it is assumed that the production function is $y_{\mathrm{T}} = \tilde{A}z^{\frac{\beta}{2}}L^{\frac{1}{2}}$, where L is labor and $0 < \beta < 1$. Therefore, $g(z, p) = p^2 A z^{\beta}$.

It is assumed that in the case of default, a sanction is imposed exogenously which leads to a reduction of $(1-\tilde{\alpha})\%$ of nontradable goods, with $1 \geq \tilde{\alpha} > 0$. Let q be the constant price of the nontradable good. The production problem of the nontradable good y_{NT} at time 2 is given as follows:

$$\begin{aligned}
\phi^1(k) &= \max_{y_{\mathrm{NT}}}\left\{qy_{\mathrm{NT}} - c_2(y_{\mathrm{NT}}, k)\right\} \quad \text{in case of no default} \\
\phi^2(k, \tilde{\alpha}) &= \max_{y_{\mathrm{NT}}}\left\{\tilde{\alpha}qy_{\mathrm{NT}} - c_2(y_{\mathrm{NT}}, k)\right\} \quad \text{in case of default}
\end{aligned}$$

where $c_2(y_{\mathrm{NT}}, k)$ is a twice continuously differentiable function with positive first and second derivative in y_{NT} and $c_2(0, k) = 0$. To simplify the exposition, the following production function $y_{\mathrm{NT}} = \tilde{B}k^{1-\eta}L^{\eta}$ has been considered, where $1 > \eta > 0$, and consequently $\phi^1(k) = Bk$ and $\phi^2(k, \alpha) = \alpha Bk$, with $\alpha = (q\tilde{\alpha})^{\frac{1}{1-\eta}}$, $1 \geq \alpha > 0$.

Consumption occurs in period 2. Consumers are risk-neutral and gain utility just from the consumption of the nontradable good. Thus, maximizing aggregate utility corresponds to maximizing k.

2.2. *Time 1*

At time 1 the country borrows from foreign creditors funds to purchase and use $k + z$ units of capital. Since there are only two periods, the loan has to be paid back at time 2. All debt contract has to be collateralized. Let r be the repayment price per unit of capital. Let x represent the futures position ($x > 0$ is short) and s the straddle[2] position ($s > 0$ is short) that firms take to hedge the risk associated with price uncertainty. Denote the random profit of the economy at time 1 by:

$$\pi(p) = p^2 A z^\beta - rz - rk + (f - p)x + (t - v)s \tag{1}$$

where $f = E(p)$, $t = E(|p - p^*|)$, and $v = |p - p^*|$, where p^* is the strike price. Then, the collateral constraint requires $\pi(p) \geq 0$. Notice that for $s > 0$, i.e., a short position in straddles, the economy increases its financial resources available for investment in the first period at the cost of reducing them in the second period, while for $s < 0$, i.e., a long position in straddles, the opposite occurs. Since in the present model the economy has no initial endowments, for $s > 0$ straddles are used for financing purposes since shortening straddles reduces financial constraints in the first period where investment decisions have to be taken. For $s < 0$ straddles are used for speculative purposes since financial resources are reduced when these are needed for investment purposes, while financial constraints are alleviated in the second period when repayments are due. The same argument holds true for short and long positions in futures.

Given the collateral constraint, at time 1 when the price uncertainty has not been solved yet, the problem is specified as follows:

$$\max_{k,z,x,s} \Omega(k, \alpha, \chi) \equiv Bk[1 - (1 - \alpha)(1 - \chi)] \tag{2}$$

[2] A long/short straddle is a portfolio which consists of a long/short put and a long/short call on the same asset with the same strike price and exercise time.

where $\chi = \int_P I_{\pi(p)\geq 0}\psi^*(p)\, dp$, $I_{\pi(p)\geq 0}$ is an indicator function, $\psi^*(p)$ is the probability density function of the price of y_T, defined over the set P. For simplicity,[3] $p = \overline{p} + \varepsilon$ is defined, where $E(\varepsilon) = 0$ and assume that $\varepsilon \in [-\overline{p}, \overline{p}]$ and is symmetrically and uniformly distributed, with probability density function $\psi(\varepsilon) = \frac{1}{2\overline{p}}$. It is assumed that $p^* = \overline{p}$. Thus, $f = \overline{p}, t = \frac{\overline{p}}{2}$, and $v = |\varepsilon|$.

2.3. *Benchmark*

Consider the case where the price of the collateral is known with certainty, and equal to its average value, i.e., $p = \overline{p}$, where $\overline{p} = E(p)$. The problem reduces to:

$$\max_z \{\overline{p}^2 A z^\beta - rz\}$$

From the first-order condition $z^0 = \left(\frac{\beta_p^{-2} A}{r}\right)^{\frac{1}{1-\beta}}$ is obtained and thus, optimal k is obtained from condition $\pi(\overline{p}) = 0$ which yields $k^0 = \frac{1-\beta}{\beta} z^0$.

3. Optimal Hedging

Since $g(z, p)$ is quadratic in p,

$$\pi(\varepsilon) = \overline{p}^2 A z^\beta - r(z+k) + [2\overline{p} A z^\beta - x]\varepsilon + A z^\beta \varepsilon^2 + \left(\frac{\overline{p}}{2} - |\varepsilon|\right) s$$

Since ε is symmetrically distributed over the set $[-\overline{p}, \overline{p}]$, π can be rewritten considering only positive values of ε. Thus, for $\varepsilon \geq 0$,

$$\pi(\varepsilon) = \overline{p}^2 A z^\beta - rz - rk + \frac{\overline{p}}{2} s + [2\overline{p} A z^\beta - x - s]\varepsilon + A z^\beta \varepsilon^2$$

$$\pi(-\varepsilon) = \overline{p}^2 A z^\beta - rz - rk + \frac{\overline{p}}{2} s - [2\overline{p} A z^\beta - x - s]\varepsilon + A z^\beta \varepsilon^2$$

The following result can be obtained.

Proposition 1. *A short futures position* $x = g_p(z, \overline{p}) = 2\overline{p} A z^\beta$ *is optimal.*

---

[3]The assunptions of a symmetric distribution of prices and of a profit function quadratic in price is also in Moschini and Laplan (1992, 1995), where they show that futures and options have a role in hedging price risk.

Optimality requires a short position in futures equal to $2\overline{p}Az^{\beta}$. Thus, a short futures position increases the funds available at time 1 for investment purposes. Moreover, the future position does not depend on the cost of default α.

For $x = 2\overline{p}Az^{\beta}$, $\pi(-\varepsilon) = \pi(\varepsilon)$ is obtained, where:

$$\pi(\varepsilon) = \overline{p}^2 Az^{\beta} - r(z+k) + \left(\frac{\overline{p}}{2} - \varepsilon\right)s + Az^{\beta}\varepsilon^2$$

$\pi(\varepsilon) \geq 0$ for values external to the two roots:

$$\varepsilon_{1,2} = \frac{s \pm \sqrt{s^2 - 4[\overline{p}^2 Az^{\beta} - r(z+k) + \frac{\overline{p}}{2}s]Az^{\beta}}}{2Az^{\beta}} \tag{3}$$

where $\delta = \frac{s}{s^*}$ and $s^* = \overline{p}Az^{\beta}$. It is assumed that only a finite amount of straddles are available on the market. This corresponds to imposing upper and lower bounds on δ, i.e., $|\delta| \leq \overline{\delta}$. To find a solution to problem (2) it proceeds in two steps. First, using the first-order condition for z, the optimal level of capital k which yields a given probability of default c is found, where $c \in [0, 1]$. In this way k is obtained as a function of c and δ. The payoff function in (2) can be rewritten as:

$$\Omega(c, \delta) = k(c, \delta)[1 - (1-\alpha)c] \tag{4}$$

In the second step, the optimal position in straddles and the optimal probability of default $c \in [0, 1]$ are found. From (4) it is observed that maximizing the payoff function with respect to δ reduces to maximizing $k(c, \delta)$ over appropriate values of δ, for each given c. Subsequently, it can be shown (see the Appendix) that $k(c, \delta^*)$, where δ^* is the optimal value of δ, is an increasing function of c. Thus, in maximizing the payoff function with respect to c, the economy has to trade-off a larger expected punishment due to default against larger values of k. The size of the expected punishment depends on the value of α. The larger this value is, the lower is the punishment in the case of default. Consequently, the solution to this trade-off depends on the size of α.

The following result can be obtained.

Proposition 2. *There exists a critical level* $\alpha^*(\beta, \overline{\delta})$ *such that for* $0 \leq \alpha < \alpha^*(\beta, \overline{\delta})$ *the optimal choice is* $\delta = 1$ *and* $c = 0$, *while for* $\alpha^*(\beta, \overline{\delta}) < \alpha \leq 1$ *the optimal choice is* $\delta = -\overline{\delta}$ *and* $c \in (\frac{1}{2}, 1]$, *where* $\alpha^*(\beta, \overline{\delta})$ *is a decreasing* *function of β and $\overline{\delta}$ and is strictly positive for $\beta < \beta(\overline{\delta})$ and* 0 *otherwise, where $\beta'(\overline{\delta}) < 0$.*

Proposition 2 states that optimality requires nonlinear hedging. For sufficiently low values of α, i.e., sufficiently large costs of default, optimality requires a short position of $s^* \equiv \overline{p}Az^{\beta}$ straddles. Moreover, in this regime, the economy is induced never to default. The intuition for this result is as follows. Short selling straddles increases financial resources available for investment in the first period while it increases financial constraints in the second period. Thus, if default costs are sufficiently large, borrowing constraints are tighter, and thus the economy uses straddles to reduce these constraints in the first period and chooses not to default. Thus, in this regime straddles are used for financing purposes. For sufficiently large values of α, i.e., sufficiently low costs of default, optimality requires a long position of $s = -\overline{\delta}\overline{p}Az^{\beta}$. Moreover, in this regime, the economy is induced to default with a probability larger than $\frac{1}{2}$. In this regime default costs are low and consequently financial constraints in the first period and borrowing constraints are loose. Thus, in this regime straddles are employed for speculative motives and furthermore the country will default with a probability larger than $\frac{1}{2}$.

Thus, the event of default can be avoided for $\beta < \beta(\overline{\delta})$, chosing an α lower than $\alpha^*(\beta, \overline{\delta})$.

Corollary 1. *The optimal investment in k is an increasing function of α.*

The above mentioned optimal hedging strategies have direct implication in terms of resource allocation for the economy. It is straightforward to prove the following.

Corollary 2. *There is overinvestment in k, z with respect to the benchmark case.*

4. Conclusion

This chapter shows how financially constrained economies should hedge. It thus extends the literature on risk management that shows why firms hedge and which are the optimal hedging instruments, and the contributions on emerging markets, which point out that if collateral is endogenous, then the debt capacity of an economy is altered.

Within a sovereign debt model with default risk and endogenous collateral, the optimal choice of hedging instruments is studied when both futures and nonlinear derivatives are available. It is shown that in addition to futures, optimality requires either concave or convex hedging, depending on the size of the default cost. If this latter is sufficiently large, then optimality requires a short position in straddles and furthermore, the economy is induced never to default. If the default cost is sufficiently low, then optimility requires a long position in straddles and the economy is induced to default with a probability larger than $\frac{1}{2}$.

Appendix

Proof of Proposition 1. $\pi(\varepsilon) \geq 0$ for values external to the two roots

$$\varepsilon^{+}_{1,2} = \frac{-(2\overline{p}Az^{\beta} - x - s) \pm \sqrt{\begin{array}{c}(2\overline{p}Az^{\beta} - x - s)^2 - 4[\overline{p}^2 Az^{\beta} \\ -r(z+k) + \frac{\overline{p}}{2}s]Az^{\beta}\end{array}}}{2Az^{\beta}}$$

while $\pi(-\varepsilon) \geq 0$ for values external to the two roots

$$\varepsilon^{-}_{1,2} = \frac{2\overline{p}Az^{\beta} - x + s \pm \sqrt{\begin{array}{c}(2\overline{p}Az^{\beta} - x + s)^2 - 4[\overline{p}^2 Az^{\beta} \\ -r(z+k) + \frac{\overline{p}}{2}s]Az^{\beta}\end{array}}}{2Az^{\beta}}$$

Maximizing[4] (2) with respect to x yields:

$$\frac{\partial \varepsilon^{+}_{1}}{\partial x} - \frac{\partial \varepsilon^{+}_{2}}{\partial x} + \frac{\partial \varepsilon^{-}_{1}}{\partial x} - \frac{\partial \varepsilon^{-}_{2}}{\partial x} = 0 \tag{5}$$

Expression (5) is satisfied if $x = 2\overline{p}Az^{\beta}$.

[4]For simplicity of exposition the case where all roots exists and are included in the interval $[-\overline{p}, \overline{p}]$ has been considered here. The result remains the same also in the other cases.

Proof. Three cases arise. Case 1: $\overline{p} \geq \varepsilon_{1,2} \geq 0$; case 2: $\overline{p} \geq \varepsilon_1 \geq 0$ and $\varepsilon_2 < 0$; case 3: $\overline{p} \geq \varepsilon_2 \geq 0$ and $\varepsilon_1 > \overline{p}$. Using the definition of δ, (3) and the probability of default c, these conditions can be redefined as: case 1: $c \leq \delta \leq 2 - c$; case 2: $-\overline{\delta} \leq \delta < c$; and case 3: $\overline{\delta} \geq \delta > 2 - c$. □

Case 1

Result A1. *Given the probability of default* $c \in [0, 1]$, *for each* $c \leq \delta \leq 2-c$, *the optimal strategy is* $\delta = 1$, $k = \frac{1-\beta}{\beta}z$ *and*:

$$k(c, 1) = \frac{1-\beta}{\beta}\left(\frac{\beta A}{r}\overline{p}^2\frac{5+c^2}{4}\right)^{\frac{1}{1-\beta}} \tag{6}$$

Using the definition of δ, the first-order condition for z requires:

$$k(\delta) = \frac{1-\beta}{\beta}z - \frac{\delta(\delta-1)}{2r}\overline{p}^2 A z^\beta \tag{7}$$

Now by holding the probability of default constant, the optimal strategy δ can be found. Using (3) and (7), the probability of default $c = \frac{\varepsilon_1 - \varepsilon_2}{\overline{p}}$ yields $z(c, \delta) = \left(\frac{\beta A}{r}\frac{4+\delta^2+c^2}{4}\overline{p}^2\right)^{\frac{1}{1-\beta}}$. Thus, for $z(\delta)$ and the corresponding value of $k(7)$ the probability of default is c. The maximum payoff, subject to the condition of a constant probability of default, is obtained maximizing k as in (7) over values of δ, i.e.,

$$\max_\delta k(c, \delta) = \left(\frac{\beta A}{r}\overline{p}^2\frac{4+\delta^2+c^2}{4}\right)^{\frac{1}{1-\beta}}\left[\frac{1-\beta}{\beta} - \frac{\delta(\delta-1)}{\beta}\frac{2}{4+\delta^2+c^2}\right]$$

which yields $\delta = 1$.

Thus, the problem is reduced to find the optimal level of c,

$$\max_{c\in[0,\frac{1}{2}]} \Omega(c, 1) \equiv B\frac{1-\beta}{\beta}\left(\frac{\beta A}{r}\overline{p}^2\frac{5+c^2}{4}\right)^{\frac{1}{1-\beta}}[1-(1-\alpha)c] \tag{8}$$

Case 2

Result A2. *For each given* $c \leq \frac{1}{2}$, $-\overline{\delta} \leq \delta < c$ *is never optimal, while for* $c > \frac{1}{2}$ *it is optimal to choose* $\delta = -\overline{\delta}$ *and the corresponding capital level is*

$$k(c, -\overline{\delta}) = \left[\frac{\beta A}{r}\overline{p}^2(c^2+1)\right]^{\frac{1}{1-\beta}}\left(\frac{1-\beta}{\beta} + \frac{\overline{\delta}}{\beta}\frac{c-\frac{1}{2}}{c^2+1}\right) \tag{9}$$

From the first-order conditions of z:

$$k_{1,2} = z\frac{1-\beta}{\beta} + \frac{s}{r}\left(\frac{\overline{p}}{2} \pm \sqrt{\frac{r}{\beta A} z^{1-\beta} - \overline{p}^2}\right) \tag{10}$$

For a given probability of default c, simple algebra shows that:

$$k_1(c, \delta) = \left\{\frac{\beta A}{r}\overline{p}^2[(c-\delta)^2 + 1]\right\}^{\frac{1}{1-\beta}} \left[\frac{1-\beta}{\beta} + \frac{\delta}{\beta}\frac{\frac{1}{2} + c - \delta}{(c-\delta)^2 + 1}\right]$$

$$k_2(c, \delta) = \left[\frac{\beta A}{r}\overline{p}^2(c^2 + 1)\right]^{\frac{1}{1-\beta}} \left[\frac{1-\beta}{\beta} + \frac{\delta}{\beta}\frac{\frac{1}{2} - c}{1 + c^2}\right]$$

For $c \leq \frac{1}{2}$, inspection shows that $k_1(c, \delta) < k_2(c, \delta)$ and further $k_2(c, \delta)$ is increasing in δ and thus the maximum is achieved in $\delta = c$. Furthermore $k_2(c, c)$ is increasing in c, and thus $k(\frac{1}{2}, \frac{1}{2}) = \frac{1-\beta}{\beta}\left(\frac{\beta A}{r}\overline{p}^2\frac{5}{4}\right)^{\frac{1}{1-\beta}} < \frac{1-\beta}{\beta}\left(\frac{\beta A}{r}\overline{p}^2\frac{5+c^2}{4}\right)^{\frac{1}{1-\beta}} = k(c, 1)$. Consequently, if $c \leq \frac{1}{2}$ is optimal, then $\delta = 1$ is optimal. For $c > \frac{1}{2}$ it is observed that $\frac{\partial}{\partial\delta}k_2(c, \delta) < 0$ and further that $k_2(c, -\overline{\delta}) > k_1(c, \delta)$ for each $\delta \in [-\overline{\delta}, c]$.

Case 3

Result A3. *For each given* $0 \leq c \leq \frac{1}{2}$, $\overline{\delta} \geq \delta > 2 - c$ *is never optimal, while for* $c > \frac{1}{2}$ *it is optimal to choose* $\delta = \overline{\delta}$ *and the corresponding capital level is*

$$k(c, \overline{\delta}) = \left\{\frac{\beta A}{r}\overline{p}^2[1 + (1-c)^2]\right\}^{\frac{1}{1-\beta}} \left[\frac{1-\beta}{\beta} + \frac{\overline{\delta}}{\beta}\frac{c - \frac{1}{2}}{1 + (1-c)^2}\right] \tag{11}$$

From the first-order conditions of z, (10) is obtained and consequently, for a given probability of default c, simple algebra shows that

$$k_1 = \left\{\frac{\beta A}{r}\overline{p}^2[1 + [\delta - (1-c)]^2]\right\}^{\frac{1}{1-\beta}} \left[\frac{1-\beta}{\beta} + \frac{\delta}{\beta}\frac{\frac{3}{2} - \delta - c}{1 + [\delta - (1-c)]^2}\right]$$

$$k_2 = \left\{\frac{\beta A}{r}\overline{p}^2[1 + (1-c)^2]\right\}^{\frac{1}{1-\beta}} \left[\frac{1-\beta}{\beta} + \frac{\delta}{\beta}\frac{c - \frac{1}{2}}{1 + (1-c)^2}\right]$$

For each given $c \leq \frac{1}{2}$, $\frac{\partial}{\partial\delta}k_{1,2} \leq 0$ and consequently the maximum value of $k_{1,2}$ is obtained in $\delta = 2 - c$. Simple inspection shows that for each $c \leq \frac{1}{2}$,

$k_2(c, 2-c) \geq k_1(c, 2-c)$. Furthermore, $k(c, 2-c)$ is increasing in c and $k\left(\frac{1}{2}, 2-\frac{1}{2}\right) = \frac{1-\beta}{\beta}\left(\frac{\beta A}{r}\overline{p}^2\frac{5}{4}\right)^{\frac{1}{1-\beta}} < \frac{1-\beta}{\beta}\left(\frac{\beta A}{r}\overline{p}^2\frac{5+c^2}{4}\right)^{\frac{1}{1-\beta}} = k(c, 1)$.

For $c > \frac{1}{2}$ it is observed that $\frac{\partial}{\partial\delta}k_2 > 0$ and further that $k_2(c, \overline{\delta}) > k_1(c, \delta)$, for each $\delta \in [2-c, \overline{\delta}]$. It is now possible to prove Proposition 2. First, notice that as for each $\overline{\delta} \geq 1$, $k(c, -\overline{\delta}) > k(c, \overline{\delta})$. Consequently the country prefers to buy straddles instead of shortening them, i.e., $\Omega(-\overline{\delta}, 1) > \Omega(1, 1)$. Furthermore observe that, applying the envelope theorem, $\frac{\partial}{\partial\overline{\delta}}\Omega(-\overline{\delta}, c) > 0$, $\frac{\partial}{\partial\alpha}\Omega(-\overline{\delta}, c) > 0$ and $\frac{\partial}{\partial\alpha}\Omega(1, c) > 0$.

Consider the case of $\alpha = 1$ where no punishment occurs in the case of default. Since the optimal amount of capital $k(c, \delta)$ is increasing in c, it is always optimal to choose $c = 1$. Since $\Omega(-\overline{\delta}, 1) > \Omega(1, 1)$ for each $\overline{\delta} \geq 0$, a long position in straddles is optimal.

Consider the case of $\alpha = 0$. Since $\Omega(-\overline{\delta}, 1) = 0$ and $\frac{\partial}{\partial c}\Omega\left(-\overline{\delta}, \frac{1}{2}\right) > 0$, the optimal value of c is obtained in $c \in \left(\frac{1}{2}, 1\right)$. Let $c_{\mathrm{L}} = \arg\max_c \Omega(-\overline{\delta}, c)$ and $c_{\mathrm{S}} = \arg\max_c \Omega(1, c)$, then for $\beta \to 0$ and $\overline{\delta} = 2$, $\Omega(-\overline{\delta}, c_{\mathrm{L}}) < \Omega(1, 0)$. Furthermore, computing $\Omega(-\overline{\delta}, c_{\mathrm{L}})$ and $\Omega(1, c_{\mathrm{S}})$ for all possible values of α, it is observed that there exists a critical level of α such that for all values below this level it is optimal to short straddles ($\delta = 1$), while for values of α above this level it is optimal to buy straddles ($\delta = -\overline{\delta}$). Notice that $\Omega(-\overline{\delta}, c_{\mathrm{L}})$ is increasing in $\overline{\delta}$ and thus the larger $\overline{\delta}$ is, the lower is this critical level.

For $\alpha = 0$, $\frac{\partial}{\partial\beta}\Omega(-\overline{\delta}, c_{\mathrm{L}}) > \frac{\partial}{\partial\beta}\Omega(1, c_{\mathrm{S}})$, for each value of β, and since for $\beta \to 1$, $c_{\mathrm{L}}, c_{\mathrm{S}} \to 1$, and $\frac{\Omega(-\overline{\delta}, c_{\mathrm{L}})}{\Omega(1, c_{\mathrm{S}})} \to \infty$ there exists a critical level of $\beta(\overline{\delta})$ below which $\frac{\Omega(-\overline{\delta}, c_{\mathrm{L}})}{\Omega(1, c_{\mathrm{S}})} < 1$ and above which $\frac{\Omega(-\overline{\delta}, c_{\mathrm{L}})}{\Omega(1, c_{\mathrm{S}})} > 1$. Since $\Omega(-\overline{\delta}, c_{\mathrm{L}})$ is increasing in $\overline{\delta}$, this critical level is decreasing in β. It is known that for $\alpha = 1 \frac{\Omega(-\overline{\delta}, c_{\mathrm{L}})}{\Omega(1, c_{\mathrm{S}})} > 1$ and thus computing for each $\beta < \beta(\overline{\delta})$ and each value of α the payoffs $\Omega(-\overline{\delta}, c_{\mathrm{L}})$ and $\Omega(1, c_{\mathrm{S}})$ it is observed that there exists a critical value of α where $\frac{\Omega(-\overline{\delta}, c_{\mathrm{L}})}{\Omega(1, c_{\mathrm{S}})} = 1$. Since $\Omega(-\overline{\delta}, c_{\mathrm{L}})$ is increasing in $\overline{\delta}$ this critical value is decreasing in $\overline{\delta}$.

Proof of Corollary 1. The result follows from Proposition 2, (9), and from the fact that c_{L} is increasing in α.

Proof of Corollary 2. From Proposition 2 it follows that for $\alpha < \alpha^*$ the equilibrium is $\delta = 1$ and $c = 0$ and thus optimal investment in z is $z = \left(\frac{\beta A}{r}\overline{p}^2\frac{5}{4}\right)^{\frac{1}{1-\beta}} > z^0$. Furthermore, since $k = \frac{1-\beta}{\beta}z$, it follows from $k(c, 1)$ that $k(0, 1) > k^0$. For $\alpha > \alpha^*$ the equilibrium is $\delta = -\overline{\delta}$ and $c \in \left(\frac{1}{2}, 1\right]$ and thus optimal investment in z is $z = \left[\frac{\beta A}{r}\overline{p}^2(1+c^2)\right]^{\frac{1}{1-\beta}} > z^0$. Furthermore, from (9) it follows that $k(c, -\overline{\delta}) > k^0$.

References

Adam, TR (2002). Risk management and the credit risk premium. *Journal of Banking & Finance*, 26, 243–269.

Adam, TR (2004). *Why Firms Use Nonlinear Hedging Strategies*. Mimeo: Hong Kong University of Science & Technology.

Albuquerque, R (2003). Optimal currency hedging. Mimeo, University of Rochester.

Batterman, HL, M Braulke, U Broll and J Schimmelpfennig (2000). The preferred hedge instrument. *Economics Letters*, 66, 85–91.

Bulow, J and K Rogoff (1989). A constant recontracting model of sovereign debt. *Journal of Political Economy*, 96, 155–178.

Bulow, J and K Rogoff (1991). Sovereign debt repurchases: No cure for overhang. *Quarterly Journal of Economics*, 106, 1219–1235.

Caballero, R (2003). The future of the IMF. *American Economic Review, Papers and Proceedings*, 93, 31–38.

Caballero, R and S Panageas (2003). Hedging sudden stops and precautionary recessions: A quantitative approach, Mimeo, MIT.

Eaton, J and M Gersovitz (1981). Debt with potential repudiation, *Review of Economic Studies*, 48, 289–309.

Eichengreen, B (2003). Restructuring sovereign debt. *Journal of Economic Perspectives*, 17, 75–98.

Froot, KA, DS Scharfstein and JC Stein (1993). Risk management: Coordinating corporate investment and financing policies. *The Journal of Finance*, 48, 1629–1658.

Kiyotaki, N and J Moore (1997). Credit cycles, *Journal of Political Economy*, 105, 211–248.

Krishnamurthy, A (2003). Collateral constraints and the amplification mechanism. *Journal of Economic Theory*, 11, 277–292.

Moschini, G and H Lapan (1992). Hedging price risk with options and futures for the competitive firm with production flexibility. *International Economic Review*, 33, 607–618.

Shiller, RJ (2003). *The New Financial Order*. Princeton: Princeton University Press.

Smith, CW and RM Stulz (1985). The determinants of firm's hedging policies. *Journal of Financial and Quantitative Analysis*, 4, 391–405.

Thomas, JP (2004). Default costs, willingness to pay, and sovereign debt buybacks. *Journal of Restructuring Finance*, 1, 35–47.

Wong, KP (2003). Currency hedging with options and futures. *European Economic Review*, 47, 833–839.

This page intentionally left blank

Chapter 2

A Concave Quadratic Programming Marketing Strategy Model with Product Life Cycles

Paul Y. Kim and Chin W. Yang
Clarion University of Pennsylvania, USA

Cindy Hsiao-Ping Peng
Yu Da College of Business, Taiwan

Ken Hung
National Dong Hua University, Taiwan

As a more general approach, the authors formulate a concave quadratic programming model of the marketing strategy (QPMS) problem. Due to some built-in limitations of its corresponding linear programming version, the development of the QPMS model is necessary to further improve the research effort of evaluating the profit and sales impact of alternative marketing strategies. It is the desire of the authors that this study will increase the utilization of programming models in marketing strategy decisions by removing artificially restrictive limitations necessary for linear programming solutions, which preclude the study of interaction effects of quantity and price in the objective function. The simulation analysis of the QPMS and its linear counterpart LPMS indicates that the solutions of the QPMS model are considerably more consistent with *a priori* expectations of theory and real world conditions.

Keywords: Marketing strategy model; concave quadratic programming.

1. Introduction

One of the marketing strategic decisions may involve the optimal allocation of sales force and advertising effort in such a way that a firm maximizes its profit or sales. Efforts designed to evaluate the profit and sales impact of alternative sales force and advertising effort are particularly useful in today's highly competitive marketing environment. The purpose of this chapter is threefold. First, the conventional linear programming marketing strategy (LPMS) model is examined to identify several limitations in marketing strategy problems. Second, a quadratic programming model was formulated to extend and complement the LPMS model of the existing literature in marketing strategy.

Finally, results obtained from both models were compared and critical evaluations are made to highlight the difficulty embedded in the marketing strategy problem. A brief review of the well-known linear programming marketing strategy model is provided prior to describing the quadratic programming model of marketing strategy problem.

In the wake of growing globalization and bubbling electronic commerce, how to match products to market is of primary importance, especially in terms of gaining greater dominance in a market. For example, the Coca-Cola Company attempted to increase its market share from 42% to 50% of the US soft drink market by 2000. The mix of marketing strategies includes lower prices, expanding distribution capacity, and heavier promotional efforts in extolling the products (Frank, 1996). Needless to say, positioning strategies are intended to deliver the value proposition of a product or group of products in the eyes and minds of the targeted buyers. The value requirements are exclusively derived from the buyers. It is said that the success of Dell Computer Corporation can be traced to Michael Dell's strategic vision of high-performance, low-priced personal computers marketed directly to end-users (Kerwin and Peterson, 2001). Another important marketing strategy is the development and management of product life cycle. In the stage of introduction and growth, the emphasis is on trial purchasers and price is typically higher. As the product moves into maturity-saturation stage, the focus is on repeat purchasers with lower prices as sales volume reaches its peak. Regardless of the reasons, be it a market niche or product life cycle, pricing of a product holds the key to the success of a business organization.

2. The Linear Programming Marketing Strategy Model

As is well known, the objective of a marketing manager is often focused on profit maximization[1] given the various constraints such as availability of sales force, advertising budget, and machine hours. Granted that the total profit level after deducting relevant costs and expenses may not increase at a constant rate, however, in a very short time period, profit per unit of output or service facing a firm may well be constant, i.e., the unit profit level is independent of the sales volume. Thus, the manager can solve the conventional linear programming

[1]Other objectives of a firm, other than profit maximization, may be found in the works by Shleifer and Vishny (1988), Navarro (1988), Winn and Shoenhair (1988), and Boudreaux and Holcombe (1989).

marketing strategy (LPMS) model from the following profit-maximization problem:

$$\underset{x_i}{\text{Maximize}} \quad \pi = \sum_{i \in I} P_i x_i \tag{1}$$

subject to

$$\sum_{i \in I} a_i x_i \le A \tag{2}$$

$$\sum_{i \in I} s_i x_i \le S \tag{3}$$

$$\sum_{i \in I} x_i \le k \tag{4}$$

$$\sum_{i \in I} x_i \ge 1_j \quad \text{for some } j \in J \tag{5}$$

$$x_i \ge 0 \tag{6}$$

where $I = \{1, 2, \ldots, n\}$ is an integer index set denoting n different markets or media options; and $J = \{1, 2, \ldots, m\}$ is an integer index set denoting m constraints for some or all different markets.

x_i = unit produced for the ith market or sales volume in the ith distribution channel
P_i = unit profit per x_i
a_i = unit cost of advertising per x_i
A = total advertising budget
s_i = estimated sales force effort per x_i
S = total sales force available
k_i = capacity constraint of all x_i's
l_j = minimum target sales volume of the jth constraint for $j \in J$

We can rewrite Eqs. (1) through (6) more compactly as:

$$\text{Maximize} \quad P'X \tag{7}$$

$$\text{subject to} \quad UX \le V \tag{8}$$

$$X \ge 0 \tag{9}$$

where $P \in R^n$, $U \in R^{m \times n}$, $V \in R^m$, and $X \in R_+^n$, and R_+^n is nonnegative orthant of the Euclidean n-space (R^n), and $R^{m \times n}$ is a class of real m by n matrices. As is well known, such linear programming marketing strategy model contains at least one solution if the constraint set is bounded and convex. The solution property is critically hinged on the constancy of the unit profit level P_i for each market. That is, the assumption of a constant profit level per unit gives rise to a particular set of solutions, which may be inconsistent with the *a priori* expectations of theory and real world situations.

To illustrate the limitations of the LPMS model, it is necessary to perform some simulation based on the following parameters[2]:

$$\text{Maximize } (55, 70, 27, 37)\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \tag{10}$$

subject to

$$\begin{bmatrix} 15 & 20 & 10 & 9 \\ 2 & 8 & 3.5 & 1 \\ 1 & 1 & 1 & 1 \\ -1 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \leq \begin{bmatrix} 27,500 \\ 11,000 \\ 12,500 \\ -270 \end{bmatrix} \tag{11}$$

$$X_i \geq 0 \tag{12}$$

The constraints of advertising budget, sales forces, and machine hours are 27,000, 11,000, and 12,500, respectively and minimum target for market or distribution channel 1 is 270 units. The solution for this LPMS model and its sensitivity analysis is shown in Table 1. It is evident that the LPMS model has the following three unique characteristics.

Table 1. Sensitivity analysis of the LPMS model.

Optimum solution	Original LPMS model $\Delta = 0$	$\Delta P_1 = -5$	$\Delta P_1 = 5$	$\Delta A = -200$	$\Delta A = 200$
π	109,200	107,850	110,550	108,377.8	110,022
x_1	270	270	270	270	270
x_2	0	0	0	0	0
x_3	0	0	0	0	0
x_4	2,250	2,250	2,250	2,527.8	2,572.2
y_1	4.11	4.11	4.11	4.11	4.11
y_2	0	0	0	0	0
y_3	0	0	0	0	0
y_4	6.67	11.67	1.67	6.67	6.67

Note: The simulation is performed using the software package LINDO by Schrage (1984).

[2]The LPMS example is comparable to that by Anderson *et al.* (1982).

First of all, the number of positive-valued decision variables ($x_i > 0$ for some $i \in I$) cannot exceed the number of constraints in the model (Gass, 1985). The lack of positive x_i's (two positive x_i's in our model) in many cases may limit choices of markets or distribution channels to be made by the decision-makers. One would not expect to withdraw from the other two markets or distribution channel (2 and 3) completely without having a compelling reason. This result from the LPMS model may be in direct conflict with such objective as market penetration or market diffusion. For instance, the market of Coca Cola is targeted at different markets via all distribution channels, be it radio, television, sign posting, etc. Hence, an alternative model may be necessary to circumvent the problem.

Second, the optimum x_i's are rather irresponsive to changes in unit profit margin (P_i). For instance, a change in P_1 by 5 units does not alter the primal solutions at all (see Table 1). As a matter of fact, increasing the profit margin of market 1 significantly does not change the optimum x_i's at all. From the most practical point of view, however, management would normally expect that the changes in unit profit margin be highly correlated with changes in sales volumes. In this light, it is evident that the LPMS model may not be consistent with the real-world marketing practice in the sense that sales volumes are irresponsive to the changes in unit profit contribution.

Last, the dual variables (y_j's denote marginal profit due to a unit change in the jth right-hand side constraint) remain unchanged as the right-hand side constraint is varied. It is a well-known fact that incremental profit may very well decrease as, for instance, advertising budget increases beyond some threshold level due to repeated exposure to the consumers (e.g., where is the beef?). If the effectiveness of a promotional activity can be represented by an inverted u curve, there is no compelling reason to consider unit profit to be constant. In the framework of the LPMS model, these incremental profits or y's are irresponsive to changes in the total advertising budget (A) and the profit per unit (P_i) within a given base. That is, $i \in I$ remains unchanged before and after the perturbations on the parameter for some $X_i > 0$ as can be seen from Table 1.

3. A Concave Quadratic Programming Model of the Marketing Strategy Problem

In addition to the three limitations mentioned above, LPMS model assumes average profit per x_i remains constant. This property may not be compatible in most market structures in which the unit profit margin is a decreasing function of sales volumes, i.e., markets of imperfect competitions. As markets are gradually saturated for a given product or service (life cycle of a product), the unit profit would normally decrease. Gradual decay in profit as the market matures seems to be consistent with many empirical observations. Greater profit is normally expected and typically witnessed with a new product. This being the case, it seems that ceaseless waves of innovation might have been driving forces that led to myriad of commodity life cycles throughout the history of capitalistic economy. Besides, positioning along a price-quality continuum is subject to changes of business environment. As competition toughens, positioning may well change. Hewlett-Packard priced its personal computer below Compaq and IBM in an attempt to position firmly among corporate buyers. On the other hand, Johnson & Johnson's Baby Shampoo was repositioned to include adults and the result is a fivefold increase in market share. A change in competitive environment may very well lead to different pricing strategy. For instance, Procter & Gamble began losing sales of its consumer products in the late 1990s. Kimberly-Clark's Scott brand cut into P & G's Bounty market share via cost control, pricing, and advertising (Business Week, 2001). Not until late 2000, did P & G reduce its price increase. As expected, Bounty experienced strong sales increases. It is to be noted that pricing decision is not made solely on the basis of profit maximization. Other objectives such as adequate cash flow play an important role too (Cravens and Piercy, 2003). When a product loyalty is entrenched in consumers' minds, managers would have much more flexibility in setting prices. Gillette's consumers indicated that there was little reduction in quantity demanded for a 45% price increase of MACH 3 above that of SensorExcel (Maremont, 1998). Paired-pricing is yet another example in which price does not stay constant: Toyota Camry and Lexus-ES 300 were priced in relation to each other with the ES 300 targeting the semi-luxury market (Flint, 1991) whereas Camry had much lower prices. For this reason, we would like to formulate an alternative concave quadratic programming (QPMS) model as shown below:

$$\begin{aligned} &\text{Maximize } Z = \sum_{i \in I} (c_i + d_i x_i) x_i = \sum_{i \in I} c_i x_i + \sum_{i \in I} d_i x_i^2 \\ &\text{subject to (2), (3), (4), (5), and (6)} \end{aligned} \tag{14}$$

Or more compactly:

$$\begin{aligned}
\text{maximize } Z &= C'X + X'DX \\
\text{subject to } UX &\leq V \\
X &\geq 0 \\
\text{with } C \in R^n,\ x \in R^n,\ &\text{and } D \in R^{n\times n}
\end{aligned}$$

where D is a diagonal matrix of n by n with each diagonal component $d_i < 0$ for all $i \in I$.

Since the constraint is a convex set bounded by linear inequalities, the constraint qualification is satisfied (Hadley, 1964). The necessary (and hence sufficient) conditions can be stated as follows:

$$\nabla_x L(x*, y*) = \nabla_x Z(x*) - y * \nabla_x U(x*) \leq 0 \tag{15}$$

$$\nabla_x L(x*, y*) \bullet x* = 0 \tag{16}$$

$$\nabla_y L(x*, y*) \geq 0 \tag{17}$$

$$\text{and} \qquad \nabla_y L(x*, y*) \bullet y* = 0 \tag{18}$$

where $L(x*, y*) = Z + Y(V - UX)$ is the Lagrangian equation, and $\nabla_x L$ is the gradient of the Lagrangian function with respect to $x_i \in X$ for all $i \in I$, the * denotes optimum values, and y_j is the familiar Lagrangian multipliers associates with the jth constraint (see Luenberger, 1973, Chap. 10). For example, the first component of (15) would be $c_1 + 2d_1x_1 - a_1y_1 = 0$ for $x_1 > 0$. It implies that marginal profit of the last unit of x_1 must equal the cost of advertising per unit times the incremental profit due to the increase in the total advertising budget. Conditions (15) and (16) imply that equality relations hold for $x_i^* > 0$ for some $i \in I$. Conversely, for some $x_i^* = 0$, i.e., a complete withdrawal from the ith market or distribution channel, this equality relation may not hold. Conditions (17) and (18) imply that if $y_j^* > 0$, then the corresponding advertising, sales force, physical capacity, and minimum target constraints must be binding.

The QPMS model clearly has a strictly concave objective function if D (a diagonal matrix) is negatively definite ($d_i < 0$ for all $i \in I$). With the nonempty linear constraint set, the QPMS model possesses a unique global maximum (Hadley, 1964, Chapters 6 and 7). This property holds as long as the unit profits decrease ($d_i < 0$) as more and more units of outputs are sold through various distribution channels or markets, a phenomenon consistent with empirical findings.

4. Critical Evaluations of the Marketing Strategy Models

To test the property of the QPMS model, the following parameter values[3] were assumed for sensitivity purposes.

$$\begin{aligned} C' &= (5000, 5700, 6600, 6900) \\ X' &= (x_1, x_2, x_3, x_4) \\ D &= \begin{bmatrix} -3 & & & 0 \\ & -2.7 & & \\ 0 & & -3.6 & \\ & & & -4 \end{bmatrix} \end{aligned}$$

The total profit function $C'X + X'DX$ is to be maximized, subject to the identical constraints (11) and (12) in the LPMS model. By doing so, both LPMS and QPMS models can be evaluated on the comparable basis. The optimum solution to this QPMS model is presented in Table 2 to illustrate the difference.

First, with the assumption of a decreasing unit profit function, the number of markets penetrated or the distribution channels employed ($x_i > 0$) in

Table 2. Sensitivity analysis of the QPMS model.

Optimum solution	Original QPMS model $\Delta = 0$	$\Delta C_1 = -100$	$\Delta C_1 = 100$	$\Delta A = -200$	$\Delta A = 200$
z	9,048,804	9,009,478	989,336	9,013,933	9,083,380
x_1	399.3	387.2	411.3	395.6	403
x_2	412.5	419.4	405.7	407.1	418
x_3	675.5	678.1	673	673.5	677.6
x_4	667.2	669.3	665.1	665.5	668.8
y_1	173.6	171.8	175.5	175.1	172.1
y_2	0	0	0	0	0
y_3	0	0	0	0	0
y_4	0	0	0	0	0

Note: Simulation results are derived from using GINO (Liebman *et al.*, 1986).

[3]These parameters are arbitrary, but the constraints remain the same as in the LPMS model.

the optimum solution set is more than that under the LPMS model. In our example, all four markets or distribution channels are involved in the marketing strategy problem. In a standard quadratic concave maximization problem such as QPMS model (e.g., Yang and Labys, 1981, 1982; Irwin and Yang, 1982, 1983; Yang and McNamara, 1989), it is not unusual to have more positive x's than the number of independent constraints. Consequently, the QPMS model can readily overcome the first problem of the LPMS model.

Second, as c_1 (intercept of the profit function of market or distribution channel #1) is varied by 100 units or only 2%, all the optimal x's have undergone changes (see Table 2). Consequently, the sales volumes through various distribution channels in the QPMS model are responsive to changes in the unit profit. This is more in agreement with theoretical as well as real-world expectations, i.e., change in profit environments would lead to adjustment in marketing strategy activities.

Last, as the total advertising budget is varied by \$200 as is done in the LPMS model, the corresponding dual variable y_1 (marginal profit due to the changes in the total advertising budget) assumes different values (see Table 2). The changing dual variable in the QPMS model possesses a more desirable property than the constant y's (marginal profits) in the LPMS model while both models are subject to the same constraints. Once again, the QPMS model provides a more flexible set of solutions relative to the *a priori* expectations of both theory and practice.

Care must be exercised that estimated regression coefficients more often than not, have some probability distributions, notably normal distribution. It remains an interesting topic in the future to incorporate stochastic programming in the marketing strategy model. That is, can normally distributed coefficients in the price equation give rise to a more systematic solution pattern in x's? It seems that there is no theory in this regard to indicate a hard and fast answer. In the absence of an answer, a simulation approach has been recommended using plus and minus two standard errors.

5. Conclusions

A quadratic programming model is proposed and applied in the marketing strategy problem. The solution to the QPMS problem may supply valuable information to management as to which marketing strategy or advertising mix is most appropriate in terms of profit while it meets various constraints.

The results show that once data are gathered conveniently and statistically estimated via the regression technique or other methods, one can formulate an alternative marketing strategy model. Then these estimated regression parameters can be fed into a quadratic programming package (e.g., Cutler and Pass, 1971; Schrage, 1986) to obtain a set of unique optimum solutions. The question of how to determine the alternative marketing strategies has important implications for the field of marketing management and marketing managers as well. By accommodating imperfect competition with a decreasing unit profit function, the QPMS model extends and complements its linear version significantly. Many limitations disappear as witnessed in the computer simulations.

More specifically, the model assists in examining the relative importance of different marketing mix variables, e.g., allocation of advertising effort and sales force. Furthermore, with such a more generalized QPMS model, the manager can examine the relative importance of the different levels of the variables involved. The profit and sales impacts of alternative marketing strategies can be determined with little cost incurred in the market place.

A model that provides information of this type should be invaluable to the marketing manager's efforts to plan and budget future marketing activities, particularly when it relieves the marketing manager of making a set of artificially restrictive assumptions concerning linearity and independence of the variables that are necessary to utilize linear programming marketing strategy LPMS models.

Finally, since the QPMS model removes the most restrictive assumptions of the LPMS models (in particular the assumptions that price, quantity and all cost and effort variables per unit must be constant and independent of each other) the utilization of the programming models may become more palatable to marketing managers. Our study has indicated that the QPMS model is considerably more consistent with *a priori* theoretical and practical expectations. Perhaps this combination will increase real world-applications of the QPMS model for solving marketing strategy problems. That is the authors' motivation for this study.

References

Anderson, DR, DJ Sweeney and TA Williams (1982). *An Introduction to Management Science*, 3rd Ed. New York: West Publishing Company.

Boudreaus, DJ and RG Holcombe (1989). The Coasian and Knightian theories of the firm. *Managerial and Decision Economics*, 10, 147–154.

Business Week (12 March 2001). Can Procter & Gamble clean up its act? 80, 82–83.

Cravens, DW and NF Piercy (2003). *Strategic Marketing*, 7th Ed. Boston, MA: McGraw-Hill & Irwin, 2003.

Cutler, L and DS Pass (1971). *A Computer Program for Quadratic Mathematical Models Involving Linear Constraints*. Rand Report R-516-PR. San Diego, CA: Rand Corporation

Flint, J (November 1991). Alfred Sloan spoken here. *Forbes*, 96 & 101.

Frank, R (1996). Soft-drink prices may be softening on shelves. *Wall Street Journal*, 12 September 1996, 1310.

Gass, SI (1985). *Linear Programming: Methods and Applications*, 5th Ed. New York: McGraw-Hill Book Company.

Hadley, G (1964). *Nonlinear and Dynamic Programming*. Reading, MA: Addison-Wesley Publishing Company, Inc.

Irwin, CL and CW Yang (1982). Iteration and sensitivity for a spatial equilibrium problem with linear supply and demand functions. *Operation Research*, 30(2), 319–335.

Irwin, CL and CW Yang (1983). Iteration and sensitivity for a nonlinear spatial equilibrium problem. In *Lecture Notes in Pure and Applied Mathematics*, A. Fiacco (ed.), pp. 91–107. New York: Marcel Dekker.

Kerwin, RA and RA Peterson (2001). *Strategic Marketing Problems: Cases and Comments*, 9th Ed. Upper Saddle River, NJ: Prentice Hall.

Liebman, J, L Lasdon, L Schrange and A Waren (1986). *Modeling and Optimization with GINO*. Palo Alto, CA: The Scientific Press.

Luenberger, DG (1973). *Introduction to Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley Publishing Company, Inc.

Maremont, M (1998). How Gillette brought its MACH 3 to market. *Wall Street Journal* 15 April 1998, B1.

Navarro, P (1988). Why do corporations give to charity? *Journal of Business*, 61, 65–93.

Schrage, L (1986). *Linear, Integer, and Quadratic programming with LINDO*. Palo Alto, CA: Scientific Press.

Shleifer, A and RW Vishny (1988). Value maximization and the acquisition process. *Journal of Economic Perspectives*, 2, 7–20.

Winn, DN and JD Shoenhair (1988). Compensation-based (DIS) incentives for revenue-maximizing behavior: A test of the 'revised' Baumol hypothesis. *Review of Economics and Statistics*, 70, 154–157.

Yang, CW and WC Labys (1981). Stability of appalachian coal shipments under policy variations. II *The Energy Journal*, 2(3), 111–128.

Yang, CW and WC Labys (1982). A sensitivity analysis of the stability property of the QP commodity model. *Journal of Empirical Economics*, 7, 93–107.

Yang, CW and J McNamara (1989). Two quadratic programming acquisition models with reciprocal services. In *Lecture Notes in Economic and Mathematical Systems*, TR Gulledge and LA Litteral (eds.), pp. 338–349. New York: Springer-Verlag.

Chapter 3

Evaluating the Robustness of Market Anomaly Evidence

William D. Brown, Jr.
Stonehill College, USA

Erin A. Moore
Lehigh University, USA

Ray J. Pfeiffer, Jr.
University of Massachusetts, USA

This study investigates two ways that sample selection can impact inferences about market efficiency: (1) unintentional, nonrandom exclusion of observations because of lack of available data for some firm-years ("passive deletion"); and (2) the effects of extreme return observations. The analysis proposes and illustrates a set of simple diagnostic tests designed to assess the sensitivity of estimated hedge portfolio returns to sample composition. These diagnostics are applied to the accrual anomaly and the forecast-to-price anomaly and the results indicate that the forecast-to-price anomaly is not robust to the effects of passive deletion. Moreover, extreme returns — as few as 100 firm-year observations — appear to generate the observed abnormal returns to both strategies.

Keywords: Market anomalies; market efficiency; bootstrapping; accruals anomaly; forecast-to-price anomaly.

1. Introduction

Increased attention to the equity markets in general, and increasing scrutiny over the past decade of the degree of market efficiency have led many researchers to search for anomalous stock returns in the available data. Judging by the large number of purported anomaly variables in the recent literature, one can say that the search appears to be bearing fruit.

However, some authors have cast doubt on whether some findings may reflect research design issues or are the inevitable result of "mining" the available data (e.g., Kothari, 2001; Fama, 1998; Kothari *et al.*, 2005). For example, Kothari *et al.* (2005) highlight the potentially significant impact of survival and data trimming on inferences regarding market efficiency. Specifically, they demonstrate that the measured relation between information variables and abnormal returns can be significantly biased as a result of which firms and firm-years end up in the examined sample. More generally, Bamber *et al.* (2000) raise concerns about over-generalization from evidence that (necessarily) is derived from subjective research design choices.

This study also investigates the sensitivity of conclusions regarding market efficiency to sample composition. However, unlike Kothari *et al.* (2005), which focuses on the effects of data restrictions on regression-based investigations of market efficiency, this study focuses on the widely used hedge portfolio technique. Moreover, the analysis focuses on two dimensions of the sample selection problem: (1) the effects of unintentional, nonrandom exclusion of observations because of data availability; and (2) the effects of extreme return observations in the resulting sample. Both dimensions of the problem have the potential to lead to incorrect inferences regarding the existence and/or magnitude of a given anomaly.

This study makes four important contributions to the literature: (1) provides a set of simple diagnostic tools that can be used to assess the robustness of apparent stock market anomalies, thereby reducing the incidence of potentially misleading evidence in the literature; (2) documents evidence that reinforces the recent work by Kraft *et al.* (2006) and others which finds that the accruals anomaly is driven by a very small subset of firms with unusual characteristics and extreme returns; (3) illustrates that the abnormal returns related to the forecast-to-price-based ratio are both sample-specific and driven by a small subset of firms on the "short" side of the strategy and shows how random sampling can expose such sample-specific phenomena; and (4) reinforces the notion (noted, e.g., by Sloan, 1996) that the hedge portfolio technique is an incomplete test for mispricing.

The results indicate significant sensitivity of the forecast-to-price anomaly to the effects of passive deletion. In fact, no evidence of a forecast-to-price anomaly was found in a large, unrestricted sample of firm-years. Moreover, when 1,000 random samples of the size used by Elgers *et al.* (2001) are drawn and the hedge strategy is repeated in each sample, a mean return of 7.6% is documented. Additionally, there is substantial variation in the hedge returns across the 1,000 samples, with minimum and maximum returns of −8.6% and 19.2% respectively. Returns as large as the 17.6% returns reported by Elgers *et al.* occur in only 8 of the 1,000 samples. This suggests that the forecast-to-price anomaly may only be present in certain samples with specific characteristics similar to those in the Elgers *et al.* sample.

In addition, the results indicate that the removal of the most extreme returns observations from the long and short portfolios is sufficient to displace both the accruals and forecast-to-price anomalies. Interestingly, it appears that the anomalies are driven by fewer than 2% (1%) of the observations in the forecast-to-price (accruals) samples.

2. Background

The recent literature in accounting and finance has seen a large number of studies documenting apparent patterns in security prices that violate traditional notions of semi-strong market efficiency. Examples include delayed price responses to earnings (Bernard and Thomas, 1990) investors over weighting of the accrual component of earnings (Sloan, 1996), patterns in returns related to "value" and "glamour" stocks (Lakonishok *et al.*, 1994), mispricing of foreign components of earnings (Thomas, 1999), failure to appreciate the information in fundamental components of earnings (Abarbanell and Bushee, 1998), and others.

The literature, however, is also subject to unintentional bias. Because of the long-standing null hypothesis of efficient equity markets, studies that document non-zero abnormal returns are considered more informative to the literature than those that confirm the efficiency of markets. Given this configuration of priors, studies that fail to find convincing evidence of market inefficiency are judged less informative and thus face larger hurdles to publication if indeed authors elect to submit them to academic journals.

On the other hand, studies that discover "large" abnormal returns are likely to be submitted to and published by academic journals in disproportionate numbers. Thus, in equilibrium, published anomaly studies are more likely to be those that are based on samples of unusual firm-years with unusually large abnormal returns. In other words, published anomaly studies represent a truncated portion of the population distribution of anomaly studies — those studies with smaller or zero abnormal returns are in the truncated portion of the distribution. There do not appear to be previous attempts in the literature to systematically address this issue.

Several prior studies have raised concerns that some or all of these findings may be attributable to specified and unspecified research design problems. For example, it is plausible that magnitudes of documented abnormal returns in anomaly studies are affected by the many implicit restrictions placed on samples (e.g., variables required to form the portfolios, to control for alternative explanations, and to measure stock price performance subsequent to portfolio formation). Such data restrictions have the possible unintended consequence of excluding a disproportionate number of firms with large negative returns (i.e., firms that fail and exit the databases). If the excluded firm-years would have otherwise been included in the long (short) portfolio of a hedge strategy, the exclusion biases reported hedge returns upward (downward). Using simulations, Kothari *et al.* (2005) show that *ex post* deletion of "extreme" observations from either tail of the distribution biases regression coefficients and may produce spurious regression/correlation-based inferences about market efficiency.

The next sections describe the proposed diagnostic tests as well as the two studies selected as exemplars of the approach.

3. Description of the Research Design

The investigations in this chapter focus specifically on the hedge portfolio technique. This widely used approach involves creating hypothetical and offsetting long and short investment positions in portfolios that are formed based on a variable that is hypothesized to predict future returns.[1] Given that a zero-investment portfolio has an expected return of zero (given appropriate controls for risk), any statistically significant non-zero returns to the net position (the hedge) are judged as evidence of market inefficiency with respect to the variable under study.[2]

Furthermore, the study specifically focuses on two research design issues: (1) passive deletion of observations from the sample that occur when researchers restrict their samples to require complete data for a set of variables; and (2) the effects of influential observations within the resulting sample. Both factors have the potential to create bias in the measured abnormal returns to

[1] Of course, there are trading, institutional, and practical restrictions to the short selling of securities. For example, short sales cannot be made until there is an up tick in the stock price, therefore the prices observed in the research portfolios in this study are likely not the prices at which an actual trader could short the securities. For these reasons, "anomalies" that derive most of their abnormal returns from the short side of the strategy are especially troublesome.

[2] All evidence of anomalous returns is conditional on appropriate risk adjustment. The issue of risk adjustment is a separate matter not addressed in this study that nevertheless has potential import for the evaluation of market efficiency.

a hedge strategy. The forecast-to-price anomaly (Elgers *et al.*, 2001) and the accruals anomaly (Sloan, 1996), respectively, are used to demonstrate the consequences of each of these research design issues.

Underlying both of the research design issues of concern — passive deletion and influential observations — is the notion that when measuring returns to a hedge strategy, it is consequential which observations from the population end up in the sample and which sample observations end up in the long and short portfolios. Thus the intuition behind the approach employed in this study is to ask: What if different (specifically, less restrictive) samples are selected from the population; and what if samples from within the resulting samples are selected? Would either of these impact measured abnormal returns?

3.1. *The effects of passive deletion*

This study illustrates two simple diagnostic tests that can be used to assess the sensitivity of estimated hedge portfolio returns to variation in sample membership. The diagnostic technique involves first replicating the findings in a given study. Given that an exact replication involves passive deletion of firm-years because of the lack of complete data for some required variables, the technique involves re-examining the sample selection criteria of a given study to identify sample restrictions that, in the end, turn out to be inconsequential to the study (e.g., requiring control variables that are not significant in multivariable regressions). Any such restrictions are then relaxed to create a larger, more general sample.[3] The important insight that is well known by researchers — but often overlooked in published anomaly studies — is that requiring additional variables to have complete data has potentially important consequences for measurement of anomalous returns, for firm-years with

[3]Of course, it is well known that increasing the number of required variables in a study will decrease the number of the available observations and potentially impound a survivorship bias into the sample. The intention here is to illustrate the importance of "working backwards." That is, once a control variable is determined to have an insignificant incremental partial correlation with the anomaly under study and the inclusion of the factor results in a significant loss of sample size, the factor can be excluded from the analysis in favor of increasing the size — and generalizability — of the sample. Researchers working on anomaly studies, however, are held to a very high bar in proving to editors and reviewers that their purported anomaly is incremental and distinct from previously documented priced factors.

missing data on certain variables may have returns that are significantly different from the firm-years with complete data. The returns to a hedge portfolio approach are estimated in the more general sample and compared with those in the original study. Any difference in the returns is attributed to the effects of passive deletion of observations.

In the second diagnostic, a sub-sample is randomly drawn from a sample of firm-years where the anomaly is present and statistically significant. Returns are then measured to the given hedge strategy in the sample. This process is repeated 1000 times (with replacement) to construct an empirical distribution of the hedge returns. Next, the returns obtained using the full sample are compared to the empirical distribution. To the extent that a given anomalous-return finding is robust to sample selection issues, the documented hedge returns should be near the center of the empirical distribution. On the other hand, if a purported anomalous return is much larger or smaller than most of the observations in the empirical distribution, this suggests that the result is potentially driven by inclusion/exclusion of firm-years with nonrandom characteristics. It does not appear that this random sampling technique has been employed elsewhere in the literature.

The generation of an empirical distribution of hedge returns provides the researcher with valuable incremental information about the candidate anomaly under study. Specifically, it locates the hedge returns obtained by a given researcher in the universe of hedge returns that would be obtained by researchers using the same data with different data selection criteria.

These diagnostics are applied to the findings in two specific prior studies: the forecast-to-price anomaly documented by Elgers *et al.* (2001), and the accrual anomaly documented by Sloan (1996). To provide a sense of the extent of passive deletion in these two studies, it is noted that the 17.6% average annual hedge returns to a strategy based on the forecast-to-price ratio between 1982 and 1998 reported in Elgers *et al.* (2001, Table 4, p. 626) are based on 7724 low-analyst-following firm-years, which represents approximately 4% of the available Compustat population. The sample used by Elgers *et al.* is relatively restrictive because of the requirement that nine control variables be present for all sample firm-years.

In contrast, Sloan (1996) finds the accrual anomaly in a relatively general sample — the 10.4% average annual hedge returns between 1961 and 1991 reported in Sloan (1996, Table 6, p. 307) come from a sample of 40,679 firm-years selected with *relatively* few restrictions: availability of accruals, average total assets, and returns. These 40,679 firm years represent approximately 57% of the available cases.

Based on these differences in sample selection, the forecast-to-price anomaly is used to illustrate the effects of passive deletion and both anomalies are used to illustrate the effects of extreme returns in the resulting sample.

3.2. *The effects of extreme returns*

Two tests are performed to illustrate the sensitivity of each documented anomaly to extreme returns. The presence of observations with extreme values is, of course, a problem widely known to researchers. However, in the anomaly literature, the solutions to the problem are less clear. For example, deletion of extreme cases eliminate real economic returns from the sample and subject the study to hindsight bias and make the hedge strategies unimplementable. Their inclusion, on the other hand, may help to "prove" an underlying theory about an information construct that does not hold for the average — or even a vast majority — of firms.

For example, Elgers *et al.* (2001) contend that investors in firms with less-developed information environments under-weight financial analysts' forecasts of earnings, resulting in predictable returns as investors become aware of their errors. However, if the Elgers *et al.* empirical result does not hold for the average firm — even in the low information environment partition, then what can one say about the underlying theory?

Kraft *et al.* (2006) conduct analyses surrounding the accruals anomaly in the same spirit as the above discussion. They find that it is approximately 200 firms in their sample that are driving the accruals anomaly. In addition, they find that the theory advanced by Sloan (1996) is not reflected in even the 200 observations where the anomaly is strongest. Specifically, they find no evidence of the clustering of abnormal returns around the earnings announcements for the sample of 200 firms, without which there is no accruals anomaly.

The analysis begins by constructing a sample where the anomaly is present and significant — that is where forming hedge portfolios and taking long and short positions consistent with the theory of the anomaly produces non-zero risk-adjusted hedge returns. The purpose of the next step is to examine the extent to which the deletion of a very small number of cases from the long and short portfolios affects the magnitude of the returns. The deletion of observations is achieved by removing the *n*% of firm-years with the largest (most positive) and smallest (most negative) returns from the long and short portfolios, respectively. This asymmetric deletion within the long and short portfolios will drive the returns toward zero by construction. Since the mean of the long (short) portfolio will shift toward zero it is not surprising that the hedge returns will do likewise. The point here however, is to examine *how many* firm-year observations need to be deleted in order to fully dissipate the returns to a proposed anomaly. The tests to follow document that it is a surprisingly small number.

Next, each anomaly is reexamined in the original decile portfolios, and the following questions are asked: What if offsetting positions are taken in portfolios 1 and 8 instead of 0 and 9, would the anomalous returns persist? If the anomaly is robust and the mispricing theory under examination holds for the average firm, then one would expect to find significant, albeit smaller, abnormal returns in the intervening portfolios as well. In addition, what is the effect of forming portfolios based on quintiles of the distribution instead of deciles? Does the ad hoc grouping choice help to drive the presence or magnitude of documented anomalous returns? This is yet another choice that may influence the inferences and results of factors under study.[4]

3.3. *The forecast-to-price anomaly*

Elgers *et al.* (2001) report a large and statistically significant positive relation between the ratio of analysts' earnings forecasts and share price (both measured early in the earnings year) and subsequent returns. They argue that investors appear to under-weight the price-relevant information in analysts' forecasts. Taking long and short positions based on the largest and smallest forecasts, respectively, yields positive hedge returns in the 12 months following portfolio formation (17.6% average hedge returns from the annual application of the strategy between 1982 and 1998 [Elgers *et al.*, 2001, p. 626, Table 4]).

Because of their concern of the possibility that the forecast-to-price ratio might be related to other documented return covariates, Elgers *et al.* require nine additional control variables (used in subsequent tests) in constructing their sample. These data requirements lead to significant loss of firm-years,

[4]For example, Frankel and Lee (1998) use quintiles in their test of mispricing with respect to expectations embedded in residual income valuation.

i.e., "passive deletion." Elgers *et al.* do not disclose the size of their population other than to disclose that they start with all firms in the 1999 Compustat CD-ROM Research and Active databases. The 2005 edition of Compustat Research Insight contains 165,617 Research and Active firm-years between 1985 and 1998, which indicates a total available population of approximately 190,000 firm-years for the 1982–1998 sample period used by Elgers *et al.* Elgers *et al.*'s resulting sample consists of 16,247 firm-years between 1982 and 1998, representing approximately 9% of the population. Significant abnormal returns only exist in the low analyst following sub-sample of 7,724 firm-years, which is approximately 4% of the population.

3.4. *The accruals anomaly*

Sloan (1996) wrote a seminal article in the anomaly literature that reports that future stock returns appear to be negatively related to the magnitude of the accrual component of earnings. The apparent explanation is that investors fixate on aggregate earnings and do not fully realize that accruals tend to have less persistence for future performance than do cash flows. As a result, long positions in firms with relatively small accruals and short positions in firms with relatively large accruals yield positive and significant hedge returns in the 2 years subsequent to portfolio formation (approximately 10.4% and 4.8% in years 1 and 2, respectively [Sloan, 1996, p. 307, Table 6]). Sloan's reported hedge returns represent average returns to the strategy applied annually between 1962 and 1991.

Sloan (1996) requires relatively few variables to have complete data in order for a firm-year to enter the sample. Firm-years with sufficient data to compute accruals, returns, average total assets, and subsequent year's earnings comprise the final sample. Nevertheless, his sample of 40,679 firm-years represents those that survive from an initial population of 71,732 NYSE and AMEX firm-years in Compustat, representing approximately 57% of the available population (Sloan, 1996, p. 293, fn. 7).

4. Results

4.1. *Investigating the effects of passive deletion*

The analysis begins by replicating the forecast-to-price anomaly findings in Elgers *et al.* (2001). As mentioned above, this anomaly is selected because the relatively significant sample restrictions make this an exemplar for the potentially misleading effects of passive deletion.

To construct a more general sample, all possible observations from Compustat and IBES that meet a minimum set of restrictions are identified. These requirements are the presence of size-adjusted returns, analysts' forecasts, the number of analysts underlying the IBES consensus forecasts, and share price. All variables are defined as in Elgers *et al.* (2001). Specifically, size-adjusted return is the raw return for each firm-year, less the average return within the sample for all firms in the same market-capitalization decile at the start of the cumulation period in each year. Analysts' forecasts are the IBES consensus (mean) earnings forecasts in March of the earnings year. Share price (the scaler of the forecasts) is measured at the start of the year. A total of 42,414 firm-years have complete data on these variables between 1978 and 2001. This compares with 16,247 firm-years underlying the original Elgers *et al.* (2001) findings between 1982 and 1998. As noted above, the smaller sample in Elgers *et al.* is the result of requiring available data for several control variables used in some of their tests.

As Elgers *et al.* (2001) find significant abnormal returns only in the subset of firms with relatively low analyst following, this analysis likewise partitions the sample in each year based on the median analyst following. Firm-years exactly at the median are arbitrarily assigned to the high following partition. Panels A and B of Table 1 compare key descriptive statistics in the Elgers *et al.* sample and in the more general sample.[5]

For the low analyst following partition, the mean size-adjusted returns are (by construction) approximately zero for both samples. The standard deviation of size-adjusted returns, however, is significantly larger in the current study's general sample: 124.0% versus 43.3%. The inter-quartile range of *SAR* in the general sample is 56.6% and 45.9% in the original Elgers *et al.* sample. Overall, these characteristics demonstrate the implicit truncation that occurs when imposing restrictions on the sample. As the sample is whittled down from 42,414 toward the 16,247 cases that end up in the Elgers *et al.* sample, extreme returns appear to drop out of the sample as well. The exclusion of extreme returns from a sample has potentially important effects on hedge returns because the statistics generally used — means of portfolios returns — are disproportionately affected by extreme values. For example,

[5]The authors thank Pieter Elgers for providing the data used in Elgers *et al.* (2001).

Table 1. Descriptive statistics.

Variable	$\overline{x}$	σ	Min	25%	Median	75%	Max
Panel A: General sample — all firm-years with available data (1978–2001, n = 42,414)							
Lower analyst following ($n = 19{,}510$)							
F_t/P_t	0.066	1.23	−26.38	0.042	0.081	0.119	107.50
P_t	25.24	702.69	0.02	5.25	9.50	15.81	62,600.00
SAR_{t+1}	−0.011	1.24	−2.02	−0.366	−0.074	0.2	134.94
Higher analyst following ($n = 22{,}904$)							
F_t/P_t	0.067	0.403	−54.96	0.050	0.075	0.104	1.66
P_t	24.90	453.11	0.12	10.31	17.53	27.15	68,400.00
SAR_{t+1}	0.009	0.549	−1.90	−0.241	−0.028	0.189	11.21
Panel B: Sample from Elgers et al. (2001) (1982–1998, n = 16,247)							
Lower Analyst Following ($n = 7{,}724$)							
F_t/P_t	0.064	0.088	−3.136	0.043	0.078	0.097	0.65
P_t	14.30	11.92	0.18	7.25	12.00	18.25	291.67
SAR_{t+1}	0.001	0.433	−1.225	−0.267	−0.042	0.192	3.15
Higher analyst following ($n = 8{,}523$)							
F_t/P_t	0.068	0.061	−2.654	0.046	0.076	0.091	0.85
P_t	24.26	45.39	0.43	12.75	20.00	28.75	2,031.25
SAR_{t+1}	−0.001	0.341	−1.193	−0.200	−0.018	0.149	3.05
Panel C: Accruals anomaly sample (1963–2001, n = 32,493)							
Accruals_a	−0.033	0.091	−1.721	−0.071	−0.035	0.000	1.689
P_t	23.92	24.93	0.03	8.94	18.63	31.75	627.00
SAR_{t+1}	0.001	0.509	−3.171	−0.236	−0.037	0.173	24.255

Variables are defined as follows (Compustat mnemonics are in parenthesis): F_t is the median IBES analyst's forecast of year t earnings reported in March of year t; P_t is IBES share price reported in January t; and SAR_{t+1} is size-adjusted return for the period April 1, t through March 31, $t+1$, computed as the raw return (TRT1Y) for the same period less the mean return for all firms in the same (within sample) December $t-1$ market value of equity (MKVALF) decile. Accruals_a is accruals in year t, defined as in Sloan (1996), scaled by average total assets in year t. For the accruals anomaly sample, P is share price measured at the end of March of year t. Note Elgers *et al.* (2001, Table 1, p. 618) report only averages of annual medians and interquartile ranges. To provide for more detailed comparisons, means, standard deviations, and various percentiles of the distributions are presented using the original Elgers *et al.* sample.

if just one firm with an extremely negative return ends up being located in a portfolio where a positive return is expected, that can turn the portfolio return negative and thus completely change the inference regarding the anomaly in that year.

Sizable differences also exist in measures of the magnitudes of the forecasts across the two samples. For example, while average F/P ratios across samples are quite similar (0.066 versus 0.064 for lightly followed firms), the standard deviation of F/P is 1.23 in the general sample versus 0.088 in the Elgers *et al.* sample. Furthermore, examination of the minimum and maximum values of each of the variables in both samples clearly demonstrates that the implicit sample restrictions most certainly have the effect of eliminating extreme cases from the sample.

To assess the impact of the sample restrictions (passive deletion) on inferences about anomalous returns to the forecast-to-price ratio, the diagnostic described in Section 3 of the chapter is employed. First, the strategy is replicated using essentially the same approach as in Elgers *et al.* (2001).

Panel A of Table 2 presents the results of applying the hedge portfolio analysis in each of the years from 1978 to 2001. In this and all subsequent analyses of the forecast-to-price anomaly, only results for the low-analyst-following sub-sample are reported, as Elgers *et al.* find no returns for relatively heavily followed firms. The average annual hedge return across the 24 years is −8.9%.[6] There is considerable variation in the success of the strategy across years — in 10 years the strategy yields negative returns, and returns range from −193.2% in 1999 to 21.7% in 1988. Note that while there is no basis, *ex ante*, for excluding the large negative return in 1999, the mean return for the remaining 23 years is −0.9%.[7] The median return is close to 0 (1.6%), and using the Fama and MacBeth (1973) t-statistic to measure the statistical significance of the mean return, the −8.9% return is not statistically significant ($t = -1.04$) (the −0.9% average return excluding 1999 is also not statistically significant [$t = -0.29$]).

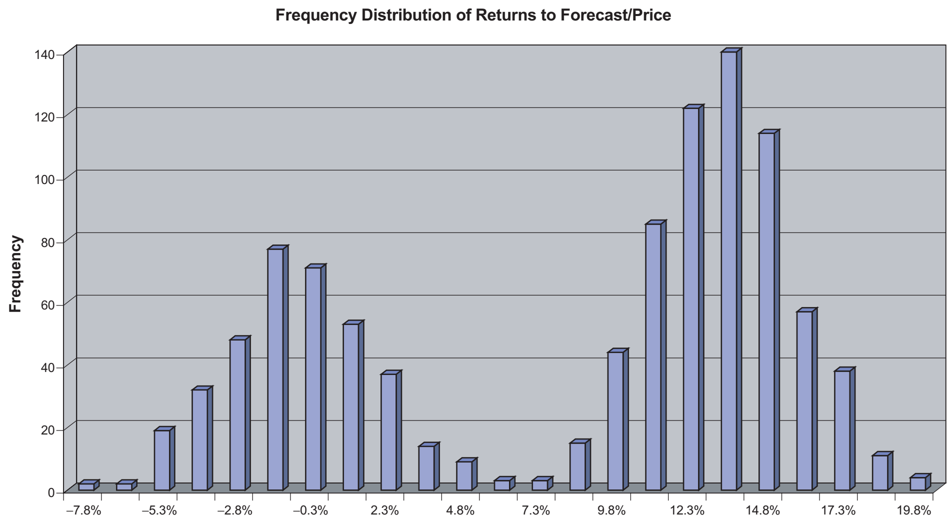
These results strongly suggest that the forecast-to-price anomaly is not present in this larger, more general sample. Moreover, the results suggest that passive deletion may account for the existence of returns in the original Elgers *et al.* data.

[6] The average annual hedge return in this sample for the years that overlap the Elgers *et al.* (2001) sample, 1982–1998, are −10.3% ($t = -0.09$). Thus, the difference in sample period does not appear to affect inferences regarding the anomaly and further highlights the effects of passive deletion.

[7] As others have documented, the relations among information variables and returns are dramatically difficult in 1999 from other years.

Table 2. Testing the effects of passive deletion: size-adjusted hedge returns to forecast-to-price-based strategy $n = 42{,}414$ firm-years between 1978 and 2001.

Mean	σ	Min	5%	10%	25%	50%	75%	90%	95%	Max
Panel A: Sample statistics of 24 annual hedge returns in unrestricted sample of 19,510 low-analyst-following firm-years										
−8.9%	41.9%	−193.2%	−47.2%	−13.1%	−9.3%	1.6%	6.8%	12.7%	13.8%	21.7%
Panel B: Sample statistics from 1,000 random samples of 16,247 firm-years drawn (with replacement) from unrestricted sample — approximately 8,000 low analyst following firm-years per sample										
7.6%	7.2%	−8.6%	−4.1%	−2.8%	−0.3%	10.9%	13.3%	14.9%	16.1%	19.2%



To gain further insights about the substantial difference in results between the two samples, Panel B of Table 2 presents the results of annual hedge returns of 1,000 random samples drawn from the general sample. Since sample years in this study surround those used by Elgers *et al.*, the Elgers *et al.* sample could conceivably be one of the 1,000 random samples. The purpose of this analysis is to investigate the variability of the returns in different subsamples.

The mean annual hedge return across the 1,000 iterations is 7.6%. Returns range from −8.6% to 19.2%. The distribution (pictured in Panel B of Table 2) appears to be somewhat unusual in that it is bimodal, with large frequencies centered around −1.5% and +13.5% returns. Given that these are randomly constructed samples, there is no basis for expecting or explaining the shape of the distribution.[8] Using the Fama and MacBeth (1973) t-statistic, the 7.6% mean return is significant ($t = 33.23$), contradicting the present inferences using the full 42,414 firm-year sample. These results suggest that indeed there may be a forecast-to-price anomaly, although its magnitude is substantially less than that documented by Elgers *et al.* In fact, the 17.6% hedge return reported by Elgers *et al.* is greater than the return obtained in 992 of the 1,000 random samples. Stated differently, if 100 researchers were to independently draw samples and conduct investigations of the forecast-to-price anomaly, less than 1 of the 100 researchers would find returns as strong as those reported by Elgers *et al.*

In sum, the evidence presented in Table 2 suggests that while there may indeed be a forecast-to-price anomaly, its magnitude depends critically upon which sample of firm-years is used in the estimation.

4.2. *Investigating the effects of extreme returns*

The first test of the effects of influential observations uses the original Elgers *et al.* data. As discussed in Section 3 above, the test is designed to reduce the potential impact of any extreme returns. This is accomplished by eliminating the most extreme returns from the long and short portfolios and examining the

[8]The entire procedure was repeated using a new set of 1,000 random samples. A similar bimodal pattern is documented; however, the results are substantially smaller: mean hedge return is 0.85% ($t = 3.56$), returns range from −15.25% to 13.89%, median return is 4.13%. The two collections of large frequencies are around −10.5% and +7.5%.

Table 3. Testing the effects of extreme returns: size-adjusted returns to portfolios based on F/P and accruals.

	Cases deleted	Mean hedge return	Median hedge return	Mean Low F/P Portfolio	Mean High F/P Portfolio
Panel A: F/P: Returns using original Elgers et al. (2001) sample; 1,881 firm-years in highest and lowest F/P deciles out of 7,724 total low-analyst-following firm-years (full sample: 16,247 firm-years between 1982 and 1998)					
No deletion	0	**17.6%**	16.4%	**−14.0%**	3.6%
1% deletion	41	**13.5%**	12.3%	**−12.5%**	1.0%
5% deletion	141	**7.1%**	5.8%	**−9.6%**	−2.5%
10% deletion	280	0.5%	0.5%	**−6.5%**	**−6.0%**
Panel B: Accruals: Returns from 6,466 firm-years in highest and lowest asset-scaled accrual deciles (full-sample: 32,493 firm-years between 1963 and 2001)					
No deletion	0	**8.4%**	5.4%	1.1%	**−7.3%**
1% deletion	100	1.7%	−0.6%	**−3.9%**	**−5.6%**
5% deletion	358	**−5.5%**	−6.8%	**−8.5%**	**−3.0%**
10% deletion	676	**−12.8%**	−12.2%	**−12.6%**	0.2%

Notes: Deletion of cases was based on the extremity of a given firm-year's return within the annual distributions of returns in the highest and lowest portfolios. For example, for the 1% deletion row in Panel A above, 1% of firm-years with the smallest (most negative) size-adjusted returns in the lowest F/P decile are excluded, and 1% of the firm-years with the largest size-adjusted returns in the highest F/P decile in each year are excluded. t-statistics are computed as the mean hedge return divided by the standard errors of the annual hedge returns. Mean returns in the table above indicated in boldface type are significant below the 0.05 level based on two-tailed tests.

returns to the forecast-to-price strategy as a greater percentage of observations are incrementally eliminated from these portfolios.

Table 3 provides the results. In Panel A, the Elgers *et al.* (2001) result is reproduced using the original data. The mean return of 17.6% is relatively large and is significant based on the Fama–MacBeth t-statistic ($t = 4.99$). There is, however, fairly substantial variation across the years, from −12.3% in 1998 to +38.3% in 1985 (not tabulated). Eliminating 1% of the most extreme cases from the yearly long and short portfolios reduces the mean hedge return by nearly one-fourth to 13.5%. This result is made even more dramatic when one considers that just 41 observations — 2/10 of 1% of the sample — are responsible for one-quarter of the observed returns.

Deleting 5% of the cases results in the mean hedge return dropping to just over 7%. While the return is still significant, note that most of the returns come from the short side of the investment strategy. As discussed in Footnote 1, due to the restrictions and frictions on short sales it is questionable as to the extent these returns would be realizable.

Eliminating 10% of the firms from the top and bottom portfolio — or less than 2% of the overall sample — eliminates the returns to the forecast-to-price strategy. Interestingly, since the anomaly is driven by the short side of the strategy, it is likely that the returns would have dissipated even if only firms from the short side had been eliminated —resulting in a much smaller loss of sample. This observation foreshadows the tests with the intervening portfolios in Table 4.

Panel B presents the results of eliminating extreme returns on the accruals anomaly. First, the Sloan (1996) result is replicated in the present data. Following the procedures outlined in Sloan, a sample of all available firm years with sufficient data is collected and asset-scaled accruals and size-adjusted returns are computed. Because Fama and MacBeth (1973) t-statistics are used, all of the annual return cumulation periods must be aligned in calendar time; note that this can be seen as an additional cause of passive deletion of observations. There is no reason to expect, however, that the accruals anomaly differs for firms with different fiscal year ends. These selection criteria yield 32,387 firm-years between 1963 and 2001.

Applying the hedge strategy annually based on the magnitude of asset-scaled accruals (a long position in the lowest decile of asset-scaled accruals and an offsetting short position in the highest decile), reveals average returns in the 12 months subsequent to portfolio formation of 8.4%. Table 3, Panel B presents the details of these findings. The 8.4% documented in this analysis compares with 10.4% for the 1962–1991 period (for 40,679 firm-years irrespective of fiscal year end) documented in Sloan (1996). There is substantial variation across years in the present data, ranging from −8.0% to 54.8%. Based on Fama and MacBeth (1973) t-statistics, the 8.4% return is significant ($t = 3.9$).

While the main purpose in this section is to investigate the effects of influential observations, the difference between the 8.4% return documented in this study and the 10.4% reported in Sloan (1996), together with the wide range of returns across years, are evidence of the effects of passive deletion as well. The two samples differ only with respect to years (Sloan uses 1962–1991 while this study examines 1963–2001) and the calendar-year-end restriction in this study. Either set of results arguably supports the *existence* of an accruals anomaly. However, Panel B seems to indicate some degree of sensitivity of the *magnitude* of the anomaly to sample composition.

Next, observations with the most extreme returns are deleted from the top and bottom portfolio in precisely the same way as in Panel A for the forecast-to-price anomaly. In the case of the accruals anomaly, deletion of just 1% of the firm-years erases the returns to asset-scaled accruals. That is, deletion of fewer than 100 observations — or only 3/10 of 1% of the sample — is sufficient to cause the returns to accruals to dissipate.[9] Note that further trimming of the data (the 5% and 10% deletion results) causes the returns to become significantly *negative*. This certainly contradicts the story underlying the accruals anomaly. However, these returns are subject to the interpretation endemic to all hedge portfolio tests: in the absence of controls for other factors likely related to subsequent returns, hedge returns are likely to reflect risk factors other than the variable used to form the portfolios. For this reason, these results are not viewed as evidence of mispricing.

In sum, the results of the tests in Table 3 indicate that both anomalies examined are highly sensitive to the inclusion/exclusion of a relatively small number of firm-years with extreme returns. The results with respect to the accruals anomaly confirm those of Kraft *et al.* (2006). The results with respect to the forecast-to-price anomaly are new to this study.

The analyses in Table 3 demonstrate the effects of eliminating extreme returns from the data, but do not take into consideration the extremity of the factor under study. That is the purpose of the next set of tests. Table 4 examines the effects of alternatively reforming portfolios in quintiles or simply taking positions in less extreme portfolios. In these tests, extremity is defined based on the information variable under study (accruals or F/P) rather than on returns. The effects are likely overlapping, however, as firms that are experiencing extreme performance or returns are likely experiencing both.

The first column of Table 4 simply replicates the anomalies under investigation and provides a benchmark against which returns to alternative strategies are measured in the remaining columns. The three columns in the middle of the table present the results of taking positions in intermediate portfolios where Portfolios 0 and 9 represent the endpoints. The final column to the far right displays the results of taking hedge positions in portfolios formed on the

[9]This result is consistent with Kraft *et al.* (2006) as they find that less than 200 observations in their larger sample are driving the accruals anomaly.

Table 4. Testing the effects of extreme returns: size-adjusted hedge returns to alternative F/P and accruals strategies.

	Deciles 0 and 9	Deciles 1 and 8	Deciles 0 and 8	Deciles 1 and 9	Quintiles
Panel A: F/P: Statistics of 17 annual hedge returns using original Elgers et al. sample; 1,881 firm-years in highest and lowest F/P deciles out of 7,724 total low-analyst-following firm-years (full sample: 16,247 firm-years between 1982 and 1998)					
Mean hedge return	17.6%	4.0%	17.7%	3.9%	11.9%
Standard error	3.5%	4.3%	2.8%	4.7%	3.4%
t-statistic	5.0	0.9	6.4	0.8	3.5
Panel B: Accruals: Statistics of 39 annual hedge returns from 6,466 firm-years in highest and lowest asset-scaled accrual deciles (full sample: 32,493 firm-years between 1963 and 2001)					
Mean hedge return	8.4%	3.2%	2.7%	8.9%	4.8%
Standard error	2.2%	2.0%	1.9%	2.1%	1.5%
t-statistic	3.9	1.6	1.4	4.2	3.8

Notes: The first column — "Deciles 0 and 9" represents the hedge strategies involving the lowest and highest deciles, respectively. The 17.6% mean hedge return in Panel A is that reported by Elgers *et al.* (2001, p. 626). The 8.4% return in Panel B is from the replication of the accrual anomaly, reported in the text. The remaining columns of the table display the results of applying the hedge strategy with different (less-extreme) portfolios — both less extreme deciles and substituting quintiles for deciles. t-statistics are computed as the mean hedge return divided by the standard errors of the annual hedge returns.

quintiles of the distribution of the forecast-to-price ratio and (asset-scaled) accruals, respectively.

For the forecast-to-price anomaly — presented in Panel A — the results are quite sensitive to the elimination of the portfolio on the short side of the strategy. Hedge returns calculated based on selling Portfolio 1 and buying Portfolio 8, are not statistically significant. As discussed earlier, if the theory underlying the forecast-to-price anomaly held, then why would one not find returns in these portfolios? Conversely, when Portfolio 0 is sold and Portfolio 8 is bought the returns are nearly identical to the returns generated by Portfolios 0 and 9 indicating that the anomaly is almost completely driven by the short side of the strategy.

When the hedge portfolio analysis is conducted within quintiles, the returns remain significant but the magnitude is muted. In fact, the returns to quintiles are nearly one-third smaller than the returns to deciles. This is clear evidence that the ad hoc choice of quintile or decile portfolios has a dramatic impact on the magnitude of the anomaly.

The results in Panel B for the accruals anomaly mirror those in Panel A. The anomaly is sensitive to dropping the original extreme short portfolio but not sensitive to changing to an intermediate long portfolio. Changing from deciles to quintiles reduces that magnitude of the accruals anomaly to 4.8% (nearly half of the original returns) and although still statistically significant, 4.8% is likely within transaction costs and other market frictions, especially given the increased costs and frictions of short-selling.

In sum, Table 4 presents compelling evidence that the two anomalies under study are not represented in intermediate portfolios, are highly sensitive to the elimination of the extreme portfolio on the short side, and are severely muted when switching from deciles to quintiles. Moreover, the results presented above in Table 3 show that even if the returns documented in the original data were achievable through implementable trading strategies, these analyses call into question the original theories underlying these constructs as they seem to only hold for a small number of firm-years.

5. Summary and Conclusions

This analysis explores the implications of sample selection and influential observations for evidence in the literature regarding anomalous stock returns. This investigation is motivated by several factors, including the heightened interest in this topic in the recent accounting and finance literatures and the concomitant exploration of so-called "behavioral finance," attention paid to research design issues in related research, including Kothari *et al.* (2005), Barber and Lyon (1997), Fama (1998), and others; and the more general concern about overall conclusions drawn regarding the functioning of the securities markets from this literature. In particular, it is important to reiterate the general concern expressed by Bamber *et al.* (2000) about the tendency in the literature to over generalize from the results of a single study and to emphasize their embrace of the value of replications in accounting research.

This study illustrates diagnostic procedures that can be used to subject evidence of anomalous returns to further scrutiny. The techniques advocated in this chapter shed light on the possible sensitivity of estimated hedge returns to sample composition, both passive deletion of observations and the presence of potentially influential observations. To demonstrate the use of these diagnostics, the evidence in two studies from the recent literature is reevaluated: the forecast-to-price anomaly documented by Elgers *et al.* (2001), and the accruals anomaly documented by Sloan (1996).

The evidence from this analysis is consistent with the interpretation that the forecast-to-price anomaly is highly sensitive to sample selection and the presence of extreme returns and does not appear to exist in many of the test samples. Likewise, inferences about the existence, magnitude, and sign of the accruals anomaly depend on the inclusion or exclusion of very few firm-year observations.

The accounting and finance literatures can benefit from application of the diagnostic approaches presented in this chapter to other documented anomalies. Doing so may potentially simplify the current understanding of investor behavior with respect to accounting information by reducing the number of verifiable patterns in security prices.

Acknowledgments

The authors gratefully acknowledge Robert Nakosteen, an anonymous reviewer for helpful comments, and I/B/E/S for providing analyst forecast data at an academic rate.

References

Abarbanell, J and B Bushee (1998). Abnormal returns to a fundamental analysis strategy. *The Accounting Review*, 73(1), 19–45.

Bamber, L, T Christensen and K Gaver (2000). Do we really ‘know’ what we think we know? A case study of seminal research and its subsequent overgeneralization. *Accounting, Organizations, and Society*, 25(2), 103–129.

Barber, B and J Lyon (1997). Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. *Journal of Financial Economics*, 43(3), 341–372.

Bernard, V and J Thomas (1990). Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics*, 13, 305–340.

Elgers, P, M Lo and R Pfeiffer (2001). Delayed security price adjustments to financial analysts’ forecasts of annual earnings. *The Accounting Review*, 76(4), 613–632.

Fama, E and J MacBeth (1973). Risk, return and equilibrium: Empirical tests. *Journal of Political Economy*, 81(May/June), 607–636.

Fama, E (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49(3), 283–306.

Frankel, R and CMC Lee (1998). Accounting valuation, market expectation, and cross-sectional stock returns. *Journal of Accounting and Economics*, 25, 283–319.

Kraft, A, A Leone and C Wasley (2006). An analysis of the theories and explanations offered for the mispricing of accruals and accrual components. *Journal of Accounting Research*, 44(2), 297–339.

Kothari, SP (2001). Capital markets research in accounting. *Journal of Accounting and Economics*, 31(1–3), 105–231.

Kothari, SP, J Sabino and T Zach (2005). Implications of survival and data trimming for tests of market efficiency. *Journal of Accounting and Economics*, 39(1), 129–161.

Lakonishok, J, A Shleifer and RW Vishny (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance*, 49(5), 1541–1578.

Sloan, R (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review*, 71(3), 289–315.

Thomas, W (1999). A test of the market's mispricing of domestic and foreign earnings. *Journal of Accounting and Economics*, 28(3), 243–267.

This page intentionally left blank

Chapter 4

Why is the Value Relevance of Earnings Lower for High-Tech Firms?

B. Brian Lee
Prairie View A&M University, USA

Eric Press
Temple University Philadelphia, USA

B. Ben Choi
Victoria, Canada

This chapter examines the value relevance of earnings in high- and low-technology industries, a topic about which unresolved questions remain. The focus is on high-tech firms, given assertions that financial reporting is least effective for their industries. It is found that adjusting for expense mismatching is more effective in low-tech industries, while diversifying noise in high-tech industries substantially increases the association of earnings and stock returns. The value relevance of earnings for either type is indistinguishable when noise and expense mismatching are jointly controlled. It is thus concluded that noise arising from uncertainty of realizing future R&D benefits is the main factor explaining the weaker association of earnings and stock returns for high-tech firms.

Keywords: Expense mismatching; noise; R&D; high-tech firms; value relevance.

1. Introduction

This chapter investigates the value relevance of earnings for high- and low-technology firms. The increased use of intangible assets in technology-intensive industries is one of the main explanations for the decline in the value relevance of earnings. If systematic problems in recording intangibles create differences between the information investors use to set stock prices versus information provided in annual earnings, the relevance of firm earnings decreases as the proportion of intangibles increases.[1] Thus, it is unsurprising

[1] Sougiannis (1994), Lev and Sougiannis (1996), Lev and Zarowin (1999), and Danielson and Press (2003), among others, note shortcomings in the expensing and balance sheet valuation of intangible items under GAAP.

if the contemporaneous relation between earnings and returns is lower in intangible-intensive firms.

Nonetheless, previous research examining intangibles intensity as an explanation for the diminished value relevance of earnings generates mixed evidence. Lev and Zarowin (1999) report the informativeness of earnings decreases as firms increase expenditures on intangible investments. In contrast, findings in Collins *et al.* (1997) and Francis and Schipper (1999) imply that firms with high intangibles do not experience lower contemporaneous associations between their earnings and stock prices.

The conflicting evidence motivates the present work. It begins with an empirical assessment of value relevance between 1984 and 1998, regressing stock returns on earnings changes, and levels for all firms in the *Compustat* files. It is observed that earnings are less informative for high-tech firms, regardless of how firms are categorized as high-tech or low-tech. Collins *et al.* (1997) note that over time, as the value relevance of earnings declines, the relevance of book values increases. However, GAAP precludes one of the most critical intangible assets in high-tech firms — the future benefit from R&D investments — from appearing on the balance sheet, an alternative information source to the income statement.

Given that balance sheet items are not likely to offer incremental explanatory power for high-tech firms, a model has been developed to address the errors-in-variables problem in the earnings–returns relation. Previous research was extended by examining factors identified by Collins *et al.* (1994) that could diminish the relevance of earnings: earnings' lack of timeliness driven by mismatching revenues and expenses, including R&D expenditures, and noise caused by uncertainty of future benefits from R&D and other intangibles.

Intangibles affect the value relevance of earnings in two ways. Under GAAP, firms expense most of their intangible investments — such as R&D, information technology, human capital, and brand identity — in the period of acquisition, but the benefits are realized and recorded in earnings in subsequent periods.[2] Investors, on the other hand, view intangibles expenditure as

[2]For most intangible expenditures, firms expense the investment immediately. Research and development is expensed as incurred (Financial Accounting Standard No. 2, 1997), as are investments in advertising and employee training [i.e., investments that in the first case create brand-name capital (Klein and Leffler, 1981), and in the second, human capital]. Other intangible assets are omitted from the balance sheet if they are internally generated, or if acquired in a merger treated as a pooling transaction (American Institute of Certified Public Accountants, 1970).

investment creating future benefits (Ben-Zion, 1978; Hirschey and Weygandt, 1985; Bublitz and Ettredge, 1989; Shevlin, 1991; Sougiannis, 1994; Lev and Sougiannis, 1996; Chambers *et al.*, 1999; Healy *et al.*, 2002). Therefore, the earnings of firms with larger expenditures on intangibles could be more seriously affected by expense mismatching.

Using the S&P 500 composite and its subsets, and 136 individual firms, Collins *et al.* (1994) report that low contemporaneous association between stock returns and earnings arises primarily from the timeliness problem, not from noise. Thus, if expensing intangibles in high-tech industries is the main factor explaining the lower value relevance of their earnings, correcting expense mismatching should strengthen the association for high-tech firms more than for firms in low-tech industries.

The second factor that could affect value relevance is noise. If the benefits from intangible investments are uncertain — given no obvious relation between intangible investments and their future benefits — investors face added difficulty in estimating firm value in high-tech industries (Kothari *et al.*, 2002). Higher uncertainty about future benefits increases information asymmetry between investors and managers, thus inducing more noise in the estimation of high-tech firms' values, in comparison to low-tech firms (Aboody and Lev, 2000; Boone and Raman, 2001). Therefore, reducing noise may be more effective in improving the association between stock returns and earnings for high-technology firms. To diversify noise, returns and earnings are aggregated by a combination of years and two-digit SIC codes, as implied by Collins *et al.* (1994) and Beaver *et al.* (1980).

When stock returns are regressed on earnings changes and levels, high-tech firms exhibit lower R^2 than their low-tech counterparts. When expense mismatching is corrected using the present model, high-tech firms still display lower R^2 than low-tech firms, using both pooled and yearly data. These findings are inconsistent with an expectation of expense mismatching being a greater problem for high-tech firms.

When noise is diversified using industry-level aggregation methods, the present results imply that earnings of high-tech firms contain more noise than earnings for low-tech firms. Indeed, when corrected for both noise and earnings' lack of timeliness, the difference in R^2 between high- and low-tech firms vanishes.

Based on this evidence, it is concluded that noise, not expense mismatching, attenuates the association between returns and earnings for high-tech firms. This can be interpreted to imply that market participants are able to infer reasonable depreciation schedules to counteract expense mismatching, but have more difficulty in estimating the level of future benefits in current R&D and other intangible investments.

The present study thus makes two contributions to the accounting literature. First, two factors — expense mismatching and noise — that vitiate the contemporaneous association between earnings and returns have been segregated, and previous work has been extended by disentangling the two effects. Second, the present study provides evidence that noise — arising from uncertainty with respect to future benefits from R&D and other intangible investments — is the dominant factor contributing to the reduced association between returns and earnings for high-tech firms. The GAAP treatment of expensing intangible investments — without disclosing expected future benefits — does not seem to provide investors with enough information to mitigate uncertainty surrounding future cash flows associated with intangibles.

The remainder of this chapter is organized as follows. Section 2 examines the earnings–return relation of high- and low-technology firms, comparing their value relevance. Section 3 presents a theoretical framework, and derives a model for testing earnings' timeliness. Section 4 contains additional empirical analyses, and concludes the chapter.

2. Contemporaneous Association between Returns and Earnings

The authors begin with an assessment of value relevance between 1984 and 1998. Recent studies investigating the association of earnings and returns in high-tech industries (e.g., Francis and Schipper, 1999) employ the earnings-change and level specification from Easton and Harris (1991), and is adopted as well. The model assumes earnings changes comprise transitory and permanent components, rather than only a permanent component, and Ali and Zarowin (1992) show the model dominates the earnings-change model for explaining the association between stock returns and earnings.

$$R_t = \alpha + \beta_1 \Delta X_t / P_{t-1} + \beta_2 X_t / P_{t-1} + \varepsilon_t. \qquad (1)$$

The returns–earnings association was investigated for all firm-year observations from *Compustat* Primary, Supplemental, Tertiary, Research, and Full Coverage files from 1984 to 1998 meeting data requirements. The change in a firm's market value is measured by annual returns.[3] Changes in current and future earnings ($\Delta X_t / P_{t-1}$) are computed by deflating current and future primary earnings per share by the beginning-of-year stock price [(X_t(Item 58)$-X_{t-1}$)$/P_{t-1}$)]. X_t / P_{t-1} (the inverse of price-earnings, EP_{t-1}) is computed by deflating X_{t-1} by P_{t-1}, and annual returns (R) are computed from period t to period $t+3$ for changes in future stock prices. In subsequent tests (described later), each valid firm-year observation also must be associated with earnings and returns ran over a future 3-year period. After eliminating firm-year observations that lack data, the sample comprises 46,223 firm-year observations. In addition, to control for outliers, the guidelines of Belsley *et al.* (1980) are followed, reducing the sample to 43,259.[4]

To assess value relevance for high-tech and low-tech firms, a classification method is required. Firms were categorized as high- or low-tech from select industries based on their use of technology in the production process, following the SIC approach in Francis and Schipper (1999, Table 5). Firms are also categorized based on the relative level of R&D usage, following Lev and Zarowin (1999), who argue that R&D intensity — the ratio of R&D expense (Compustat Item No. 46) to Sales (Compustat Item No. 12) — represents a prime driver of innovation in firms.

Table 1, Panel A reports statistics on the variables R_t, $\Delta X_t / P_{t-1}$, EP_{t-1}, and R&D INTENSITY (RDI) for the entire sample, and Panel B divides the data into quintiles. Quintiles 1 and 2 in Panel B represent low-tech firms, and 4 and 5 contain high-tech firms. Panel C comprises the 11,724 firms categorized as high- or low-technology using the SIC method. Panel D separates high- and low-technology firms using RDI, combining quintiles 1 and 2 versus 4 and 5. It is observed in Panel D that high-tech firms perform better than their low-tech counterparts, as higher mean and median returns (R_t) obtain for them

---

[3]In addition, equal-weighted adjusted returns are alternatively used. No qualitative change in empirical results (described *infra*) is observed.

[4]Belsley *et al.* (1980) suggest three statistics to identify influential observations: RSTUDENT, COVRATIO, and DFFITS. The RSTUDENT statistic computes the change in error variance by deleting the ith observation. Observations with RSTUDENT larger than 2 in absolute value are considered to be outliers. The COVRATIO statistic computes the effect of the ith observation on the determinant of the covariance matrix of the estimates. Observations with the absolute value (COVRATIO $-$ 1) greater than $3 * p/n$ are considered to be outliers, where p is the number of parameters in the model and n is the number of observations. The DFFITS statistic indicates a change in the predicted value by deleting the ith observation. The general cutoff for DFFITS is 2. All three statistics were used to eliminate influential observations.

Table 1. Descriptive statistics for selected variables, from 1984 to 1998 *Compustat* primary, supplemental, tertiary, research, and full coverage files.

Sample	Number obs.	Variable	Mean	STD	Q1	Median	Q3
Panel A: Full sample							
All	43,259	R_t	0.130	0.431	−0.153	0.086	0.342
		$\Delta X_t/P_{t-1}$	0.004	0.138	−0.021	0.004	0.029
		EP_{t-1}	0.032	0.131	0.024	0.054	0.091
		R&D INTENSITY$_t$	0.060	0.402	0.000	0.000	0.019
Quintile	**Number obs.**	**Variable**	**Mean**	**STD**	**Q1**	**Median**	**Q3**
Panel B: Quintiles based on R&D intensity, based on Lev and Zarowin (1999)							
1	8,651	R_t	0.101	0.439	−0.186	0.058	0.331
		$\Delta X_t/P_{t-1}$	0.001	0.149	−0.037	0.003	0.030
		EP_{t-1}	0.028	0.147	0.011	0.049	0.091
		R&D INTENSITY$_t$	0.000	0.000	0.000	0.000	0.000
2	8,652	R_t	0.126	0.427	−0.101	0.109	0.352
		$\Delta X_t/P_{t-1}$	0.003	0.129	−0.023	0.003	0.029
		EP_{t-1}	0.051	0.118	0.036	0.065	0.095
		R&D INTENSITY$_t$	0.000	0.000	0.000	0.000	0.000
3	8,652	R_t	0.159	0.443	−0.113	0.116	0.377
		$\Delta X_t/P_{t-1}$	0.003	0.133	−0.023	0.004	0.027
		EP_{t-1}	0.048	0.117	0.033	0.060	0.096
		R&D INTENSITY$_t$	0.000	0.000	0.000	0.000	0.000
4	8,652	R_t	0.140	0.446	−0.121	0.103	0.361
		$\Delta X_t/P_{t-1}$	0.008	0.134	−0.025	0.006	0.034
		EP_{t-1}	0.037	0.116	0.030	0.062	0.084
		R&D INTENSITY$_t$	0.011	0.014	0.004	0.010	0.019
5	8,652	R_t	0.130	0.414	−0.267	0.069	0.381
		$\Delta X_t/P_{t-1}$	0.003	0.139	−0.031	0.004	0.026
		EP_{t-1}	−0.008	0.140	−0.034	0.036	0.075
		R&D INTENSITY$_t$	0.291	0.799	0.039	0.083	0.146

(*Continued*)

Table 1. (*Continued*)

Sample	Number obs.	Variable	Mean	STD	Q1	Median	Q3
Panel C: The SIC method, following Francis and Schipper (1999, Table 5)							
HITECH	8,030	R_t	0.128	0.409	−0.261	0.049	0.400
		$\Delta X_t/P_{t-1}$	0.005	0.128	−0.031	0.006	0.037
		EP_{t-1}	−0.004	0.133	−0.026	0.031	0.051
		R&D INTENSITY$_t$	0.299	0.711	0.017	0.077	0.129
LOWTECH	3,694	R_t	0.119	0.469	−0.143	0.089	0.326
		$\Delta X_t/P_{t-1}$	0.004	0.131	−0.030	0.005	0.036
		EP_{t-1}	0.039	0.124	0.039	0.052	0.088
		R&D INTENSITY $_t$	0.008	0.025	0.000	0.000	0.014
Quintile	Number obs.	Variable	Mean	STD	Q1	Median	Q3
Panel D: The R&D INTENSITY method							
HITECH	17,304	R_t	0.135	0.427	−0.171	0.087	0.371
		$\Delta X_t/P_{t-1}$	0.006	0.135	−0.027	0.006	0.028
		EP_{t-1}	0.015	0.126	0.009	0.049	0.079
		R&D INTENSITY$_t$	0.146	0.411	0.010	0.031	0.083
LOWTECH	17,303	R_t	0.114	0.431	−0.179	0.074	0.343
		$\Delta X_t/P_{t-1}$	0.002	0.137	−0.031	0.003	0.029
		EP_{t-1}	0.039	0.131	0.021	0.058	0.092
		R&D INTENSITY$_t$	0.000	0.000	0.000	0.000	0.000
Sample	Number obs.	Variable	Mean	STD	Q1	Median	Q3
Panel E: The R&D and depreciation-sales method							
HITECH	15,308	R_t	0.126	0.461	−0.204	0.070	0.378
		$\Delta X_t/P_{t-1}$	0.005	0.144	−0.031	0.005	0.031
		EP_{t-1}	0.017	0.129	−0.003	0.051	0.077
		R&D INTENSITY$_t$	0.139	0.406	0.004	0.045	0.058

(*Continued*)

Table 1. (*Continued*)

Sample	Number obs.	Variable	Mean	STD	Q1	Median	Q3
LOWTECH	15,749	R_t	0.096	0.367	−0.165	0.085	0.311
		$\Delta X_t/P_{t-1}$	0.001	0.121	−0.029	0.003	0.026
		EP_{t-1}	0.031	0.116	0.030	0.067	0.089
		R&D INTENSITY$_t$	0.002	0.001	0.000	0.000	0.000

Definitions: STD, standard deviation; Q1, first quartile; Q3, third quartile; R_t, annual returns in period t; $\Delta X_t/P_{t-1}$, changes in primary earnings per share deflated by the beginning-of-period stock price (Compustat item no. 58 $(X_t - X_{t-1})/P_{t-1}$); EP_{t-1}, the ratio of X_{t-1} over P_{t-1}; Obs., number of firm-year observations from 1984 to 1998; R&D INTENSITY$_t$, the ratio of R&D (Compustat item no. 46) to sales (Compustat item no. 12) for period t.

Panel A: The total sample comprises 43,259 firm-year observations, after eliminating firm-year observations without required data. Outliers are removed by referring to the guidelines of Belsley *et al.* (1980).

Panel B: The full sample partitioned into quintiles based on R&D INTENSITY, defined as in Lev and Zarowin (1999).

Panel C: Firm-year observations in high-technology (HITECH hereafter) and low-technology industries (LOWTECH) are selected from the full sample following the procedures in Francis and Schipper (1999, Table 5).

Panel D: HITECH includes 17,304 firm-year observations in Quintiles 4 and 5 from Panel A; LOWTECH includes 17,303 firm-year observations in Quintiles 1 and 2 from Panel A.

Panel E: HITECH includes 15,308 firm-year observations with R&D INTENSITY greater than or equal to 0.5%, as in Lev and Zarowin (1999). The remaining 27,951 firm-year observations are sorted in ascending order by the ratio of Book (Compustat item no. 60)-to-Market (Compustat item no. 199, fiscal year-end stock price* Compustat item no. 25, shares outstanding), and also by the ratio of Depreciation (Compustat item no. 14)-to-Sales (DS). We categorize 15,749 low-tech firms from the bottom three quartiles in both ways of sorting. These firms have the highest book-to-market ratios (suggesting relatively more tangible assets), and depreciation to sales (relatively more tangible assets).

(0.135 and 0.087) compared to low-tech firms (0.114 and 0.074). The better performance supports their higher $\Delta X_t/P_{t-1}$ (0.006), compared to that of low-tech firms (0.002). Because high-tech firms have more intangibles, lower earnings-price ratios EP_{t-1} are expected, and the RDI partition yields lower mean EP_{t-1} for high-tech firms (0.015 versus 0.039). Statistics on these key variables are consistent with RDI affording a reasonable separation of firms based on their use of technology, and its implications for valuation.

Panel E includes 31,057 firm-year observations. High-tech firms (15,308) are categorized based on whether RDI equals or exceeds 0.5%, following Lev and Zarowin (1999). A portfolio of low-tech firms is also formed for tests to diversify noise (to be described below). In such procedures, test power derives from the number of observations aggregated (Beaver *et al.*, 1980). Since the number of high-tech firms is fixed by the criterion RDI $\geq$ 0.5%, to avoid introducing bias a sample of similar size must be collected to administer equally powerful tests on the high- and low-tech firms. Thus, the balance of 27,951 firm-year observations is sorted in ascending order by the ratio of Book-to-Market value (an alternate measure of R&D intensity [Francis and Schipper, 1999]). Then, another sort is made on the ratio of Depreciation-to-Sales (DS), as firms with higher DS scores tend to have more tangible assets, and thus depreciation expense, leading to higher DS scores. Then, the 15,749 firms from the bottom three quartiles are categorized on both sortings as low-tech. These firms have higher Book-to-Market ratios (suggesting relatively more tangible assets), and higher DS (suggesting relatively more tangible assets). Panel E thus provides another approach to classifying high- and low-tech firms.

In Table 2, regression results have been reported for the full sample and its partitions. For the full sample (Panel A), the model's explanatory power is 0.081. When the sample is divided using the SIC method (Panel B), high-tech firms' adjusted-R^2 of 0.051 is significantly less than low-tech firms' adjusted-R^2 of 0.077 (at the 0.01 level; z-statistic is 2.40). In Panel C, R&D intensity is used to categorize firms. Again, high-tech firms display lower adjusted-R^2 (0.049) than their low-tech comparators (0.113), at the 0.01 level. In Panel D, high- and low-tech firms are partitioned using the RDI and DS method (cf. Panel E, Table 1). High-tech firms' earnings exhibit significantly lower explanatory power (0.047) than their counterparts (0.099), at the 0.01 level (z-statistic = 9.49). Using a variety of approaches to categorize high-technology firms, it is observed that their earnings possess less value relevance compared to low-tech firms, consistent with Lev and Zarowin (1999), but ostensibly at odds with findings in Collins *et al.* (1997) and Francis and Schipper (1999).

It is considered that whether our findings might arise from differences in sample periods. Collins *et al.* (1997) note that intangible intensity increased from 1950s to 1990s, and Francis and Schipper (1999) echo this finding. The measure of R&D intensity was used in Francis and Schipper, R&D expense

Table 2. Empirical results from the earnings-change and levels model (Easton and Harris, 1991): $R_t = \alpha + \beta_1 \Delta X_t/P_{t-1} + \beta_2 X_t/P_{t-1} + \varepsilon_t$, using pooled data from 1984 to 1998.

Sample	Obs.	β_1 (t-statistic)	β_2 (t-statistic)	Adjusted R^2
Panel A: Full sample				
All	43,259	0.91	0.40	0.081
		(50.67)	(48.73)	
Panel B: The SIC method				
Total	11,724	0.79	0.33	0.057
		(21.14)	(18.57)	
HITECH	8,030	0.79	0.33	0.051
		(16.08)	(14.21)	
LOWTECH	3,694	1.09	0.36	0.077
		(14.47)	(12.51)	
Difference[a] in R^2				0.026
z-statistic				2.40
Panel C: The R&D INTENSITY method				
Total	34,607	0.94	0.42	0.086
		(47.26)	(45.88)	
HITECH	17,304	0.74	0.29	0.049
		(22.74)	(18.82)	
LOWTECH	17,303	1.05	0.49	0.113
		(42.96)	(45.02)	
Difference[a] in R^2				0.064
z-statistic				11.80
Panel D: The R&D and depreciation-sales method				
Total	31,057	0.93	0.39	0.076
		(41.93)	(38.19)	
HITECH	15,308	0.73	0.29	0.047
		(20.89)	(17.73)	
LOWTECH	15,749	1.10	0.47	0.099
		(39.49)	(37.79)	
Difference[a] in R^2				0.052
z-statistic				9.49

Variables and partitioning methods are as defined in Table 1. Two-tailed t-statistics are employed to assess the statistical significance of estimates on individual coefficients; critical values are $t_{\alpha=0.10} = 1.64$; $t_{\alpha=0.05} = 1.96$; $t_{\alpha=0.01} = 2.57$.

[a]The difference in adjusted-R^2 between HITECH and LOWTECH.

divided by total assets (RDA), to compare the present data (from 1984 to 1998) to theirs (from 1952 to 1994), using their SIC approach to categorize technology-intensity in firms. Using all data available from *Compustat*, the mean of annual averages of the RDA ratio for high-tech firms from 1952 to 1994 is 5.9% (results not tabulated; Francis and Schipper report 9.2%). For low-tech firms, the mean is 0.68% (Francis and Schipper report 0.8%). Next, the means of annual RDA ratio averages were computed from 1984 to 1998: high-tech is 15.67%, and low-tech is 1.27%. Last, these computations were repeated for 1952 to 1983, and RDA levels of 3.09% for high-tech firms, and 0.46% for low-tech firms were observed. Based on these results, it is plausible that lower R&D intensity in the 32 years from 1952 to 1983 — included in Francis and Schipper's sample, but excluded in the present study — led to their conclusion of no difference in value relevance. Thus, the differential, lower informativeness of earnings for high-tech firms observed may originate from the present sampling over a period in which the temporal increase in R&D investment — documented in both Collins *et al.* (1997) and Francis and Schipper (1999) — is substantial enough to alter the returns–earnings relation.[5]

Even though the present findings can be reconciled relative to Collins *et al.* (1997) and Francis and Schipper (1999), left unanswered is what aspect of earnings leads to lower value relevance. Collins *et al.* (1997) point to temporal changes in the information content of balance sheet versus income statement items. However, this is unlikely to help in explaining earnings of high-technology firms (Amir and Lev, 1996). Except for R&D expense, high-tech firms do not report the costs of intangibles expenditures such as brand development, customer-base creation, and human capital in either their financial statements (Danielson and Press, 2003). *A fortiori*, GAAP precludes one of the most important intangible assets in high-tech firms — the future benefits from R&D investments — from appearing on the balance sheet. Thus, balance sheet items are not likely to offer incremental explanatory power for high-tech firms.

Accordingly, the response to understanding the reason for the differential value relevance documented is to develop a model to address the errors-in-variables problem in the earnings–returns relation. Previous research

[5]Lev and Zarowin (1999, p. 376) present evidence consonant with the present conclusion, noting that the rate of business change has increased, which is associated with an increase in the intensity of R&D, and a decline in earnings' explanatory power.

was extended by examining two factors identified by Collins *et al.* (1994) that could diminish the value relevance of earnings: earnings' lack of timeliness driven by mismatching expenditures on R&D against revenues, and noise caused by uncertainty about future benefits from R&D.

3. Background and Model Development

To investigate the competing explanations of expense mismatching and noise, a model has been derived that corrects the expense-mismatching problem, and consider how noise arising from uncertainty surrounding realization of benefits from R&D and other intangibles can affect investors' value estimations.

3.1. *Expense mismatching (earnings' lack of timeliness) versus noise for high-technology firms*

Collins *et al.* (1994) argue that lack of timeliness in earnings arises from several sources. GAAP generally requires firms to recognize earnings only after all transactions are completed. In contrast, investors revise estimates of firm value based on expected future benefits, potentially even before realization of benefits. Thus, a timing difference exists between earnings and stock prices in reflecting the consequences of a given economic event. In general, stock prices are determined based on a firm's prospects, whereas earnings represent a summary measure of past activity. Accordingly, stock returns are associated with current as well as future earnings. Collins *et al.* (1994) show that including information about future earnings' growth and future stock returns substantially improves the earnings and stock-returns relation.

The timeliness problem is potentially more serious for high-tech firms because they expend relatively more on intangibles that are expensed currently. Benefits from the current expenditures yield future earnings, causing returns to be associated with current as well as future earnings.[6] Thus, it is posited that high-tech firms could be seriously affected by expense mismatching.

[6]Ben Zion (1978) and Hirschey and Weygandt (1985) report that R&D outlays increase a firm's market value relative to its book value. Shevlin (1991) indicates that the market value of a parent firm is positively associated with in-house R&D activities. Sougiannis (1994) provides similar evidence that a firm's market value is positively related with its current and past R&D outlays, and Bublitz and Ettredge (1989) conclude similarly. Chambers *et al.* (1999) and Healy *et al.* (2002) show that the explanatory power of earnings can increase by capitalizing and then amortizing R&D outlays. Chambers *et al.* (1999) report that — using the same accounting rule of capitalizing and amortizing R&D outlays for firms in all industries — one can better explain the distribution of stock prices than by expensing R&D.

Collins *et al.* (1994, p. 294) define noise as an earnings component that is "unrelated to price (and returns), not only in the current period but in all lead and lag periods as well". He considers noise as the difference in estimations by investors and managers with respect to accrual accounts such as loan-loss reserves in the banking industry, warranty expenditures, and pension and other post-retirement obligations.[7] Collins *et al.* (1994) mainly focus on the difference in estimations between investors and managers with respect to future cash flows.

In contrast, the focus is on noise resulting from uncertainty perceived by investors about future cash flows related to investment in R&D and other intangibles.[8] No significant differences in the noise related to the accrual accounts are assumed for high- and low-tech firms. Higher uncertainty about future benefits leads to more information asymmetry between investors and managers, and may induce more noise in the estimation of firm value in high-tech in comparison to low-tech industries.

Indeed, previous research documents high uncertainty associated with intangibles. Kothari *et al.* (2002) report the uncertainty of future benefits with respect to R&D expenditures is higher than for investments in plant and equipment. Lev and Sougiannis (1996) report investors do not appear to incorporate intangible investments in stock prices on a timely basis. Facing difficulty in estimating firm value, investors seek other information. Thus, Amir and Lev (1996) report financial information alone cannot explain the behavior of stock prices in the intangible-intensive cellular-telephone industry. Behn and Riley (1999) report a similar finding for airlines, where nonfinancial performance information is useful in predicting customer satisfaction and accounting measures such as revenues, expenses, and operating income.

Barth *et al.* (2001) show that analysts expend greater efforts to follow firms with substantial intangibles, and share prices in such firms reflect their values

---

These findings support the assertion that expenditures on intangibles have investment value, although benefits may be uncertain.

[7]It is unclear how investors could have a better understanding of a firm's future obligations than its managers. Managers with expertise in their field through formal education and work experience have a better grasp of their firms' current and future financial status than do investors (Myers and Majluf, 1984; Healy and Palepu, 1993).

[8]The model adopted by Collins *et al.* (1994) makes logical and intuitive arguments to correct the errors-in-variables problem, in contrast to the present analytical approach. Despite the different processes followed, both models are qualitatively similar. Use of the Collins *et al.* model is limited, however, since it can be applied only to firms that report positive earnings during the observation period.

less precisely. If this is true, the association between stock returns and earnings could be lower for high-tech firms. However, if noise is the primary factor contributing to the low association between returns and earnings, correcting expense mismatching will not strengthen the association. Investors would still be unable to properly incorporate uncertain future benefits into valuations.[9]

Nonetheless, except for Collins *et al.* (1994), previous accounting studies have not distinguished between expense mismatching and noise in investigating the weak association between earnings and stock returns. For example, Beaver *et al.* (1980) focus on noise, whereas Easton *et al.* (1992), and Warfield and Wild (1992), attempt to control earnings' lack of timeliness to improve the explanatory power of earnings.

Collins *et al.* (1997), and Francis and Schipper (1999), examine the effect of intangible intensity on the explanatory power of earnings, and conclude that intangible intensity does not necessarily lower the association between earnings and market values. On the other hand, Lev and Zarowin (1999) report decreases in the informativeness of earnings as firms increase their expenditures on intangible assets. No prior study has jointly investigated noise from uncertainty of benefits versus expense mismatching as factors contributing to the reduced association between earnings and stock returns for high-technology firms. Both hypotheses were evaluated on the reasons for lessened informativeness of high-tech firms' earnings.

3.2. *Model development*

An attenuated association between earnings and returns can arise from differences in measurement windows of stock returns and earnings. A model has been derived that mitigates the effect of the errors-in-variables problem caused by earnings' lack of timeliness. Included in the model are changes in future earnings, to capture the delayed portion of the current period's events

[9]It is not possible to test empirically if the current GAAP practice of expensing intangibles hinders investors in estimating future benefits, relative to alternative approaches such as capitalizing all intangible assets or capitalizing only successful expenditures. Neither is available to firms. Rather, the relative amount of noise between high- and low-tech firms was compared to obtain indirect evidence with respect to the current accounting practice for intangible assets. Alternatively, Healy *et al.* (2002) examine this issue by developing a simulation model for a pharmaceutical firm. His simulation results shows a higher association between stock returns and earnings when managers are allowed to exercise their discretion over the writing-off for unsuccessful projects for which costs were capitalized, rather than either expensing or capitalizing all of them.

that is reflected in future earnings as well as current earnings. Beginning-of-year earnings and future price changes were included to control for changes in future earnings triggered by prior and future events that are irrelevant to stock returns during the current period. All variables in the model are deflated by the beginning-of-year stock price.

To begin, price is modeled as the present value of expected future dividends:

$$P_t = \sum_{n=1}^{\infty} (1 + r_{t+n})^{-n} E_t(\tilde{d}_{t+n}) \tag{2}$$

where P_t is the firm's stock price at time t, r_t the discount rate at time t, $E_t[.]$ the expected value operator based on the information set at time t, and d_t is the dividend at time t. The firm subscript is suppressed for notational convenience.

Economic earnings (X^*) represent periodic changes in the firm's real economic value, which can differ from the reported accounting earnings (X) generated by the firm's accounting system. Economic earnings at time t summarize the market's expectations about future dividends:

$$P_t = \phi_t^* X_t^* \tag{3}$$

where ϕ_t^* is the theoretical price-to-earnings ratio.

Value-relevant events trigger investors' revisions of expectations about future dividends. If economic earnings are the only information that affects stock prices, then the change in a firm's economic earnings is directly related to the corresponding variation in its stock price (i.e., stock returns).

$$R_t = \phi_t^* \frac{\Delta X_t^*}{P_{t-1}} \tag{4}$$

where R_t refers to continuously compounded returns for period t, ϕ_t^* refers to the theoretical price-to-earnings ratio, and ΔX_t^* refers to the change in the firm's economic earnings for period t. In an earnings and stock-returns model, a change in accounting earnings (ΔX_t) is adopted as a proxy for ΔX_t^*. As discussed in the Appendix, ΔX_t induces the errors-in-variables problem because of the delayed recognition of economic events and noise, demonstrated in Eqs. (A8) and (A9). Thus, Eq. (5) — identical to Eq. (A15) in the appendix — is derived from expanding Eq. (2), and mitigates the errors-in-variables problem:

$$R_t = \alpha + \sum_{m=0}^{\infty} \beta_m * \frac{\Delta X_{t+m}}{P_{t-1}} + \gamma * \mathrm{EP}_{t-1} + \sum_{m=1}^{\infty} \lambda_m * \frac{\Delta P_{t+m}}{P_{t-1}} + \varepsilon_t \tag{5}$$

If earnings are the only information affecting prices, the estimated coefficients in the expense-mismatching (or lack-of-timeliness) model from Eq. (5) are:

$$\begin{aligned}
\hat{\alpha} &= \frac{-\phi_t^*}{\phi_{t-1}^*}, \\
\hat{\beta}_m &= \left[\frac{\sigma^2_{\Delta X_{t+m}}}{\sigma^2_{\Delta X_{t+m}} + \sigma^2_{\eta\infty}}\right] * \phi_t^* * (1 + r_{t+m})^{-m}, \\
\gamma &= \left[\frac{\sigma^2_{X_{t-1}}}{\sigma^2_{X_{t-1}} + \sigma^2_{\eta\infty}}\right] * \phi_t^* * \theta, \quad \text{and} \\
\lambda_m &= \left[\frac{\sigma^2_{\Delta P_{t+m}}}{\sigma^2_{\Delta P_{t+m}} + \sigma^2_{\eta\infty}}\right]\left(\frac{-\phi_t^*}{\phi_{t+m}^*}\right)
\end{aligned} \tag{6}$$

Coefficients on current and future earnings (β_m) are expected to be positive and to attenuate with m, as all measurement biases (e.g., earnings' lack of timeliness) are not completely corrected. The errors-in-variables problem becomes more serious as earnings realizations obtain further into the future. The expected positive sign of γ, the coefficient on EP_{t-1}, is identical to that reported by Collins *et al.* (1994), who show a negative partial correlation between EP_{t-1} and expected earnings growth. The expected negative sign of λ_m, the coefficient on the stream of future returns, is also consistent with what Collins *et al.* report.[10]

[10] Although Eq. (5) is similar to the model used by Collins *et al.*, some differences are noteworthy. For example, the expense-mismatching model deflates earnings changes by the beginning-of-year stock price, whereas Collins *et al.* compute continuously compounded annual growth in earnings and continuously compounded annual growth in investment. Collins *et al.* indicate that alternative measures of annual growth in earnings do not change the tenor of the findings; thus it can be presumed that the present earnings changes deflated by beginning-of-year stock price are comparable to the continuously compounded annual growth in earnings calculated by Collins *et al.* Still, Eq. (5) in this chapter is similar to Eq. (6) of Collins *et al.* (1994), denoted as C6:

$$R_t = b_0 + b_1 X_t + \sum_{k=1}^{3} b_{k+1} X_{t+k} + b_5 \text{EP}_{t-1} + b_6 \text{INV}_t + \sum_{k=1}^{3} b_{k+6} R_{t+k} + e_t \tag{C6}$$

where R_t is the continuously compounded annual return, X_t the continuously compounded annual growth in earnings before extraordinary items, discontinued operations, and special items, EP_{t-1}, the ratio of earnings for year $t-1$ to price at the end of year $t-1$, INV_t the continuously compounded growth rate of investment in year t ($\log[\text{Investment}_t/\text{Investment}_{t-1}]$), and e_t is the error term.

The present analysis, however, does not justify the inclusion of growth in investment to correct earnings' lack of timeliness. More important, the expressions in the present model are derived analytically to resolve the errors-in-variables problem, whereas Collins *et al.* (1994) resort to intuition to select alternative proxies to mitigate the measurement errors arising from the inclusion of future earnings growth in the model. Starting with theory, our analysis confirms much of the intuition underlying the Collins' *et al.* model.

3.3. *Noise from uncertain benefits*

Assume that $\mathrm{LX}_{t,t+n}$ in Eq. (A2) in the Appendix is characterized as $N[E(\mathrm{LX}_{t,t+n}), \sigma^2_{\mathrm{LX}_{t,t+n}}]$, where $\mathrm{LX}_{t,t+n}$ is defined as an incremental change in accounting earnings in period $t+n$ arising from a value-relevant event that takes place in period t. Because of uncertainty about future economic benefits of intangibles, the variance of $\mathrm{LX}_{t,t+n}(\sigma^2_{\mathrm{LX}_{t,t+n}})$ is expected to be greater for high-tech firms. The higher the variance of $\mathrm{LX}_{t,t+n}$, the lower is the R^2 for the association of stock returns with current and future earnings.

The expense-mismatching model assumes investors are able to estimate a stream of future earnings arising from a value-relevant event in a given year.

Collins *et al.* (1994) justify the inclusion of EP_{t-1} in C6 as a proxy for: (1) the market's forecast of earnings growth, (2) the expected earnings growth based on the higher-order negative serial correlation in annual earnings, and (3) the expected return on equity. In addition, they indicate that lagged return is a sub-optimal substitute for EP_{t-1}. INV is included as another proxy for expected future earnings growth. Both C6 and Eq. (5) in this chapter are motivated by mitigation of the errors-in-variables problem in the earnings and returns model by confronting a major cause, the timing difference between returns and earnings measurements. Current earnings do not contain expected financial performance in the future, which is included in returns; however, current earnings do include financial performance realized from past events, which is not included in returns. The expanded earnings and returns models [C6 and Eq. (5)] include future earnings and EP to control the former and latter, respectively.

Nonetheless, C6 and Eq. (5) are motivated differently. As shown above, Collins *et al.* include EP in C6 by referring to expected future growth, but then opens the door for further expansion. The inclusion of another variable, INV, is an example. In other words, they ponder omitted variables in the earnings–returns model by intuition and reference to prior studies. In contrast, the errors-in-variables problem associated with the use of accounting earnings was addressed as a proxy for economic earnings, assuming that economic earnings are the only variable that determines stock prices. For example, Eq. (A10) in the appendix demonstrates the need for adjustments to align accounting earnings with economic earnings. Then, Eq. (A10) is simplified to derive the final model, Eq. (A15). The contribution of the present model beyond C6 is to transform general intuition into an analytical model, and to identify the shortcomings of accounting earnings as a proxy for economic earnings in the earnings–returns model.

If high uncertainty about expected future benefits is associated with a given value-relevant event, and if information to mitigate uncertainty is unavailable, realized future earnings are different from earlier estimates. The deviation between estimated and realized amounts lessens the explanatory power of the expense-mismatching model. Thus, the greater the uncertainty, the more noise investors incorporate in their estimation of firm value. If this is the case, the efficacy of the model drops.

An industry-level aggregation method was adopted to diversify away noise, as suggested by Collins *et al.* (1994), and use it in conjunction with the portfolio-grouping approach in Beaver *et al.* (1980). Since the variance of each incremental future earnings change $\mathrm{LX}_{t,t+n}$ is assumed to be firm-specific, noise induced by high uncertainty regarding future benefits can be diversified by aggregating variables within the same industry. Thus, if high-tech firms face more uncertainty associated with future benefits of a current economic event, diversifying away noise should enhance their earnings/return associations relatively more than for their low-tech counterparts.

4. Empirical Results

4.1. *Expense mismatching*

One explanation of the weaker association between their earnings and returns is high-tech firms' earnings contain more value-irrelevant components. The present investigation was expanded by correcting lack of timeliness. And begin by estimating the expense-mismatching model — Eq. (5), *supra* — on the full sample.[11]

Table 3 displays results from testing the expense-mismatching model based on the R&D INTENSITY method in Panel A, and the R&D and Depreciation Expense-Sales method in Panel B. When expense mismatching is corrected, the explanatory power of earnings substantially increases. The adjusted-R^2 for the pooled sample in Panel A is 0.144, a 67% increase relative to the earnings-change and levels model (0.086 in Table 2, Panel C). The pooled sample in Panel B shows a 72% increase (0.131, Table 3, Panel B versus

[11]We include three future years' earnings changes to investigate earnings' lack of timeliness, consistent with prior studies (Warfield and Wild, 1992; Collins *et al.*, 1994). The model was also examined with up to six future years' earnings changes.

Table 3. Empirical results of the expense-mismatching model, using pooled data from 1984 to 1998, $R_t = \alpha_0 + \beta_1 \Delta X_t / P_{t-1} + \sum_m \beta_{m+1} \Delta X_{t+m/P_{t-1}} + \gamma X_{t-1}/P_{t-1} + \sum_m \lambda_m R_{t+m} + \varepsilon_t$, where $m = 1$, 2 and 3

Sample	β_1	β_2	β_3	β_4	γ	λ_1	λ_2	λ_3	Obs.	Adjusted R^2	Increase in R^2
Panel A: The R&D INTENSITY method											
ALL	1.86	0.97	0.33	0.12	0.84	−0.06	−0.03	−0.06	34,607	0.144	0.058
(t-statistic)	(73.11)	(55.38)	(25.10)	(11.61)	(71.89)	(−13.22)	(−6.89)	(−13.14)			
HITECH	1.64	0.88	0.29	0.11	0.70	−0.06	−0.02	−0.02	17,304	0.081	0.032
(t-statistic)	(36.96)	(30.08)	(13.07)	(6.10)	(33.96)	(−8.94)	(−2.37)	(−2.42)			
LOWTECH	1.97	1.01	0.34	0.12	0.92	−0.06	−0.05	−0.05	17,303	0.219	0.106
(t-statistic)	(65.57)	(48.04)	(22.19)	(10.21)	(67.32)	(−9.83)	(−8.30)	(−8.93)			
Difference[a] in R^2											0.074
z-statistic											14.65
Panel B: The R&D and depreciation-sales method											
ALL	1.86	0.95	0.35	0.13	0.81	−0.05	−0.03	−0.03	31,057	0.131	0.055
(t-statistic)	(65.46)	(50.49)	(24.00)	(11.51)	(62.34)	(−11.11)	(−6.04)	(−13.24)			
HITECH	1.66	0.89	0.29	0.11	0.72	−0.06	−0.01	−0.02	15,308	0.101	0.054
(t-statistic)	(34.62)	(28.44)	(12.18)	(5.88)	(32.26)	(−8.33)	(−1.30)	(−1.83)			
LOWTECH	2.03	0.99	0.41	0.15	0.89	−0.05	−0.06	−0.06	15,749	0.201	0.102
(t-statistic)	(54.84)	(44.02)	(22.74)	(10.87)	(58.28)	(−7.5)	(−8.94)	(−9.87)			
Difference[a] in R^2											0.048
z-statistic											8.49

Variables and partitioning methods are as defined in Table 1. Increase in R^2: The incremental explanatory power computed by subtracting adjusted-R^2 of the earnings-change and levels model reported in Table 2 from the corresponding R^2 of the expense-mismatching model (Eq. [5]) model. z-statistic: A z-statistic, based on Cramer (1987), for testing whether or not the increase in R^2 from the earnings-change and levels model to the expense-mismatching model for firms in LOWTECH is greater than those for HITECH (cf. Sec. 4.2 for further elaboration). Two tailed t-statistics are employed to assess the statistical significance of estimates on individual coefficients; critical values are $t_{\alpha=0.10} = 1.64$; $t_{\alpha=0.05} = 1.96$; $t_{\alpha=0.01} = 2.57$.

[a]The difference in adjusted-R^2 from the earnings-change and levels model (cf. Table 2) versus the expense-mismatching model (Eq. [5]), between HITECH and LOWTECH.

0.076 in Table 2, Panel D). All coefficients are significant, and have signs consistent with the model's predictions.

Coefficients for changes in current and future earnings and EP_{t-1} are positive, and those for all future Rs are negative. In addition, the magnitude of coefficients on changes in future earnings gradually attenuates as they progress further into the future. For example, in Panel A of Table 3, β_1 (the coefficient on $\Delta X_t/P_{t-1}$) is 1.86, and β_4 (the coefficient on $\Delta X_{t+3}/P_{t-1}$, three-period-ahead earnings) is 0.12. In other words, the errors-in-variable problem is larger for future earnings that are also affected by future periods' value-relevant events. Panel A provides evidence on expense mismatching for the sample divided into high- and low-tech firms using R&D INTENSITY; explanatory power for HITECH (adjusted-R^2 of 0.081) is lower than that for LOWTECH (0.219). In addition, the adjusted-R^2 for high-tech firms increases 0.032 between the earnings-change and levels model and the expense-mismatching model, but increases by 0.106 for the low-tech firms.

To assess the statistical significance of the changes in adjusted-R^2, Cramer's (1987) procedure was invoked. A z-statistic was estimated as $[E(\Delta R_T^2) - E(\Delta R_K^2)] \Big/ \sqrt{\sigma^2(\Delta R_T^2) + \sigma^2(\Delta R_K^2)}$, where $E(\Delta R_i^2)$ and $\sigma^2(\Delta R_i^2)$ represent the expected value and variance of ΔR^2 for sample i. This differs slightly from Cramer (1987), who tests the difference of R^2 between two different samples, and whose method is adopted in studies such as Nwaeze (1998).

However, both R^2 and ΔR^2 follow the same probability-density function proposed by Cramer (Eq. 11a). For instance, the R^2 from the expense-mismatching model (XM) with nine parameters (0.081 for HITECH firms) comprises two components: R^2 (0.049) associated with the earnings-change and levels model with three parameters, and ΔR^2 (0.032) arising from the inclusion of six additional parameters in the model. Thus, if both R^2s from the XM and earnings-change and levels models follow the Cramer probability-density function, ΔR^2 follows suit. Accordingly, R^2 was replaced with ΔR^2 to estimate the z-statistic. Applying Cramer's approach, it is found that the 0.074 difference between the samples is significant at the 0.01 level (z-statistic is 14.65).

Panel B displays similar results. The adjusted-R^2 for HITECH firms increases 0.054 between the earnings-change and levels model and the expense-mismatching model, but increases by 0.102 for the LOWTECH firms.

Cramer's test indicates that the 0.048 difference between the two samples is significant at the 0.01 level (z-statistic $= 8.49$). Thus, correcting the expense-mismatching problem for low-tech firms has more impact on increasing the explanatory power of earnings. This is an important outcome in light of prior studies that report no substantial difference in the relevance of financial information between high- and low-technology firms (Collins *et al.*, 1997; Francis and Schipper, 1999).[12] Apparently, correcting the distortion that arises in depreciating investment bases comprising a higher percentage of tangible assets enhances the earnings–return relation more in low-tech firms. As well, empirical support obtains for the present analytical model, because correcting expense mismatching enhances the explanatory power of earnings in the full sample, and for both the HITECH and LOWTECH sub-samples.

The present analysis shows that the association between earnings and returns is weaker for high-tech firms, even after correcting expense mismatching. Little support was found for the hypothesis that, because high-tech firms spend more on intangible assets, they should realize a greater increase in the explanatory power of financial information by correcting expense mismatching.

This finding can be further examined. First, only three future years were included in the testing, following Collins *et al.* (1994), but the timing problem may encompass a broader horizon (Healy *et al.*, 2002). Thus, the expense-mismatching model was tested by including earnings changes and stock returns in six future years. The longer horizon of observations reduces the size of sample, which declines to 23,835 firm-year observations. However, the present findings are unchanged qualitatively. High-tech firms for which RDI is greater or equal to 0.5% display lower adjusted-R^2 (0.103) than low-tech counterparts (0.151), in which RDI is lower than 0.5% (untabulated). Therefore, no empirical support was found for the argument that expenditures on intangible assets affect future earnings for a horizon greater than 3 years.

---

[12]A direct comparison between the present findings and those of prior studies, in particular, Francis and Schipper (1999) should not be made without caution. First, the present sample is more restricted, requiring firms to report earnings in three more future years. Second, the most recent period of 15 years was focused, whereas Francis and Schipper cover a longer period, from 1952 to 1994. Further, Francis and Schipper show that the value relevance of earnings has declined in the recent period, especially for high-tech firms. If this decline were associated with an increase in a firm's expenditures on intangibles, the present findings can be construed as consistent with Francis and Schipper.

Second, market participants with rational expectations could make their own schedule for amortizing R&D investments, yet the explanatory power of the model for high-tech firms remains lower than for low-tech firms because investors face greater uncertainty (noise) with respect to future benefits for high-tech firms.[13] In the next section, the effect of noise, manifest as uncertain benefit realizations, was investigated.

4.2. *Noise*

If investors face higher uncertainty associated with future benefits from intangibles, the estimation of future benefits is less accurate for high-tech firms compared to LOWTECH — i.e., HITECH is noisier. Thus, inclusion of future earnings growth in the returns and earnings model (i.e., correcting lack of timeliness) for high-tech firms is not as effective in improving the explanatory power of the model.

Noise is assumed to follow not only a random-walk process with an expected value of 0, but also to be firm specific, with no systematic correlation between firms (Beaver *et al.*, 1980). Collins *et al.* (1994) use this assumption on the S&P 500 composite and its subsets, to diversify away noise, assuming noise in each firm occurs independently within the industry. However, they also allows for the possibility that noise has cross-sectional and serial correlations.

The approaches taken in the two papers — group techniques (Beaver *et al.*) and industry controls (Collins *et al.*) — were combined by aggregating earnings and stock returns using a combination of year and two-digit SIC codes. For each year, the medians for current and future Rs and $\Delta X_t/P_{t-1}$s are selected from each SIC two-digit portfolio to form an aggregated sample. As Table 4 shows, after aggregating the annual observations based on R&D Intensity by two-digit SIC codes, sample sizes across years range from 39 to 66 industries, comprising 808 to 1,428 firms. The smaller number of two-digit industries in HITECH is consistent with a concentration of R&D expenditures in selected

[13]This interpretation is consistent with results from prior studies. Kothari *et al.* (2002) indicate that future earnings resulting from R&D expenditures are more uncertain and less reliable than those from capital expenditures. Thus, there are more opportunities for informed investors to earn abnormal returns from R&D-intensive firms than from other firms (Aboody and Lev, 2000). Cf. AAA Financial Accounting Standards Committee (2003) for further discussions regarding the systematic undervaluation of R&D-intensive firms.

industries. Using the aggregated sample for both HITECH and LOWTECH partitions, the earnings-change and levels, and expense mismatching, models were ran, and the results were reported in Table 4.[14]

Results on yearly data are lower than those from the pooled data. The mean of adjusted-R^2s from yearly data is 0.038 for firms in HITECH (Table 4, Panel A) when R is regressed on $\Delta X_t/P_{t-1}$, and X_t/P_{t-1} and 0.078 for LOWTECH (Panel B). When the timeliness problem is corrected using the expense-mismatching (XM) model, the mean of annual adjusted-R^2s increases to 0.11 for firms in HITECH (Panel A), and 0.176 for LOWTECH (Panel B). Thus, results of yearly data are consistent with those based on the pooled data. When noise is dispersed using the two-digit SIC industry-level aggregation technique [Columns 6 and 7, labeled CL(C) and XM(D)], the mean of adjusted-R^2s for firms in HITECH (LOWTECH) is 0.235 (0.217), representing an increase of 0.197 (0.139) in comparison to the earnings-change and levels model.

Table 4, Panel C summarizes the effects on adjusted-R^2 when expense mismatching and noise are jointly corrected. The mean difference in adjusted-R^2 between the earnings-change and levels model for HITECH versus LOWTECH is -0.04 — obtained by subtracting the average annual adjusted-R^2 of 0.078 (for LOWTECH) from 0.038 (HITECH) — and the t-statistic for the test of differences of -3.08 is statistically significant. When earnings' lack of timeliness is controlled, higher mean changes in adjusted-R^2 (0.098) obtain for firms in LOWTECH than for HITECH counterparts (0.071), and the mean annual difference -0.027 is significant at the 0.05 level (t-statistic $= -2.04$). Correcting expense mismatching is more effective for LOWTECH, consistent with results from Table 3. The evidence is consistent with correction of expense timing having a greater impact for low-tech firms, because the amortization for their future benefits can be predicted with more certainty.

When the noise in earnings is controlled, high-tech firms exhibit greater increases in adjusted-R^2 than do low-tech firms (0.197 versus 0.139), but the difference (0.058) is insignificant (t-statistic $= 1.46$). Nonetheless, in 10 out

---

[14]Table 4 displays empirical results using samples categorized by the R&D and Depreciation-Sales method. No qualitative difference in empirical results exists between the R&D INTENSITY method and the R&D and DS method. Since the latter in part attempts to control the effect of other intangibles in addition to R&D, the R&D and DS method is chosen for Table 4.

Table 4. Adjusted-R^2s of yearly cross-sectional regression models from 1984 to 1998, the earnings-change and levels model (CL): $R_t = \alpha_0 + \beta_1 \Delta X_t / P_{t-1} + \beta_2 X_t / P_{t-1} + \varepsilon_t$, and the expense-mismatching model (XM): $R_t = \alpha_0 + \beta_1 \Delta X_t / P_{t-1} + \sum_m \beta_{m+1} \Delta X_{t+m} / P_{t-1} + \gamma X_{t-1} / P_{t-1} + \sum_m \lambda_m R_{t+m} + \varepsilon_t$, where $m = 1, 2$ and 3.

Year	No. of firms[a]	Raw data[b]		No. of SIC two-digit industries[c]	SIC two-digit portfolio[d]		Incremental R^2	
		CL	XM		CL	XM	Raw[e] XM–CL	CL[f] SIC2-raw
Panel A: HITECH sample								
1984	894	0.106	0.227	43	0.331	0.542	0.121	0.225
1985	884	0.084	0.175	41	0.207	0.435	0.091	0.123
1986	857	0.007	0.083	39	0.275	0.562	0.076	0.268
1987	896	0.033	0.137	41	0.315	0.619	0.104	0.282
1988	963	0.054	0.108	42	0.243	0.552	0.054	0.189
1989	979	0.039	0.171	41	0.275	0.435	0.132	0.236
1990	977	0.039	0.132	44	0.325	0.412	0.093	0.286
1991	1008	0.019	0.065	41	0.063	0.247	0.046	0.044
1992	1039	0.051	0.097	42	0.243	0.329	0.046	0.192
1993	1145	0.004	0.049	44	0.162	0.387	0.045	0.158
1994	1182	0.044	0.088	41	0.199	0.309	0.044	0.155
1995	1161	0.014	0.058	40	0.087	0.185	0.044	0.073
1996	1137	0.043	0.107	39	0.273	0.211	0.064	0.230
1997	1142	0.015	0.048	43	0.189	0.135	0.033	0.174
1998	1044	0.021	0.100	39	0.337	0.046	0.079	0.316
Mean		0.038	0.110		0.235	0.360	0.071	0.197
Panel B: LOWTECH sample								
1984	858	0.009	0.094	63	0.152	0.532	0.085	0.143
1985	824	0.008	0.097	62	0.314	0.492	0.089	0.306
1986	808	0.076	0.148	64	0.059	0.569	0.072	−0.017
1987	938	0.119	0.190	65	0.437	0.454	0.071	0.318
1988	957	0.025	0.039	65	0.303	0.441	0.014	0.278
1986	936	0.109	0.276	66	0.226	0.314	0.167	0.117
1990	1053	0.114	0.261	65	0.386	0.304	0.147	0.272
1991	1020	0.067	0.221	65	0.342	0.509	0.154	0.275
1992	976	0.068	0.135	66	0.317	0.604	0.067	0.249
1993	1087	0.055	0.134	65	0.076	0.268	0.079	0.021
1994	1428	0.098	0.233	64	0.192	0.230	0.135	0.094
1995	1367	0.073	0.149	64	0.071	0.356	0.076	−0.002

(*Continued*)

Table 4. (*Continued*)

Year	No. of firms[a]	Raw data[b] CL	Raw data[b] XM	No. of SIC two-digit industries[c]	SIC two-digit portfolio[d] CL	SIC two-digit portfolio[d] XM	Incremental R^2 Raw[e] XM–CL	Incremental R^2 CL[f] SIC2-raw
1996	1291	0.129	0.220	66	0.166	0.331	0.091	0.037
1997	1090	0.141	0.259	65	0.164	0.332	0.118	0.023
1998	1116	0.080	0.189	63	0.053	0.118	0.109	−0.027
Mean		0.078	0.176		0.217	0.390	0.098	0.139

Panel C: Tests for mean differences between HITECH and LOWTECH industries — Panels A versus B

	Comparisons of mean values of adjusted-R^2 (Panel A–Panel B)	*Differences in means (t-statistics)*
Raw data (CL): The earnings-change and levels model	(0.038–0.078)	−0.04 (−3.08)
Raw data (XM–CL): Incremental R^2 by correcting expense mismatching	(0.071–0.098)	−0.027 (−2.04)
CL (SIC2-Raw): Incremental R^2 by diversifying away noise	(0.197–0.139)	0.058 (1.46)
SIC two-digit portfolio (XM): The model correcting both expense mismatching and noise	(0.360–0.390)	−0.03 (−0.53)

The aggregated variables are computed by partitioning yearly firm observations into SIC two-digit industries and then selecting median dependent and independent variables from each partitioned group; all other variables are as defined in Table 1. Two-tailed t-statistics assess the significance of estimates; critical values of $t_{\alpha=0.10} = 1.64$; $t_{\alpha=0.05} = 1.96$; $t_{\alpha=0.01} = 2.57$.
[a]Number of firms, by year.
[b]R^2s from CL and XM models, using yearly firm observations.
[c]Number of SIC two-digit industries.
[d]R^2s from CL and XM models, using yearly data comprised of industry-level aggregated variables.
[e]Increase in R^2 by correcting expense mismatching — XM versus CL using raw data.
[f]Increase in R^2 by diversifying away noise-SIC two — digit portfolio versus raw data using the CL model.

of 15 years, high-tech firms show greater increases in adjusted-R^2 than do low-tech firms. Finally, when we jointly control expense mismatching and noise (by using both XM and the SIC two-digit industry data), the difference of adjusted-R^2s, −0.03 (0.36 less 0.39) is insignificant (t-statistic = −0.53).

The finding was interpreted to imply that noise — not expense mismatching — is the reason for the reduced association between earnings and stock returns for firms in high-tech industries. Elevated levels of noise for high-technology firms arise from uncertainty about the realization of benefits from their investments in R&D and intangibles.

The present result is consistent with the findings of Kothari *et al.* (2002), who observe that future benefits from investments in R&D are less certain than those from capital expenditures. Since many capital expenditures and their related amortization processes are disclosed in, or can be inferred from, details in financial statements, investors use these figures to estimate firm value. Future benefits from intangible investments, however, are more difficult to estimate because of lack of information. Indeed, in industries such as cellular telephone and airlines, financial data alone are scarcely relevant for valuation (Amir and Lev, 1996; Behn and Riley, 1999).

5. Conclusions

In this chapter, the value relevance of earnings was assessed for high-tech firms by investigating two factors — expense mismatching that causes earnings' timeliness problems, and noise arising from uncertain benefits from R&D and intangibles — that prior research suggests diminish the contemporaneous association between earnings and returns. The weaker association between stock returns and earnings for firms in high-technology industries — relative to firms in low-technology industries — primarily were attributed to noise. This conclusion differs from that in Collins *et al.* (1994), who find lack of timelines is the main factor weakening the association between stock returns and earnings.

To explain the difference, it is noted that Collins' *et al.* finding could be affected by his sample, profitable firms that survive for long periods (at least 39 years). In that setting, investors may predict future performance more reliably from the past time series. Further, the prediction of positive earnings is much easier than that of negative earnings (Collins *et al.* include only firms with positive earnings, but the present sample includes negative and positive earnings). Thus, Collins' *et al.* conclusion that expense mismatching is the main factor vitiating the association between returns and earnings may stem from the low uncertainty investors perceive in predicting future performance in the type of firms studied.

A broader sample was studied — all firms for which data are available from Compustat and CRSP between 1984 and 1998. This chapter contributes to the accounting literature by separately evaluating the two factors that weaken the association between earnings and returns. The expense-mismatching problem is unavoidable because of conservatism underlying the accounting-measurement process. Our results indicate firms in high-tech industries are not as severely affected by expense mismatching as firms in low-tech industries. Rather, the problem for investors is more uncertainty about the valuation of future benefits — noise — arising from expenditures on intangible assets. After correcting for both expense mismatching and noise, no reliable differences were found in the explanatory power of their earnings.

The present model and evidence provide new insights into the process by which earnings are incorporated into prices, especially in light of accounting conservatism that excludes intangibles from financial statements. To mitigate the uncertainty associated with future benefits and improve the quality of investors' decisions, high-tech firms could provide investors with more information about their intangible expenditures (Boone and Raman, 2001). The practice of expensing all R&D outlays does not appear to serve this purpose; managers have no means to communicate private information regarding their expectations of future investment value either through the balance sheet or income statement. The present recommendation is consistent with results in Lundholm and Myers (2002), who provide empirical evidence that a firm's stock returns better reflect future earnings news when its level of voluntary disclosure increases. For example, a firm could accumulate R&D expenditures incurred for each of the major research projects it has undertaken. General disclosures of research projects and patents or other tangible consequences that flow from past efforts would assist investors in projecting current R&D expenditures onto future successful outcomes, yet need not entail any revelation of proprietary secrets. Such incremental efforts could aid in reducing the uncertainty of estimating future benefits.

Thus, the present findings can support capitalization of intangibles and continual restatement of financial statements to reflect the evolving resolution of uncertainty, as proposed by Lev and Zarowin (1999). Alternatively, Healy *et al.* (2002) indicate investors would receive more relevant information by allowing managers to exercise discretion over the measurement of intangible investments. This process may lead to a trade-off between earnings management and communication of relevant information. Whether the expected benefits from providing additional private information outweigh the costs from potential earnings management is an important question left for future research.

APPENDIX

Both R_t and ΔX_t^* represent a change in a firm's value resulting from value-relevant events during period t. Since current earnings represent the current period's economic benefits resulting from value-relevant events in past as well as current periods, X_t^* can be expressed as the accumulated balance of ΔX^*s from the firm's inception to time t as follows:

$$X_t^* = \sum_{k=1}^{t} \Delta X_k^* \tag{A1}$$

ΔX_t is defined as the change in accounting earnings for period t. Let $\mathrm{LX}_{t,t+n}$ be the change in accounting earnings during period $t+n$ arising from value-relevant events during period t. The change in stock price during period t can be expressed in terms of a stream of $\mathrm{LX}_{t,t+n}$s:

$$\Delta P_t = \sum_{n=0}^{\infty} [(1 + r_{t+n})^{-n} E_t(\mathrm{LX}_{t,t+n})] \tag{A2}$$

Combining Eq. (3) in the text and Eq. (A2):

$$\Delta X_t^* = \frac{\Delta P_t^*}{\phi_t^*} = \frac{\sum_{n=0}^{\infty} [(1 + r_{t+n})^{-n} E_t(\mathrm{LX}_{t,t+n})]}{\phi_t^*} \tag{A3}$$

If earnings are not subject to lack of timeliness, LX_t is constant from time t to time $t+n$, and ΔX_t^* is equal to ΔX_t. In the presence of earnings' lack of timeliness, however, a portion of ΔX_t^* will gradually be reflected in a stream of future earnings until its effect is exhausted. An incremental increase in economic earnings during period $t+n$ resulting from a value-relevant event during period t is expressed as:

$$\omega_{t,t+n} \Delta X_t^* = \mathrm{LX}_{t,t+n} - \mathrm{LX}_{t,t+n-1} \quad n = 0, 1, \ldots, \infty \tag{A4}$$

where $\omega_{t,t+n}$ represents an incremental amount of ΔX_t^*, which is recognized during period $t+n$. $\sum_{n=0}^{\infty} \omega_{t,t+n} \Delta X_t^*$ is greater than ΔX_t^* because of time value of money with respect to delayed recognition of the series of $\omega_{t,t+n} \Delta X_t^*$s in the future. Thus, the ratio of ΔX_t^* to $\sum_{n=0}^{\infty} \omega_{t,t+n} \Delta X_t^*$ measures the severity of a firm's earnings' lack of timeliness. Combining the above two expressions, the following is obtained

$$\theta_t \equiv \frac{\Delta X_t^*}{\sum_{n=0}^{\infty} \omega_{t,t+n} \Delta X_t^*} \tag{A5}$$

By intuition, θ_t is inversely related to the severity of earnings' lack of timeliness and a firm's discount rate.

Next, it is shown that the earnings change (ΔX_t) contains information about economic earnings in prior as well as current periods. Thus, earnings are expressed using the notation in Eq. (A4), as:

$$\begin{aligned} X_t &= \mathrm{LX}_{1,t} + \mathrm{LX}_{2,t} + \cdots + \mathrm{LX}_{t-1,t} + \omega_{t,t} \Delta X_t^* + \eta_t \\ &= \sum_{k=1}^{t-1} \left(\sum_{l=k}^{t} \omega_{k,l} \Delta X_k^* \right) + \omega_{t,t} \Delta X_t^* + \eta_t \end{aligned} \tag{A6}$$

where X_t is the reported earnings at time t and η_t is the value-irrelevant noise. The first difference of X_t in (A6) is:

$$\begin{aligned} \Delta X_t &= \sum_{k=1}^{t} \omega_{k,t} \Delta X_k^* + (\eta_t - \eta_{t-1}) \\ &= \sum_{k=1}^{t-1} \omega_{k,t} \Delta X_k^* + \omega_{t,t} \Delta X_t^* + (\eta_t - \eta_{t-1}) \end{aligned} \tag{A7}$$

When the change in accounting earnings is adopted as a proxy for unexpected earnings, ΔX_t is garbled with components that are not incorporated in stock returns. For example, ΔX_t in (A7) includes three components: (1) changes in the firm's economic earnings in period t triggered by value-relevant events occurring in the past, (2) a fraction of changes in the firm's economic earnings in period t triggered by the value-relevant event in the same period, and (3) value-irrelevant noise. The first component is value-irrelevant with respect to current stock returns. The second component is value relevant with respect to current stock returns, but reflects only a fraction of the change in a firm's value resulting from the current period's value-relevant events. Its remainder is deferred to future periods. Accordingly, using ΔX_t as a proxy for ΔX_t^* in the contemporaneous association between earnings and returns induces an errors-in-variables problem. Equations (A5) and (A7) have been combined to highlight the inconsistency in measurement between accounting earnings (ΔX_t) and a change in a firm's economic value (ΔX_t^*):

$$\begin{aligned}\mu_t &= \Delta X_t - \Delta X_t^* \\ &= \sum_{k=1}^{t-1} \omega_{k,t} \Delta X_k^* + \left[\omega_{t,t} \Delta X_t^* - \theta_t \sum_{n=0}^{\infty} \omega_{t,t+n} \Delta X_t^* \right] + (\eta_t - \eta_{t-1}) \quad \text{(A8)}\end{aligned}$$

Earnings' lack of timeliness implies that current value-relevant events affect earnings in current as well as future periods. To capture the effect of value-relevant events in the current period on future earnings, changes in current and future earnings are combined, using Eqs. (A5) and (A7):

$$\begin{aligned}\sum_{m=0}^{\infty} \Delta X_{t+m} &= \sum_{k=1}^{t-1} \sum_{l=t}^{\infty} \omega_{k,l} \Delta X_k^* + \sum_{n=0}^{\infty} \omega_{t,t+n} \Delta X_t^* \\ &\quad + \sum_{m=1}^{\infty} \sum_{n=m}^{\infty} \omega_{t+m,t+n} \Delta X_{t+m}^* + [\eta_\infty - \eta_{t-1}] \\ &= \sum_{k=1}^{t-1} \sum_{l=t}^{\infty} \omega_{k,l} \Delta X_k^* + \frac{\Delta X_t^*}{\theta_t} \\ &\quad + \sum_{m=1}^{\infty} \sum_{n=m}^{\infty} \omega_{t+m,t+n} \Delta X_{t+m}^* + [\eta_\infty - \eta_{t-1}] \quad \text{(A9)}\end{aligned}$$

Changes in both current and future earnings comprise four components: (1) changes in both current and future earnings resulting from past value-relevant events; (2) changes in both current and future earnings resulting from value-relevant events in the current period; (3) changes in future earnings arising from value-relevant events that will occur in the future; and (4) noise. When Eq. (A9) is rearranged with respect to ΔX_t^*:

$$\begin{aligned}\Delta X_t^* &= \theta_t * \left\{ \sum_{m=0}^{\infty} \Delta X_{t+m} - \sum_{k=1}^{t-1} \sum_{l=t}^{\infty} \omega_{k,l} \Delta X_k^* \right. \\ &\quad \left. - \sum_{m=1}^{\infty} \sum_{n=m}^{\infty} \omega_{t+m,t+n} \Delta X_{t+m}^* - [\eta_\infty - \eta_{t-1}] \right\} \quad \text{(A10)}\end{aligned}$$

The first block within the bracket on the right-hand side of Eq. (A10) represents changes in both current and future earnings. The second block refers to effects of prior value-relevant events on changes in both current and future earnings, which are not value-relevant with respect to current returns. The third block indicates the effects of future value-relevant events on changes in future earnings. The last block within the bracket refers to value-irrelevant noise. Using Eq. (A5), Eq. (A10) is simplified as:

$$\Delta X_t^* = \theta_t * \left\{ \left[\sum_{m=0}^{\infty} \Delta X_{t+m} \right] - \left[\sum_{k=1}^{t-1} \frac{\Delta X_k^*}{\theta_k} - X_{t-1} \right] - \left[\sum_{m=1}^{\infty} \frac{\Delta X_{t+m}^*}{\theta_{t+m}} \right] - \eta_\infty \right\} \tag{A11}$$

Assuming θ is constant over time, $\sum_{k=1}^{t-1} \frac{\Delta X_k^*}{\theta_k}$ is reduced to X_{t-1}^*/θ. Eq. (A11) is then

$$\Delta X_t^* = \theta * \left\{ \left[\sum_{m=0}^{\infty} \Delta X_{t+m} \right] - \left[\frac{X_{t-1}^*}{\theta} - X_{t-1} \right] - \left[\sum_{m=1}^{\infty} \frac{\Delta X_{t+m}^*}{\theta} \right] - \eta_\infty \right\} \tag{A12}$$

The expression X_{t-1}^*/θ in Eq. (A12) refers to the sum of changes in firm value arising from all value-relevant events that had taken place over the period from the firm's inception to time $t-1$. Thus, subtracting X_{t-1} from X_{t-1}^*/θ eliminates changes in a firm's value associated with a stream of accounting earnings up to time $t-1$, leaving only effects of prior periods' value-relevant events on changes in earnings during current and future periods.

Eq. (3) in the text is restated by replacing ΔX_t^* with the terms that are defined in (A12):

$$R_t = \left[\sum_{m=0}^{\infty} \phi_t^* \frac{(1+r_{t+m})^{-m} \Delta X_{t+m}}{P_{t-1}} \right] - \left[\phi_t^* \frac{X_{t-1}^*}{P_{t-1}} \right] + \left[(\theta * \phi_t^*) \frac{X_{t-1}}{P_{t-1}} \right] - \left[\sum_{m=1}^{\infty} \phi_t^* \frac{\Delta X_{t+m}^*}{P_{t-1}} \right] - \phi_t^* \frac{\eta_\infty}{P_{t-1}} \tag{A13}$$

Since $\frac{X^*_{t-1}}{P_{t-1}} = \frac{1}{\phi_t}^*$ (see Eq. [2]) and $\Delta X^*_{t+m} = \Delta P_{t+m}/\phi^*_{t+m}$ (see Eq. [3]), Eq. (A13) is simplified as:

$$R_t = \left[\sum_{m=0}^{\infty} \phi^*_t \frac{(1+r_{t+m})^{-m}\Delta X_{t+m}}{P_{t-1}}\right] - \left[\frac{\phi^*_t}{\phi^*_{t-1}}\right] + \left[\left(\theta * \phi^*_t\right)\mathrm{EP}_{t-1}\right] - \left[\sum_{m=1}^{\infty} \frac{\phi^*_t}{\phi^*_{t+m}} \frac{\Delta P_{t+m}}{P_{t-1}}\right] - \phi^*_t \frac{\eta_\infty}{P_{t-1}} \quad \text{(A14)}$$

where EP_{t-1} is an earnings-to-price ratio at time $t-1$. Finally, Eq. (A14) is converted into a form suitable for empirical testing (which is identical to Eq. (5)):

$$R_t = \alpha + \sum_{m=0}^{\infty} \beta_m * \frac{\Delta X_{t+m}}{P_{t-1}} + \gamma * \mathrm{EP}_{t-1} + \sum_{m=1}^{\infty} \lambda_m * \frac{\Delta P_{t+m}}{P_{t-1}} + \varepsilon_t \quad \text{(A15)}$$

References

AAA Financial Accounting Standards Committee (2003). Implications of accounting research for the FASB's initiative on disclosure of information about intangible assets. *Accounting Horizons*, 17, 175–185.

Aboody, D and B Lev (2000). Information asymmetry, R&D, and insider gains. *Journal of Finance*, 55, 747–766.

Ali, A and P Zarowin (1992). The role of earnings levels in annual earnings-returns studies. *Journal of Accounting Research*, 30, 286–296.

American Institute of Certified Public Accountants (1970). Business combinations. APB Opinion No. 16.

Amir, E and B Lev (1996). Value-relevance of nonfinancial information: The wireless communications industry. *Journal of Accounting and Economics*, 22, 3–30.

Barth, ME, R Kasznik and MF McNichols (2001). Analyst coverage and intangible assets. *Journal of Accounting Research*, 39, 11–34.

Beaver, W, R Lambert and D Morse (1980). The information content of security prices. *Journal of Accounting and Economics*, 2, 3–28.

Behn, B and RA Riley (1999). Using nonfinancial information to predict financial performance: The case of the US airline industry. *Journal of Accounting, Auditing & Finance*, 14, 29–56.

Belsley, DA, E Kuh and RE Welsch (1980). Regression diagnostics. New York, NY: Wiley.

Ben Zion, U (1978). The investment aspect of nonproduction expenditures: An empirical test. *Journal of Economics and Business*, 30, 224–229.

Boone, JP and KK Raman (2001). Off-balance sheet R&D assets and market liquidity. *Journal of Accounting and Public Policy*, 20, 97–128.

Bublitz, B and M Ettredge (1989). The information in discretionary outlays: Advertising, and research and development. *The Accounting Review*, 64, 108–124.

Chambers, D, R Jennings and RB Thompson II (1999). Evidence on the usefulness of capital expenditures as an alternative measure of depreciation. *Review of Accounting Studies*, 4, 169–195.

Collins, DW, EL Maydew and IS Weiss (1997). Changes in the value-relevance of earnings and book values over the past forty years. *Journal of Accounting and Economics*, 24, 39–67.

Collins, DW, SP Kothari, J Shanken and RG Sloan (1994). Lack of timeliness and noise as explanations for the low contemporaneous return-earnings association. *Journal of Accounting and Economics*, 18, 289–324.

Cramer, JS (1987). Mean and variance of R^2 in small and moderate samples. *Journal of Econometrics*, 35, 253–266.

Danielson, M and E Press (2003). Accounting returns revisited: Evidence of their usefulness in estimating economic returns. *Review of Accounting Studies*, 8, 493–530.

Easton, PD and TS Harris (1991). Earnings as an explanatory variable for returns. *Journal of Accounting Research*, 29, 19–37.

Easton, PD, TS Harris and JA Ohlson (1992). Aggregate accounting earnings can explain most of security returns. *Journal of Accounting and Economics*, 15, 119–142.

Financial Accounting Standards Board (1997). Statement of financial accounting standards no. 2: Accounting for research and development costs.

Francis, J and K Schipper (1999). Have financial statements lost their relevance? *Journal of Accounting Research*, 37, 319–352.

Healy, PM and KG Palepu (1993). The effect of firms' financial disclosure strategies on stock prices. *Accounting Horizons*, 7, 1–11.

Healy, PM, SC Myers and CD Howe (2002). R&D accounting and the tradeoff between relevance and objectivity. *Journal of Accounting Research*, 40, 677–710.

Hirschey, M and J Weygandt (1985). Amortization policy for advertising and research and development expenditures. *Journal of Accounting Research*, 23, 326–335.

Klein, B and K Leffler (1981). The role of market forces in assuring contractual performance. *Journal of Political Economy*, 89, 615–641.

Kothari, SP, TE Laguerre and AJ Leone (2002). Capitalization versus expensing: Evidence on the uncertainty of future earnings from capital expenditures versus R&D outlays. *Review of Accounting Studies*, 7, 355–382.

Lev, B and T Sougiannis (1996). The capitalization, amortization and value-relevance of R&D. *Journal of Accounting and Economics*, 21, 107–138.

Lev, B and P Zarowin (1999). The boundaries of financial reporting and how to extend them. *Journal of Accounting Research*, 37, 353–385.

Lundholm, R and LA Myers (2002). Bringing the future forward: The effect of disclosure on the returns-earnings relation. *Journal of Accounting Research*, 40, 809–839.

Myers, SC and NS Majluf (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics*, 13, 187–221.

Nwaeze, ET (1998). Regulation and the valuation relevance of book value and earnings: Evidence from the United States. *Contemporary Accounting Research*, 15, 547–573.

Shevlin, T (1991). The valuation of R&D firms with R&D limited partnerships. *The Accounting Review*, 69, 1–21.

Sougiannis, T (1994). The accounting based valuation of corporate R&D. *The Accounting Review*, 69, 44–68.

Warfield, TD and JJ Wild (1992). Accounting measurement error and the pricing of interim earnings. *The Accounting Review*, 67, 821–842.

Chapter 5

Thirty Years of Canadian Evidence on Stock Splits, Reverse Stock Splits, and Stock Dividends

Vijay Jog and PengCheng Zhu
Carleton University, Canada

Thirty years of Canadian evidence has been used to shed light on the motivation and implications of stock splits, stock dividends, and reverse splits. The maximum 5-year period before and after the "event" month was focused and the changes in stock return, earnings per share (EPS), beta, trading volume, number of transactions, price to earning ratio (P/E), valuation, and corporate governance characteristics were examined. Strong information signaling effect of stock splits was found, as well as persistent superior performance and a favorable change in relative valuation are observed in the post-stock split period. The total trading volume increases after stock split ex-date while the trading volume per transaction decreases considerably, signifying a possible change in investor composition. However, no accompanying change was found in the governance environment. The overall evidence supports the signaling hypothesis as well as the optimum price and the relative valuation hypotheses. Stock dividend and reverse stock split firms have significantly weaker stock performance and operating performance than stock split firms. The negative trend does not improve in a long-term after the ex-date.

Keywords: Stock split; stock dividend; reverse stock split; valuation.

1. Introduction

Stock splits, stock dividends, and reverse splits are sometimes referred to as "cosmetic events" with no direct valuation implications as they simply represent a change in the number of outstanding shares. The reason for the interest is therefore to understand why managers would undertake such (potentially costly) cosmetic decisions. Two strands of explanations have been proposed: the first explanation relies on information signaling whereas the second resorts to possible valuation implications due to reasons such as the establishment of an "optimal" price range that changes the shareholder composition and increased liquidity and changes in the return distribution characteristics (e.g., variance).

In this chapter, 30 years of Canadian evidence was used to investigate these hypotheses. This chapter is unique in three aspects. It provides long-term historical documentation in trends of stock splits, reverse stock splits, and stock dividends in the Canadian context. In contrast to several other Canadian studies in the literature, for example, Dipchand (1977), Kryzanowski and Zhang (1991, 1993, 1996), Masse *et al.* (1997), Elfakhani and Lung (2003), this chapter covers the longest observation period (i.e., 32 years from 1970 to 2002) and compares the long-term trend for the three events altogether. Given the dominance of US-based studies in the literature, the present research provides a comprehensive out-of-sample test to reexamine the stock split and dividend topics. Second, it analyzes long-term performance and trading patterns by using multiple testing periods around the "event" month, namely the exercise date of the three events. The length of the testing periods is from 6 months up to 5 years before and after the ex-date. Thus, it provides the opportunity to examine the consistency and persistency of the event impact across different time horizons. Even in the latest articles on US evidence, a comparable long-term event window was rarely found as in the present chapter (Akhigbe *et al.*, 1995; Desai and Jain, 1997; Byun and Rozeff, 2003). Third, this is one of a few articles which compare the three sequential events all together. Masse *et al.* (1997) used short-term event study method to observe the announcement impact of the three events on firm values. It is believed that no other study has simultaneously examined these three events within the context of long-term firm performance, trading patterns, and other firm characteristics. In addition, the chapter also provides an indirect evidence to test the hypothesis that managers may use stock splits to deliberately change the shareholder composition to reduce potential monitoring by large shareholders. It is believed that further specific tests of this hypothesis, in addition to the signaling, optimal price range, and valuation hypotheses may provide an explanation as to why managers undertake a seemingly "cosmetic" event such as stock splits.

This chapter is organized as follows: the next section provides the context for this chapter by reviewing the pertinent literature, including the existing Canadian evidence. Then the sample and methodology are described, followed by results. The chapter ends with conclusions by comparing and contrasting the present results with those found in the US and Canada and presents a potentially testable hypothesis for future research.

2. Literature Review

The literature in this area has used various definitions of the stock split and stock dividend. Most research focused on stock splits where the split ratio is at least 1.25 to 1; those with ratios less than 1.25 are considered as stock dividends. Reverse stock splits have the split ratios of less than 1. Some other articles suggested using accounting treatment to distinguish between the stock split and stock dividend sample, for example, Peterson *et al.* (1996) and Foster and Scribner (1998). This chapter will follow the traditional way of the 1.25 split factor to define the stock split sample. Since the literature on stock dividend and reverse stock split is much less than that of stock split, the present work focus mostly on the articles that deal with stock splits and only refer to the other two events when the articles show asymmetric evidence or different conjectures.

Broadly speaking, the stock split literature can be split along three categories: the first category deals with the potential theoretical reasons that can explain why managers may resort to stock splits.[1] The second category consists of articles that are predominantly of empirical nature and those that investigate and document the reaction of the stock market around the announcement (and/or the ex-date) of the decision to split the stock; this literature is termed as event analysis literature since it follows a classical event analysis methodology.[2] The third category of articles deals with the long-term implications of stock split and compares variables such as rate of returns, variance, short interest, beta, bid-asked spread, volume, liquidity, ownership structure, and valuation across the pre- and the post-stock split periods.[3] The present

[1] For example, see Ross (1977), Grinblatt *et al.* (1984), Brennan and Copeland (1988a, 1988b), Peterson *et al.* (1996), Rankine and Stice (1997), Ye (1999), Kadiyala and Vetsuypens (2002), Crawford *et al.* (2005) for the signaling aspects of stock splits and Leland and Pyle (1977), Arbel and Swanson (1993) for information asymmetry rationale and Baker and Gallagher (1980), McNichols and Dravid (1990), Baker and Powell (1993) for management rationale for stock splits.

[2] See, for example, Bar-Yosef and Brown (1977), Woolridge (1983), Grinblatt *et al*. (1984), Elfakhani and Lung (2003), and Ariff (2004).

[3] These papers include Brennan and Copeland (1988a,b), Wiggins (1992) on changes in beta; Kadiyala and Vetsuypens (2002) on short interest; Desai *et al.* (1998), Schultz (2000), Dennis (2003), Gray *et al.* (2003) on changes in liquidity and transaction costs; Szewczyk and Tsetsekos (1993), Mukherji *et al.* (1997), Han and Suk (1998), Mason and Shelor (1998), Easley *et al.* (2001), Dennis and Strickland (2003) on changes in shareholder composition or ownership structure; Dubosfsky (1991), Kryzanowski and Zhang (1993), Park and Krishnamurti (1995), Koski (1998), Angel *et al.* (2004) on changes in volatility; Akhigbe *et al.* (1995), Ikenberry *et al.* (1996), Desai and Jain (1997), Byun and Rozeff (2003) on postsplit long-term performance.

chapter falls in the third category and thus, it focus on reviewing more thoroughly the pertinent literature that falls in this third category. To provide the context, the literature was reviewed chronologically by focusing first on the US evidence followed by the review of the available Canadian evidence.

While the early evidence on stock splits can be traced back to the seminal article by Fama *et al.* (1969) followed by Bar-Yosef and Brown (1977), Charest (1978), Baker and Gallaghar (1980), and Grinblatt *et al.* (1984), one of the most extensive analysis of stock splits and stock dividends was conducted by Lakonishok and Lev (1987), Lakonishok and Lev, hereafter.

Using a sample of over 1,000 stock splits and stock dividends, each during the period 1963 and 1982 in the US, Lakonishok and Lev claim that stock splits occur to bring the stock in an "optimal" price range and during periods of extraordinary growth in stock prices. They also show that there is some evidence that the splitting firms exhibit a somewhat higher growth in earnings and dividends during the post-split period. Lakonishok and Lev further report that there is no industrial concentration in stock splits and stock dividend firms and that stock split firms are bigger (in terms of market capitalization) than the population, while stock dividend firms are smaller than the population. They do not observe any noticeable increase in volume of trading but state that there may be changes to investor composition and that further research is required.

Many articles have appeared after the Lakonishok and Lev article that investigate a specific hypothesis about stock splits. Brennan and Copeland (1988a, 1988b) and Doran (1995) claim that managers use stock splits as a costly signal to convey private information. McNichols and Dravid (1990) provide evidence that management selects split factor to signal private information. Szewczyk *et al.* (1992) and Szewczyk and Tsetsekos (1993) find support for the signaling hypothesis by observing the inverse relationship between the share price reaction to stock splits and the degree of institutional ownership (and managerial ownership). Another version of signaling is the "attention-getting" hypothesis. Grinblatt *et al.* (1984), McNichols and Dravid (1990), Arbel and Swanson (1993), and Ye (1999) suggest that managers use stock splits to attract attention from institutional investors and financial analysts to re-evaluate their undervalued company value. More recent articles also provide supportive evidence to the signaling evidence, for example, Peterson *et al.* (1996), Mukherji *et al.* (1997), and Louise and Robinson (2005). However, using short interest data, Kadiyala and Vetsuypens (2002) put the signaling hypothesis in doubt. They show the short interests do not decline around stock splits, and thus is contrary to the positive signaling effect of stock split.

Additional evidence that questions the signaling hypothesis includes Rozeff (1998) who investigates mutual fund stock splits and Reboredo (2003) who uses Spanish market data. Ma *et al.* (2000) observe the increased insider selling around the split announcement date, which suggests insider trading before stock split announcement is not motivated by the information content of the announcement.

Another strand in the literature is to investigate liquidity (as measured by bid-ask spread, trading volume, frequency, or turnover) changes after stock splits (Dennis, 2003). Conroy *et al.* (1990) document an increase in relative bid-asked spread and true variance in the post-split period, which is confirmed by Dubosfsky (1991), Kryzanowski and Zhang (1993), Park and Krishnamurti (1995), Desai *et al.* (1998) and more recently by French and Foster (2002) and Gray *et al.* (2003). Koski (1998) shows the increased variance cannot be completely explained by the bid-ask spread error or price discreetness. By using the when-issued shares data, Angel *et al.* (2004) provide new evidence to attribute the higher post-split variance to the increase in small-volume traders (or noise traders). The evidence of wider bid-ask spreads (i.e., higher transaction costs) and more small-volume investors helps to develop a new hypothesis, namely "broker promotion" hypothesis. Schultz (2000), Gray *et al.* (2003), and Kadapakkam *et al.* (2005) argue that market makers have strong incentive to promote the stocks to small-volume traders as stock split induces wider spread that brings excess profit to them.

Optimal price range hypothesis is another widely accepted explanation to the rational of stock split in both academic literature and professional community. Two surveys conducted by Baker and Gallagher (1980) and Baker and Powell (1993) found that the main motive for stock splits was to move the stock price into a better trading range. Conroy and Harris (1999) collected a longitudinal stock split sample and found that managers appear to design splits to return the company's stock price to the price level achieved after the last split. By observing the substantially different stock prices across countries, Angel (1997) argues that companies split stocks so that the institutionally mandated minimum tick size is optimal relative to the stock price. So and Tse (2000) extend the Lakonishok and Lev's work by using a longer time period and provide support for the optimal price hypothesis as well. However, Easley *et al.* (2001) contend that uninformed trading increases following splits and that there is no evidence to indicate that stock splits result in the arrival of new information. They report only a weak evidence for the optimal price range hypothesis.

To further investigate the shareholder composition conjecture put forward by Lakonishok and Lev and recently by Schultz (2000), Denis and Strickland (2003) attempt to directly test the changes in shareholder composition around the split event. They put forward three conclusions based on a relatively short time period of 1990–1993. First, the proportion of institutional shareholders increases significantly conditional on the level of institutional shareholders before the split; this conclusion is inconsistent with the notion that stock splits occur to bring the stock price in a trading range where it makes it attractive for small shareholders to own the stock. Second, they document a large increase in trading volume in stocks where there is a significant increase in institutional ownership; they explain it by asserting that institutions trade more often. However, they never explain why institutions would increase their holdings and then undertake a higher level of trading. Third, they find statistically positive returns around announcement date and claim that these returns are due more to the expected increase in post-split liquidity rather than information signaling hypothesis.

Several other articles analyze the post-split long-term performance. The results based on US evidence are mixed and sometimes contradictory. Ikenberry *et al.* (1996) find significant post-split excess return of 7.93% in the first year and 12.15% in the first 3 years for a sample of two-for-one stock splits from 1975 to 1990. Desai and Jain (1997) confirm the positive post-split performance using a sample from 1976 to 1991. They also show negative post-split performance for the reverse stock split sample in the same period. However, Akhigbe *et al.* (1995) use a stock split sample from 1980 to 1988. Their result suggests that the cumulative abnormal returns are positive and statistically significant through the 11th month after a stock split. The cumulative abnormal returns then decrease nearly monotonically through the 36-month after the split. A recent article by Byun and Rozeff (2003) carefully conducts the long-run performance of stock splits during a much longer period (from 1927 to 1996) and using a more detailed robust test with different weighting and benchmarking techniques. Their overall conclusion indicates that buyers and sellers of splitting stocks do not, on average, earn abnormal return that is significantly different from zero. They also point out that the results of studies in long-run performance of stock split are sensitive to time period, method of estimation, and sampling. However, their paper only focuses on the 12 months subsequent to the stock split event. Given the mixed results upon different time horizons and different fiscal time period in the US literature, it warrants a careful reexamination of the long-term impact of stock splits using multiple testing periods and an “out-of-sample” dataset. This is one of the motivations for this chapter.

Above all, given the US evidence, there is a considerable amount of uncertainty about why firms split their stocks and about the changes that are observed in the post-split period compared to the pre-split period. Except for findings that (a) positive excess returns are observed around the announcement date; (b) splits occur after a considerable increase in both earnings and stock prices immediately preceding stock splits; and (c) there is a change in the liquidity following the post-split period; it is hard to come up with other stylized facts about stock splits that support the various competing theories and management rationale from the evidence available.

Canadian evidence on stock splits is mostly confined to the second category — event analysis around the announcement or ex-dates. Using the period 1978–1987, Kryzanowski and Zhang (1991, 1993) report an increase in mean returns, measured variance, and beta after the ex-date. Kryzanowki and Zhang (1996) find significant change in trading patterns of small traders but not of big traders after split ex-date. Masse *et al.* (1997) shows a positive stock market response to the announcement of stock splits, stock dividends, as well as reverse stock splits in Canada. The positive reaction to the reverse stock split announcement is not consistent with the US evidence. The results of finding negative long-term performance of reverse stock split in the present chapter does not confirm with their result either. The Elfakhani and Lung article (2003) using data over a period 1973–1992 show that the reaction of the stock market to stock splits is driven by a few significant individual splits and in the 2-year post-split period, they also observe increases in both earnings and trading volume. None of these articles attempt to investigate long-term consequences of the stock split events or to provide evidence on the competing hypotheses. Comparing with the existing literature on stock splits, the present chapter documents the longest historical trend of the three events altogether in the Canadian context (from 1970 to 2002). The present focus on both the long-term performance and trading pattern (i.e., up to 60 months before and after split ex-date) provides unique contribution to the Canadian literature and supplements the “out of the sample” evidence to the extensive US literature.

3. Sample Description and Methodology

The initial sample consists of 836 stock splits, 234 reverse stock splits, and 577 stock dividends for firms listed on the Toronto Stock Exchange (TSE) from 1970 to 2002. The data are collected from the Canadian Financial Markets Research Center (CFMRC) database. Particularly, the historical events, split factors, and the corresponding ex-dates are identified by using the CFMRC price adjustment data file. The stock price, industry classification, market capitalization, stock returns, earnings per share, beta, trading volume, and transactions are collected or calculated from the monthly data file of the database.

In the process of data collection, some misclassified data among stock splits, reverse splits, and stock dividends in the database were corrected. In addition and consistent with the standard in the literature, a proper classification rule between stock splits and stock dividends was used, namely split factors larger than 5-for-4 (1.25 split factor) are considered stock splits, others are classified as stock dividends; those with split factors less than 1 are classified as reverse splits.

The sample consists of some multiple stock split and dividend events for the same firm within the same year. However, in order to compare the overall evolution of the events relative to the population, multiple events per individual firm are aggregated on the annual basis. In other words, a company is counted to have one stock split in a particular year even if it actually has more than one stock split in that year. Table 1 and Fig. 1 illustrate the historical trend of these three events relative to the population of the TSE listed companies; Fig. 1 also plots the annual TSE index returns (scaled to 10%).

The results in Table 1 and Fig. 1 show that both stock splits and stock dividends demonstrate a declining trend but reverse splits show a slightly increasing trend in recent years. Comparing the trends with the TSE300 index returns, a general correlation between the number of stock splits, stock dividends, and the market index returns can be seen. A heuristic explanation may be that companies whose stock prices rise significantly in an "up" market tend to split stocks or issue stock dividends and those whose stock prices decline considerably during the "down" market tend to do reverse-splits.

Table 2 shows the 30-year Canadian ex-date sample in terms of split factors. The maximum stock split factor is 10 for 1, and the most common split factor is 2 for 1. The highest stock split frequency for a company is seven times over the entire sample period. For reverse stock split, the maximum

Table 1. Annual numbers of Canadian stock split, reverse stock splits, and stock dividends.

Year	Stock split		Reverse stock split		Stock dividend	
	Number	%[a]	Number	%[a]	Number	%[a]
1970	13	1.51	1	0.12	15	1.74
1971	25	2.91	3	0.35	21	2.33
1972	45	5.23	0	0.00	15	1.51
1973	33	3.60	1	0.12	6	0.47
1974	12	1.40	1	0.00	9	0.93
1975	12	1.28	1	0.11	9	0.85
1976	11	1.33	0	0.00	9	0.88
1977	17	1.92	1	0.11	5	0.34
1978	29	3.53	0	0.00	15	1.83
1979	36	4.38	1	0.13	30	3.63
1980	56	7.01	1	0.13	32	4.13
1981	47	5.37	1	0.12	28	3.46
1982	7	0.73	0	0.00	36	4.15
1983	43	4.70	0	0.00	38	4.24
1984	37	3.97	2	0.21	36	3.54
1985	43	4.45	1	0.10	33	3.21
1986	68	6.18	2	0.18	30	2.95
1987	57	4.47	5	0.41	30	2.57
1988	11	0.91	11	0.91	22	1.73
1989	15	1.24	8	0.66	24	1.73
1990	10	0.84	12	1.01	21	1.59
1991	7	0.53	15	1.32	23	1.85
1992	12	0.98	8	0.71	12	1.07
1993	20	1.51	18	1.34	18	1.51
1994	21	1.68	11	0.72	8	0.56
1995	10	0.79	15	1.19	5	0.32
1996	31	2.27	16	1.21	13	0.83
1997	31	2.11	17	1.20	17	1.20
1998	37	2.44	25	1.61	15	0.91
1999	16	1.10	18	1.24	10	0.48
2000	18	1.41	23	1.34	13	0.77
2001	13	0.76	13	0.99	7	0.53
2002	14	1.00	15	1.07	6	0.38

[a] Percent of the total listed companies on TSE by that particular year end.

split factor is about 1 for 0.92, the minimum factor is 25 (0.004) for 1, and the most common factor is 5 (0.2) for 1. The stock dividend sample shows that the minimum factor is 1.002 for 1, and the most common factor is 1.02 for 1.

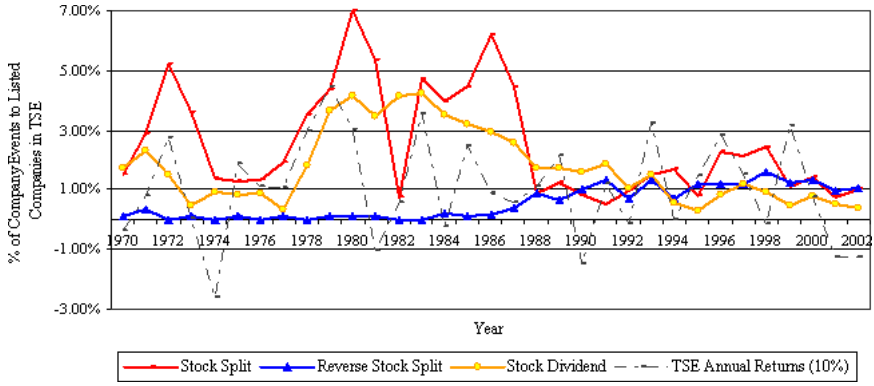


Figure 1. The historical trend of Canadian stock splits, reverse stock splits, and stock dividends.

Table 2. Split factor statistics.

	Stock split	Reverse stock split	Stock dividend
Highest rate	10 for 1	0.92 for 1	1.25 for 1
Lowest rate	1.25 for 1	0.004 for 1	1.002 for 1
Mean rate	2.43 for 1	0.22 for 1	1.05 for 1
Median rate	2 for 1	0.2 for 1	1.03 for 1
Mode rate	2 for 1	0.2 for 1	1.02 for 1
Std. Dev.	1.08	0.16	0.06
Highest event frequency	7	2	—
Full sample size	836	234	577

Table 3 shows the industry distribution of the splitting firms. Some industries, such as manufacturing, mining, and oil gas that have big presence on the TSE also represent a large number of splitting firms and therefore in relative terms, no clear industry concentration can be realized in stock split and stock dividend companies. However, compared with the other two events, reverse stock split companies are seen to be heavily concentrated in the mining and oil gas industry. The relatively weak performance of the two industries in the observation years may explain the increasing number of reverse stock splits in the sample.

To further document the characteristics of the sample firms, Table 4 provides both the pre-ex-date and post-ex-date month-end stock price and market capitalization information for the sample of the three events. Evidence on

Table 3. Industry distributions of stock split, reverse stock split, and stock dividend firms.

Industry	Stock split		Reverse stock split		Stock dividend	
	Company #	%	Company #	%	Company #	%
Manufacture	94	18.4	15	6.8	38	15.1
Oil gas	81	15.9	54	24.4	37	14.7
Finance	54	10.6	20	9.0	30	11.9
Mining	38	7.5	50	22.6	30	11.9
Publishing, paper, forest product	34	6.7	2	0.9	16	6.3
Retail, distribution	29	5.7	2	0.9	10	4.0
Consumer	28	5.5	12	5.4	5	2.0
Service	24	4.7	10	4.5	7	2.8
Transportation and communication	19	3.7	4	1.8	11	4.4
Real estate	17	3.3	11	5.0	11	4.4
High-tech	14	2.7	17	7.7	16	6.3
Telecom	11	2.2	5	2.3	2	0.8
Energy	7	1.4	0	0.0	4	1.6
Utility	7	1.4	2	0.9	5	2.0
Other	53	10.4	17	7.7	30	11.9
Total	510	100	221	100	252	100

both the nominal price and real price was provided (using 2002 Consumer Price Index to bring all prices to 2002 prices) and the post-ex-date price was adjusted by the corresponding split factors for proper comparison. In the process of data collection, some stocks are found to have no price or related information around the ex-date. Particularly, a large number of stock dividend firms are not (actively) traded. Therefore, the present sample size reduces to 740 stock splits, 229 reverse splits, and 155 stock dividends in this comparison.

Not surprisingly, stock split firms are found to have the highest stock price levels in the three groups, and reverse split stocks have the lowest prices. The real prices of the split and dividend stocks are higher than their nominal prices, which suggest that the historical prices for these stocks were higher relative to the purchasing power of the economy than the recent prices. However, the phenomenon is not so clear in reverse stock split since the reverse splits only gained attentions in recent years in the Canadian market. Also, by comparing the pre-ex-date price with the post-ex-date adjusted price,

Table 4. Price range and market value at one month before and after stock split, reverse stock split, and stock dividend.

	Stock split		Reverse stock split		Stock dividend	
	M-1 price	Adj. M price	M-1 price	Adj. M price	M-1 price	Adj. M price
Nominal price, in $						
Min price	0.2	0.3	0.01	0.02	1.2	1.0
Max price	337.0	468.0	123.5	99.8	58.3	69.3
Mean price	38.0	40.1**	2.5	2.3**	12.1	12.8
Median price	31.5	33.0	0.5	0.5	9.9	10.6
Std. Dev.	30.4	35.2	10.9	8.8	9.4	10.2
2002 real price, in $						
Min price	0.4	0.5	0.01	0.02	1.4	1.5
Max price	566.3	559.8	133.0	110.4	100.3	129.2
Mean price	70.6	73.6**	2.9	2.7	21.9	23.6
Median price	52.5	54.5	0.6	0.5	14.6	15.6
Std. Dev.	63.9	67.4	11.8	9.7	21.1	24.2
	M-1 market value	M market value	M-1 market value	M market value	M-1 market value	M market value
Nominal price, in million $						
Min MV	1.2	1.5	0.5	0.4	0.8	0.7
Max MV	2379706.6	2274151.0	9561.4	5340.8	1723.9	1954.2
Mean MV	1225.1	1283.2*	148.6	126.7**	156.6	162.7
Median MV	189.7	198.6	21.5	18.1	63.2	62.8
Std. Dev.	9163.5	8958.8	703.7	468.9	259.9	274.0
2002 real price, in million $						
Min MV	2.0	2.8	0.7	0.6	1.3	1.2
Max MV	249502.2	238472.8	10296.9	5751.7	2528.6	2701.1
Mean MV	1549.7	1622.2*	164.0	140.3**	244.8	255.2
Median MV	298.4	312.9	25.1	21.6	98.5	97.7
Std. Dev.	9674.5	9472.4	758.6	506.1	405.0	429.8

M-1 denotes month end before ex-date, M stands for month end after ex-date.
*5% significance level.
**1% significance level.

the market seems to react positively to stock splits in the time period immediately around the event month, but continues to react negatively to reverse splits in short term. Both the post-ex-date adjusted prices for stock splits and stock dividends are higher than the pre-ex-date month end prices, but this is not the case for reverse splits. More specifically, stock prices increase by 5.53% for the stock split sample in the event month. In the 36 months preceding the splits, stock prices increase by 97.58%, 81.65% in 24 preceding months, and 48.49% in 12 months prior to the event month. The prices over the next 12, 24, and 36 months increase only by 1.24%, −2.41%, and −6.1%, respectively.

It should be noted that the market capitalization data as well as many other trading data show substantial skewness in the distributions, thus it is problematic to use the simple t-test to compare the sample means. In this chapter, a well-known resampling method, namely nonparametric bootstrapping method, suggested by Efron and Tibshirani (1993) was adopted to detect the difference in either the sample mean or sample median for skewed data. The statistical test results for either mean or median were reported, unless the statistical test shows contrasting results for the two statistics.

The trend characteristics of the three events were examined, by using multiple testing periods from 6 months to 60 months before and after the ex-date with the adjustment by the split factor wherever necessary for comparison purposes for variables such as stock prices, EPS, trading volume, and trading volume per transaction. Note that the present event is not the announcement date but the split date, which means there is no information content at this event; the actual split date is already known to the market participants. Typically, the announcement date is between 45 and 75 days ahead of the actual split date.

To evaluate the long-term performance of the three groups in addition to examining the trends in prices, Cumulative Abnormal Returns (CAR) is calculated using the TSE 300 index total return as the benchmark. As noted later, there is only very slight change in the pre- and post-betas of the sample firms so the results based on the market model would be similar to that reported using simple market adjusted returns.

For the CAR methodology, the market adjusted return for stock i in event month t is defined as:

$$\mathrm{AR}_{it} = R_{it} - R_{mt} \tag{1}$$

where R_{it} is the raw return of stock i in month t, and R_{mt} is the total return of TSE 300 index in month t. The average market adjusted return of each event group for event month t is the equally weighted arithmetic average of the market-adjusted returns:

$$\mathrm{AR}_t = \frac{1}{n}\sum_{i=1}^{n} \mathrm{AR}_{it} \tag{2}$$

The cumulative market-adjusted return of each group from event month q to event month s is the compounding product of the average market-adjusted returns.

$$\mathrm{CAR} = \left[\prod_{t=q}^{s}(1 + AR_t)\right] - 1 \tag{3}$$

When there are small number of missing values in the return data, the average abnormal return of the group for the month (AR_t) is calculated by the equally weighted average of the remaining firms in that group. In other words, group mean has been used to replace the missing values.

4. Empirical Results

4.1. *Stock price trend*

Figure 2 shows the relative mean stock price trend for the three events. In each month the stock prices are averaged within each group and converted into a number relative to the event month figure. In other words, a value of 2 in a particular month implies that the mean stock price for that month is two times higher than the event month price. This normalization allows us to compare the trends between pre-ex-date and post-ex-date period within the same event group. Figure 2 clearly demonstrates the differences in price growth across the three groups. Stock split firms have experienced a consistent increasing price trend before the ex-date, reverse split firms have a consistent decreasing price trend, and stock dividend firms are in the middle. However, the price trends in the three groups all level off after the ex-date, implying that stock splits on average do not show significant post-event increases in prices.

4.2. *Stock return trend*

Since the comparisons based purely on price trend do not account for changes in the overall market conditions as well as possible changes in dividend per share, now the results are shown using the Cumulative Abnormal Return

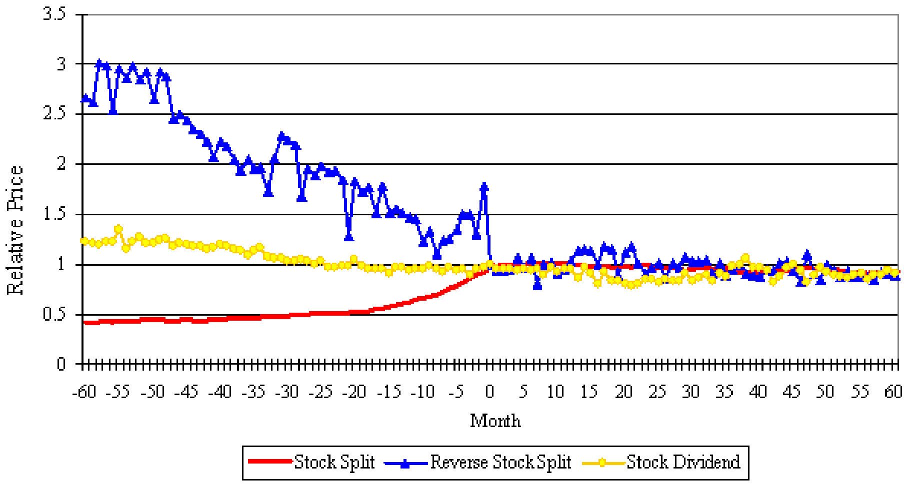


Figure 2. Relative average price trend.

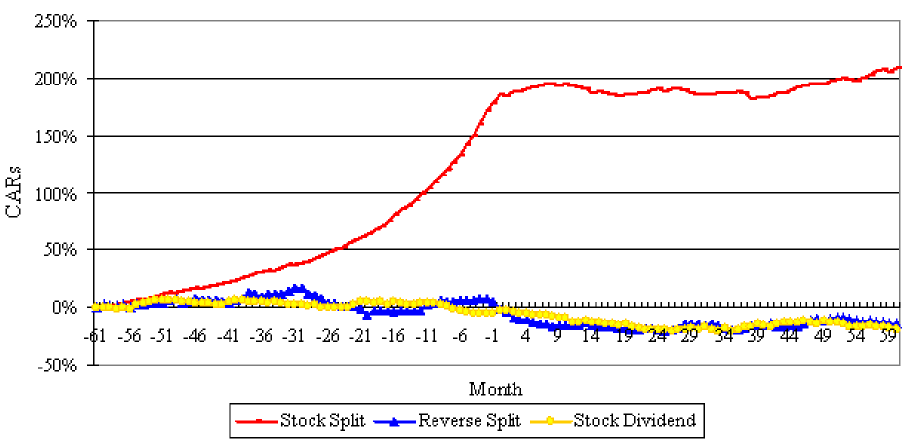


Figure 3. Cumulative abnormal return trend.

(CAR) methodology. Figure 3 illustrates the long-term performance of each group relative to the market index. The stock split firms perform much better than the other two groups before the ex-date. An obvious break point at the ex-date can also be seen for the stock split group. The superior performance levels off almost immediately after the stock split ex-date. Reverse splits and stock dividends firms show consistently weaker performance than the stock splits firms. The difference, if any, in performance between reverse splits and stock dividends is not that clear.

In Table 5, the statistical test based on multiple time windows provides a more robust conclusion of the empirical evidence. The stock performance of the stock split firms before ex-date is astonishing. Over the previous 5 years before ex-date, the cumulative abnormal returns of the stock split firms, on average, almost beat the market return by 180%. The superior performance is found to be accelerating in the 1 to 2 years before the stock splits. The positive trend continues after the ex-date and persists for a very long time (up to 5 years after stock splits). However, the magnitude of the post-split CARs is much smaller than the pre-split CARs. The criticisms of using long-term event window are noticed in the literature as the data noises are expected

Table 5. Statistical tests of cumulative abnormal returns.

	Before ex-date				After ex-date		
	CAR	Std. Dev.	*t*-stats		CAR	Std. Dev.	*t*-stats
Stock split							
CAR (−60, −54)	0.05	0.28	4.31**	CAR (0, +6)	0.05	0.31	3.86**
CAR (−60, −48)	0.11	0.46	5.45**	CAR (0, +12)	0.07	0.49	3.47**
CAR (−60, −36)	0.29	1.11	5.86**	CAR (0, +24)	0.09	0.69	2.96**
CAR (−60, −24)	0.46	1.41	7.36**	CAR (0, +36)	0.07	0.71	2.36**
CAR (−60, −12)	0.87	2.76	7.12**	CAR (0, +48)	0.08	0.78	2.30**
CAR (−60, 0)	1.84	6.72	6.20**	CAR (0, +60)	0.10	0.88	2.58**
Reverse stock split							
CAR (−60, −54)	0.03	0.56	0.68	CAR (0, +6)	−0.11	0.51	−2.83**
CAR (−60, −48)	0.03	0.69	0.55	CAR (0, +12)	−0.05	1.46	−0.44
CAR (−60, −36)	0.02	0.94	0.31	CAR (0, +24)	−0.12	0.95	−1.64
CAR (−60, −24)	−0.07	0.96	−0.95	CAR (0, +36)	−0.16	0.86	−2.51**
CAR (−60, −12)	−0.15	1.01	−1.96	CAR (0, +48)	−0.11	0.97	−1.49
CAR (−60, 0)	−0.08	1.56	−0.69	CAR (0, +60)	−0.14	1.04	−1.73
Stock dividend							
CAR (−60, −54)	0.02	0.20	1.02	CAR (0, +6)	−0.04	0.24	−1.84
CAR (−60, −48)	0.05	0.31	1.64	CAR (0, +12)	−0.09	0.35	−2.74**
CAR (−60, −36)	0.07	0.61	1.27	CAR (0, +24)	−0.11	0.56	−2.11**
CAR (−60, −24)	0.03	0.67	0.52	CAR (0, +36)	−0.08	0.63	−1.47
CAR (−60, −12)	0.09	0.97	1.00	CAR (0, +48)	−0.07	0.67	−1.21
CAR (−60, 0)	0.00	0.88	0.02	CAR (0, +60)	−0.07	0.73	−1.07

*5% significance level.

**1% significance level.

to increase with the length of the time horizon. Interestingly, a clear pattern of increasing noise was found over the 5 years before ex-date (i.e., standard deviation of CARs increases consistently with the increase in time horizon). However, the standard deviation of CARs does not noticeably change after the ex-date. Therefore, it is believed that the results in Fig. 3 and Table 5 provide robust measures to illustrate the persistency and consistency of the superior long-term performance of the stock split firms after the ex-date. This finding is important to support the signaling hypothesis in which managers use stock split to send positive signals to the market. Based on the empirical evidence, a very long persistent trend of superior stock performance of the stock split firms was also observed (i.e., up to 5 years after ex-date).

Contrary to the stock split firms, stock dividend and reverse stock split firms have fundamentally different stock performance over the testing period. The reverse stock split firms have experienced consistent negative CARs in the past 2 years before the ex-date. The negative CARs become even more significant in the short-term (i.e., 6 months) after the ex-date and the negative trend does not seem to improve in the long-term. In the post-ex-date period (as short as 6 months), the reverse stock split firms on average underperform the market return by 11%, and the underperformance trend does not seem to improve in a considerably long-term. Stock dividend firms have similar post-ex-date performance as reverse stock split.

This evidence suggests that the main motivation behind the reverse stock splits is to avoid the threat of delisting from the stock exchange and that managers may also wish to improve the confidence and reputation of the stocks traded in the public market. However, the post-split performance suggests that the consequence is contradictory to managers' initial motivation. Actually, the reverse stock splits send a strong negative signal to the investors that the underlying operating performance may not improve in the long-term and stock prices will continue their decline. Table 5 shows that stock dividend firms do not experience any abnormal performance against the market before the ex-date. However, the post-ex-date trend is significantly underperforming the market benchmark. The contrasting results between stock dividend and stock split prove that stock dividends are not "small stock splits". Actually, they are the two different "cosmetic" events which have fundamentally different value implications to the investors in both short-term and long-term periods.

4.3. *Earnings per share trend*

Table 6's Panel A shows the EPS of stocks in each event group for the pre-ex-date and post-ex-date periods adjusted for the corresponding split factors. The stock split firms display a significantly increased EPS level after the

Table 6. Statistical tests of the changes in EPS, beta, trading volume, and transaction number before and after the ex-date.

	Stock split		Reverse stock split		Stock dividend	
	Before ex-date	After ex-date	Before ex-date	After ex-date	Before ex-date	After ex-date
Panel A. EPS						
1–6 Month	1.44	1.60**	−0.03	−0.01**	−0.13	−0.20
1–12 Month	1.39	1.63**	−0.03	−0.01**	0.03	−0.20**
1–24 Month	1.25	1.63**	−0.03	−0.01**	0.22	−0.15**
1–36 Month	1.14	1.58**	−0.04	−0.01**	0.51	−0.08**
1–48 Month	1.04	1.59**	−0.04	−0.01**	0.68	−0.04**
1–60 Month	1.00	1.57**	−0.03	0.00**	0.73	0.01**
Panel B. Beta						
1–6 Month	0.88	0.89**	1.13	1.11**	1.11	1.10
1–12 Month	0.87	0.89**	1.11	1.11	1.10	1.10
1–24 Month	0.87	0.91**	1.10	1.07**	1.08	1.12**
1–36 Month	0.86	0.91**	1.09	1.06**	1.07	1.12**
1–48 Month	0.86	0.92**	1.10	1.04**	1.06	1.12**
1–60 Month	0.86	0.93**	1.12	1.02**	1.06	1.12**
Panel C. Volume						
1–6 Month	935.83	986.25	8347.25	8824.93	933.92	969.67
1–12 Month	840.63	1027.21**	7049.42	8208.62*	938.75	989.27
1–24 Month	746.10	1073.57**	5742.85	8265.60**	882.90	1056.34**
1–36 Month	687.17	1123.00**	5133.65	8679.47**	812.26	1215.56**
1–48 Month	644.01	1173.63**	4656.35	9403.45**	783.53	1360.32**
1–60 Month	612.39	1200.65**	4245.91	10600.69**	748.12	1493.21**
Panel D. Transaction number						
1–6 Month	152.08	200.50**	189.17	147.75	88.42	114.67
1–12 Month	132.50	205.13**	164.54	134.96	92.38	106.33
1–24 Month	114.46	211.52**	143.08	123.65**	93.25	108.77
1–36 Month	106.19	215.50**	138.03	120.81*	93.74	119.15*
1–48 Month	101.06	221.55**	135.92	120.81*	96.75	124.09*
1–60 Month	97.65	221.68**	134.63	128.50	101.38	130.79

*5% significance level.
**1% significance level.

ex-date. And the companies are able to maintain the EPS level in the long run. This empirical result clearly justifies for the signaling hypothesis in the literature. The average EPS in the post-split period is permanently higher than that in the pre-split period. It shows that managers have quite accurate expectations about the company's future performance and they use stock split as an instrument to send the positive signal to the outsider investors. Combined with the long-term performance documented before, it is argued that the signaling effect is significant and persistent in the long run.

Stock dividend firms actually experience a reverse trend compared to the stock split firms in terms of EPS. The EPS of these firms decreases significantly in the immediate post-ex-date period, even though the EPS slightly increases in a long period. This finding illustrates the poor operating performance of the stock dividend firms in the short-term around the ex-date. Thus, the firms have to use stock dividend as a substitution to cash dividend to pay the investors. Stock dividends imply poor operating performance and have negative impact on shareholders' wealth in the short to medium time period.

Reverse split firms almost have a constant negative EPS values over the pre- and post-ex-date periods. Although, it shows statistical significance in the pre- and post-ex-date comparison, the EPS difference is not economically significant. This result is consistent with the finding in Vafeas (2001) which also suggests the earning performance does not change significantly after reverse stock split in US. However, the present results further suggest that even though the operating performance in reverse stock split firms does not deteriorate in the post-ex-date, the stock performance is still consistently underperforming the market benchmark, which suggests psychological lack of confidence of the investors in the short to medium time period.

4.4. *Systematic risk (beta) trend*

Figure 4 plots the trend of the systematic risk measure, beta, collected from the monthly data file in CFMRC database. Table 5's Panel B shows the statistical test results of the changes in the system risk in the three groups. US evidence suggests mixed beta trend after stock split. Lamoureux and Poon (1987) and Brennan and Copeland (1988a, 1988b) find beta increases after stock split, while Wiggins (1992) finds no change in the beta trend by using a different estimation method. The present chapter uses the traditional estimation method of beta, which is provided by the CFMRC database. The results provide the support for the increasing beta conjecture. Besides, the beta trend

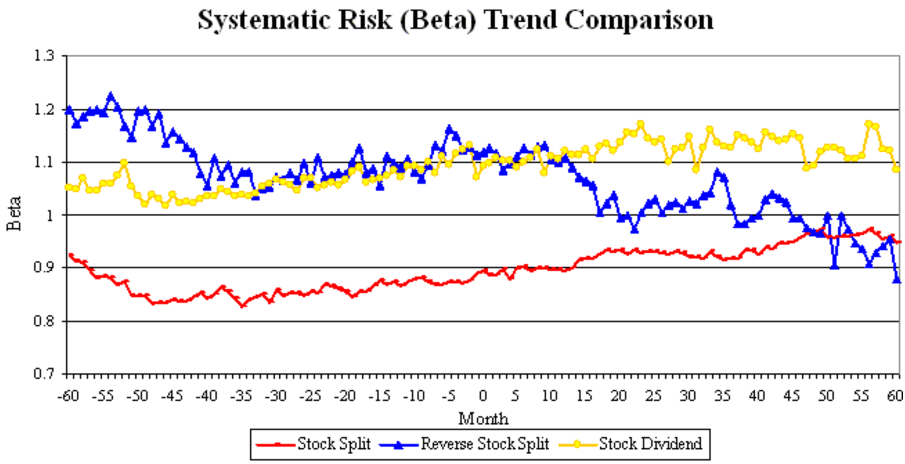


Figure 4. Changes in systematic risks.

of stock dividend events is almost parallel to the stock split trend, and also has a consistently higher level. The difference in the firm size and EPS volatility between the two groups may provide a heuristic explanation for the difference in the systematic risk. Stock split firms are much bigger in market capitalization and have more stable positive EPS level than the stock dividend firms. Reverse split beta trend is on the opposite: it is decreasing over the time period. Although the empirical finding is clear, no theoretical explanation has been found in the literature.

4.5. *Trading volume trend*

Figure 5 shows the trading volume trend by converting the monthly trading volume data relative to the event month for each group as well as by adjusting the post-ex-date volumes to the pre-ex-date scale. Table 5's Panel C conducts the statistical test to compare the trading volume changes in the post-ex-date period. All the three event groups demonstrate an increasing trend in trading volume after the ex-date. It is concluded that there is an increase in liquidity following the three "cosmetic" events. The stock split results are consistent with both the US and Canadian evidence in the literature (Dennis, 2003; Elfakhani and Lung, 2003; Han, 1995). In Table 5's Panel C, it can also be seen that reverse split stocks have the highest trading volume comparing with the other two groups. The reason for this may be due to the lower price

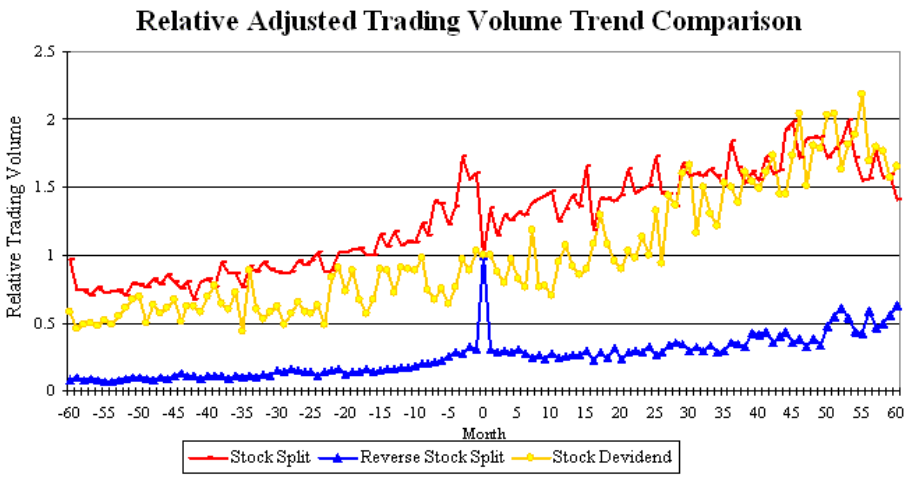


Figure 5. Trading volume trend.

level of reverse split stocks. The trading values of each group are still almost comparable.

4.6. *Transaction number trend*

Although increase in trading volume has been observed, there is no direct way to know whether there is a change in the investor composition. Since there is no information on the institutional holdings in these firms, the investor composition change hypothesis was tested indirectly. Table 5's Panel D and Fig. 6 show the transaction number changes among the three groups. In Fig. 6, the trading volume numbers are normalized by not only the event month but also the stock split group as the benchmark so that the graph demonstrates both the within-group and between-group trend differences.

A significant jump in the transaction numbers from pre-ex-date to the post-ex-date can be seen in the stock split group. Reverse split firms have an active trading period around the ex-date month. However, there is no significant transaction level shift afterwards. The trend for stock dividends firms is not as strong as that of stock split either.

Comparatively speaking, stock split firms have the highest number of transactions among the three. However, since stock split firms have been found to be much bigger than the other two, it is hard to conclude that stock split firms

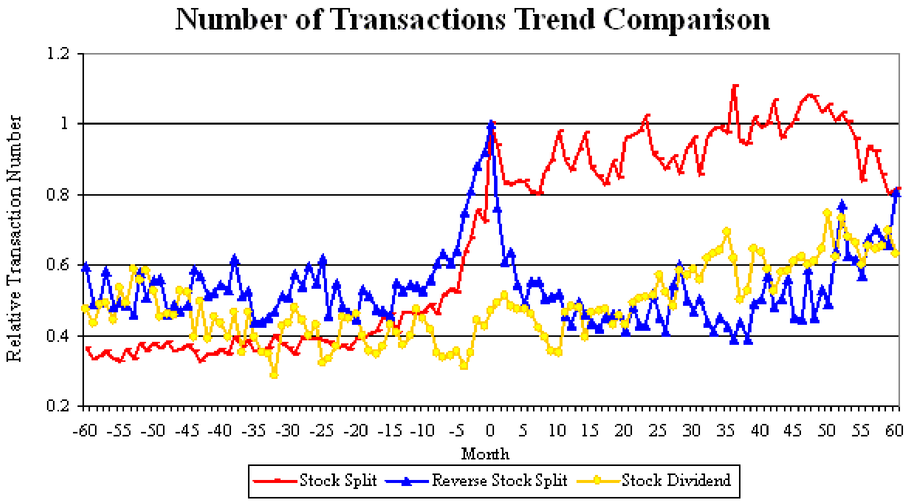


Figure 6. Transaction number trend.

have more liquidity than the other two groups. In the section below, the results are shown using changes, if any, in trading volume per transaction.

4.7. *Possible changes in shareholder composition*

A widely held belief in the literature argues that stock splits are intended to keep the price of shares within some "optimal" range so that investors with wealth constraints can continue to buy stocks in round lots. On the other hand, wealthy investors and institutions will save on brokerage costs if securities are priced high because of the fixed per-share transaction cost component. Therefore, the argument goes, there exists an optimal price range that equilibrates the preferences of these classes of investors and that managers can change the shareholders' ownership composition by means of splitting stocks (Lakonishok and Lev, 1987).

The results in Fig. 7 shed some lights on the possible changes in the shareholder composition using the trading volume per transaction as a proxy. A market dominated by institutional investors would be characterized by larger volumes per transaction, while a market dominated by small investors would result in smaller volumes per transaction. Alternatively, if there is no ownership composition change after the split, there should not be any systematic change observed in the measure of volume per transaction.

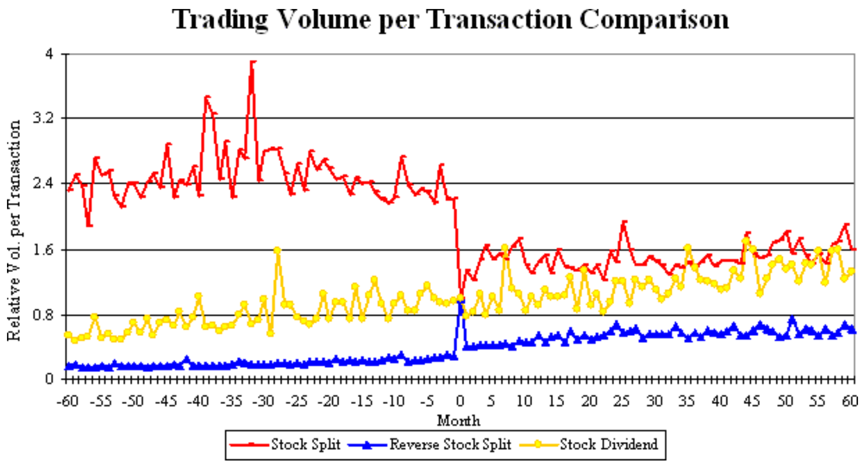


Figure 7. Trading volume per transaction trend.

However, as it can be seen in Fig. 7, there is an obvious change in the volume of trading per transaction between the pre-ex-date and post-ex-date in the stock split group. The trading volume per transaction decreases significantly after the ex-date and remains constant afterwards, signifying that the average trade is now smaller in size. One could argue that this evidence may indicate that in the post-split period, small-volume traders dominate the shareholder composition. There is no such clear trend in the reverse split and stock dividend sample. Table 7 confirms that the reduction in volume per transaction is statistically significant in the stock split sample. The present results do not support the recent US evidence in Denis and Strickland (2003) article. It is believed that an intuitive explanation for this observation is that the post-split period is characterized by smaller investors and not institutional investors. Smaller investors may look at the pre-split performance and believe that it signals superior future performance. This excess demand is met by institutional investors selling their stock. However, the results in this chapter only provide indirect and approximated evidence to support the investor composition hypothesis. Obviously, a direct test of this hypothesis is required in future research.

4.8. *Post-split dividend behavior*

While there is no reason as to why dividend policy or dividend payout might change as a result of (or following) a stock split, existing empirical research

Table 7. Statistical tests for the changes in trading volume per transaction.

Volume per transaction	Stock split	Reverse stock split	Stock dividend
1 year average pre-event	10.02	53.04	13.08
1 year average post-event	6.35	88.10	13.63
t-statistics	−2.33*	6.88**	0.17
2 year average pre-event	10.37	48.75	12.67
2 year average post-event	6.21	97.81	14.03
t-statistics	−2.88**	10.23**	0.42
3 year average pre-event	10.85	45.45	12.32
3 year average post-event	6.25	103.47	14.79
t-statistics	−2.75**	12.46**	0.73
5 year average pre-event	10.79	40.99	10.89
5 year average post-event	6.46	110.06	16.24
t-statistics	−2.42*	14.91**	1.63

*5% significance level.
**1% significance level.

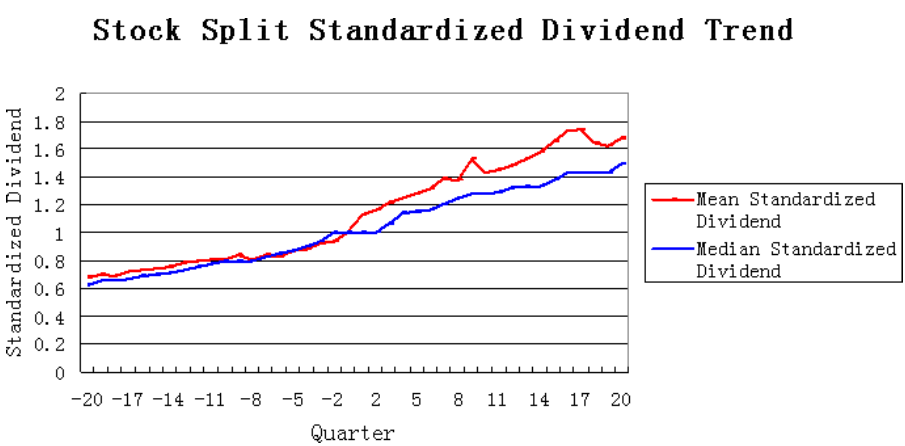


Figure 8. Trends in dividend payments.

indicates that there is a noticeable change in the dividend patterns following a stock split. In Fig. 8, this pattern is shown for the present sample of stock split firms. For the sake of brevity, the changes in dividends are shown by standardizing the quarterly dividends around the stock split (−20 to +20 quarters) by dividing the dividends by the dividend amount on “Quarter−1” (i.e., the quarterly dividend just before the stock split event).

The present sample for this analysis includes 336 companies. All the quarterly dividend information were collected for reverse stock split and stock dividend but it was found that about 95% of the reverse stock splits do not issue cash dividends. It was also found that most of the stock dividend companies do not issue cash dividends as well. Therefore, in this section, only the dividend trend in the stock split sample is discussed. In Fig. 8, the stock split firms demonstrate increasing dividend trends prior to the split event — both mean and median numbers rise from approximately 0.6 to 1.0 just prior to the split date. This trend continues in the post-split event where the standardized values rise to approximately 1.6 at Quarter +20 — a somewhat symmetrical increase around the split event. Given this symmetry, it is hard to say that there are any noticeable changes to dividend policy except to conclude that the growth trend in dividends continues after the split date. The dividend trend confirms the signaling effect again as it is found that the stock split firms are able to maintain an stable increasing dividend trend over a long time period after stock splits.

4.9. *Valuation impact*

Next the attention was turned to potential valuation impact around the stock split event. The signaling hypothesis suggests that firm value can be increased by reducing the asymmetric information gap between the inside managers and outside investors. Besides, if it is believed that the post-split period is characterized by smaller investors, it should be due to the stock price being in the "optimum" price range and this can then explain extra liquidity for the stock as well as smaller trading volume per transaction. The motivation of managers is to see an increase in valuation as a consequence of the increased liquidity as a result of the entry of small shareholders. Therefore, both of the signaling and optimum price range hypotheses should lead to the positive valuation impact on the stock split firms.

To investigate the possibility of change in valuation, the price to earnings (P/E) ratio of the stock split sample was compared during the pre- and the post-event period.[4] In Fig. 9, there is a noticeable change in the P/E ratio

---

[4]Only the results of the stock split sample are shown since most of the stock dividend paying stocks do not trade or trade actively and thus the results may not be representative. It should also be noted that the results in this section is based on about 180 stock splits due to the lack of data availability.

Figure 9. Changes in the valuation parameter of the stock split firms.

of the stock split firms; most of the change occurs prior to the split event and the P/E ratio trend climbs in the last 12 months before the event and then remains constant at a higher level in the post-event period. More specifically, the average of the median P/E ratios in the $(-60, -13)$ month period is 16.28 and the corresponding value in the $(+13, +60)$ month period is 18.15; the difference is statistically significant at 1% level. Also, although not reported here, the percentage of the firms who see an increase in their P/E ratios in the long term is significantly higher than the percentage of firms with reduced P/E ratios at 1% significant level. This evidence indicates that there is a permanent valuation impact as a result of the stock split.

Although higher P/E ratios have been found permanently (along with higher EPS ratios) in the post-split period, causality cannot be established. It might be argued that the increase in the number of shareholders leads to more noise trading and that may lead to higher valuations. Although, at this stage, this remains as a conjecture. However, the present results also suggest that managers can effectively transfer the favorable inside information and management expectations to the outside investors which in turn reduces the asymmetric information gap and results in an increase in the firm valuation in long term.

It is also possible that this change in shareholder composition produces another indirect but potentially significant benefit to managers. It is argued that as stock price increases, an increasing number of individual shareholders exit the stock and they are replaced by institutional shareholders who now have a higher incentive (and capability) to monitor management as there is a reduction in the "free rider" problem. It is argued that managers who dislike the increasing monitoring mechanism from big institutional investors may use the stock split mechanism to shift the investor composition in such a way that small investors can displace institutional investors. Similar arguments have been proposed by Brennan and Franks (1997) in the context of the underpricing of initial public offerings. However, even if the composition may change from institutional to individual investors, one must show that this change results in a relaxed corporate governance environment. To test whether this is indeed the case, an attempt was made to examine the board characteristics and ownership structures of the stock split firms.

4.10. *Post-split corporate governance environment*

The main objective of this section is to investigate whether there are significant differences in the governance characteristics of the stock split firms. Since there is a lack of readily available historical data on governance variables, it prevents us from conducting a longitudinal study in the governance changes across the pre- and the post-split periods. This study focused on analyzing governance framework of 74 firms that split stocks between the years 1998 and 2002. CFMRC database has been used as the population to randomly select 83 firms (as the control group), which did not split stocks in the past 10 years (from 1992 to 2002). The 2002 governance variables were collected for each of the stock split firms and the control firms by using SEDAR and Stock Guide database. Due to the lack of the governance information on seven stock split firms, the final sample size was reduced to 67 firms. Table 8 shows the differences in the means of the two samples on governance variables.[5] There is no significant difference in the governance characteristics between the two groups, except board size which is related to firm size, since the stock split firms in this comparison tended to be bigger than the control group.[6] Thus,

[5]The governance variables chosen are standard variables used in the governance literature; for example, see Agrawal and Knoeber (1996); Coles and Hesterly (2000); Coles, McWilliam and Sen (2000); and Core *et al.* ((1999).

[6]Also the multivariate relationships among the governance variables were considered by using some other techniques, such as logistic regression to compare the difference in governance environment. The multivariate results yield the same conclusion as the univariate ones.

Table 8. Statistical test of governance characteristics between the stock split firms and the control firms.

Governance variables	Mean difference	Std. Error	t	df	p-value
Board size	1.44	0.44	3.25	148	0.001**
Int_Director	0.12	0.17	0.68	121	0.50
PINDIR	−0.03	0.02	−1.29	148	0.20
CEO_Chair	0.03	0.08	0.33	147	0.74
Com_Owner%	−4.56	4.55	−1.00	138	0.32
Com_Vot%	2.81	5.30	0.53	138	0.60
Director_Own	−4.46	4.02	−1.11	138	0.27
Dir_Vot%	0.50	5.87	0.08	138	0.93
Block_Own%	0.12	3.56	0.03	137	0.97

Governance Variable Definitions. Board Size: Number of board directors. Int_Director: Number of inside directors. PINDIR: Percentage of inside directors to the board size. CEO_Chair: If CEO is the board chair, code as 1; otherwise, 0. Com_Owner%: Combined equity ownership by directors and block shareholders. Com_Vot%: Combined equity voting rights by directors and block shareholders. Director_Own%: Directors' ownership. Dir_Vot%: Directors' voting rights. Block Own%: Block shareholders' ownership.
**1% significance level.

although there seems to be a change in the investor composition in the post-split period away from institutional investors, it may not translate into a more relaxed governance regime for the split firms, as suggested in the empirical results from the study.

5. Summary and Conclusions

In this chapter, 30 years of Canadian evidence was used to document and provide evidence on three "cosmetic" events, namely stock split, reverse stock split, and stock dividend. The analysis of price levels around the split event dates indicates that the optimal price ranges have changed considerably over the 30-year period. Using the nominal prices, the median price of the stock around the pre-split event is \$31.50 and using real prices (adjusted for year 2002), the median price is \$52.50. Using the long-term analysis (as opposed to "the traditional event" analysis), it is shown that stock splits occur in times of bull markets where prices increase along with abnormal returns. The superior abnormal return trend of the stock split firms continues for a very long term after splits. The EPS level of the stock split firms has also permanently increased in the post-split period. However, reverse stock split and stock dividend firms have a fundamentally different value implication on shareholder wealth. In both the short-term and long-term post-split period, reverse stock split and stock dividend firms consistently underperform the market benchmark. Besides, the operating performance of these firms does not improve for quite a long time.

The stock split results strongly support the signaling hypothesis that managers use stock splits to convey favorable inside information and management expectations to the outside investors. Both the operating performance and the stock performance increase in the post-split period. The present evidence also indicates that the stock splits do bring stock into an "optimal" price range as there is a significant increase in the volume of trading and the number of transaction in the post-split period. Moreover, it is observed that the volume per transaction decreases considerably, which may be explained by a possible increase in the number of individual shareholders. It cannot be concluded that there are any changes to the dividend policy of the firms after stock splits except to note that the stock split firms continue to increase dividends in a steady fashion. In addition, permanent changes to the P/E ratios are found in the post-split period that means there is a permanent impact on the relative valuation parameters after the splits either because of the improved operating performance and improved investors' confidence or because of the increased liquidity appreciated by smaller investors.

Based on these results, one could argue that by splitting stocks, managers can actively make changes to the shareholders' ownership composition from large institutional investors to smaller investors so that firm valuation parameters can be enhanced. Indirect evidence was provided for this conjecture based on the reduction in the post-split transaction size. However, this shift toward individual investors is not accompanied by a relaxed governance environment, as there seems to be no difference in governance characteristics between the firms that split the stocks and those who did not.

Although the present results clearly support the signaling hypothesis, it is believed that the second hypothesis related to valuation changes needs more empirical evidence. The question of whether the rationale for stock split regarding optimal price range is actually related to (or can be replaced by) increases in valuation parameters as a result of noise trading by individual shareholders remains untested.

References

Agrawal, A and CR Knoeber (1996). Firm performance and mechanisms to control agency problems between managers and shareholders. *Journal of Financial and Quantitative Analysis*, 31(3), 377–397.

Akhigbe, A, J Madura and SP Zera (1995). Long-term valuation effects of stock splits. *Journal of Economics and Finance*, 19(3), 119–134.

Angel, JJ (1997). Tick size, share prices, and stock splits. *The Journal of Finance*, LII(2), 655–681.

Angel, J, RM Brooks and PG Mathew (2004). When-issued shares, small trades, and the variance of returns around stock splits. *The Journal of Financial Research*, XXVII(3), 415–433.

Arbel, A and G Swanson (1993). The role of information in stock split announcement effects. *Quarterly Journal of Business and Economics*, 32(2), 14–25.

Ariff, M (2004). Are stock splits credible signals? Evidence from the Singapore market. *The Singapore Economic Review*, 49(2), 163–177.

Baker, HK and PL Gallagher (1980). Management's view of stock splits. *Financial Management*, 9(2), 73–77.

Baker, HK and GE Powell (1993). Further evidence on managerial motives for stock splits. *Quarterly Journal of Business and Economics*, 32(3), 20–31.

Bar-Yosef, S and LD Brown (1977). A reexamination of stock splits using moving betas. *The Journal of Finance*, 32(4), 1069–1080.

Brennan, MJ and TE Copeland (1988a). Stock splits, stock prices, and transaction costs. *Journal of Financial Economics*, 22(1), 83–101.

Brennan, MJ and TE Copeland (1988b). Beta changes around stock splits: A note. *The Journal of Finance*, 43(4), 1009–1013.

Brennan, MJ and J Franks (1997). Underpricing, ownership and control in initial public offerings of equity securities in the UK. *Journal of Financial Economics*, 45, 391–413.

Byun, J and MS Rozeff (2003). Long-run performance after stock splits: 1927 to 1996. *The Journal of Finance*, LVIII(3), 1063–1085.

Charest, G (1978). Split information, stock returns and market efficiency-I and -II. *Journal of Financial Economics*, 6(2–3), 265–330.

Coles, JW and WS Hesterly (2000). Independence of the chairman and board composition: Firm choices and shareholder value. *Journal of Management*, 26(2), 195–214.

Coles, JW, VB McWilliams and N Sen (2000). An examination of the relationship of governance mechanisms to performance. *Journal of Management*, 27, 23–50.

Conroy, RM, RS Harris and BA Benet (1990). The effects of stock splits on bid-ask spreads. *The Journal of Finance* 45(4), 1285–1296.

Conroy, RM and RS Harris (1999). Stock splits and information: The role of share price. *Financial Management*, 28(3), 28–40.

Core, JE, RW Holthausen and DF Larcker (1999). Corporate governance, chief executive officer compensation, and firm performance. *Journal of Financial Economics*, 51, 371–406.

Crawford, D, DR Franz and GJ Lobo (2005). Signalling managerial optimism through stock dividends and stock splits: A reexamination of the retained earnings hypothesis. *Journal of Financial and Quantitative Analysis*, 40(3), 531–561.

Dennis, P (2003). Stock splits and liquidity: The case of the Nasdaq-100 index tracking stock. *The Financial Review*, 38, 415–433.

Dennis, P and D Strickland (2003). The effect of stock splits on liquidity and excess returns: Evidence from shareholder ownership composition. *Journal of Financial Research*, 26(3), 355–370.

Desai, H and PC Jain (1997). Long-run common stock returns following stock splits and reverse splits. *Journal of Business*, 70(3), 409–433.

Desai, A, M Nimalendran and S Venkataraman (1998). Changes in trading activity following stock splits and their effect on volatility and the adverse-information component of the bid-ask spread. *The Journal of Financial Research*, XXI(2), 159–183.

Dipchand, CR (1977). The Canadian experience with inherent stock splits of rights issues. *Financial Management*, Summer, 6(2), 34–41.

Doran, DT (1995). Stock splits and false signaling cost within a management reputation framework. *Journal of Applied Business Research*, 11(3), 115–127.

Dubosfsky, DA (1991). Volatility increases subsequent to NYSE and AMEX stock splits. *The Journal of Finance*, 46, 421–431.

Easley, D, M O'Hara and G Saar (2001). How stock splits affect trading: A microstructure approach. *Journal of Financial & Quantitative Analysis*, 36(1), 25–52.

Efron, B and R Tibshrani (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Elfakhani, S and T Lung (2003). The effect of split announcements on Canadian stocks. *Global Finance Journal*, 14(2), 197–216.

Fama, EF, L Fisher, M Jensen and R Roll (1969). The adjustment of stock prices to new information. *International Economic Review*, 10(1), 1–21.

Foster III, TW and EA Scribner (1998). Accounting for stock dividends and stock splits: Corrections to textbook coverage. *Issues in Accounting Education*, 13(1), 1–13.

French, DW and TW Foster III (2002). Does price discreteness affect the increase in return volatility following stock splits? *Financial Review*, 37(2), 281–294.

Gray, SF, T Smith and RE Whaley (2003). Stock splits: Implications for investor trading costs. *Journal of Empirical Finance*, 10(3), 271–303.

Grinblatt, MS, RW Masulis and S Titman (1984). The valuation effects of stock splits and stock dividends. *Journal of Financial Economics*, 13(4), 461–490.

Han, KC (1995). The effects of reverse splits on the liquidity of the stock. *Journal of Financial Quantitative Analysis*, 30(1), 159–169.

Ikenberry, DL, G Rankine and EK Stice (1996). What do stock splits really signal? *Journal of Financial and Quantitative Analysis*, 31(3), 357–375.

Kadapakkam, P, S Krishnamurthy and Y Tse (2005). Stock splits, broker promotion, and decimalization. *Journal of Financial and Quantitative Analysis*, 40(4), 875–895.

Kadiyala, P and MR Vetsuypens (2002). Are stock splits credible signals? Evidence from short-interest data. *Financial Management*, 31(1), 31–49.

Koski, JL (1998). Measurement effects and the variance of returns after stock splits and stock dividends. *The Review of Financial Studies*, 11(1), 143–162.

Kryzanowski, L and H Zhang (1991). Valuation effects of Canadian stock split announcements. *Economics Letters*, 36(3), 317–322.

Kryzanowski, L and H Zhang (1993). Market behaviour around Canadian stock-split ex-dates. *Journal of Empirical Finance*, 1(1), 57–81.

Kryzanowski, L and H Zhang (1996). Trading patterns of small and large traders around stock split ex-dates. *The Journal of Financial Research*, XIX(1), 75–90.

Lakonishok, J and B Lev (1987). Stock splits and stock dividends: Why, who and when. *The Journal of Finance*, 42(4), 913–932.

Lamoureux, CG and P Poon (1987). The market reaction to stock splits. *Journal of Finance*, 42(5), 1347–1370.

Leland, HE and DH Pyle (1977). Informational asymmetries, financial structure, and financial intermediation. *The Journal of Finance*, 32(2), 371–387.

Louis, H and D Robinson (2005). Do managers credibly use accruals to signal private information? Evidence from the pricing of discretionary accruals around stock splits. *Journal of Accounting and Economics*, 39, 361–380.

Ma Y, H Sun and J Yur-Austin (2000). Insider trading around stock split announcement. *The Journal of Applied Business Research*, 16(3), 13–25.

Mason, HB and RM Shelor (1998). Stock splits: An institutional investor preference. *The Financial Review*, 33(4), 33–46.

Masse, I, JR Hanrahan and J Kushner (1997). The effect of Canadian stock splits, stock dividends, and reverse splits on the value of the firm. *Quarterly Journal of Business & Economics*, 36(4), 51–62.

McNichols, M and A Dravid (1990). Stock dividends, stock splits, and signaling. *The Journal of Finance*, 45(3), 857–879.

Mukherjin, S, YH Kim and MC Walker (1997). The effect of stock splits on the ownership structure of firms. *Journal of Corporate Finance*, 3(2), 167–188.

Park, J and C Krishnamurti (1995). Stock splits, bid-ask spreads, and return variances: An empirical investigation of NASDAQ stock. *Quarterly Journal of Business and Economics*, 34(4), 39–45.

Peterson, CA, JA Millar and JN Rimbey (1996). The economic consequences of accounting for stock splits and large stock dividends. *The Accounting Review*, 71(2), 241–253.

Rankine, G and EK Stice (1997). The market reaction to the choice of accounting method for stock splits and large stock dividends. *Journal of Financial and Quantitative Analysis*, 32(2), 161–182.

Reboredo, JC (2003). How is the market reaction to stock splits? *Applied Financial Economics*, 13(5), 361–368.

Ross, SA (1977). The determination of financial structure: The incentive signalling approach. *Bell Journal of Economics*, 8(1), 23–40.

Rozeff, MS (1998). Stock splits: Evidence from mutual funds. *The Journal of Finance*, LIII(1), 335–349.

Schultz, P (2000). Stock splits, tick size, and sponsorship. *The Journal of Finance*, 55(1), 429–451.

So, RW and Y Tse (2000). Rationality of stock splits: The target-price habit hypothesis. *Review of Quantitative Finance & Accounting*, 14(1), 67–85.

Szewczyk, SH, GP Tsetsekos and R Varma (1992). Institutional ownership and the liquidity of common stock offerings. *The Financial Review*, 27(2), 211–225.

Szewczyk, SH and GP Tsetsekos (1993). The effect of managerial ownership on stock split-induced abnormal returns. *The Financial Review*, 28(3), 351–370.

Vafeas, N (2001). Reverse stock splits and earnings performance. *Accounting and Business Research*, 31(3), 191–202.

Wiggins, JB (1992). Beta changes around stock splits revisited. *Journal of Financial and Quantitative Analysis*, 27(4), 631–640.

Woolridge, JR (1983). Ex-date stock price adjustment to stock dividends: A note. *The Journal of Finance*, 38(1), 247–255.

Ye, J (1999). Excess returns, stock splits, and analyst earnings forecasts. *Journal of Portfolio Management*, 25(2), 70–76.

This page intentionally left blank

Chapter 6

Intraday Volume — Volatility Relation of the DOW: A Behavioral Interpretation

Ali F. Darrat
Louisiana Tech University, USA

Shafiqur Rahman
Portland State University, USA

Maosen Zhong
The University of Queensland, Australia

In a recent article, Darrat *et al.* (2003) report results for the DJIA in which higher volume causes higher volatility without significant feedback. These empirical results have interesting behavioral interpretations. It is argued that the observed positive causality from volume to volatility supports overconfidence hypothesis over other alternatives, including Andreassen's (1990) salience hypothesis. The evidence suggests that investors suffer from a psychological error (overconfidence), inciting them to trade too aggressively and drive prices away from their fundamental values.

Keywords: Trading volume; return volatility; behavioral finance; overconfidence; DJIA.

1. Introduction

The temporal relation between trading volume and return volatility in stock markets has attracted a great deal of research. The literature provides several alternative views on this relation, and two are particularly interesting. There is first the contention that volume and volatility are related contemporaneously; while others believe that the two variables exhibit a causal (lead/lag) relation. The contemporaneous relation is based on the mixture of distribution hypothesis (MDH) of Clark (1973), Harris (1987), among others. The MDH argues that the arrival of information impels simultaneous changes in volume and volatility and new market equilibrium is reached. Accordingly, volume and volatility should possess a contemporaneous correlation. The causal relation between volume and volatility is predicted by the sequential information arrival hypothesis (SIAH) of Copeland (1976), Smirlock and Starks (1988), and others. The SIAH assumes that investors react to new information differently. Therefore, the formation of new equilibria in the market is not immediate but requires some time (perhaps minutes), giving rise to a lead/lag relation between volume and volatility. Using intraday trading data for the 30 stocks of the DJIA, Darrat *et al.* (2003) recently report empirical results showing strong support for SIAH, that is, high volume causing high volatility.

This chapter uses behavioral insights to explain the empirical results of Darrat *et al.* (2003). It is argued in this chapter that Darrat *et al.*'s evidence of a positive causal effect from volume to volatility may not always be attributable to the arrival of sequential information. Certain plausible behavioral patterns, even in the absence of new information, can explain the intraday relation between volume and volatility in the US stock market.

The chapter is organized as follows: Section 2 discusses alternative interpretation of lead/lag relation between volume and volatility reported in Darrat *et al.* (2003). Section 3 reports the empirical results. Section 4 concludes the chapter.

2. A Behavioral Interpretation

Behavioral finance suggests that investors are generally quasirational. Several studies address the implications of quasirational behavior for trading in stock markets. For example, Kyle and Wang (1997), Daniel *et al.* (1998), Odean (1998), Gervais and Odean (2001), and Glaser and Weber (2003) argue that investors are often overconfident and tend to trade too aggressively. Investors' gains from frequent trading generally fall short of their expectations and, at times, may not even be sufficient to offset trading costs. Odean (1999) examines this overconfidence hypothesis by analyzing trading activities of individual investors at a large discount brokerage firm. He finds that investors do trade more aggressively than fully rational behavior would suggest, even after controlling for relevant factors like liquidity demands, portfolio diversification, risk management, and tax implications.[1] Statman

[1] In Odean (1998), overconfident traders cause excess volatility, while overconfident market makers dampen it. Thus, unidirectional causality from volume to volatility could imply that overconfident traders' effect dominates that of the market makers. For further discussion of potential behavioral patterns in the volume/volatility relation, see Bessembinder *et al.* (1996), Chen *et al.* (1999), Brown and Hartzell (2001), and Chordia *et al.* (2001). Other examples of overconfidence of investors in the financial markets can be found in Barber and Odean (2000, 2001).

et al. (2006) attempt to explain high observed trading volume in terms of investors' overconfidence about their valuation and trading skills. They find strong positive lead–lag relationship between returns and trading volume, consistent with the prediction of the overconfidence hypothesis.

Therefore, overconfidence may incite higher trading volumes. In addition to being overconfident about the precision of information signals they receive, investors may also be overconfident about the interpretation of such signals (Odean, 1998). When overconfident investors take larger positions than justified by rational behavior, prices tend to drift further apart from their true values, leading to higher volatility. Hence, overconfidence hypothesis suggests the presence of a positive causal link flowing from trading volume to return volatility.

In competitive markets, the conditional mean price of a given stock is approximately equal to the true fundamental value plus a random error. Conditional volatility measures the degree of time-varying deviations of actual prices from their conditional mean. Under perfectly rational expectations, there should be no deviation of actual returns from their expected values in equilibrium. If a discrepancy exists for some stocks, arbitrage would ensue to exploit any possible profits, causing higher volume. As rational traders take larger positions in the market, stock prices converge back to their fundamental values, and return volatility should then fall. Therefore, under market rationality, an increase in trading volume causes a subsequent decrease in market volatility, which is opposite to the overconfidence prediction.

Reverse causality from volatility to volume is also possible according to the "salience" behavioral theory — because stock prices are salient stimuli (Berlyne, 1970), higher market volatility could excite some investors. Additionally, higher return volatility often attracts further media coverage and demands alternative explanations (Andreassen, 1990). In this environment, it is possible that aroused public excitement triggers higher volatility even in the absence of genuinely new information.[2] Andreassen also suggests that media coverage may bias investors' forecasts by increasing the relative salience of information on price volatility. With a large number of investors, increased

[2]Huberman and Regev (2001) report an interesting case. They find that the stock price of EnterMed, a biotechnology firm, quadrupled over a weekend after a *New York Times* Sunday article examined a potential development of new cancer-curing drugs. Interestingly, other popular media had already reported this apparent breakthrough in cancer treatment more than 5 months earlier.

salience causes belief dispersions, which in turn encourage excessive trading (Blume *et al.*, 1994). Therefore, the salience behavioral hypothesis provides a third possibility where higher volatility triggers excessive trading.

3. Empirical Results

Darrat *et al.* (2003) use intraday (5-minute) transaction prices and trading volumes from the NYSE Trade and Quote database for the 30 DJIA stocks over the period April 1, through June 30, 1998. They omit the initial two 5-minute return observations of each day to mitigate the effect of stale price information. Given the clustering attribute of stock returns, Darrat *et al.* (2003) employ the exponential generalized autoregressive conditional heteroscedasticity (EGARCH) model to measure conditional return volatility. For comparability, the same data set has been used as in Darrat *et al.* (2003) and the distributional properties of intraday returns (log relatives of prices) and trading volumes for all 30 DJIA stocks were displayed in Table 1. The skewness and kurtosis are highly significant across all stocks, suggesting that the 5-minute returns do not follow a normal distribution. The return distributions are highly peaked around zero and exhibit longer and thicker tails than the corresponding normal variant with the same mean and variance.

Table 2 summarizes the empirical results discussed in Darrat *et al.*'s (2003) for each of the 30 DJIA stocks, along with the signs of the summed causal relations between volume and volatility. The conditional volatility h_t^2 of each DJIA stock's intraday returns is derived from a univariate EGARCH in mean model with 12 autoregressive lags of returns in the mean equation. A vector regressive (VAR) model is used to test the Granger-causality between the conditional volatility and log trading volume (V_t). As is clear from the table, the results support the presence of positive causal effects from volume to volatility in almost all 30 DJIA stocks, where such causal effects are statistically significant in a large number (12) of these stocks. As another check, we use Gibbons (1985) binomial pooled z-test to assess the overall causal effect between volume and volatility in the DJIA as a group. The binomial pooled z-test statistic (8.37) is indeed highly significant with a large and positive summed coefficient (2.15). Such a finding is consistent with the overconfidence prediction that higher trading volume causes higher return volatility. Moreover, the overwhelming evidence against negative causality from volume

Table 1. Summary statistics for intraday stock returns and trading volume of the 30 DJIA stocks.

	Stock returns				Volume (in 1,000)			Stock returns				Volume (in 1,000)	
Stock	Mean (%)	Std. Dev. (%)	Skewness	Kurtosis	Mean	Std. Dev.	Stock	Mean (%)	Std. Dev. (%)	Skewness	Kurtosis	Mean	Std. Dev.
AA	−0.0037	0.2160	1.94*	35.95*	13.22	20.61	IP	−0.0019	0.2963	0.90*	10.93*	15.18	19.67
ALD	0.0019	0.3545	0.69*	21.34*	17.14	23.75	JNJ	0.0005	0.2378	−0.13*	12.40*	26.78	28.20
AXP	0.0027	0.2489	1.00*	20.00*	19.36	21.90	JPM	−0.0056	0.1974	0.11*	5.20*	10.42	12.09
BA	−0.0045	0.4475	0.54*	429.04*	45.51	49.48	MCD	0.0036	0.2232	0.48*	36.44*	28.81	32.30
CAT	−0.0033	0.2620	0.33*	7.65*	15.83	21.88	MMM	−0.0015	0.2428	−0.16*	114.09*	11.35	14.35
CCE	−0.0002	0.3220	1.06*	13.49*	6.58	13.03	MO	−0.0021	0.3921	0.77*	34.00*	109.05	149.62
CHV	0.0006	0.2053	0.38*	14.30*	17.02	22.40	MRK	0.0014	0.2360	0.05*	61.49*	34.97	44.29
DD	0.0016	0.3093	9.14*	303.74*	34.34	38.45	PG	0.0025	0.2311	−0.22*	6.51*	25.78	28.35
DIS	0.0024	0.2416	2.15*	45.92*	26.45	34.89	S	0.0020	0.2510	0.09*	7.02*	15.72	23.25
EK	0.0018	0.2329	0.69*	12.95*	17.90	24.69	T	−0.0041	0.3305	−2.10*	113.05*	69.68	92.19
GE	0.0015	0.2389	0.36*	26.69*	49.67	42.25	TRV	−0.0025	0.4610	0.74*	478.29*	50.06	66.82
GM	−0.0016	0.2281	1.32*	17.13*	34.42	42.95	UK	−0.0013	0.3073	0.44*	28.91*	10.06	19.48
GT	−0.0046	0.2191	−0.24*	11.54*	7.47	14.94	UTX	0.0003	0.2027	0.01*	4.73*	9.69	15.43
HWP	−0.0014	0.4041	−12.06*	462.43*	54.17	84.55	WMT	0.0047	0.2715	0.49*	10.98*	35.30	35.98
IBM	0.0025	0.2260	1.08*	22.11*	50.65	73.31	XON	0.0011	0.2207	0.34*	4.23*	40.76	46.01

Notes: The data represent 5-minutes continuously compounding stock returns and trading volumes (in 1,000) of the 30 Dow Jones Industrial Average stocks. The first column contains stock symbols, and they are: Aluminum Co of America (AA), Allied Signal Inc (ALD), American Express Co (AXP), Boeing Co (BA), Caterpillar Inc (CAT), Coca-Cola Co (CCE), Chevron Corp (CHV), E.I. Du Pont de Nemours & Co (DD), WALT Disney Co (DIS), Eastman Kodak (EK), General Electric Co (GE), General Motors (GM), Goodyear Tire & Rubber Co (GT), Hewlett-Packard (HWP), International Business Machine (IBM), International Paper Co (IP), Johnson & Johnson (JNJ), JP Morgan Chase (JPM), McDonalds Corp (MCD), Minnesota Mining & Manufacturing Co (MMM), Philip Morris (MO), Merck & Co (MRK), Procter & Gamble (PG), Sears Roebuck (S), AT & T Corp (T), Travelers Group Inc (TRV), Union Carbide Co (UK), United Technology (UTX), Wal-Mart Stores (WMT), Exxon Corp (XON). The skewness and kurtosis statistics test the normality of stock returns.

*Indicates rejection of return normality.

Table 2. Likelihood ratio tests of Granger-causality between return volatility (h^2) and trading volume (V) in the 30 DJIA stocks.

$$h_t^2 = \gamma_1 + \sum_{k=1}^{12} a_k h_{t-k}^2 + \sum_{k=1}^{12} b_k V_{t-k} + \varepsilon_{1t}$$
$$V_t = \gamma_2 + \sum_{k=1}^{12} c_k V_{t-k} + \sum_{k=1}^{12} d_k h_{t-k}^2 + \varepsilon_{2t}$$

Stock	Null: volume $\not\rightarrow$ volatility $[b_1 = b_2 = \cdots b_{12} = 0]$		Null: volatility $\not\rightarrow$ volume $[d_1 = d_2 = \cdots d_{12} = 0]$		Bayesian-adjusted χ^2 critical values
	$\sum_{k=1}^{12} b_k$	LR (χ^2)	$\sum_{k=1}^{12} d_k$	LR (χ^2)	
AA	0.50	31.67	0.70	4.94	48.46
ALD	0.93*	52.38	0.57	15.21	48.85
AXP	0.32*	54.55	0.73	12.04	49.64
BA	41.99	9.61	0.61	7.45	48.69
CAT	0.39*	66.09	0.67	5.16	48.76
CCE	0.20	10.95	0.48	6.39	47.84
CHV	0.29	42.47	0.66	10.28	49.02
DD	0.12	8.32	0.63	8.10	49.65
DIS	0.37*	50.58	0.69*	72.38	49.75
EK	0.13*	79.86	0.75	8.76	48.93
GE	0.19	24.65	0.46	15.33	49.50
GM	0.32	42.84	0.57	14.25	48.74
GT	0.28	28.43	0.54	8.27	48.49
HWP	13.09	15.42	0.71*	51.85	49.51
IBM	0.68*	99.94	0.60	14.67	49.81
IP	0.19*	120.65	0.72	16.58	48.80
JNJ	0.33*	49.16	0.62	6.09	49.15
JPM	0.15*	52.56	0.70	3.27	49.69
MCD	0.09	10.57	0.63	11.43	48.67
MMM	0.09	9.61	0.69	8.05	49.03
MO	0.26*	56.49	0.50	8.34	48.46
MRK	0.38	26.29	0.71	12.66	49.87
PG	0.55*	77.58	0.68	13.29	49.56
S	0.25	27.17	0.66	7.66	49.03
T	0.14*	50.79	0.63	28.16	49.02
TRV	1.24	38.07	0.69	4.97	49.12
UK	0.15	14.54	0.66	10.95	48.47
UTX	−0.10	48.58	0.67	6.94	49.03
WMT	0.21	24.96	0.57	8.72	48.92
XON	0.72	43.31	0.54	10.26	49.29

Notes: See notes to Table 1. The last column reports the χ^2 critical values adjusted by Zellner's posterior odds ratio statistic.
*Indicates statistical significance.

to volatility in the vast majority of the DJIA stocks (all except one) is contrary to the rational-expectations model.

As to reverse causality from volatility to volume, the results reveal a positive link across all DJIA stocks as the salience hypothesis predicts. However, the estimated causal effects fail to achieve significance in almost all stocks. A pooled z-test provides a similar verdict and indicates that reverse causality from volatility to volume in the DJIA as a group carries no statistical significance. Of course, the consistently positive causality from volatility to volume does not contradict the salience behavioral hypothesis, although the general lack of significant causality casts some doubts on its validity.

4. Concluding Remarks

Behavioral finance suggests that trading volume and return volatility should follow a positive casual pattern in either direction. Behavioral insights were used to explain the results reported recently by Darrat *et al.* (2003) for the 30 DJIA stocks. The results show that higher trading volume causes higher return volatility in almost all DJIA stocks, and that the causal effects are also statistically significant in a large number of the stocks. These findings support overconfidence hypothesis, but contradict the rational-expectations prediction of a negative causal effect from volume to volatility. On the other hand, reverse causality from volatility to volume, although positive for all stocks, fails to achieve any statistical significance in the vast majority of the stocks. Clearly, such results cast some doubts on the validity of Andreassen's (1990) salience hypothesis.

Therefore, the evidence seems to suggest that investors suffer from an overconfidence psychological error, inducing them to trade too aggressively, driving prices away from their fundamental values, and triggering higher volatility. This could partly explain why market volatility appears too high to be compatible with rational behavior (Shiller, 1981), and could also explain why stock prices temporarily deviate from their fundamental values (Zhong *et al.*, 2003).

Finally, the analysis indicates that trading volume could be a useful predictor of future movements in stock prices. It does appear that "it takes volume to move prices". Yet, caution should be exercised since some components of the trading volume might be due to investors' heuristic-driven biased predictions rather than to genuinely new information about market fundamentals.

Excessive trading adds noise to the information dispersal process, which could disrupt the price-setting mechanism and provoke vicious volatility cycles.

References

Andreassen, PB (1990). Judgmental extrapolation and market overreaction: On the use and disuse of news. *Journal of Behavioral Decision-Making*, 3, 153–174.

Barber, B and T Odean (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance*, 55, 773–806.

Barber, B and T Odean (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics*, 16, 262–292.

Berlyne, DE (1970). Novelty, complexity and hedonic value. *Perception and Psychophysics*, 8, 279–286.

Bessembinder, H, K Chan and PJ Seguin (1996). An empirical examination of information, differences of opinion, and trading activities. *Journal of Financial Economics*, 40, 105–134.

Blume, L, D Easley and M O'Hara (1994). Market statistics and technical analysis, the role of volume. *Journal of Finance*, 49, 153–181.

Brown, GW and JC Hartzell (2001). Market reaction to public information: The atypical case of the Boston Celtics. *Journal of Financial Economics*, 60, 333–370.

Chen, CR, NJ Mohan and TL Steiner (1999). Discount rate changes, stock market returns, volatility, and trading volume: Evidence from intraday data and implications for market efficiency. *Journal of Banking & Finance*, 23, 897–924.

Chordia, T, A Subrahmanyam and VR Anshuman (2001). Trading activity and expected stock returns. *Journal of Financial Economics*, 59, 3–32.

Clark, P (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41, 135–155.

Copeland, T (1976). A model of asset trading under the assumption of sequential information arrival. *Journal of Finance*, 31, 1149–1168.

Daniel K, D Hirshleifer and A Subrahmanyam (1998). Investor psychology and security market under- and over-reaction. *Journal of Finance*, 53, 1839–1885.

Darrat, AF, S Rahman and M Zhong (2003). Intraday trading volume and return volatility of the DJIA stocks: A note. *Journal of Banking and Finance*, 27, 2035–2043.

Gervais, M and T Odean (2001). Learning to be overconfident. *Review of Financial Studies*, 14, 1–14.

Gibbons, JD (1985). *Nonparametric Methods for Quantitative Analysis*, 2nd Ed. Columbus, OH: American Science Press.

Glaser, M and M Weber (2004). Over confidence and trading volume. Working Paper, Center for Economic Policy, University of Mannheim, Germany.

Harris, L (1987). Transaction data test of the mixture of distribution hypothesis. *Journal of Financial and Quantitative Analysis*, 22, 127–141.

Huberman, G and T Regev (2001). Contagious speculation and a cure for cancer: A nonevent that made stock prices soar. *Journal of Finance*, 56, 387–396.

Kyle, AS and FA Wang (1997). Speculation duopoly with agreement to disagree: Can overconfidence survive the market test? *Journal of Finance*, 52, 2073–2090.

Odean, T (1998). Volume, volatility, price, and profit when all traders are above average. *Journal of Finance*, 53, 1887–1934.

Odean, T (1999). Do investors trade too much? *American Economic Review*, 89, 1279–1298.

Shiller, RJ (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, 71, 421–436.

Smirlock, M and Starks L (1988). An empirical analysis of the stock price–volume relationship. *Journal of Banking and Finance*, 21, 31–41.

Statman, M, S Thorley and K Vorkink (2006). Investor overconfidence and trading volume. *Review of Financial Studies*, 19, 1531–1565.

Zhong, M, AF Darrat and D Anderson (2003). Do US stock prices deviate from their fundamental values? Some new evidence. *Journal of Banking and Finance*, 27, 673–697.

Chapter 7

The Pricing of Initial Public Offerings: An Option Approach

Sheen Liu
Washington State University–Vancouver, USA

Chunchi Wu
Singapore Management University, Singapore

Peter Huaiyu Chen
Youngstown State University, USA

This chapter proposes a theoretical model of initial public offering by taking into account the uncertainty in security pricing and the underwriting process. Three major factors are shown to affect the IPO pricing: (1) the uncertainty of stock price, (2) the legal requirement that restricts underwriters from selling any part of the issue above the fixed offering price, and (3) the underwriters' risk tolerance. It is shown that underpricing is negatively related to the underwriter's risk tolerance, and positively related to the length of premarket period and market price volatility.

Keywords: Initial public offerings (IPO); options; IPO pricing.

1. Introduction

Underpricing of IPOs, measured as the percent difference between the closing day price and the initial offer price over the first few days of trading,[1] is well documented in the finance literature. Underwriters historically have underpriced the IPOs (Ibbotson, 1975; Smith, 1986; Habib and Ljungqvist, 2001), and the extent of underpricing has reached an astonishing level in some recent years (see Ritter and Welch, 2002). The size and persistence of underpricing in the US and foreign markets has long puzzled financial researchers.

Numerous explanations for IPO underpricing puzzle have been offered (see Ritter and Welch, 2002, for a detailed review). Several theories are based

[1] The number of days of trading over which the initial offer price is defined varies in different studies. For example, Habib and Ljungqvist (2001) define the offer price is the closing share price on the first day of trading, extracted from the daily CRSP tapes.

on the premise of asymmetric information among issuers, investment bankers, and investors. Under information asymmetry, it is costly to assess the fair value of an IPO. Rock (1986) shows that uninformed investors face a winner's curse when they submit an order for IPO shares. If the offer price is set at the expected market-clearing price, uninformed investors will systematically earn below normal returns. If an issue is underpriced, informed investors will submit bids and the issue is rationed. Conversely, if the issue is overpriced, informed investors will not submit bids and the issue is more likely to be undersubscribed. Hence, uninformed investors systematically receive more of overpriced issues and less of underpriced issues. This adverse selection problem for the issuers leads to underpricing on average because firms are forced to underprice their IPOs to compensate uninformed investors' losses.

Beatty and Ritter (1986) extends Rock's model to show that underpricing is higher for issues for which there is greater uncertainty about their value. They argue that the mechanism by which the underpricing equilibrium is enforced is via the investment banking industry. In order for investment bankers to find it in their interest to maintain the underpricing equilibrium, three conditions must be met. First, the underwriters are not perfect forecasters of the aftermarket price. Second, each underwriter must have nonsalvageable reputation capital at stake on which it earns a return. Third, any underwriter who cheats by pricing "off the line" must lose clients on average. If underwriters overprice the issue, investors who buy the IPOs will lose both money and trust in the investment banker. If they underprice the issue, the issuing firm will lose both capital and trust in the investment banker. Beatty and Ritter interpret their empirical findings as supporting their argument that investment bankers enforce the underpricing equilibrium.

Investment bankers may also seek to underprice IPOs either to compensate investors for revealing information during the preselling period to facilitate the pricing of the issue (Benveniste and Spindt, 1989), or to "leave a good taste in investors' mouths" so that future underwriting from the same issuer could be sold at attractive prices (Allen and Faulhaber, 1989; Grinblatt and Hwang, 1989; Welch 1989). Signaling theories suggest that IPO underpricing signals the quality of the issuers. IPO underpricing imposes severe costs on low-quality firms to preclude them from mimicking high-quality firms and therefore, leads to a separating equilibrium where firms voluntarily reveal their quality. Book-building theories (Benveniste and Spindt, 1989; Sherman, 2000; Sherman and Titman, 2001) suggest that investment bankers underprice IPOs to effectively extract information from their investors to facilitate the pricing of the new issue. To induce investors to reveal their information, investment bankers must offer them incentives, typically with some combination of more IPO allocations and underpricing. The allocational discretion given to underwriters in the book-building method enables underwriters to establish a long-term relationship with regular investors to reduce information cost and increase pricing efficiency.

The "deliberate underpricing" theory suggests that either investment bankers or issuers deliberately underprice IPOs in the premarket. The underpricing is deliberate either because offer prices are discounted to reflect the fundamental risk of the firm or because of the characteristics of the preoffering process. Hunt-McCool *et al.* (1996) provide empirical evidence supporting the deliberate underpricing theory. They develop a measure of deliberate underpricing using the information available only in the premarket period without reliance on the aftermarket price.[2]

Deliberate underpricing theory offers a plausible explanation for the underpricing of IPOs.[3] However, it does not provide a convincing argument to show that underpricing is an optimal strategy for underwriters. By underpricing IPOs, the underwriter benefits investors at the expense of the issuer. This imbalance begs a question: Is the underpricing consistent with market efficiency, or is the level of underpricing optimal for market participants?

In this chapter, an alternative model has been proposed to explain the underpricing of IPOs. The present model formalizes the argument by Beatty and Ritter (1986) that underpricing occurs because of uncertainty about IPO values. Under information uncertainty, underwriters face substantial risk of possible losses. If an IPO is overpriced, the deal may fail. On the other hand, if it is underpriced, the underwriter may lose their reputation and clients.

---

[2]Booth and Smith (1986), and Habib and Ljungqvist (2001) view promotional activities and underpricing as substitutes. Issuers choose an optimal promotion strategy, which involves deciding which underwriter and auditor to choose and how much to spend on advertising, as well as other promotional activities that may help reduce underpricing. Tinic (1988) attributes the underpricing in IPOs to the risk of potential lawsuits that might arise. An offering that starts trading at a higher price than the offering price is less likely to be sued.

[3]Other studies do not assume that the underpricing is deliberately undertaken. Ritter (1991) investigates the aftermarket efficiency of IPOs. His study shows that typically there are no positive abnormal returns for investors purchasing in the aftermarket. Ruud (1993) argues that the positive initial day returns are due to underwriter price support. The Securities and Exchange Commission allows investment bankers to intervene in the IPO aftermarket to stabilize prices. The effect of this price stabilization is to significantly reduce the number of negative initial returns that would be observed in the IPO aftermarket.

Two sources of pricing uncertainty were considered in the present model. The first is market volatility. The market can be volatile because of uncertain future economic or financial conditions. This external market condition is beyond the firm's control. A more volatile market increases the price uncertainty of IPOs. The second source of uncertainty is that the underwriter has incomplete information and is thus unsure about the value of the firm or the price the market is willing to bear. In this circumstance, pricing accuracy depends on underwriters' experience and the information gathering process. A well-established underwriter with more experienced analysts tends to generate more accurate price predictions. Alternatively, an underwriter may actively solicit information from outside investors through a "road show" or other information acquisition schemes to purchase information. To the extent that large underwriters have more capital to engage in information acquisition, they tend to possess better information and incur smaller errors in IPO pricing.

Another factor considered in our study is the risk tolerance of the underwriter. Underwriting is highly cyclical and only big firms can weather high fluctuations in volume and profitability.[4] In addition, underwriters often have to bear substantial risk in the aftermarket price support. Larger firms with more capital to sustain losses would have higher risk tolerance in IPO underwriting. It is shown that the underwriter's risk tolerance affects the magnitude of IPO underpricing and gross spread.

The remainder of this chapter is organized as follows. Section 2 sets up the environment, describes the underwriter's optimization problem, and presents an alternative model of IPO pricing. The model generates the optimal underpricing in IPOs and underwriting spread. Section 3 provides numerical examples to assess the impacts of pricing uncertainty, the length of the premarket period, and the underwriter's risk tolerance on IPO underpricing and the optimal spread. Finally, Section 4 concludes the chapter.

2. The Model

The model consists of two dates. At time 0, the underwriter sets the offering price. At time t_1, shares are allocated by the underwriter to investors and

[4]The Wall Street often alludes to the investment banking business as the MGM industry, which refers to the three giant investment bankers: Morgan Stanley, Goldman Sachs, and Merrill Lynch. Historically, few small investment banks have survived.

trading begins. At time T, the underwriter's obligation to support the price ends. The length of the period between setting the offering price (0) and actual trading (t_1) depends on the practice of investment bankers. In the US, this period is typically very short (e.g., the night before the trading in the morning) but in some countries, this period could be much longer. The underwriter's primary revenue source is from the gross spread a. The gross spread a should cover the underwriting cost C. When the offering price K is higher than the market-clearing price at time T, S_T, the underwriter bears the entire cost $(K - S_T)$ due to the obligation for price support. Although for simplicity, it is assumed that T is the end of the price support period, it can be interpreted as the period of price support itself and S_T is simply the weighted price in this period calculated at time T. The basic idea here is to set a time to calculate the cost that the underwrite bears in order to support the price after trading begins. For convenience, define T as the expiration date of the underwriter's price support and S_T is the weighted price over the price support period between t_1 and T throughout this chapter. It is assumed that the underwriter bears the entire cost of $(K - S_T)$. This assumption is reasonable when the underwriter enters a firm commitment agreement to support the aftermarket price. It is straightforward to extend the analysis to the case that the underwriter only bears a fraction of this cost (e.g., best-effort price stabilization).

The payoff for the underwriter of an IPO at the date when the shares start trading can hence be written as:

$$\min(a - C, S_T - K + a - C) \tag{1}$$

The underwriter sets the gross spread so that their losses can be fully compensated.

The underwriter has his own risk tolerance, characterized by the total value loss allowance V (in absolute value) and the maximum probability of the loss of V, α, that he can tolerate. The underwriter's risk tolerance on a per share basis can be formally expressed as:

$$P\left(S_T < K - a + C - \frac{V}{n}\right) < \alpha \tag{2}$$

where n is the total number of IPO shares.

There is no market price before the shares start trading. It is assumed that the unobserved true price follows a stochastic process in the period between time 0 and T. An underwriter does not have a perfect foresight of the market-clearing price of IPOs; instead, he obtains the best price estimate from available information. Two sources of errors contribute to the discrepancy of the underwriter's estimated price from the market-clearing price (before the end of the price support period). The first is the volatility of true price during the pre-market period. True price may change between the time the underwriter sets the offering price and the time the securities are sold to the public. This price uncertainty is determined by the nature of the state beyond the underwriter's control. The second source of price uncertainty is due to the underwriter's imperfect information. At the time 0, the underwriter estimates the market clearing price S_0 based on the information available to him at that time. The accuracy of price estimation depends on the underwriter's experience and the quality of his information.

Assume that the true stock price follows a stochastic process:

$$dS = \mu S dt + \sigma S dW \tag{3}$$

where μ is the mean and σ is the standard deviation of returns, and W is a Wiener process. The solution of (3) for the price at time T is:

$$S_T = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma W(T)} \tag{4}$$

where S_0 is the unobserved true price at time 0.

The underwriter needs to estimate the clearing price at time T. Since the underwriter has to set the price at time 0, their estimate of the clearing price is based on their information at time 0, F_0. The expected market clearing price at T, conditional on the underwriter's information at time 0 is:

$$\begin{aligned} E(S_T|F_0) &= E\left(S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma W(T)} \middle| F_0 \right) \\ &= S_0 e^{\left(\mu + \frac{\sigma^2}{2}\right)T} \end{aligned} \tag{5}$$

where E is the expectation operator. The true price S_0 at time 0 is unknown. Denote S_E as the underwriter's estimate of the unobserved price S_0 at time 0.

The fact that the underwriter does not have a perfect foresight of S_0 adds to another price uncertainty. The underwriter's price estimate can be characterized as:

$$\ln S_\mathrm{E} = \ln S_0 + \sigma_\mathrm{E} Z \tag{6}$$

where S_E is the underwriter's mean estimate of S_0 and $\sigma_E Z$ is the error term. The standard deviation, σ_E, reflects how accurately the underwriter estimates the unobserved true price. The error of pricing generally depends on the underwriter's experience, efforts to gather information and relationship with informed investors. For ease of notation, denote $\mu_E = \ln S_E$. Then, (6) can be rewritten as:

$$S_0 = e^{\mu_E - \sigma_E Z} \tag{7}$$

where S_0 follows a lognormal distribution, Z has a standard normal distribution, μ_E is the mean and σ_E is the standard deviation of $\ln S_0$. It is assumed that Z is independent of the Wiener process W in (3). It is reasonable to postulate in (6) that S_0 has a lognormal distribution because the security price can never be negative.

Substituting (7) into (4), we obtain the relationship between the clearing price and the price estimated by the underwriters:

$$S_T = e^{\mu_E - \sigma_E Z + \left(\mu - \frac{\sigma^2}{2}\right)T + \sigma W(T)}$$

The above equation includes two types of uncertainty, σ and σ_E, faced by the underwriter. The first is the price volatility of the market, and the second is the pricing dispersion due to the underwriter's imperfect information. The underwriter's risk tolerance in (2) can then be written as:

$$P\left(e^{\mu_E - \sigma_E Z + \left(\mu - \frac{\sigma^2}{2}\right)T + \sigma W(T)} < K - a + C - \frac{V}{n}\right) < \alpha \tag{2a}$$

Recall that Z and W are independently normally distributed. A new random variable X has been defined as:

$$X = \mu_E - \sigma_E Z + \left(\mu - \frac{\sigma^2}{2}\right)T + \sigma W(T) \tag{8}$$

X has a normal distribution with a mean:

$$\overline{\mu} = \mu_E + \left(\mu - \frac{\sigma^2}{2}\right)T \tag{9}$$

and a variance:

$$\overline{\sigma}^2 = \sigma_E^2 + \sigma^2 T \tag{10}$$

where the mean and variance of X include parameters of both the true price distribution and the underwriter's subjective price estimates. Using (8), (9), and (10), the underwriter's risk tolerance in (2a) can be rewritten as:

$$\frac{1}{\overline{\sigma}\sqrt{2\pi}}\int_{-\infty}^{\ln\left(K-a+C-\frac{V}{n}\right)} e^{-\frac{1}{2}\left(\frac{X-\overline{\mu}}{\overline{\sigma}}\right)^2} dX < \alpha \tag{11}$$

Note that:

$$\begin{aligned}\frac{1}{\overline{\sigma}\sqrt{2\pi}}&\int_{-\infty}^{\ln\left(K-a+C-\frac{V}{n}\right)} e^{-\frac{1}{2}\left(\frac{X-\overline{\mu}}{\overline{\sigma}}\right)^2} dX \\ &= N\left[\frac{\ln\left(K-a+C-\frac{V}{n}\right)-\overline{\mu}}{\overline{\sigma}}\right]\end{aligned} \tag{12}$$

Therefore, the underwriter's risk tolerance can be expressed as:

$$N\left[\frac{\ln\left(K-a+C-\frac{V}{n}\right)-\overline{\mu}}{\overline{\sigma}}\right] < \alpha$$

Setting the left side equal to the right side value α, the maximum offering price K conditional can be solved on the underwriter's risk tolerance:

$$K = \exp[\overline{\sigma}Q(\alpha) + \overline{\mu}] + a - C + \frac{V}{n} \tag{13}$$

where $Q(p)$ is the inverse normal distribution function.

Since the expected clearing price is $E(S_T) = e^{\overline{\mu}+\frac{\sigma_E^2}{2}}$, the underpricing of the IPO in percentage terms can be expressed as:

$$\frac{E(S_T)-K}{E(S_T)} = \left\{1 - \exp\left[\overline{\sigma}Q(\alpha) + \frac{\sigma_E^2}{2}\right] - \frac{a-C+V/n}{E(S_T)}\right\} \tag{14}$$

The underpricing in percentage depends on the uncertainty, $\overline{\sigma}^2$, the gross spread, a, and the risk tolerance of the underwriter V. The gross spread is unknown. The underwriter sets the value of gross spread to compensate for their potential losses. In the following the optimal gross spread was derived.

The payoff function for the underwriter in (1) at time T can be rewritten as a function of a put option on the underlying stock:

$$\begin{aligned}&\min(a - C, S_T - K + a - C)\\&\quad = a - C - \max(0, K - S_T)\\&\quad = a - C - p_T \qquad (15)\end{aligned}$$

where p_T is the put option price. The underwriter entering a firm commitment agreement and engaging in price pegging in effect sells a put option to guarantee the share price of the issuer in the aftermarket. If the market price of the new issue falls below the offering price, the put option is in the money and the underwriter incurs a loss. On the other hand, when the market price of the new issue is above the offering price, price stabilization is not needed. The put option is worthless and the underwriter incurs no loss. Note that the objective function in (15) does not rule out the possibility that underwriters can generate positive profit while stabilizing IPOs (see Ellis *et al.*, 2000). Here, the underwriter focuses on minimizing the cost by preventing the offer price to be substantially below the market clearing price.

Applying the Black-Scholes option model, the present value of the payoff can be expressed for the underwriter at time 0 for a given initial stock price S_0 as:

$$(a - C)e^{-rT} + S_0 N(-d_1) - e^{-rT} K N(-d_2) \qquad (16)$$

where $d_{1,2}$ is given by:

$$d_{1,2} = \frac{\ln\frac{S_0}{K} + \left(r \pm \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$$

Recall that $\ln S_0$ has a normal distribution with mean μ_{E} and standard deviation σ_{E}. It follows that the expected present value of the payoff for the risk neutral underwriter is:

$$\begin{aligned}E_0(a - C - p_T) = (a - C)e^{-rT}&\\ + \frac{1}{\sigma_{\mathrm{E}}\sqrt{2\pi}}\int_{-\infty}^{\infty}&[S_0 N(-d_1) - e^{-tT} K N(-d_2)]\\ \times e^{-\frac{1}{2}\left(\frac{\ln S_0 - \mu_{\mathrm{E}}}{\sigma_{\mathrm{E}}}\right)^2}&\, d\ln S_0 \qquad (17)\end{aligned}$$

where E_0 is the expectation taken at time 0. The breakeven gross spread is given by:

$$a = C - \frac{e^{tT}}{\sigma_{\mathrm{E}}\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[S_0 N(-d_1) - e^{-tT} K N(-d_2) \right] e^{-\frac{1}{2}\left(\frac{\ln S_0 - \mu_{\mathrm{E}}}{\sigma_{\mathrm{E}}}\right)^2} d \ln S_0 \tag{18}$$

Combining (13), the gross spread and the offering price K can be solved for a given floatation cost C. Since the offering price is a function of a, the above equation is an implicit function of a. The closed-form solution of a is not available and it needs to be solved numerically. In the next section, the implications of the model and present numerical examples are explored.

3. Numerical Analysis

The model shows that the underpricing of IPOs and the gross spread are determined by several factors. The first important factor is the underwriter's risk tolerance. A smaller amount of underpricing increases the probability of both the need for price support in the aftermarket and the potential loss in dollars and reputation for the underwriter. It can be shown from (14) that the magnitude of underpricing depends on the underwriter's risk tolerance. Other things being equal, the magnitude of the underpricing decreases with the loss allowance V, which is measured in absolute value.

Figure 1 shows the effect of the relative risk tolerance $\frac{V}{nE(S_T)}$ on the underpricing and gross spread. As the underwriter's tolerance of per share loss increases, the price discount (in percentage) decreases, while the gross spread (in percentage) increases. For example, as $\frac{V}{nE(S_T)}$ increases from 2% to 15%, the underpricing decreases from 19.31% to 3.21%. Thus, the per share loss tolerance has a strong effect on the percentage of underpricing. On the other hand, the gross spread moves in opposite direction, increasing from 0.62% to 4.21%. Note that the result here does not necessarily imply that gross spread is inversely related to the extent of underpricing. To determine the relationship between gross spread and underpricing, it is necessary to consider not only the risk tolerance but also other factors that affect these two variables. Recently, Chen and Ritter (2000) have found that there is not

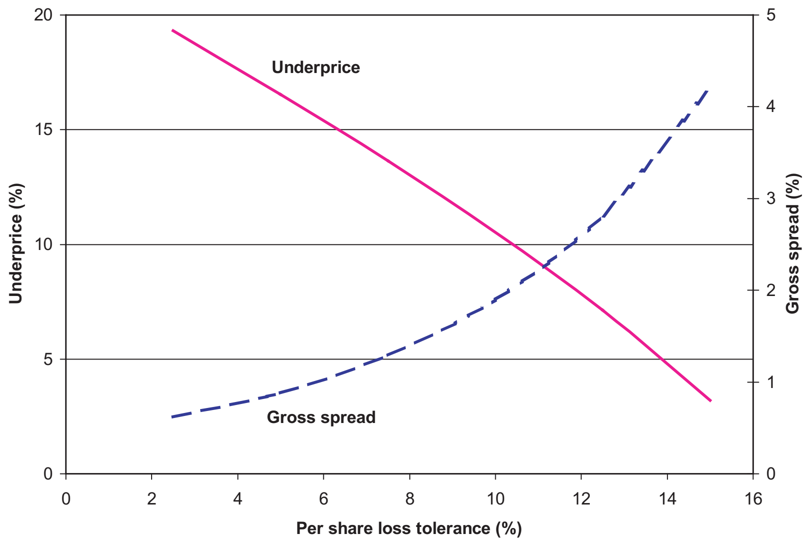


Figure 1. The effects of V/n on underpricing and gross spread. $S_{\mathrm{E}} = 100$, $r = 0.05$, $\sigma = 0.15$, $\sigma_{\mathrm{E}} = 0.15$, $\alpha = 0.05$, $\mu = 0$, $V/n = 10$, $T = 14$ days, and $C = 0.2$.

much relationship between gross spread and underwriting, using an extensive international IPO sample.

Furthermore, the model predicts that the greater the value of σ_{E}, the larger the underpricing. This confirms Beatty and Ritter's argument that underpricing is higher for issues that have greater uncertainty about their value. The higher the price uncertainty, the greater the chance that the aftermarket price will drop below the offering price and the higher the probability of exceeding the underwriter's risk tolerance. Therefore, the magnitude of underpricing is higher.

The present model predicts that underpricing is smaller for underwriters who can tolerate or allow for a bigger loss. Whether bigger or smaller underwriters have a higher level of risk tolerance is an empirical question. Recently, Loughran and Ritter (2004) and Beatty and Welch (1996) find that larger underwriters actually have greater underpricing compared to smaller, less prestigious underwriters. Their finding suggests that smaller underwriters may not necessarily have lower risk tolerance.

In general, larger reputed underwriters are more competitive in IPO pricing and they are able to absorb bigger losses. In the past, many small investment

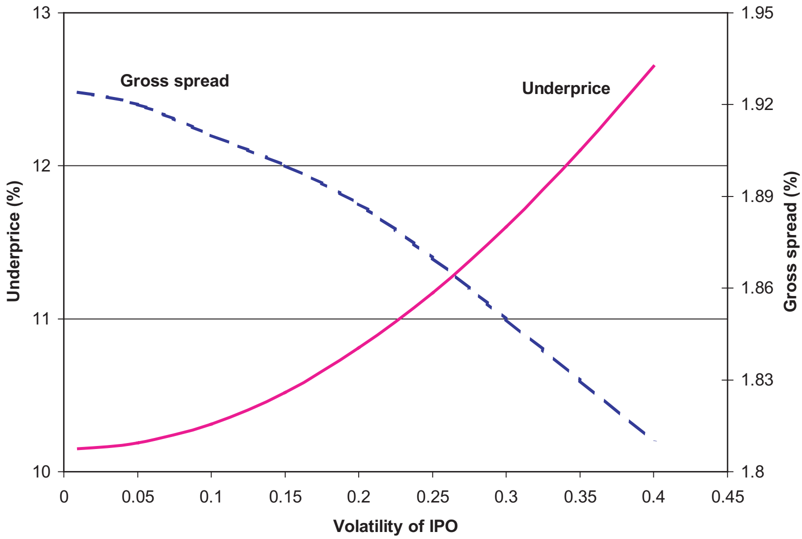


Figure 2. The effects of the lockup period on the underpricing and gross spread. $S_E = 100$, $r = 0.05$, $\sigma = 0.15$, $\sigma_E = 0.15$, $\alpha = 0.05$, $\mu = 0$, $V/n = 10$, and $C = 0.2$.

banks have faded away. Today, the underwriting business is dominated by a few large investment banking firms. Our model provides an explanation for the dominance of a few large investment banks in the underwriting business.

The model also predicts that the extent of underpricing depends on the length of the price support period of the IPO. The uncertainty increases with the length of the period between the time the offering price is set and the time that the price support period ends. Again, higher price uncertainty leads to greater underpricing. Figure 2 shows the effect of T on the underpricing and the gross spread. As T increases from 1 day to 28 days, the percentage underpricing increases from 10.12% to 10.94%. On the other hand, the gross spread decreases from 1.97% down to 1.84%.

Price volatility also plays an important role in the determination of IPO underpricing. An increase in volatility is expected to have a similar effect as an increase in the length of the premarket period. In Fig. 3, the expiration date of the price support period T is fixed at 14 days. It shows that an increase in volatility results in a greater underpricing. As the volatility increases from 0.01 to 0.4, the underpricing widens from 10.15% to 12.65%. On the other hand, the gross spread moves in opposite direction. It drops from 1.92% down to 1.81%.

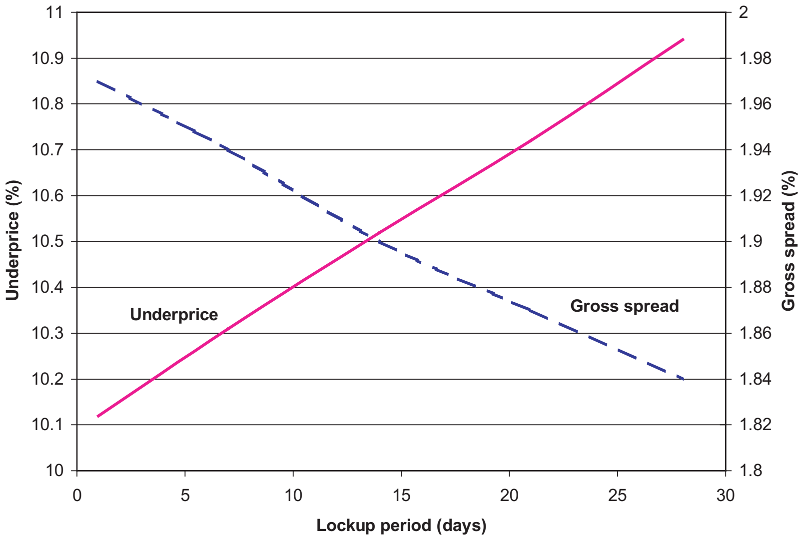


Figure 3. Effect of volatility on underprice and gross spread. $S_E = 100$, $r = 0.05$, $T =$ 14 days, $\sigma_E = 0.05$, $\alpha = 0.05$, $\mu = 0$, $V/n = 10$, and $C = 0.2$.

4. Conclusions

In this chapter, an IPO pricing model consistent with rational economic theory has been proposed to explain the underpricing and underwriters' spreads. The model takes into account the uncertainty in the price of the new issue and the underwriter's risk tolerance. Two major factors affect the uncertainty of the IPO price. First, underwriters have imperfect information and the information obtained from a limited pool of customers who may be subjective and biased. Second, unexpected events may occur between the time the offering price is set and the time when shares begin trading and thus change the fundamental price of the underwritten security. Higher price uncertainty leads to greater IPO underpricing. In addition, the underwriter's risk tolerance affects the extent of IPO underpricing. It is shown that the greater the loss tolerance, the lower the underpricing and the higher the gross spread. Furthermore, the length of the price support period and the volatility of price have a positive effect the underpricing of IPOs. The model quantifies the effects of these factors on both underpricing and the gross spread.

The model is useful for predicting IPO pricing. It is shown that the IPO pricing can be evaluated by a parsimonious framework that incorporates important factors related to information uncertainty, risk tolerance implied by the underwriter's risk management, and marketing and flotation costs. This model would be particularly useful for valuing the IPOs of small firms which are typically young and have strong growth opportunity but do not have much asset in place. The valuation of such securities is challenging because these firms often have negative earnings or net cash flows.[5] High uncertainty in the future growth of these firms makes it difficult to estimate their future earnings. The conventional discounted cash flow approach will be hard to apply not only because future earnings and sales estimates are quite subjective but also it is difficult to measure risk precisely to determine a proper discount rate. Under this circumstance, the option approach becomes a superior choice for valuing IPO firms with negative earnings but positive future growth.

The model can also be used to analyze the components of underwriters' spreads. Traditional underwriting spread analyses have relied on *ad hoc* regression models with little theoretical justification. The present model offers a different approach based on rational economic theory. The model provides a consistent framework to estimate underwriting spread and to decompose it into costs of floatation and market maintenance.

The present model can be refined to accommodate the unique features of small businesses using different underwriting venues and from different industries. It can also be generalized to account for the effects of business cycle and changing market conditions. The performance of the model can be evaluated using either numerical or empirical analyses. The model generates several interesting hypotheses which can be tested empirically. This empirical scrutiny has been left for future work.

References

Allen, F and GR Faulhaber (1989). Signaling by underpricing in the IPO market. *Journal of Financial Economics*, 23, 303–324.

Beatty, RP and JR Ritter (1986). Investment banking, reputation, and the underpricing of initial public offerings. *Journal of Financial Economics*, 15, 213–232.

Beatty, RP and I Welch (1996). Issuer expenses and legal liability in initial public offerings. *Journal of Law and Economics*, 39, 545–602.

---

[5]The percentage of firms with negative earnings in the 12 months before going public increases substantially recently. The percentage increases from 19% of the total IPO firms in 1980s, to 37% by 1995 and 1998 and reaches 79% during the period of Internet bubble in early 2000. See also Ritter and Welch (2002).

Benveniste, LM and PA Spindt (1989). How investment bankers determine the offer price and allocation of new issues. *Journal of Financial Economics*, 24, 343–362.

Booth, JR and RL Smith (1986). Capital raising, underwriting and the certification hypothesis. *Journal of Financial Economics*, 15, 261–281.

Chen, H and J Ritter (2000). The seven percent solution. *Journal of Finance*, 55, 1105–1131.

Ellis, K, R Michaely and M O'Hara (2000). When the underwriter is the market maker: An examination of trading in the IPO aftermarket. *Journal of Finance*, 55, 1039–1074.

Grinblatt, M and CY Hwang (1989). Signaling and the pricing of new issues. *Journal of Finance*, 44, 393–420.

Habib, MA and AP Ljungqvist (2001). Underpricing and entrepreneurial wealth losses in IPOs: Theory and evidence. *Review of Financial Studies*, 14(2), 433–458.

Hunt-McCool, J, SC Koh and BB Francis (1996). Testing for deliberate underpricing in the IPO premarket: A stochastic frontier approach. *Review of Financial Studies*, 9(4), 1251–1269.

Ibbotson, RG (1975). Price performance of common stock new issues. *Journal of Financial Economics*, 2, 235–272.

Loughran, T and J Ritter (2004). Why has IPO underpricing changed over time? *Financial Management*, 33, 5–37.

Ritter, J (1991). The long-run performance of initial public offerings. *Journal of Finance*, 46, 3–27.

Ritter, J and I Welch (2002). A review of IPO activity, pricing and allocations. *Journal of Finance*, 57, 1795–1828.

Rock, K (1986). Why new issues are underpriced? *Journal of Financial Economics*, 15, 187–212.

Ruud, J (1993). Underwriter price support and the IPO underpricing puzzle. *Journal of Financial Economics*, 34, 135–151.

Sherman, AE (2000). IPOs and long-term relationships: An advantage of book building. *Review of Financial Studies*, 13, 697–714.

Sherman, AE and S Titman (2001). Building the IPO order book: Underpricing and participation limits with costly information. Working Paper, University of Notre Dame.

Smith, CW (1986). Investment banking and the capital acquisition process. *Journal of Financial Economics*, 15, 3–29.

Tinic, SM (1988). Anatomy of initial public offerings of common stock. *Journal of Finance*, 43, 789–822.

Welch, I (1989). Seasoned offerings, imitation costs, and the underpricing of initial public offerings. *Journal of Finance*, 44, 421–449.

Chapter 8

Determinants of Winner–Loser Effects in National Stock Markets

Ming-Shiun Pan
Shippensburg University, USA

In this study, the determinants of profits to momentum/contrarian strategies were examined when applied to national stock market indexes. Using monthly stock market index data of 16 countries from December 1969 to December 2000, it is found that momentum strategies are profitable over horizons from 3 to 12 months, while contrarian strategies are profitable for long horizons such as 2 years or longer. However, the profit is statistically significant for only the 6-month horizon. The present decomposition analysis indicates that international momentum/contrarian profits are mainly due to the autocorrelations in these national market index returns, not to cross-serial correlations or to cross-sectional variation in their mean returns. Consistent with the trading strategy results, it is also found that most of the stock market indexes follow a mean-reverting process, implying positive autocorrelations in short-horizon returns and negative autocorrelations in long lags.

Keywords: International winner-loser effect; momentum; contrarian.

1. Introduction

Numerous studies have uncovered return anomalies based on trading strategies in the US stock market. For example, DeBondt and Thaler (1985) and others find a long-horizon return reversal effect that shows winners over the past 3 to 5 years having the tendency to become future losers, and *vice versa*. Furthermore, Jegadeesh and Titman (1993) show that momentum strategies that buy past 6- to 12-month winners and short sell past losers earn significant profits over the subsequent 6- to 12-month period. Similar price momentum also exists in other countries (e.g., Rouwenhorst, 1998).

Similar trading strategies when applied to national stock market indexes are also found profitable. Richards (1995) and Balvers *et al.* (2000) find that there are long-term winner–loser reversals in national stock market indexes of well-developed capital markets. Richards (1997) shows that the reversals cannot be explained by the differences in riskiness between loser and winner countries or adverse economic states of the world. Chan *et al.* (2000) provide evidence of momentum profits based on individual stock market indexes, especially for short holding periods. They also find that the profits to momentum trading strategies remain statistically significant after controlling for world beta risk as well as excluding emerging markets in the analysis. Fong *et al.* (2005) further show that international momentum strategies are profitable using a stochastic dominance approach. Fong *et al.*'s finding suggests that standard asset pricing models such as the CAPM and the three-factor model proposed by Fama and French (1993) cannot explain the momentum effect. Rather, it is an indication of market inefficiency.

One interpretation for the existence of international trading strategy effects is that returns on different countries' stock indexes are influenced by some global return factors, suggesting that they are cross-sectionally correlated. For instance, Bekaert and Harvey (1995) document the predictability for world equity markets using a set of global information variables. Their finding suggests that an international winner–loser effect could be attributed to the predictability of relative returns related to a common world component. Richards's (1995) finding indicates that the existence of a winner–loser effect is due to the predictability of relative returns in national equity indexes. His analysis indicates that the relative return predictability is associated with the existence of a common world component in each national return index.

Another possible interpretation is the time-series predictability in national equity markets. As Lo and MacKinlay (1990) demonstrate, profits from trading strategies can be from return autocorrelations, cross-serial correlations among securities, or cross-sectional variation in securities' unconditional mean returns. Specifically, they show that momentum profits are positively related to return autocorrelations, while contrarian profits are negatively related to return autocorrelations. Thus, international trading strategy effects could be contributed by positive autocorrelations in short-horizon stock returns and negative autocorrelations in long-horizon returns. Indeed, the mean-reverting property (i.e., positive autocorrelations in short lags and negative autocorrelations in long lags) that Poterba and Summers (1988) document in many national equity indexes could well explain why a winner–loser effect would prevail in national equity index returns. Richards (1995) claims that the long-run winner–loser effect in national equity indexes is mainly due to a mean-reverting component contained in national equity indexes. Similarly, Chan *et al.* (2000) suggest that international momentum profits arise from time-series predictability in stock market indexes.

Finally, profits to international trading strategies could arise because there is variation in unconditional mean returns across national equity markets. Lo and MacKinlay (1990) show that the variation in unconditional mean returns contributes positively (negatively) to the profit of trading strategies that long (short) winners and short (long) losers. Intuitively, if realized returns are strongly correlated to expected returns, then past winners (losers) that have higher (lower) returns tend to yield higher (lower) expected returns in the future. Consequently, momentum strategies that buy past winner countries and short sell past loser countries will gain from the cross-sectional dispersion in the mean returns of those winner and loser national equity indexes. On the other hand, the profit of buying losers and shorting winners will be affected by the variation in mean returns negatively.

While prior research documents international trading strategy effects, it is not clear what determinants of the profits are. In this study, an attempt was made to determine the determinants of profits from applying trading strategies to national equity market indexes. To explore possible causes for the international trading strategy effect, Lo and MacKinlay (1990) have been followed and the profits have been decomposed into three components, including (1) time-series predictability (autocovariance) in individual stock market indexes, (2) cross-sectional predictability (cross-serial covariance) between countries, and (3) variation in national equity markets' mean returns.[1] The present empirical results indicate that international momentum strategies yield profits over horizons from 3 to 12 months, while contrarian strategies generate profits for long horizons such as 2 years or beyond. Nevertheless, the profit is statistically significant for only the 6-month horizon. More importantly, the present results show that the international momentum and contrarian profits are mainly driven by individual stock markets' time-series predictability, not by the other two components.

The present results imply that national equity indexes in general follow a mean-reverting process — namely, positive autocorrelations in short-horizon stock returns and negative autocorrelations in long lags (e.g., see Fama and French, 1988; Poterba and Summers, 1988). To further examine this issue, Lo and MacKinlay's (1988) variance ratio analysis was employed. The present variance ratio results indicate mean reversion in most of the national equity

[1] A similar decomposition method is used in Conrad and Kaul (1998), Jegadeesh and Titman (2002), and Lewellen (2002).

indexes. Nevertheless, statistically speaking, the evidence against the random walk null is weak.

The rest of the chapter is organized as follows. Section 2 describes the strategies followed in formulating trading rules and discusses the decompositions of profits into various determinants using the Lo and MacKinlay (1990) method. Section 3 presents the profitability to international trading strategies and examines each individual stock market's time-serial dependence. The conclusion is in the final section.

2. Trading Strategies and Determinants of Profits

In this chapter, the Lo and MacKinlay (1990) method has been employed to formulate momentum (contrarian) strategies that buy (short sell) national stock market indexes at time t that were winners in the previous k periods and short sell (buy) national stock market indexes at time t that were losers in the previous k periods. Specifically, trading strategies portfolios are constructed with investment weights in stock index i determined as:

$$w_{i,t-1}(k) = \pm(1/N)[R_{i,t-1}(k) - R_{m,t-1}(k)] \tag{1}$$

where N is the number of national stock market indexes available, $R_{i,t-1}(k)$ is the return for stock index i at time $t-1$, and $R_{m,t-1}(k) = (1/N)\sum_{i=1}^{N} R_{i,t-1}(k)$ is the return for an equal-weighted portfolio of the stock market indexes at time $t-1$, and k is the return interval between time $t-1$ and t. Equation (1) shows that the investment weights are determined based on the performance of stock indexes against an equal-weighted world stock index. That is, the trading rules will buy or sell winner stock indexes at time $t-1$ that have higher returns than the average over the previous k periods and sell short or buy loser stock indexes at time $t-1$ that underperform the average in the previous k periods. The positions will be held for a horizon of k. By construction, the investment weights lead to a zero-cost, arbitrage portfolio since weights sum to 0, i.e., $\sum_{i=1}^{N} w_{i,t-1}(k) = 0$. Furthermore, bigger winners and losers will receive greater weights, as can be seen clearly from Eq. (1). Also, momentum strategies are implemented in an exactly opposite way as contrarian strategies. That is, a positive sign in the investment weight is for momentum strategies, while a negative sign is for contrarian strategies. In other words, a profitable momentum strategy implies that a same return-horizon contrarian strategy will yield a loss. Since the profit (loss) of a contrarian strategy exactly equals to the loss (profit) of a momentum strategy, the analyses in what follows assume that only momentum strategies are implemented.

The profit that a momentum strategy will realize at time t, $\pi_t(k)$, is:

$$\begin{aligned}
\pi_t(k) &= \sum_{i=1}^{N} w_{i,t-1}(k)R_{i,t}(k) \\
&= \frac{1}{N}\sum_{i=1}^{N}[R_{i,t-1}(k) - R_{m,t-1}(k)]R_{i,t}(k) \\
&= \frac{1}{N}\sum_{i=1}^{N}[R_{i,t-1}(k)R_{i,t}(k)] - R_{m,t-1}(k)R_{m,t}(k)
\end{aligned} \tag{2}$$

Assuming that unconditional mean returns of individual national stock markets are constant, the expected profits of momentum strategies can be decomposed into various components by taking expectations on both sides of Eq. (2):

$$\begin{aligned}
E[\pi_t(k)] &= \frac{1}{N}\sum_{i=1}^{N} E[R_{i,t-1}(k)R_{i,t}(k)] - E[R_{m,t-1}(k)R_{m,t}(k)] \\
&= \frac{1}{N}\sum_{i=1}^{N}(\mathrm{Cov}[R_{i,t-1}(k), R_{i,t}(k)] + \mu_i^2(k)) \\
&\qquad - (\mathrm{Cov}[R_{m,t-1}(k), R_{m,t}(k)] + \mu_m^2(k)) \\
&= -\mathrm{Cov}[R_{m,t-1}(k), R_{m,t}(k)] \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}\mathrm{Cov}[R_{i,t-1}(k), R_{i,t}(k)] \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}[\mu_i(k) - \mu_m(k)]^2
\end{aligned} \tag{3}$$

where μ_i and μ_m are the unconditional mean returns of stock market index i and the equal-weighted portfolio, respectively. Equation (3) indicates that the expected profits of momentum strategies come from three determinants: (1) the negative of the first-order autocovariance of the k-period returns for the equal-weighted world market portfolio, (2) the average of the first-order autocovariances of the k-period returns for national market indexes, and (3) the variance of the mean returns of stock indexes. Note that if each stock market index follows a random walk and also the equal-weighted world stock portfolio, then the expected momentum profit equals to the cross-sectional variation in these stock markets' mean returns.

We can further rewrite Eq. (3) as[2]:

$$\begin{aligned} E[\pi_t(k)] &= -\left\{ \mathrm{Cov}[R_{m,t-1}, R_{m,t}] - \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Cov}[R_{i,t-1}, R_{i,t}] \right\} \\ &\quad + \frac{N-1}{N^2}\sum_{i=1}^{N}\mathrm{Cov}[R_{i,t-1}, R_{i,t}] + \frac{1}{N}\sum_{i=1}^{N}(\mu_i - \mu_m)^2 \\ &= -C_1 + O_1 + \sigma^2(\mu) \end{aligned} \tag{4}$$

Equation (4) shows that the profitability of the international momentum strategy depends not only on the time-series predictability of individual stock markets, measured by the first-order autocovariance O_1, but also on the cross-serial predictability measured by the first-order cross-serial covariance C_1 and on the cross-sectional variations in mean returns of these stock markets. Thus, the decomposition shows that international momentum profits result from three determinants. First, stock index returns are negatively cross-serially correlated, implying that an equity market with a high return today is associated with low returns in the future for other equity markets. Second, individual stock indexes might be positively serially correlated, implying that an equity market with a high return today is expected to have high returns in the future. The final source arises because international momentum strategies tend to buy equity markets with a high mean return and short sell others with a low mean return.

For a contrarian strategy, the signs of the three terms on the right-hand side of Eq. (4) become just the opposite compared to a momentum strategy. Thus, time-series predictability and the variation of mean returns both contribute to contrarian profits negatively, while cross-serial predictability leads to positive contrarian profits.

[2] To simplify the expression, the return interval notation k is omitted from the right-hand side of equation.

3. Empirical Results

3.1. *Data*

Data employed in this study are monthly stock market indexes of Australia, Austria, Canada, Denmark, France, Germany, Hong Kong, Italy, Japan, the Netherlands, Norway, Spain, Sweden, Switzerland, the UK, and the US. The present study focuses on these countries because Richards (1995, 1997) examines international winner–loser effects using the stock market indexes of these 16 countries. Monthly Morgan Stanley Capital International (MSCI) total return indexes (capital gains plus dividends) from December 1969 to December 2000 are used.[3] The sample data consist of 373 monthly stock index returns. The analyses were conducted on return index data in both US dollars and local currency units.

3.2. *Profits to international trading strategies*

Table 1 reports expected profits to momentum strategies implemented on the 16 stock market index data in local currency units for five different horizons, with k equals 3, 6, 12, 24, and 36 months. Consistent with Richards (1997), momentum strategies appear to be profitable for horizons up to 1 year.[4] For horizons longer than 2 years, contrarian strategies that buy losers and short sell winners become profitable. However, the z-statistics,[5] which are asymptotically standard normal under the null hypothesis that the "true" profits equal to 0, suggest that the expected profits are significantly different from 0 at the 10% level for only the 6-month horizon.

Table 1 also contains the three components that make up the average momentum profits: the negative of the first-order cross-serial covariance $-C_1$,

---

[3]The data were retrieved from http://www.mscibarra.com.

[4]Note that unlike most prior studies that often use overlapping data to increase the power of statistical tests, the present study uses nonoverlapping data. Using overlapping data will yield overestimated autocovariance and cross-serial covariance and also too high z-statistics as Valkanov (2003) demonstrates. Valkanov shows that long-horizon regressions that are based on overlapping sums of the original series will always yield spurious results. He demonstrates that the ordinary least squares estimator in long-horizon regressions is not consistent and, furthermore, that t-statistics explode as the overlap increases, resulting in increasingly significant t-values when the horizon increases.

[5]The z-statistics are corrected for heteroskedasticity and autocorrelations up to eight lags based on the adjustments outlined in Newey and West (1987).

Table 1. Profits to momentum strategies for national stock market indexes in local currency.

Horizon	$E[\pi_t(k)]$	$-C_1$	O_1	$\sigma^2(\mu)$	$\%[-C_1]$	$\%[O_1]$	$\%[\sigma^2(\mu)]$
3-Month	0.1038	−0.1411	0.1958	0.0491	−135.93	188.63	47.30
	(0.595)	(−0.240)	(0.307)				
	$p = 0.302$	$p = 0.340$	$p = 0.265$				
6-Month	1.7831	−0.4857	2.0725	0.1963	−27.24	116.23	11.01
	(2.660)	(−0.277)	(0.960)				
	$p = 0.001$	$p = 0.318$	$p = 0.038$				
12-Month	1.7520	4.5881	−3.6213	0.7852	261.88	−206.70	44.82
	(1.475)	(0.805)	(−0.577)				
	$p = 0.152$	$p = 0.225$	$p = 0.357$				
24-Month	−5.3159	6.3968	−15.2221	3.5094	−120.33	286.35	−66.02
	(−1.113)	(0.621)	(−1.631)				
	$p = 0.182$	$p = 0.173$	$p = 0.103$				
36-Month	−14.1181	0.4758	−22.4900	7.8961	−3.37	159.30	−55.93
	(−1.590)	(0.016)	(−0.607)				
	$p = 0.071$	$p = 0.160$	$p = 0.035$				

This table contains the decomposition of average profits of trading strategies that long past winner stock market indexes and short past loser stock market indexes. Stock index data in local currency are used. The sample period is from December 1969 to December 2000. Expected profit is given by $E[\pi_t(k)] = -C_1 + O_1 + \sigma^2(\mu)$, where C_1 mainly depends on the average first-order cross-serial covariance of the returns for the 16 stock market indexes, O_1 depends on the average first-order autocovariance, and $\sigma^2(\mu)$ is the cross-sectional variance of the mean returns. The numbers in parentheses are z-statistics that are asymptotically $N(0, 1)$ under the null hypothesis that the relevant parameter is 0, and are robust to autocorrelation and heteroskedasticity. The p-value reports the probability that the 1,000 bootstrap average profits from the bootstrap distribution are less (larger) than the sample average profit if the sample value is less (larger) than the median of the bootstrap distribution. All profit estimates are multiplied by 1,000.

the first-order autocovariance O_1, and the variation in mean returns $\sigma^2(\mu)$. It is noteworthy that for all horizons, the first two components, $-C_1$ and O_1, are opposite in their signs, implying that positive (negative) autocorrelation is associated with positive (negative) cross-serial correlation. Comparing these two components suggests that the autocorrelations in the national stock market index returns are more important than the cross-serial correlations in determining the profitability of trading strategies. For instance, at the 6-month horizon, the autocovariance component counts for about 116% of the momentum profit, while the cross-serial covariance component contributes a negative 27% to the profit. For long horizons that contrarian strategies are profitable, the profits arise because the autocovariances are negative for long-horizon returns and they contribute to contrarian profits negatively. Nevertheless, based on the z-statistics, none of the auto- and cross-serial covariance components is statistically significant at any conventional levels.

The cross-sectional variation in the mean returns of these stock market indexes appears not to affect the winner–loser effect that much (see Table 1). For the momentum effect, the variation in these markets' returns counts for a small percentage of the profit when compared to that of the autocovariance component. For the contrarian effect, the variation in mean returns indeed contributes negatively to the profits.

It is quite clear from Table 1 that the own time-series predictability in each stock market index is the main driving force for the international winner–loser effect. However, it should be noted that due to small sample bias the statistical power of the z test might be low, especially for long horizons. For example, at the 24-month horizon, the effective sample size (i.e., the number of independent pieces of information) is only 15 for the present data. To remedy this problem, a bootstrap test was performed. For the bootstrap test, the monthly stock index returns of 16 countries were shuffled (without replacement)[6] simultaneously so that both auto- and cross-serial correlations are eliminated. The expected profits and the profit components of C_1 and O_1 were calculated for each bootstrap sample. A total of 1,000 replications are implemented. The results from the bootstrap analysis are also provided in Table 1. The p-value, which is the probability that the 1,000 bootstrap average profits from the bootstrap distribution are less (larger) than the sample average profit if the sample value is less (larger) than the median of the bootstrap distribution, has been focused for statistical inference. Based on the p-values, the momentum strategy at the 6-month horizon generates significant profit at the 1% level, while the contrarian strategy at the 3-year horizon yields significant profit at the 10% level. The autocovariance estimate for both of these two cases is significant at the 5% level, but not the cross-serial covariance

[6]Usually bootstrap experiments are done with replacement (e.g., see Conrad and Kaul, 1998). However, as Jegadeesh and Titman (2002) show, bootstrap experiments where returns are drawn with replacement will overstate the cross-sectional variation in stock mean returns for small samples. Thus, even though shuffling has eliminated any time- and cross-serial dependence in stock returns, bootstrapping with replacement may still generate significant profits because of the small sample bias.

Table 2. Profits to momentum strategies for national stock market indexes in US dollar.

Horizon	$E[\pi_t(k)]$	$-C_1$	O_1	$\sigma^2(\mu)$	$\%[-C_1]$	$\%[O_1]$	$\%[\sigma^2(\mu)]$
3-Month	0.1355	−0.2169	0.3123	0.0401	−160.07	230.48	29.59
	(0.587)	(−0.270)	(0.357)				
	$p = 0.246$	$p = 0.301$	$p = 0.228$				
6-Month	1.5304	−1.5679	2.9398	0.1604	−102.57	192.09	10.48
	(2.388)	(−0.658)	(1.184)				
	$p = 0.008$	$p = 0.095$	$p = 0.011$				
12-Month	1.1811	1.5566	−1.0171	0.6416	131.79	−86.11	54.32
	(1.102)	(0.292)	(−0.187)				
	$p = 0.204$	$p = 0.407$	$p = 0.411$				
24-Month	−3.5281	−0.6602	−5.7864	2.9185	18.71	164.01	−82.72
	(−0.791)	(−0.045)	(−0.405)				
	$p = 0.391$	$p = 0.404$	$p = 0.377$				
36-Month	−14.7218	13.0964	−34.3849	6.5667	−88.96	233.57	−44.61
	(−2.057)	(0.426)	(−0.944)				
	$p = 0.131$	$p = 0.094$	$p = 0.027$				

This table contains the decomposition of average profits of trading strategies that long past winner stock market indexes and short past loser stock market indexes. Stock index data in US dollar are used. The sample period is from December 1969 to December 2000. Expected profit is given by $E[\pi_t(k)] = -C_1 + O_1 + \sigma^2(\mu)$, where C_1 mainly depends on the average first-order cross-serial covariance of the returns for the 16 stock market indexes, O_1 depends on the average first-order autocovariance, and $\sigma^2(\mu)$ is the cross-sectional variance of the mean returns. The numbers in parentheses are z-statistics that are asymptotically $N(0, 1)$ under the null hypothesis that the relevant parameter is 0, and are robust to autocorrelation and heteroskedasticity. The p-value reports the probability that the 1,000 bootstrap average profits from the bootstrap distribution are less (larger) than the sample average profit if the sample value is less (larger) than the median of the bootstrap distribution. All profit estimates are multiplied by 1,000.

components. Thus, the statistical significance of the international winner–loser effects is apparently due to the autocovariance component.

Table 2 shows the results from applying the analysis on the data in US dollars. Note that the US dollar return is approximately equal to the sum of the local currency return and the rate of change in exchange rates with respect to US dollars. Therefore, in addition to time- and cross-serial predictability of equity returns and variation of mean returns, the expected profit in US dollars also contains components that are attributable to time- and cross-serial predictability of exchange rates. The results from the US dollar returns analysis are qualitatively similar to those of the local currency returns analysis.

Specifically, the z-statistics indicate that only the 6-month momentum and the 36-month contrarian strategies earn significant profits. Also, the major source of the momentum/contrarian profits arises from the autocovariance component, not from the cross-serial covariance or from the cross-sectional variation in mean returns. The bootstrap analysis reconfirms the importance of the autocovariance in contributing to the profits. For instance, for the 6- and 36-month horizons that the profits are significant, the autocovariances are highly statistically significant as well while the cross-serial covariances are only marginally significant. Comparing Tables 1 and 2, the profits and the component estimates in Table 2 are in general larger (sign ignored) than those in Table 1, suggesting that the time- and cross-serial predictability of exchange rates increases the momentum/contrarian profits. However, the predictability of exchange rates appears not to affect the profits that much in a statistical sense.[7]

3.3. *Variance ratio analysis*

The present international winner–loser analysis thus far indicates that intermediate-term momentum and long-term contrarian strategies are profitable. Moreover, this international momentum/contrarian effect is primarily due to time-series predictability of individual equity markets; i.e., positive autocorrelations in intermediate-horizon stock returns and negative autocorrelations in long-horizon returns. Thus, the present finding suggests that the international winner/loser effect is related to the time-series predictability in national stock market index returns. Fama and French (1988) and Poterba and Summers (1988) have investigated this pattern of autocorrelations in the United States and national equity markets, respectively. Both of these two studies focus on testing whether stock prices follow a mean-reverting behavior. Mean reversion implies that return autocorrelations are positive at intermediate horizons and negative at long horizons. In this study, Lo and MacKinlay's (1988) variance ratio analysis has been employed to test the significance of the autocorrelations of returns. The variance ratio analysis is appealing because it has greater power in testing the random walk null against the mean-reverting

[7] Chan *et al.* (2000) also document that short-term momentum profits in global equity markets are mainly contributed by time-series predictability in individual equity markets, not by predictability in exchange rate movements.

alternative compared with other commonly used statistics (see Richardson and Smith, 1994). The variance-ratio test exploits the fact that the variance of the increments in a random walk is linear in the sampling interval. That is, if a stock price series follows a random walk process, the variance of the k-period returns would be k times the variance of the single-period returns. In this study the variance ratio at lag k, $VR(k)$, is defined as:

$$VR(k) = \frac{Var[R_i(k)]}{Var[R_i(1)] \times k} \tag{5}$$

where $Var[\bullet]$ is the unbiased estimation of the variance of $[\bullet]$, $R_i(1)$ is return at the monthly interval for stock market i, and $R_i(k)$ is return at a k-month interval.

It can be shown that the variance ratio estimate at lag $k+1$ is approximately equal to one plus weighted sum of the first k autocorrelation coefficients of the single-period returns. Under the random walk hypothesis, all the autocorrelations should be 0 and hence $VR(k)$ equals 1. Thus, if $VR(k)$ is significantly different from unity, the null hypothesis of a random walk process should be rejected. It is also noticed that a variance ratio of less than 1 implies negative dependence and a larger than 1 variance ratio implies positive dependence. Furthermore, variance ratios for a mean reverting process will increase until some lags and then decline, while variance ratios for mean aversion increase along with the lags.

The variance ratio estimates for lags up to 40 are plotted in Fig. 1. Since most prior studies perform trading strategies on data in US dollars, the data in US dollars were focused. As can be seen, most variance ratios increase initially and start to decline at lags between 9 and 19 (e.g., Denmark, France, Norway, and Sweden, among others). For these stock index series, their return autocorrelations at short- and intermediate-lag are positive and become negative at long lags. Thus, apply momentum/contrarian strategies to these stock market indexes could yield profits. However, some market indexes, such as Austria, Italy, Japan, and Spain, seem to exhibit a mean-averting behavior because their variance ratio estimates increase almost monotonously. Nevertheless, only 4 out of 16 stock indexes show mean aversion and hence the national stock market indexes are dominated by mean reversion, which is consistent with results of trading strategies reported above.

Table 3 shows the variance ratio test statistics for the 16 national stock market index returns and the associated heteroskedasticity-robust standard

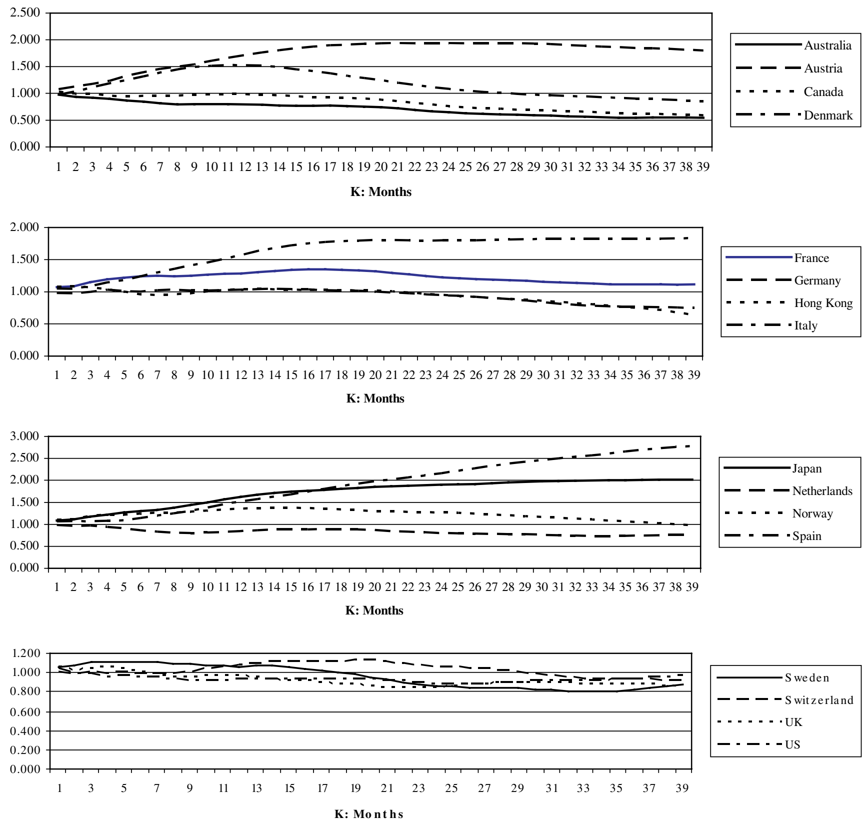


Figure 1. Variance ratio estimates of monthly returns on the 16 national stock market indexes in US dollar.

normal z-statistics. To save space, only the variance ratio test statistic are reported for lags of 4, 7, 13, 25, and 37 months, which are corresponding to the five horizons examined in the trading strategies. Again, most stock indexes show an increase in the variance ratio estimates initially and then a decrease after about lag 13 when the lag order increases. Interestingly, for the stock indexes that contain a mean-reverting feature (e.g., Denmark, Germany, and Sweden), the variance ratios are in general statistically insignificant. On the other hand, for the stock indexes that the variance ratios suggest mean aversion (e.g., Austria, Italy, Japan, and Spain), most estimates are statistically significant, especially for the estimates after lag 13. Thus, while the present variance ratio analysis shows that most stock market indexes follow a mean-reverting behavior, their deviations from random walks appear not to be statistically significant.

Table 3. Variance-ratio analysis of the random walk hypothesis for national stock market indexes.

Country	Variance ratio test statistic at lag				
	4	7	13	25	37
Australia	0.917	0.845	0.795	0.646	0.551
	(−0.74)	(−0.94)	(−0.89)	(−1.11)	(−1.18)
Austria	1.177	**1.393**	**1.719**	**1.940**	**1.842**
	(1.20)	(1.97)	(2.58)	(2.45)	(1.86)
Canada	0.997	0.956	0.986	0.764	0.618
	(−0.03)	(−0.27)	(−0.06)	(−0.77)	(−1.03)
Denmark	0.998	1.000	1.031	0.947	0.765
	(−0.02)	(0.01)	(0.14)	(−0.17)	(−0.63)
France	1.147	1.238	1.284	1.224	1.115
	(1.37)	(1.52)	(1.28)	(0.72)	(0.30)
Germany	0.998	1.001	1.031	0.947	0.765
	(−0.02)	(0.01)	(0.14)	(−0.17)	(−0.63)
Hong Kong	1.071	0.957	1.043	0.950	0.741
	(0.50)	(−0.21)	(0.16)	(−0.14)	(−0.60)
Italy	1.087	1.239	**1.574**	**1.802**	**1.826**
	(0.81)	(1.57)	(2.66)	(2.65)	(2.25)
Japan	1.173	**1.300**	**1.625**	**1.901**	**2.011**
	(1.60)	(1.95)	(2.85)	(2.94)	(2.72)
The Netherlands	0.973	0.864	0.845	0.805	0.746
	(−0.27)	(−0.95)	(−0.75)	(−0.66)	(−0.70)
Norway	**1.194**	1.247	1.358	1.269	1.050
	(1.76)	(1.57)	(1.64)	(0.87)	(0.14)
Spain	1.067	1.142	**1.520**	**2.158**	**2.698**
	(0.60)	(0.91)	(2.40)	(3.83)	(4.63)
Sweden	1.110	1.117	1.063	0.864	0.824
	(1.01)	(0.75)	(0.29)	(−0.46)	(−0.50)
Switzerland	0.997	0.988	1.068	1.061	0.928
	(−0.03)	(−0.08)	(0.32)	(0.21)	(−0.20)
The United Kingdom	1.039	1.000	0.965	0.850	0.874
	(0.27)	(0.01)	(−0.12)	(−0.39)	(−0.28)
The United States	0.981	0.954	0.924	0.885	0.937
	(−0.16)	(−0.28)	(−0.34)	(−0.37)	(−0.17)

This table reports Lo and MacKinlay's (1988) variance-ratio statistics for testing the significance of serial correlation. Stock index data in US dollar are used. The sample period is from December 1969 to December 2000. One-month is taken as a base observation interval. The variance ratio estimates are given in the main rows, with the absolute values of heteroskedasticity-robust z-statistics given in parentheses. Under the hypothesis that returns are serially uncorrelated, the variance ratio estimate is 1, and the test statistics are asymptotically $N(0,1)$. Bold denotes estimates are significant at the 10% level.

Clearly, the present finding that most stock indexes follow random walks in a statistical sense is inconsistent with the winner/loser effect reported above. One possible explanation is that the trading strategies are not designed for testing random walks and hence the results are not directly comparable with those of the variance ratio test.

4. Conclusions

In this study, the determinants of profits to momentum/contrarian strategies were examined when applied to national stock market indexes. Using monthly stock market index data of 16 countries from December 1969 to December 2000, it is found that momentum strategies are profitable for intermediate horizons, while contrarian strategies are profitable for long horizons such as 2 years or longer. However, the profit is statistically significant for only the 6-month horizon. The decomposition analysis suggests that the international momentum/contrarian profits are mainly due to the autocorrelations in these national market index returns. The other two profit components, cross-serial correlation and cross-sectional variation in mean returns, affect the profits not so significantly compared to the autocorrelation component. Consistent with the decomposition result, the variance ratio test results indicate that most stock market indexes follow a mean-reverting process, meaning that short-term return autocorrelations are positive and long-term return autocorrelations are negative. Moreover, it is found that for horizons that momentum is present the autocovariance component is positive, whereas for the horizons that contrarian presents the autocovariance component is negative. Since autocovariances contribute to momentum (contrarian) profits positively (negatively), coupled with a larger percentage contribution from this profit component, the present findings imply that the profitability to international trading strategies is mainly due to the time-series predictability that each individual stock market exhibits.

References

Balvers, R, Y Wu and E Gilliland (2000). Mean reversion across national stock markets and parametric contrarian investment strategies. *Journal of Finance*, 55, 745–772.

Bekaert, G and CR Harvey (1995). Time-varying world market integration. *Journal of Finance*, 50, 403–444.

Chan, K, A Hameed and W Tong (2000). Profitability of momentum strategies in the international equity markets. *Journal of Financial and Quantitative Analysis*, 35, 153–172.

Conrad, J and G Kaul (1998). An anatomy of trading strategies. *Review of Financial Studies*, 11, 489–519.

DeBondt, WFM and RH Thaler (1985). Does the stock market overreact? *Journal of Finance*, 40, 793–808.

Fama, EF and KR French (1988). Permanent and temporary components of stock prices. *Journal of Political Economy*, 96, 246–273.

Fama, EF and KR French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3–56.

Fong, WM, WK Wong and HH Lean (2005). International momentum strategies: A stochastic dominance approach. *Journal of Financial Markets*, 8, 89–109.

Jegadeesh, N and S Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48, 65–91.

Jegadeesh, N and S Titman (2002). Cross-sectional and time-series determinants of momentum returns. *Review of Financial Studies*, 15, 143–157.

Lewellen J (2002). Momentum and autocorrelation in stock returns. *Review of Financial Studies*, 15, 533–563.

Lo, AW and AC MacKinlay (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, 1, 41–66.

Lo, AW and AC MacKinlay (1990). When are contrarian profits due to stock market overreaction? *Review of Financial Studies*, 3, 175–205.

Newey, WK and KD West (1987). A simple, positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55, 703–708.

Poterba, JM and LH Summers (1988). Mean reversion in stock returns: Evidence and implications. *Journal of Financial Economics*, 22, 27–59.

Richards, AJ (1995). Comovements in national stock market returns: Evidence of predictability, but not cointegration. *Journal of Monetary Economics*, 36, 631–654.

Richards, AJ (1997). Winner–loser reversals in national stock market indices: Can they be explained. *Journal of Finance*, 52, 2129–2144.

Richardson, M and T Smith (1994). A unified approach to testing for serial correlation in stock returns. *Journal of Business*, 67, 371–399.

Rouwenhorst, KG (1998). International momentum strategies. *Journal of Finance*, 53, 263–284.

Valkanov, R (2003). Long-horizon regressions: Theoretical results and applications. *Journal of Financial Economics*, 68, 201–232.

Chapter 9

Earnings Management in Corporate Voting: Evidence from Antitakeover Charter Amendments

Chun-Keung Hoi
Rochester Institute of Technology, USA

Michael Lacina
University of Houston-Clear Lake, USA

Patricia L. Wollan
Rochester Institute of Technology, USA

Earnings management is examined around the time of proposals of antitakeover charter amendments. Overall, weak but statistically significant, negative abnormal accruals are found in the proposal year. The overall result appears to be driven by firms proposing restrictive charter amendments, which have strong negative abnormal accruals in the proposal year. The present interpretation is that executives of firms proposing restrictive amendments manage earnings opportunistically in that they defer income-decreasing accruals in the year prior to the amendment proposal to after the shareholders have voted on the amendment.

Keywords: Earnings management; corporate governance; corporate voting; antitakeover amendments; abnormal accruals.

1. Introduction

Shareholders have the right to vote on key managerial decisions and propose changes to corporate charters. With their vote, shareholders can potentially control and guide managerial discretion. If managers initiate proposals that are detrimental to the firm, shareholders can vote to reject them. However, if managers manipulate the voting process, either directly (via their ownership) or indirectly (via other efforts), the monitoring effects of the voting mechanism will be compromised.

In this chapter, corporate earnings surrounding proposals of antitakeover charter amendments (ATCAs) are used as a context to study indirect managerial influence of the corporate voting process. It is hypothesized that managers of firms proposing ATCAs have an incentive to manipulate earnings around the time of the amendment proposals. Opportunistic earnings management is taken to imply that managers can either accelerate the recognition of income-increasing (positive) accruals prior to the vote or postpone the recognition of income-decreasing (negative) accruals until after the vote. The present test design is based on the timing of the ATCA proposals and their corresponding vote. Year 0 is defined as the fiscal year in which the proposal and the corresponding vote take place. For 142 of the 148 firms in the sample, both the proposal and the corresponding vote occurred in the first half of the fiscal year. Therefore, if these firms accelerated the recognition of income-increasing accruals, positive abnormal accruals will be expected in Year −1. By contrast, if the firms postponed the recognition of income-decreasing accruals until after the vote, negative abnormal accruals will be expected in Year 0.

On average, in the whole sample, it is found that abnormal accruals in Year 0 (the year of the vote) are negative and statistically significant. However, in Year −1, no statistically significant abnormal accruals have been found. Since some ATCAs are more restrictive than others (i.e., with a greater potential to avert the threat of hostile takeover), the analysis was repeated using subsamples grouped by types of ATCAs. It is found that firms proposing restrictive charter amendments have negative and statistically significant abnormal accruals in Year 0, but firms proposing other types of charter amendments do not. Taken together, these findings lend credence to the notion that the managers of firms proposing restrictive amendments postpone the recognition of income-decreasing accruals until after the shareholder vote. Additional tests using quarterly abnormal accruals confirm that firms proposing restrictive amendments postpone income-decreasing accruals until after the vote takes place.

This chapter contributes to the literature in two ways. First, it provides additional evidence on managerial influence in the corporate voting process. The present results complement the findings of Bhagat and Jefferis (1991) for joint proposals, Bethel and Gillan (2002) for proxy solicitation services, and Brickley *et al.* (1994) for timing and pressure-sensitive institutional owners. Second, it adds to the strand of accounting literature that examines earnings management in episodic corporate events. In particular, the present findings on abnormal accruals extend the results of DeAngelo (1988) on proxy contests. DeAngelo examines an extreme corporate voting event, where dissident stockholders seek shareholder approval to replace incumbent managers. In contrast, a more robust methodology has been used to detect earnings management in a much less extreme corporate voting event, where incumbent managers have submitted an ATCA for shareholder approval.

The rest of the chapter proceeds as follows. Section 2 motivates the study, provides a literature review, and states the authors' hypotheses. Section 3 discusses the research methodology and describes the sample selection procedure. Section 4 presents the empirical results on abnormal accruals. Section 5 presents additional complementary evidence. Finally, the conclusion is provided in Section 6.

2. Antitakeover Charter Amendments and Abnormal Accruals

ATCAs are nonroutine management-initiated modifications to corporate charters that provide protection against hostile takeovers. Firms can use ATCAs to avert the threat of takeover. With the protection of ATCAs, managers can focus more on profitable long-term investments and less on myopic behavior (DeAngelo and Rice, 1983; Stein, 1988). However, ATCAs can also promote managerial interests at the expense of shareholders. They can insulate incumbent managers from the scrutiny of the corporate control market (DeAngelo and Rice, 1983). Moreover, ATCAs can help incumbent managers maintain above-average compensation levels (Borokhovich *et al.*, 1997).

Since ATCAs have the potential to create self-serving private benefits for managers, it is plausible that managers have an incentive to manipulate the corporate voting process in order to get them enacted. Existing findings have shown that managers can bias the vote outcome with a number of tactics: they time the ATCA proposals; they bundle the charter amendment with other issues in joint proposals (Bhagat and Jefferis, 1991); they use proxy solicitation services (Bethel and Gillan, 2002), and they put pressure on pressure-sensitive institutional owners (Brickley *et al.*, 1994).

Just as timing, bundling and using proxy solicitation services can influence voting, so can opportunistic earnings management. Hence, as a part of a package of concerted efforts to influence corporate voting, managers of firms proposing ATCAs might attempt to paint a better picture of their own performance by manipulating accounting earnings around the time of the amendment proposals. In this study, it is posited that ATCA-proposing managers will create such a temporary distortion in earnings by manipulating accounting earnings through accruals.

Previous researchers have documented earnings management in various episodic events; however, to the authors' knowledge, only DeAngelo (1988) has examined earnings management in the context of corporate voting.[1] DeAngelo uses abnormal (i.e., unexpected) accruals as a measure of earnings management. The authors assume a random walk by defining abnormal accruals as the change in total accruals. DeAngelo presents evidence that incumbent executives use abnormal accruals to inflate earnings during a proxy contest, where dissident stockholders seek election to the firm's board of directors.

The DeAngelo results are subject to criticism. Perry and Williams (1994), for instance, argue "the DeAngelo measure of earnings management may include a nondiscretionary component related to the level of activity and a nondiscretionary accrual mean reversion component." In addition, but more generally, Kothari *et al.* (2005) suggest that discretionary accrual[2] measures that fail to adjust for a performance-matched firm's discretionary accruals are unreliable and more likely to lead to incorrect inferences.

The methodological problems with the prior research provide motivation to re-examine earnings management in corporate voting. In this study, KLW (2005) has been followed and a more powerful and well-specified measurement of abnormal accruals was used to detect earnings management. In addition, earnings management is chosen for examination in the less extreme corporate voting event of the ATCA.

Opportunistic earnings management is taken as implying that managers can either accelerate the recognition of income-increasing (positive) accruals prior to the vote or postpone the recognition of income-decreasing (negative) accruals until after the vote. The authors' conjectures lead to the following hypotheses.

$\mathbf{H}_1$: Positive abnormal accruals in Year -1.
$\mathbf{H}_2$: Negative abnormal accruals in Year 0.

---

[1] Previous research has documented earnings management in events such as management buyouts (Perry and Williams, 1994), nonroutine executive changes (Pourciau, 1993), import relief investigations (Jones, 1991), seasoned equity offers (Teoh *et al.*, 1998a), initial public offers (Teoh *et al.*, 1998b; Teoh *et al.*, 1998c), stock-financed acquisitions (Erickson and Wang, 1999; Louis, 2004), and cancellations and subsequent reissuances of executive stock options (Coles *et al.*, 2006).

[2] The terms "abnormal accrual" and "discretionary accrual" have been used interchangeably.

If firms accelerate the recognition of income-increasing accruals, evidence would be expected to be consistent with H_1. Alternatively, if firms postpone the recognition of income-decreasing accruals until after the vote, then evidence would be expected to be consistent with H_2. Last, if firms do not manage earnings, then there should be no evidence found to support either H_1 or H_2.

However, not all charter amendments are the same. Fair price amendments, for instance, are arguably less likely to engender managerial entrenchment while classified board and supermajority amendments are more likely to be motivated by management's private incentives to reduce the scrutiny of the corporate control market (Jarrell and Poulson, 1987).[3] Accordingly, one might expect managers of firms proposing more restrictive amendments such as classified board and supermajority to more aggressively influence shareholder votes through earnings management, especially since these amendments are likely to receive less shareholder support. Accordingly, one would expect to observe different results by types of amendment proposals.

3. Methodology and Sample Selection

The KLW (2005) approach is based on the Jones (1991) discretionary accruals model. It augments the Jones model with a performance-matching procedure that incorporates return-on-assets as a control for earnings momentum and mean reversion in earnings.

3.1. *Jones model discretionary accruals*

Discretionary accruals (DA_{it}) is estimated as follows:

$$DA_{it} = TA_{it} - NDA_{it} \tag{1}$$

where TA_{it} is total accruals, NDA_{it} is nondiscretionary accruals, i is the firm, and t is the time period. Consistent with existing research, all measures of accruals (total, discretionary, and nondiscretionary) are scaled by total assets

[3]The supermajority amendment requires a minimum affirmative vote ranging from 66% to 90% of voting stock for a takeover to occur (Johnson and Rao, 1997). The classified board amendment staggers the election of the board of directors so that only a proportion of the board of directors can be elected at a point in time. Fair price amendments require a bidder for a company to pay a "fair" price for *all* purchased shares it has acquired. The failure to offer a fair price initiates a supermajority requirement.

from the prior period. Total accruals are defined as follows:

$$\mathrm{TA}_{it} = \frac{\Delta\mathrm{NoncashCAssets}_{it} - \Delta\mathrm{CLiabs}^{*}_{it}}{\mathrm{TAssets}_{it-1}} \tag{2}$$

where $\Delta\mathrm{NoncashCAssets}_{it}$ is the change in noncash current assets, $\Delta\mathrm{CLiabs}^{*}$ is the change in current liabilities, excluding the current portion of long-term debt, minus depreciation and amortization,[4] and $\mathrm{TAssets}_{it-1}$ is the prior-period total assets.

Following Jones (1991), cross-sectional regressions are used to estimate the annual nondiscretionary accruals component in Eq. (1).[5] This is a two-step procedure. First, the annual predictive OLS equation is obtained as follows:

$$\mathrm{TA}_{it} = \mathrm{a}_0 + \mathrm{a}_1(1/\mathrm{TAssets}_{it-1}) + \mathrm{a}_2(\Delta\mathrm{REV}_{it}) + \mathrm{a}_3(\mathrm{PPE}_{it}) + \nu_{it} \tag{3}$$

where TA_{it} is total accruals as defined above, $\mathrm{TAssets}_{it-1}$ is lagged total assets, $\Delta\mathrm{REV}_{it}$ is change in net sales scaled by $\mathrm{TAssets}_{it-1,}$ and PPE_{it} is gross property, plant, and equipment scaled by $\mathrm{TAssets}_{it-1}$. Regressions are performed on an industry-year basis. A separate predictive OLS equation is obtained for each industry-year in the sample using all available same-year, nonsample COMPUSTAT observations in the same two-digit SIC code. Second, the parameter estimates from Eq. (3) are used to estimate the annual nondiscretionary accruals (NDA) for the sample and control firms (described later) as follows:

$$\mathrm{NDA}_{it} = \mathrm{a}_1(1/\mathrm{TAssets}_{it-1}) + \mathrm{a}_2(\Delta\mathrm{REV}_{it}) + \mathrm{a}_3(\mathrm{PPE}_{it}) \tag{4}$$

Using the total accruals definition from Eq. (2) and the estimate of nondiscretionary accruals from Eq. (4), discretionary accruals were computed for both the sample firms and the control firms using Eq. (1).

[4]With respect to COMPUSTAT annual data items (A), total accruals = {[ΔA4 (current assets) – ΔA1 (cash and cash equivalents)] – [ΔA5 (current liabilities) – ΔA34 (debt in current liabilities)] – [A14 (depreciation and amortization)]} / lagged A6 (total assets). If the change in debt included in current liabilities is unavailable, it is set to 0. The remaining data items are required.

[5]A constant term has been included in the regressions because Kothari *et al.* (2005) find that failure to include a constant term in the Jones model increases the likelihood of misspecification errors.

3.2. *Performance-matched discretionary accruals*

As in Kothari *et al.* (2005), each sample firm is matched with a corresponding control firm based on two-digit SIC code, fiscal year, and Year -1 return-on-assets.[6] Performance-matched discretionary accruals are determined as follows:

$$\text{PMDA}_{it} = \text{SDA}_{it} - \text{CDA}_{it} \tag{5}$$

where PMDA_{it} is performance-matched discretionary accruals, SDA_{it} is the sample firm's discretionary accruals, and CDA_{it} is the control firm's discretionary accruals.

3.3. *Sample*

Investor Responsibility Research Center (IRRC) voting records and SEC proxy statements are used to identify firms that proposed ATCAs between October 1984 and December 1990.[7] Two firms that proposed ATCAs twice during the sample period were eliminated, resulting in an initial sample of 215 firms.

Table 1, Panel A shows the sample screening process. To be included in the final sample, a sample firm must have at least 10 counterparts in the same two-digit SIC code with adequate COMPUSTAT accounting data for all fiscal years within the 5-year period surrounding Year 0. This screening requirement ensures reliable estimation of the nondiscretionary accruals component using Equations (3) and (4). It excludes 15 firms, all with a two-digit SIC code of 60. SIC code 60 consists of banks and credit unions, which tend to have less available information on COMPUSTAT due to the nature of their financial statements. Also, 52 firms were excluded because adequate data were not present to estimate performance-matched discretionary accruals for all years in the 5-year period straddling Year 0. After these exclusions, the final sample consists of 148 firms. The amendment proposals passed for all but three firms in the final sample.

[6] Year -1 return-on-assets is computed as (COMPUSTAT number in parentheses): Income before extraordinary items and discontinued operations $(18)_{t-1}$/Total assets $(6)_{t-2}$.

[7] A firm is not included in the sample if it proposed an ATCA within a fiscal year ending before September 1985. It is realized that the sampling period is dated. However, this period represents the time in which ATCAs were most popular.

Table 1. Sample selection and distribution by year.

Panel A: Sample selection	
Initial sample	215
Two digit SIC code has < 10 nonsample firms with sufficient information to compute discretionary accruals (SIC 60)	(15)
	200
Not enough information on Compustat	(52)
Final sample	148
Panel B: Distribution by year	
Year	Number of firms
1984	2
1985	55
1986	35
1987	14
1988	10
1989	18
1990	14

Panel B shows the final sample's distribution by year of the ATCA proposal.

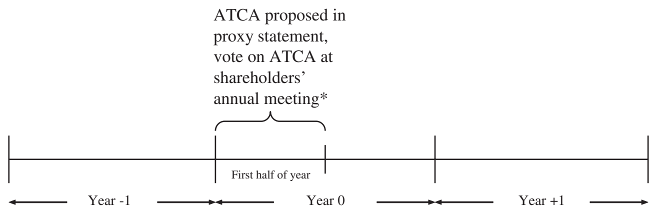
Table 1, Panel B lists the distribution of the proposal years for the sample firms. There is a significant drop in the number of observations between 1985 and 1987 because ATCA proposals became less frequent after 1985.

4. Empirical Results

The present analysis is based on the timing of the amendment proposals and their corresponding vote. Year 0 is defined as the fiscal year in which shareholders voted on the amendments. Figure 1 contains the relevant timeline. For 142 of the 148 firms in the sample, shareholders voted on the amendment proposals in the first half of Year 0. For the other six firms, the vote took place in the second half of that year.

4.1. *Firm performance*

Table 2 provides sample statistics for both unadjusted and adjusted firm performance. Return on assets is used to measure unadjusted firm performance. Return on assets (ROA) is defined as earnings divided by lagged total assets.



* The ATCA proposal and vote on the ATCA took place in the first half of Year 0 for 142 out of 148 sample firms. For the other six firms, those events took place in the second half of Year 0.

Figure 1. Timeline for 3-year window surrounding Year 0, including Year 0 events.

Table 2. Firm performance surrounding the year of the shareholder vote.

Year relative to ATCA proposal	ROA (return-on-assets)		PMROA (Performance-matched ROA)	
	Mean	Median	Mean	Median
Year −2	7.33	6.00	0.857	0.000
Year −1	6.77	6.35	0.026	0.000
Year 0	5.62	5.50	0.288	0.300
Year 1	5.20	5.05	−0.290	0.100
Year 2	5.33	5.40	1.664*	0.900**

ROA is defined as earnings divided by lagged total assets (in %). PMROA is the difference between a sample firm's ROA and the ROA of its corresponding matched control firm. A matched firm is identified based on two-digit SIC code, fiscal year, and Year –1 return-on-assets. A standard t-test (Wilcoxon signed rank test) is used to determine whether a mean (median) value is significantly different from 0.

*Significant at 0.05 (two-tailed).
**Significant at 0.01 (two-tailed).

PMROA is the adjusted firm performance measure. It is the difference between a sample firm's ROA and the ROA of its corresponding matched/control firm. A matched/control firm is identified based on two-digit SIC code, fiscal year, and Year −1 return-on-assets. As shown, return on assets declines steadily from Year −2 to Year 1. The mean ROA is 7.33% 2 years prior to the ATCA vote (Year −2) and it drops to 5.20% in the first full fiscal year after the vote (Year 1). Using a parametric t-test, the null hypothesis that this performance change is equal to 0 is rejected at the 1% significance level.[8] The mean PMROA also decreases over the same 4-year period; however, this decrease is not significant. The mean and median changes in PMROA are -1.147% ($p(t) = 0.37$) and 0.300% (signed-rank test $p = 0.66$). Therefore, although the unadjusted firm performance exhibits a significant downward momentum, its effects are somewhat attenuated by performance matching.

Nevertheless, the significant downward momentum in unadjusted firm performance could have potentially important implications on hypothesis testing. The present hypotheses are that managers can accelerate the recognition of income-increasing accruals prior to the vote (H_1) and/or postpone the recognition of income-decreasing accruals until after the vote (H_2). Since firm performance declines significantly in the years prior to the vote, it would be more difficult for managers to accelerate income-increasing accruals, creating a potential bias against H_1.

4.2. *Abnormal accruals surrounding year of shareholder vote*

Table 3 reports abnormal accruals for the full sample for the 5-year window straddling the year of the vote (Year 0). The results provide support for H_2 (negative accruals in Year 0) but fail to support H_1 (positive accruals in Year -1). The mean PMDA for Year 0 is -1.5%, which is statistically significant at the 10% level (based on a t test). The corresponding median value is -1.39%, though it is not statistically significant at conventional levels (using a Wilcoxon signed rank test). By contrast, neither the mean nor median PMDA values for Year -1 are statistically significant.[9]

The PMDAs for Year 1 and Year 2 are small and statistically insignificant, as expected. However, the PMDA for Year -2 is negative and statistically significant. The mean (median) PMDA for Year -2 is -2.07% (-1.24%), which is significant at the 10% (5%) level.

[8]Untabulated results show that over the same window, a smaller, but statistically significant, median change in ROA of -0.70% is observed. Using a nonparametric Wilcoxon signed-rank test, the null hypothesis that this median change in ROA is equal to 0 has been rejected at the 1% level.

[9]It is realized that the present finding of Year $-$ 1 abnormal accruals that are insignificant with Year 0 mean abnormal accruals that are negative and significant at the 10% level is weak. However, upcoming results show strong negative Year 0 mean and median abnormal accruals for firms that propose restrictive antitakeover charter amendments.

Table 3. Characteristics of abnormal accruals surrounding year of the shareholder vote.

Year relative to ATCA proposal			Mean		Median
Panel A: Abnormal accruals surrounding the year of the shareholder vote					
Year −2			−2.07*		−1.24**
Year −1			−0.65		−0.58
Year 0			−1.50*		−1.39
Year 1			−0.12		−0.52
Year 2			0.22		−0.06
	PMDA_{-2}	PMDA_{-1}	PMDA_0	PMDA_1	PMDA_2
Panel B: Serial correlations of abnormal accruals					
PMDA_{-2}		−0.089	0.021	0.007	0.011
PMDA_{-1}			−0.221**	−0.075	−0.168*
PMDA_0				−0.031	0.017
PMDA_1					−0.170**

Abnormal accruals are estimated using the Jones (1991) approach combined with the procedure proposed by Kothari *et al.* (2005). In Panel A, a standard t-test (Wilcoxon signed rank test) is used to determine whether a mean (median) value is significantly different from 0. Panel B reports the estimated coefficient (β) of the following regression: $\text{PMDA}_{it} = \alpha + \beta \times \text{PMDA}_{it-1} + \varepsilon_{it}$.

*Significant at 0.10 (two-tailed).

**Significant at 0.05 (two-tailed).

Abnormal accruals can be due to unexpected changes in economic reality or earnings management. One way to assess the cause is to evaluate the serial correlations of the abnormal accruals. Accruals that are due to unexpected changes in a firm's underlying economic conditions should be serially independent. On the other hand, although executives can use accounting discretion to distort earnings, such distortions are transitory. If managers postpone the recognition of income-decreasing negative accruals, then the current year's accruals and last year's accruals should be serially correlated. Alternatively, if managers accelerate the recognition of income-increasing accruals, then the current year's accruals and next year's accruals should be serially correlated.

In any given year, if earnings management causes significant negative accruals, then significant negative serial correlation in abnormal accruals between successive years should result. Therefore, to evaluate the cause of the negative accruals in Year 0 and Year −2, the serial properties of abnormal accruals have been examined in the present sample by performing the following regression:

$$\mathrm{PMDA}_{it} = \alpha + \beta * \mathrm{PMDA}_{it-\tau} + \varepsilon_{it}, \tag{6}$$

where τ is the lag. A slope coefficient of 0 indicates no serial correlation in abnormal accruals, while a negative slope indicates a reversal in accruals.

Table 3, Panel B reports our estimates (β) from Eq. (6). As shown, the strongest estimated autocorrelation coefficient is -0.221 for the $(-1, 0)$ window, and the result is statistically significant ($p = 0.015$). This autocorrelation pattern is consistent with the argument that managers of firms proposing ATCAs manipulate accruals before shareholders vote on the ATCA. By contrast, untablulated results show no correlation between Year -3 accruals and those of Year -2. Therefore, even though the PMDAs for Year -2 are negative and statistically significant, the serial correlation test gives no indication that earnings management is the cause.

The accrual results seem reasonable, given the performance characteristics of the sample firms. As noted earlier, unadjusted performance of firms in this sample is declining prior to the amendment proposals. Given the significant performance decline, it would seem more reasonable to expect to observe postponement of income-decreasing, negative accruals rather than acceleration of income-increasing, positive accruals in this sample. As a consequence, the accruals results have been interpreted as providing evidence consistent with earnings management in that managers of firms in this sample attempt to paint a better picture of their own performance prior to the shareholder vote by deferring negative accruals.

4.3. *Abnormal accruals by amendment types*

For a more refined analysis, the sample was divided into three subgroups based on ATCA type. The first subgroup includes firms proposing the more restrictive charter amendments (i.e., supermajority amendments and classified boards [SMCB]). The second subgroup includes firms proposing the less restrictive charter amendments (i.e., fair price amendments [FP]). Finally, the third subgroup includes firms proposing both types of charter amendment (i.e., supermajority amendment and/or a classified board as well as a fair price amendment [MIXED]). The number of firms in each subgroup is 64, 41, and 43, respectively.

Table 4 reports performance-matched discretionary accruals by ATCA type: more restrictive (classified board and/or supermajority [SMCB]), less restrictive (fair price [FP]), and both (MIXED). The full sample results appear to be driven by the SMCB subgroup. As with the full sample, the null hypothesis for H_1 is rejected for all subsamples across all ATCA types; the mean and median performance that matched discretionary accruals for Year –1 are either negative or insignificant. However, the results vary by subgroup for H_2. The Year 0 mean and median PMDAs for the SMCB firms are both negative (−3.73% and −3.47%, respectively) and significant (at the 1% level). By contrast, the Year 0 corresponding values for the MIXED and FP subgroups are much closer to 0 (none is farther from 0 than −1.23%) and are statistically insignificant. Moreover, the null hypothesis of equal mean and median values has been rejected across the three categories using the standard F-test ($p = 0.07$) and the Kruskal–Wallis test ($p = 0.06$), respectively. In addition, it is tested whether the mean and median Year 0 PMDA values for SMCB firms are equal to the corresponding mean and median PMDA values for FP and MIXED firms combined (results untabulated). An F-test rejects the null hypothesis of equal means ($p = 0.02$). Using a Kruskal–Wallis test, the analogous null hypothesis of equal medians ($p = 0.02$) has been rejected. These findings indicate that managers proposing restrictive ATCAs, such as supermajority and/or classified board amendments, attempt to influence the shareholder vote through earnings management, while those proposing less restrictive ATCAs do not.

5. Additional Evidence

Thus far, evidence on earnings management has been provided using annual discretionary accruals. A shortcoming of this methodology is that it provides no specific information on the sources of earnings management. Also, since annual data have been used, the results provide no evidence on the timing of earnings management. To address these issues, additional, complementary tests were performed. The results of these complementary tests are discussed in the following sections; for brevity, the results are not tabulated.

Table 4. Abnormal accruals by types of ATCAs surrounding the year of the shareholder vote.

	Mean				Median			
Year relative to ATCA proposal	SMCB ($n = 64$)	MIXED ($n = 43$)	FP ($n = 41$)	$p(F)$	SMCB ($n = 64$)	MIXED ($n = 43$)	FP ($n = 41$)	$p(\chi^2)$
Year −2	−2.05	−3.11	−1.03	0.77	−1.45*	−0.33	−2.00	0.66
Year −1	−0.78	0.32	−1.47	0.68	−0.16	0.00	−2.72*	0.37
Year 0	−3.73***	0.09	0.34	0.07	−3.47**	−1.23	0.57	0.06
Year 1	1.34	−0.00	−2.54	0.26	−0.71	0.17	−0.36	0.85
Year 2	−0.28	1.25	−0.07	0.78	0.49	0.28	−1.31	0.35

Abnormal accruals are estimated using the Jones (1991) approach and the procedure proposed by Kothari *et al.* (2005). A standard t-test (Wilcoxon signed-rank test) is used to determine whether a mean (median) value is significantly different from 0. The notation $p(F)$ indicates p-value for a standard F-test of equal means across the three groups of firms dichotomized by ATCA type. The notation $p(\chi^2)$ is the p-value for the Kruskal–Wallis test of equal medians across the three ATCA subsamples. SMCB firms propose a supermajority and/or a classified board amendment. FP firms propose only a fair price amendment. MIXED firms propose both (1) a fair price amendment and (2) a supermajority and/or a classified board amendment.

*Significant at 0.10 (two-tailed).
**Significant at 0.05 (two-tailed).
***Significant at 0.01 (two-tailed).

5.1. *Changes in nonoperating items*

The annual changes in nonoperating income (expense) have been analyzed. The goal is to explore whether the managers of firms proposing charter amendments also manipulate nonoperating items. It is assumed that nonoperating income (expense) follows a random walk process and that changes in nonoperating items are discretionary. This assumption is reasonable because nonoperating items are usually stable over time. Examples of stable nonoperating items include dividend income from a nonconsolidated subsidiary and the amortization of negative intangibles. Further, results from Ball and Watts (1972) and later research do not reject the null hypothesis that annual total net income on average follows a random walk process.[10] Nonoperating items in year t (NON_t) is defined as nonoperating income (expense) excluding interest income (COMPUSTAT item 190) divided by net sales of that year. The change in nonoperating items ($\Delta\mathrm{NON}_t$) is the difference between nonoperating items of two successive years (i.e., $\Delta\mathrm{NON}_t = \mathrm{NON}_t - \mathrm{NON}_{t-1}$). In the absence of opportunistic earnings management, no significant changes were expected in nonoperating items (i.e., average $\Delta\mathrm{NON}_t$ equal to 0).

Because the data appear to be unduly affected by outliers, the standard t-test was performed using winsorized data by setting the values in the bottom and top 1% of observations to the values of the 1st and 99th percentiles, respectively. For the sample as a whole ($n = 148$), it is found that annual changes in nonoperating items around the time of amendment proposals are small and not statistically significant, except in Year 0. For Year 0, it is found that the median value of $\Delta\mathrm{NON}$ is -0.08% of net sales. Using a nonparametric Wilcoxon signed-rank test, the null hypothesis that the median value is equal to 0 has been rejected at conventional levels ($p < 0.01$). The corresponding winsorized mean value is -0.42% of net sales and is also significant at conventional levels ($p < 0.01$). Overall, the results convey that managers delay the reporting of income-decreasing nonoperating items until after the shareholder vote on the ATCA in Year 0. These findings are in agreement with those from the discretionary accruals tests.

[10]This evidence does not refute the idea that managers may manipulate earnings (Watts and Zimmerman, 1986).

5.2. *Quarterly accruals results*

In the full sample, for 142 out of 148 firms, the shareholder vote takes place in the first 6 months of the voting year (Year 0). Therefore, quarterly data might reveal information concerning the specific timing of earnings management that annual data cannot provide. For example, if the shareholder meeting and therefore the vote on the ATCA take place in the second quarter, then negative discretionary accruals were expected for the second, third, or fourth quarter of Year 0 if management postpones income-decreasing discretionary accruals until after the vote takes place because earnings for a particular quarter are generally released during the following quarter. However, the quarterly data results can be tenuous and should be interpreted with caution. First, the Jones model is designed for annual data; there is no evidence that the Jones model works equally well for quarterly data. Second, there is a lack of information available on the quarterly COMPUSTAT files. Data limitation is quite severe in this sample as (1) the original sample is reduced by more than 50% and (2) the sample size varies from quarter to quarter due to data limitations. Nonetheless, the analyses have been repeated using quarterly data in Year −1 and Year 0.

To determine whether management postpones income-decreasing accruals until after the shareholder meeting takes place, for the quarterly accrual sample it is required that a firm's annual shareholder meeting take place: (1) in the second quarter of its fiscal year and (2) four calendar days before the first quarter earnings announcement as listed on the COMPUSTAT database or later.[11] These requirements limit the sample to firms that vote on the ATCA after the first quarter earnings are known but before the second quarter earnings are known. The remaining screens for the quarterly accrual sample are the same as the annual accrual sample. After the screening, the quarterly accrual sample consists of 69 firms. However, there are fewer observations for each quarter because, due to small sample sizes, a sample firm is not required to have enough information to compute discretionary accruals for every quarter in Year −1 and Year 0.

It is found that for the full sample, discretionary accruals are not significantly different from 0 during any quarter in Year −1 or Year 0. However,

[11] If the shareholder meeting takes place within a few days of the official COMPUSTAT first quarter earnings announcement date, it is likely that at least some preliminary indication of first quarter earnings was given during the meeting.

when the sample was divided by type of amendment (SMCB, FP, MIXED), it is found that discretionary accruals are almost always insignificant except in the second quarter of Year 0 for firms with SMCB proposals. Mean (median) abnormal accruals are -3.61% (-3.00%) and significant at the 10% (5%) level.[12] Thus, for the second quarter of Year 0, it is tested whether the mean and median PMDA values for SMCB firms are equal to the corresponding mean and median PMDA values for the combined sample of FP and MIXED firms. An F-test rejects the hypothesis of equal means ($p = 0.06$) and a Kruskal–Wallis test rejects the hypothesis of equal medians ($p = 0.08$). These results indicate that for firms in our quarterly accruals sample, managers proposing restrictive ATCAs postpone income-decreasing accruals until the second quarter earnings, which are released after the vote on the ATCA takes place. All in all, the patterns of findings from both annual accrual data and quarterly accrual data are largely consistent.

6. Conclusions

This study documents managerial influence on the corporate voting process in the context of antitakeover charter amendments. It is found that firms proposing restrictive charter amendments, such as supermajority and classified boards, tend to have negative annual discretionary accruals in the year of the shareholder vote. It is argued that this finding reveals a managerial tendency to influence corporate voting by manipulating earnings around the time of charter amendment proposals. The present interpretation is that executives of firms proposing restrictive amendments manage earnings opportunistically by deferring income-decreasing accruals in the year prior to the amendment proposal to after the shareholders have voted on the amendments. Tests using quarterly discretionary accruals confirm the present interpretation.

Acknowledgment

The authors thank Vladimir Atanasov, Mark Beasley, Jerry Curnutt, and two anonymous reviewers for helpful comments and suggestions. Also, the authors

[12]For the remaining amendment type-quarter combinations, only one other significant abnormal accrual amount has been found, which is for MIXED amendments during the second quarter of Year -1 (significant at the 10% level).

thank participants at the 2003 American Accounting Association Annual Meeting and the 2003 Financial Management Association Annual Meeting for their feedback. In addition, the authors acknowledge excellent research assistance from Nicholas Amrhein and Somit Gupta. Hoi and Wollan thank the Saunders College of Business at the Rochester Institute of Technology for summer support.

References

Ball, RJ and R Watts (1972). Some time series properties of accounting income. *The Journal of Finance*, 27, 663–682.

Bethel, JE and SL Gillan (2002). The impact of institutional and regulatory environment on shareholder voting. *Financial Management*, 31, 29–54.

Bhagat, S and RH Jefferis (1991). Voting power in the proxy process: The case of antitakeover charter amendments. *Journal of Financial Economics*, 30, 193–225.

Borokhovich, KA, KR Brunarski and R Parrino (1997). CEO contracting and antitakeover amendments. *The Journal of Finance*, 52, 1495–1517.

Brickley, JA, RC Lease and CW Smith Jr (1994). Corporate voting: Evidence from charter amendment proposals. *Journal of Corporate Finance*, 1, 5–31.

Coles, JL, M Hertzel and S Kalpathy (2006). Earnings management around employee stock option reissues. *Journal of Accounting and Economics*, 41, 173–200.

DeAngelo, H and EM Rice (1983). Antitakeover amendments and stockholder wealth. *Journal of Financial Economics*, 11, 329–360.

DeAngelo, LE (1988). Managerial competition, information costs, and corporate governance: The use of accounting performance measures in proxy contests. *Journal of Accounting and Economics*, 10, 3–36.

Erickson, M and S Wang (1999). Earnings management by acquiring firms in stock for stock mergers. *Journal of Accounting and Economics*, 27, 149–176.

Jarrell, GA and AB Poulsen (1987). Shark repellents and stock prices: The effects of antitakeover amendments since 1980. *Journal of Financial Economics*, 19, 127–168.

Johnson, MS and RP Rao. (1997). The impact of antitakeover amendments on corporate financial performance. *The Financial Review*, 32, 25–56.

Jones, JJ (1991). Earnings management during import relief investigations. *Journal of Accounting Research*, 29, 193–228.

Kothari, SP, AJ Leone and CE Wasley (2005). Performance matched discretionary accrual measures. *Journal of Accounting and Economics*, 39, 163–197.

Louis, H (2004). Earnings management and the market performance of acquiring firms. *Journal of Financial Economics*, 74, 121–148.

Perry, S and T Williams (1994). Earnings management preceding management buyout offers. *Journal of Accounting and Economics*, 18, 157–179.

Pourciau, S (1993). Earnings management and nonroutine executive changes. *Journal of Accounting and Economics*, 16, 317–336.

Stein, JC (1988). Takeover threats and managerial myopia. *Journal of Political Economy*, 96, 61–80.

Teoh, SH, I Welch and TJ Wong (1998a). Earnings management and the post-issue performance of seasoned equity offerings. *Journal of Financial Economics*, 50, 63–99.

Teoh, SH, I Welch and TJ Wong (1998b). Earnings management and the long-term market performance of initial public offerings. *Journal of Finance*, 53, 1935–1974.

Teoh, SH, TJ Wong and G Rao (1998c). Are accruals during initial public offerings opportunistic? *Review of Accounting Studies*, 3, 175–208.

Watts, RL and JL Zimmerman (1986). *Positive Accounting Theory*. Englewood Cliffs, NJ: Prentice Hall.

This page intentionally left blank

Chapter 10

Deterministic Portfolio Selection Models, Selection Bias, and an Unlikely Hero

Herbert E. Phillips
Temple University, USA

Portfolio selection models are programmed by their respective efficiency criteria to fall into a state of first-order condition love with the right sort of outliers. Nothing is changed, unfortunately, when, in general, a deterministic portfolio optimization model's inputs are stochastic rather than parametric. Distributional properties of the input estimator functions employed by four common portfolio selection models are reviewed and their solution algorithms studied in search of unique interactive effects that may mitigate the estimation error problem. Empirical and analytic support is provided for the conclusion that there is one model, an unlikely hero, that is least susceptible.

Keywords: Portfolio optimization algorithms; deterministic portfolio selection models; stochastic applications of deterministic optimization models; sampling error; selection bias.

1. Introduction

The Markowitz full covariance model was the first to provide a general framework for solving a portfolio selection problem under conditions of risk, but was, and continues to be, costly to apply. To facilitate practical applications, simplified solution procedures were subsequently developed by Sharpe, and by Elton, Gruber, and Padberg. These models, though introduced decades ago, continue to be the subject of widespread confusion as regards the framework of risk under which they were derived, and the efficacy of their solutions when applied, as is typically the case, under conditions of generalized uncertainty.

An optimization model's objective function provides a variant form of roadmap to a deterministic solution. Operating under efficiency criteria that draw no distinction between parametric signal and stochastic noise, any member of this class of deterministic optimization models may be likened to an exclusionary rule that apportions membership to an exclusive club (i.e., a solution set) among would-be applicants according to rigid criteria that reflect the objectives to be realized, while taking no account of the precision or accuracy of the information being used to rank competing candidates.

Optimization models in general, and the portfolio selection models in particular, are programmed by their respective efficiency criteria to prefer nominally efficient winners over losers, and, in effect, to fall into a state of first-order condition love with the right sort of outliers. Nothing is changed, unfortunately, when, as is the case in general, an optimization model's inputs are stochastic rather than parametric. Under these conditions that Arrow characterized as conditions of generalized uncertainty, a deterministic optimization model's solutions and forecasts must be correlated with both the signs and absolute magnitudes of estimation errors, and for this reason its solutions are subject to selection bias. The deleterious effects of this systematic bias, however, are shown in this chapter to vary widely from model to model, depending on the interactive effects of any restrictive assumptions that may be imposed, distributional properties of the input estimator functions, and, in one case, certain unique properties of the solution algorithm employed.

It is easy to see why the authors of seminal articles dealing with the derivation and/or application of normative portfolio optimization models opted to sidestep the issue of estimation error, but it is not so easy to understand why, with the passing of years and even decades, problems having to do with the stochastic nature of the inputs of these oft-cited and applied models should have received such scant attention. As no mathematically tractable solution to a stochastic optimization problem has yet been devised, or, to the author's knowledge, is in the offing, these — now aging — deterministic portfolio selection models are likely to remain in use for some time to come. It behooves us, therefore, even at this late date, to gain a better understanding of their operating characteristics when applied under conditions of generalized uncertainty — which, in a sense, is the only game in town.

Rather than assume the estimation problem away, or merely take casual note of it, this chapter attempts to break new ground: The distributional properties (i.e., error characteristics) of the input estimator functions employed by the most oft-cited portfolio selection models are carefully reviewed, and their extra mathematical assumptions are taken into account. The solution algorithms are then contrasted in search of unique properties that may mitigate the deleterious effects of estimation error. These detailed contrasts bring to light common misconceptions (plural!) about Elton, Gruber, and Padberg's single index model analog.

Empirical and analytic evidence is presented to support the chapter's main conclusion that Elton, Gruber, and Padberg's single index analog is far more general than is commonly supposed, and is stochastically more robust than competing models — that is, is less susceptible to the deleterious effects of estimation error. The Elton, Gruber, Padberg's single index model analog is called an unlikely hero because it was derived in order to solve a very different (albeit nonexistent) problem, subject to restrictive distributional assumptions that were alluded to but never imposed, and because nothing could have been further from the minds of its authors than stochastic application of the deterministic portfolio selection model — notwithstanding that, in the final analysis, this is, and has always been, the only game in town.

2. Background and Review of Previous Work

The stochastic views that Markowitz expressed in his seminal article (1952), and more recently in his Nobel lecture (1991), were decidedly Bayesian. While his mean-variance utility maxim was surely the precursor of modern finance, Markowitz did not consider, and to some extent his Bayesian nuances served to obscure, stochastic issues that arise in sample-based applications of deterministic optimization models in general, and the full covariance portfolio selection model in particular.

An optimization model's solutions are deterministic in the sense that they are uniquely determined by the inputs, in accord with the efficiency criteria. This is true under the conditions of risk that Markowitz envisaged, where the inputs are introspectively, or otherwise, known, but is equally true under conditions of generalized uncertainty, as defined by Arrow (1971), where the inputs are subject to random estimation errors.

McCall (1971) raises an interesting point in this regard by noting that there is a "widely held belief that most of the results of deterministic analysis remain basically the same when a stochastic model is employed." To understand how the optimization model's solutions differ, it is necessary to consider the nature of the interactions that take place between a deterministic optimization model's efficiency criteria, and the stochastic (i.e., sampling error) properties of the input sample estimator functions.

Blume (1971), for example, studied the performance over time of portfolios with similar sample beta coefficients, and observed that portfolio beta estimates invariably rise when estimation is based on returns in the period following the sample data selection period. An analytical explanation for Blume's observations was subsequently provided by Frankfurter *et al.* (1974), who showed them to be a logical outcome of selection bias, and demonstrated that model selection bias is a direct and inevitable result of the interaction of estimation error and a deterministic optimization model's rigid selection criteria.

Kalymon (1971) and Brown (1979) attempt to resolve the problem of estimation risk by resorting to Bayesian posterior analysis approaches, rigorously explained by Raiffa and Schlaifer (1961). Frankfurter and Lamoureux (1989, p. 181), by contrast, employ Monte Carlo techniques to dichotomize risk into estimation risk and selection risk components, and, notwithstanding selection bias, show that "portfolios selected according to the Markowitz–Sharpe maxim are superior to naive selection rules."

Elton and Gruber (1973) were not the first or last to discover that sample-based estimates of a correlation structure have very poor small sample properties. Tadjuden and Landgrebe (1999), and many others not cited here, have more recently reached the same conclusion. Elton and Gruber (1973) went further than anyone else before or since, however, by formally suggesting that a simple average of the sample correlation coefficients might provide a better indication of a correlation structure than a composite of $(n^2 - n)/2$ unique estimates. In the framework of estimation, this conclusion may be correct; but the object of the constant correlation model (CCM) that Elton *et al.* (1976) subsequently derived on the basis of this assertion is decision, and not estimation.

Sharpe's (1963) single index model (SIM) was intended to facilitate practical applications of the Markowitz full covariance model (FCM). Elton *et al.* (1976), by contrast, arguing that the FCM and SIM quadratic programming portfolio selection approaches were needlessly formalistic, introduced two alternative models with solution algorithms based on simple security-ranking procedures: CCM was offered as a substitute for FCM, notwithstanding that CCM's distributional assumption has no parametric equivalent; and a single index model analog (SIM*) was offered as a substitute for SIM, notwithstanding that SIM's distributional constraints were never actually imposed. These points are taken up in detail in the sections that follow.

The preoccupation of Sharpe (1963) and Elton *et al.* (1976) on the development of "simplified" models, that merely mimic FCM solutions under the same assumed conditions of risk, rather than address the problem of applications under generalized uncertainty, effectively buried stochastic optimization issue under successive layers of sometimes irrelevant rhetoric. As the declared purpose of each "simplified" model is simply to facilitate solution of the same portfolio selection problem that FCM addresses under the same assumed conditions of risk, it seemed to follow that FCM should be most robust and the standard against which other models should be compared, ranked, and judged. Most previous researchers missed the point, however, that stochastic inputs (i.e., sample estimates) served in place of known parameters in sample-based applications, and, thus, that this and other deterministic logic may not necessarily apply where the deterministic models are applied under conditions of generalized uncertainty.

Previous researchers, by contrast, have tended to disregard the sampling error issue altogether, and to wrongly assume that FCM is a uniquely robust model against which other models should be ranked and judged. In an early article, for example, Cohen and Pogue (1967) used FCM and SIM to plot nominally efficient frontiers on the same set of coordinate axes. Finding the curves to plot in close proximity, but with no knowledge of the security composition of the portfolios represented by the points that plotted along either frontier, Cohen and Pogue concluded that the SIM solutions must provide fair approximations of the FCM solutions. Burgess and Bey (1988, p. 162) have more recently argued, as Markowitz (1959) did long before, that FCM is difficult and expensive to apply in the case of a large universe of stocks. With scant justification, they argued that the FCM and CCM solutions must be sufficiently similar to allow CCM to be used as a filter for "screening sets of securities that would be contained in the Markowitz portfolio."

The authors of most such published works, like Cohen and Pogue (1967) and Burgess and Bey (1988), seem to suffer from a common misconception: A contrast of points that plot along a particular efficient frontier, or different frontiers, may serve to convey information about differential portfolio risk-rate of return tradeoffs, but tells nothing about the security composition or relative weightings of the portfolios that, more often than not, is really the subject of inference.

Two previous attempts have been made by the present author to correct this common misconception. Using monthly rate of return data for 800 stocks covering the period from December 1980 to January 1989 in the author's 1993 *RQFA* article (Phillips, 1993), nominally efficient FCM, SIM, SIM*, and CCM solutions corresponding to a number of prior-selected levels of *ex ante* risk were identified. Unlike previous studies, a record was kept of the securities actually contained in each nominally efficient solution set. The empirical results reported in this chapter show that, in sample-based applications, the solutions obtained by these oft-cited deterministic portfolio selection models that are supposed to solve the very same portfolio optimization problem, result in stock selection and portfolio weighting schemes that differ markedly from model to model, and that the differences are more striking at lower levels of *ex ante* portfolio risk (i.e., more highly diversified portfolios) than at higher levels.

These *ex ante* results were replicated in a subsequent *RQFA* article by Frankfurter *et al.* (1999), and then extended to consider *ex post* implications. The analysis and conclusions published in this article, unfortunately, while correct, was incomplete. As SIM* seemed such an unlikely hero, for example, the authors failed to take proper note of the fact (Frankfurter *et al.*, 1999, Table 3.1, p. 360) that in every subsample period featured in the study, given monthly *ex ante* target rate of return equal to or less than 2.5% (roughly equivalent of an annualized target return of 30%), SIM*'s *ex post* performance satisfied the *ex ante* target rate of return objectives better, and more consistently, than CCM, FCM, or SIM.
Moreover, while noting that:

> To the extent that a portfolio selection model succeeds at identifying portfolios that reduce investment risk through diversification, and, in the process, outperforms an index portfolio, its success is probably best explained in terms of the way that the model, incorporates, and exploits statistical covariation. Frankfurter *et al.* (1999, p. 365).

The authors failed to adequately exploit this observation in explaining out their own published empirical results.

The present chapter was initially motivated by a desire to report important implications suggested by the empirical results presented in the Frankfurter *et al.* (1999) article, but overlooked. In the process of cataloguing this materials however, strong analytic justifications were first called into view and then rigorously developed. These derivations, systematically developed in the sections that follow, are based on a contrast of well-known (and some not so well-known) parametric models and their stochastic counterparts.

3. The Normative Portfolio Selection Models: A Brief Outline of Parametric Forms

The parametric forms of a number of oft-cited and supposedly well-known models are introduced in this section, and some overlooked details and common misconceptions are uncovered.

3.1. *The full covariance model*

Markowitz (1952) represents a mean-variance utility maximizer's portfolio selection problem under conditions of risk as a search for a set of optimal security assignments, $\{X_{\mathrm{p}}|\lambda\}$, by constrained optimization of an objective function:

$$\Phi = \sigma_{\mathrm{p}}^2 - \lambda\mu_{\mathrm{p}} \tag{1}$$

subject to budget and trading constraints, where $\lambda_{\mathrm{p}} = \partial\mu_{\mathrm{p}}/\partial\sigma_{\mathrm{p}}^2$ is a risk/return tradeoff,

$$\mu_{\mathrm{p}} = \sum_{i=1}^{n} x_i\mu_i \tag{2}$$

is a parametric portfolio mean,

$$\sigma_{\mathrm{p}}^2 = \sum_{i=1}^{n} x_i^2\sigma_i^2 + \sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n} x_i x_j \sigma_{ij} \tag{3}$$

is a parametric variance, σ_{ij}, $i \neq j$, is a covariance term, and where $X_{\mathrm{p}} = [x_1, \ldots, x_n]$ is a vector of security weights.

Practical applications of the Markowitz full covariance model (FCM) were, and to this day are, complicated by the fact that each iterative step in the solution process involves inversion of a full and unrestricted variance–covariance matrix of order n, where n may be very large. To facilitate practical applications under assumed conditions of risk, Markowitz (1959, pp. 100–101) suggested, and Sharpe (1963) later worked out the details of an alternative solution approach based on a simple, single index, regression structure.

3.2. *The single index model*

The major characteristic of Sharpe's single-index model (SIM) is the assumption that security rates of return are independently distributed linear functions:

$$R_{it} = \alpha_i + \beta_i R_{\mathrm{m}t} + \epsilon_{it}, \quad \forall_{it} \tag{4}$$

of a single underlying common factor, R_{m}, where α_i and β_i are parametric regression coefficients, and ϵ_{it} is a residual relative to the ith regression function. It is formally and explicitly, assumed that, for any pair of regression functions, the residuals $\{\epsilon_{it} \cap \epsilon_{jt}\}$ are cross-sectionally uncorrelated, and thus:

$$\sigma(\epsilon_i, \epsilon_j) = 0, \quad \forall_{i \neq j} \tag{5}$$

SIM's portfolio mean in parametric form:

$$\mu_{\mathrm{p}} = \sum_{i=1}^{n} x_i \alpha_i + \left(\sum_{i=1}^{n} x_i \beta_i \right) \mu_{\mathrm{m}} \tag{6}$$

is derived by substituting variable definitions from Equation (4) in Equation (2). In this parametric representation, the mathematical equivalent of FCM's variance definition in Equation (3):

$$\sigma_{\mathrm{p}}^2 = \sum_{i=1}^{n} x_i^2 \sigma(\epsilon_i)^2 + \left(\sum_{i=1}^{n} x_i \beta_i \right)^2 \sigma_{\mathrm{m}}^2 + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} x_i x_j \beta_i \beta_j \sigma_{\mathrm{m}}^2 \tag{7}$$

may be obtained by similar substitutions, where:

$$\begin{aligned} \sigma_{ij} &= \sigma(\epsilon_i)^2, \quad &\forall_{i=j} \\ \sigma_{ij} &= \beta_i \beta_j \, \sigma_{\mathrm{m}}^2, \quad &\forall_{i \neq j} \end{aligned}$$

and σ_{m}^2 is the variance of the independent variable in Equation (4) — typically rates of return on a broad stock market index.

Sharpe (1963, p. 283) does not merely assume "that the returns of various securities are related only through common relationships with the same basic underlying factor," from which the condition outline in Equation (5) would follow, but, in order to obtain a diagonalized form of variance–covariance matrix, he imposes that condition by setting all terms involving $\sigma_{ij} = \beta_i \beta_j \, \sigma_{\mathrm{m}}^2$,

$i \neq j$, in Equation (7) to 0 in SIM's portfolio variance definition:

$$\sigma_{\mathrm{p}}^2 = \sum_{i=1}^{n} x_i^2 \sigma(\epsilon_i)^2 + \left(\sum_{i=1}^{n} x_i \beta_i\right)^2 \sigma_{\mathrm{m}}^2 \tag{8}$$

3.3. *The Elton, Gruber, and Padberg security-ranking approaches*

3.3.1. *The single index model analog, a left-out detail, and common misconceptions*

Elton *et al.* (1976), arguing that the FCM and SIM quadratic programming portfolio selection approaches are needlessly formalistic, introduce two alternative models that employ security-ranking procedures. One approach (SIM*) was intended merely to mimic SIM's solutions. SIM* uses the same variable definitions as SIM, outlined in Equation (4) through (7) above, but modifies the solution logic in two important respects. First, rather than a search for a set of security weights, $\{X_{\mathrm{p}}|\lambda\}$, to maximize a quadratic objective function Equation (1), SIM* instead searches for a set of security weights to maximize:

$$\delta_{\mathrm{p}} = (\mu_{\mathrm{p}} - R)/\sigma_{\mathrm{p}} \tag{9}$$

where R may be a risk-free rate assumption or operator used to vary the slope of a tangent line — which is tantamount to a first-order condition. Second, SIM*'s risk definition:

$$\sigma_{\mathrm{p}} = \left[\sum_{i=1}^{n} x_i^2 \sigma(\epsilon_i)^2 + \left(\sum_{i=1}^{n} x_i \beta_i\right)^2 \sigma_{\mathrm{m}}^2 + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} x_i x_j \beta_i \beta_j \, \sigma_{\mathrm{m}}^2\right]^{\frac{1}{2}} \tag{10}$$

is modeled according to Equation (7) rather than Equation (8).

Sharpe's primary purpose in deriving his diagonal model, which later became more commonly known as SIM, was computational efficiency. Sharpe (1963, p. 287) accomplished this by imposing the restrictive distributional assumption outlined by Equation (5), and thus by replacing FCM's unrestricted, full form, variance–covariance matrix with a diagonal form matrix. The quadratic programming optimization algorithm employed by FCM and

SIM are the same, but the matrix inversions that take place in each iteration require less computer memory and execution time when the off-diagonal elements are set to 0 or simply ignored.

By setting all terms involving $\beta_i \beta_j \, \sigma_{\mathrm{m}}^2, i \neq j$, in Equation (7) to 0, Sharpe effectively restricted SIM's valid applications to the case where the residuals in Equation (4) are, in fact, cross-sectionally uncorrelated. In an often overlooked caveat included his concluding remarks, Sharpe (1963, p. 91) was very clear about this:

> The assumptions of the diagonal model lie near one end of the spectrum of possible assumptions about the relationships among securities. The model's extreme simplicity enables the investigator to perform a portfolio analysis at very low cost .... However, it is entirely possible that this simplicity so restricts the security analyst to making his predictions that the value of the resulting portfolios analysis is also very small.

Elton *et al.* (1976, p. 1342) may have had this passage in mind when drafting SIM*'s introductory, which reads, in part, as follows:

> ... we shall assume that the standard single index model is an accurate description of reality.... The assumption implied ... is that only joint movement between securities comes about because of the common response to a market index.

These introductory remarks, unfortunately, which were repeated in kind in their 1978 article, left an impression, widely held to this day, that SIM* does, in fact, impose the same distributional restriction as SIM — as outlined in Equation (5). Contrasting Equations (7), (8), and (10), however, it is seen that this is not the case. Indeed, a detailed review of Elton *et al.*'s SIM* derivations (1976, pp. 1342–1348) would reveal that in no respect are their results dependent on the condition that $\beta_i \beta_j \, \sigma_{\mathrm{m}}^2 = 0, \forall_{i \neq j}$.

As SIM* does not employ mathematical programming methods and no matrix inversion operations are involved, it is easy to understand that Elton, Gruber, and Padberg had no reason to be concerned about the form of the variance–covariance matrix, and no need to force the issue by setting all terms involving $\beta_i \beta_j \, \sigma_{\mathrm{m}}^2$, $i \neq j$, in Equation (7) to 0 — as Sharpe found it necessary to do. It is not so easy to understand, however, why Elton, Gruber, and Padberg should have overlooked this point. In fact, the only thing that SIM and SIM* have in common is that variable definitions employed by each are derived in the context of regression, assuming a simple bivariate regression structure such as outlined by Equation (4), rather than in the more general framework of covariance. There is no condition, explicitly imposed by SIM*, however, that requires the residuals in Equation (4) to be cross-sectionally uncorrelated. Notwithstanding a commonly held misconception, that clearly reflects Elton, Gruber, and Padberg's erroneous indications to the contrary: FCM and SIM*'s parametric forms are perfectly compatible under the most general assumption that $\sigma(\epsilon_i, \epsilon_j) \neq 0, \exists_{i \neq j}$.

3.3.2. *The constant correlation model*

Elton *et al.* (1976) derived a second model that they called the constant correlation model (CCM). CCM uses the same portfolio mean definition as FCM, shown in Equation (2), but a different measure of risk. Substituting from an elementary variable definition, $\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$, in Equation (3), and exploiting the assumption that "all pairwise correlation coefficients are equal" (*ibid.*), one obtains:

$$\sigma_{\mathrm{p}}^2 = \sum_{i=1}^{n} x_i^2 \sigma_i^2 + \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} x_i x_j \sigma_i \sigma_j \bar{\rho} \tag{11}$$

where

$$\bar{\rho} = \left[\left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij} \right\} - n \right] \Big/ (n^2 - n) \tag{12}$$

is a simple weighted average of the off-diagonal elements of the correlation matrix.

Elton and Gruber (1973) were not the first or last to discover that a sample-based estimates of a correlation structure have very poor small sample properties, but went further than anyone else before or since by formally suggesting that a simple average of the sample correlation coefficients, as described by Equation (12), might provide a better indication of a correlation structure than a composite of $(n^2 - n)/2$ unique estimates. In the framework of estimation, this conclusion may be correct. Estimation is not the central objective of normative portfolio theory, however; the purpose of normative theory is to serve as a guide to action, not estimation.

The correlation structure is key to explaining how a parametric model diversifies under conditions of risk, but is also key to explaining how a deterministic portfolio optimization model may go wrong when its assumptions about the nature, form, and composition of the covariance or correlation matrix are wrong — as when, under conditions of generalized uncertainty, stochastic inputs miscarried as known parameters. The chaotic behavior of standard statistical procedures for estimating a correlation structure that Elton and Gruber (1973) observed (and which are more rigorously explained below), for example, would not prevent FCM from finding correlation effects to exploit in sample space, even if none existed in parameter space. In sample-based applications, would CCM fare any better? Obviously not.

Even in the unlikely event that the true, but unknown, parameters $\rho_{ij} = 0\,\forall_{i\neq j}$: in sample space, the probability that any, much less that all, the sample correlation coefficients would be equal to 0 is 0. Thus, even in the unlikely event that $\rho_{ij} = 0\ \forall_{i\neq j}$, as in the general case where $\rho_{ij} \neq 0\ \exists_{i\neq j}$, the average sample correlation coefficient will be affected by sampling error and by outliers. As such, CCM may fare even worse than FCM because the very same average estimation error is replicated $(n^2 - n)$ times as the average sample correlation coefficient is substituted for each of the $(n^2 - n)$ off-diagonal elements of a sample correlation matrix. As the sampling distributions of the correlation coefficients are skewed to the right, moreover, the sampling errors cannot average out, but, to the contrary, must be additive; that is, the average sampling error must be a linear function of the number of stocks under consideration — and thus matrix order, n, of the correlation matrix.

4. On the Deleterious Effects of Estimation Error in Sample-Based Applications

In sample-based applications, estimates produced by one or more stochastic estimation processes replace the known parameters represented in each equation, Equation (1) through (12), above. In the context of unbiased estimation, purely random estimation error averages out. In the framework of optimization, on the other hand — whether accomplished by quadratic programming, Lagrangian multipliers, or security rankings — the so-called law of large numbers cannot apply. In each iteration of an optimization model's deterministic solution process, inputs, of whatever kind, are systematically reviewed by the program as it searches for relationships that, individually or in combination, contribute best to achieving the algorithm's objectives in that iteration. As the deterministic efficiency criteria make no distinction between underlying parametric effects and purely random effects, however, random estimation error does not average out, but is, rather, a primary source of systematic selection bias.

4.1. *Optimization model selection bias*

Suppose, for example, that a stock expected rate of return maximizer plans to select a single stock, subscripted i or j depending on the method of selection, from each of K different industrial groupings, and that once selected, the stocks are to be combined into an equally weighted portfolio, where $x_k = 1/K, \forall_k$. Suppose further that two alternative stock selection strategies are under active consideration: the first is an expected rate of return optimizing approach and the second is a purely random approach.

If the stocks in each group k are ranked according to their historical average rates of return, $E(R_{ik})$, where $E(R_{ik}) \sim N(\mu_{ik}, \sigma_{ik}^2/T)$, and a single "winner" from each group is selected for inclusion in an equally weighted portfolio on this basis, then the expected rate of return optimizing strategy's portfolio mean is defined and denoted as follows:

$$E(R_{\text{opt}}) = \sum_{k=1}^{K} E(R_{ik})/K$$

If, on the other hand, a single stock j is selected at random from each group k, then the random stock selection strategy's portfolio mean is defined and written as follows:

$$E(R_{\text{ran}}) = \sum_{k=1}^{K} E(R_{jk})/K$$

The random strategy may not be very scientific, but nevertheless $E(R_{\text{ran}})$ is a clearly defined and unbiased estimator — albeit uninteresting — of an unknown parametric mean, μ_{ran}. The expected rate of return optimizing strategy, by contrast, which is more appealing at first glance, carries with it at least one unfortunate implication. To see this implication, let $E(R_{ik}) = \mu_{ik} + \delta_{ik}$, where $\delta_{ik} \sim N(0, \sigma_{ik}^2/T)$, $k = 1, K$ is a random error. Regardless of degrees of freedom, the larger is the parametric variance, σ_{ik}^2, of the stock selected from group k, the larger the absolute value of the "winner's" error, $|\delta_{ik}|$, is likely to be. All other things equal, therefore, this optimizing approach, like any other, will avoid any stock i in group k, when $\delta_{ik} < 0$, favor it when $\delta_{ik} > 0$, and fall into a state of first-order condition love with any stock i, in any group, $k = 1, \ldots, K$, when, simultaneously, σ_{ik}^2 is large and $\delta_{ik} > 0$.

The deterministic portfolio selection models currently available suffer from the same amorous tendencies especially in the presence of right sort of outlier(s), but their tendencies toward selection bias are multidimensional, and therefore the linkage is more complex than the situation illustrated in this simple unidimensional example. This simple example is sufficient to show, however, that, under conditions of generalized uncertainly, an optimization model's selections and its predictions must be correlated with both the signs and absolute magnitudes of estimation errors. The deleterious effects of this selection bias are not uniform across the various models however, as will be seen below.

4.2. *Lessons from a previous study*

The empirical results reported in the author's 1999 *RQFA* article (Frankfurter *et al.* 1999) are based on an unusually rich, multi-period, sample design that would be extremely difficult and costly to duplicate. Monthly rates of return data were obtained from the Center for Research in Security Prices (CRSP) NYSE-AMEX and NASDAQ tapes for the period from January 1964 through March 1994. The sample was first divided in to three 120-month subsample periods ending December 1973, 1983, and 1993, respectively, which were used for *ex ante* sample estimations in the experiment, each followed by one additional monthly observation that was used for *ex post* evaluations in January, following the last December observation. Table 1 provides a summary of a portion of the empirical results reported in the author's study, reordered here to better illustrate the points at issue in this chapter. As pointed out above, the author's previous review of these empirical results was incomplete. The findings and interpretations reported in this section, and their analytic explanations that follow, where overlooked.

Selected target expected rates of return are shown in Column 1 of Table 1. The remaining columns are then divided into three blocks, each representing one of the three, nonoverlapping, subsample periods; empirical results obtained during the first subsample period are reported in Block 1, those

Table 1. Number of stocks included in *ex ante* efficient solution sets

	1964–1973				1974–1983				1984–1993			
$E(R_p)$	CCM	FCM	SIM*	SIM	CCM	FCM	SIM*	SIM	CCM	FCM	SIM*	SIM
4.00	1	1	1	1	20	18	44	44	21	30	47	51
3.50	4	2	6	5	25	22	46	47	27	74	67	74
3.00	9	7	11	12	23	23	52	53	34	48	84	95
2.50	14	14	23	29	24	24	62	63	45	59	96	114
2.00	16	21	34	41	22	27	61	62	49	66	121	138
1.50	21	22	44	49	20	33	75	75	57	61	117	135
1.00		25		50		25				66		119
0.50		27		39								

obtained during the second subsample period are reported in Block 2, and the results obtained for the third subsample period are reported in Block 3. Each block contains four columns, headed by CCM, FCM, SIM*, and SIM, respectively. For each target rate of return shown in Column 1, and for each subsample period represented by a unique block in the table, the number of stocks, N_{CCM}, N_{FCMI}, $N_{\mathrm{SIM}*}$, and N_{SIM} contained, respectively, in the CCM, FCM, SIM*, and SIM efficient solution sets are recorded in the appropriate blocks, columns, and rows of the table.

4.2.1. *How the models diversify*

Points relatively high on an efficient frontier are associated with relatively high target expected returns in the table, and less risky — more diversified — portfolios are represented by lower target portfolio rates of return in the table. Starting at relatively high feasible target rate of return, that corresponds to a point relatively high along some efficient frontier, as a model attempts to diversify it searches for portfolio composition and security-weighting combinations that optimize a risk/return tradeoff. It is seen from the table, however, that in no subsample period considered do the models accomplish these tradeoffs in precisely the same way. That is, as the models diversify, accepting lower target rates of return as the price of risk reduction, a systematic relationship emerges between the target rates of return shown in Column 1 of the table, and number of stocks contained in nominally efficient CCM, FCM, SIM*, and SIM solution sets.

4.2.2. *Systematic relationships between target rate of return and portfolio size*

This systematic relationship between portfolio target rate of return and portfolio size can be seen by reading across the rows of the table that correspond to reasonably well-diversified portfolios. For every target monthly rate of return equal to 2.5% (roughly equivalent to a 30% annualized rate of return) or less in the table, for example, it is seen that: $N_{\text{CCM}} \leq N_{\text{FCM}} < N_{\text{SIM}^*} < N_{\text{SIM.}}$. If CCM were excluded from consideration, moreover, then the even stronger result would be: $N_{\text{FCM}} < N_{\text{SIM}^*} < N_{\text{SIM.}}$ for every target monthly return equal of 3% or less in the table.

4.2.3. *Diversification opportunities seen and not seen*

Contrasting the relative sizes of the CCM and FCM nominally efficient portfolios for target monthly rates of return, reported in Column 1 of the table, equal to or less than 2%, it is seen that CCM is able to achieve any particular target rate of return with fewer stocks than the corresponding FCM solution in that block requires. In every subsample period, therefore, CCM, it would appear, is systematically able to find nominally efficient diversification opportunities that FCM cannot find. For any target rate of return equal to or less than 3%, by contrast, the data shown in each block of the table suggest that FCM is systematically able to find nominally efficient diversification opportunities that SIM* does not recognize, and that SIM*, in turn, is able to find nominally efficient opportunities that SIM does not see. To what do we owe one model's systematic ability to find nominally efficient, risk-reducing, diversification opportunities (real or imagined) that another model cannot find?

4.3. ***Econometric logic***

It will be seen in this section that how a portfolio selection model diversifies under conditions of generalized uncertainty is determined, at least in part, by three primary factors: (1) statistical properties of the sample variance–covariance (or equivalently, correlation) estimates, (2) whether or not the sample covariance and/or correlation information is considered, and, if account is taken, (3) how the information is incorporated into the model's solution algorithm.

4.3.1. *The full covariance model*

FCM diversifies by searching for securities to include in an investment portfolio that are less than perfectly correlated. Thus, the covariance effect, while not acting alone, is key to explaining FCM's analytic operations, security selections, and real or imagined ability to diversify.

In sample-based applications, FCM's sample portfolio mean:

$$E(R_{\mathrm{p}}) = \sum_{i=1}^{n} x_i E(R_i) \tag{13}$$

and sample portfolio variance:

$$\mathrm{Var}(R_{\mathrm{p}}) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \mathrm{Cov}(R_i, R_j) \tag{14}$$

operators are obtained by replacing the parametric variable definitions shown in Equations (2) and (3) with sample theory definitions, where $E(R_i)$, is a sample mean and $\mathrm{Cov}(R_i, R_j)$ is an element of an n-by-n sample variance–covariance matrix $\boldsymbol{C}$. Regardless of the distributional form of actual rate of return distributions, $f(R_i)$, it is known from the central limit theorem that $E(R_i) \sim N\ (\mu_i, \sigma_i^2/df)$. The sampling distribution of a sample variance–covariance matrix $\boldsymbol{C}$, on the other hand, as explained by Johnson and Wichern (1988, p. 143), is described by a Wishart distribution, formally; $\boldsymbol{C} \sim \boldsymbol{W}_{\mathrm{m}}(\cdot|\boldsymbol{\Sigma})$.

The sample properties of a Wishart distribution are poor in general, and are especially so when the sample size is small. But how small is small? Johnson and Wichern (1988, p. 143) show that a "density does not exist unless the sample size . . . is greater than the number of variables." The empirical results reported in Table 1, for example, are based on samples of 120 monthly rate of return observations on each of 790, 1,467, or 2,212 stocks, depending on subsample period. In this empirical work, as in general, it follows, therefore, that there is no mathematically tractable way to mark percentiles of the relevant covariance distributions. All it is known is that $\boldsymbol{C} \sim \boldsymbol{W}_{\mathrm{m}}(\cdot|\boldsymbol{\Sigma})$.

Some conclusions about the operating characteristics of the statistical generating functions can be drawn, however, from the empirical results reported in Table 1, and from Elton and Gruber's (1973) previous observation that a simple average of sample correlation coefficients may provide a better indication of a correlation structure than a composite of $(n^2 - n)/2$ individual estimates.

FCM's sample portfolio variance operator in Equation (14) is simply a weighted average taken over all the elements of a sample variance–covariance matrix, $\boldsymbol{C} \sim \boldsymbol{W}_{\mathrm{m}}(\cdot|\boldsymbol{\Sigma})$. Elton and Gruber (1973) confirm, and Tadjuden and Landgrebe (1999) more rigorously show, that a Wishart distributed random variable behaves erratically. Thus, it should be expected that any sample variance–covariance matrix to contain many, and sometimes extreme, covariance outliers. It should come as no surprise to learn, therefore, that FCM might find covariance relationships to exploit in a sample variance–covariance matrix, $\boldsymbol{C}$, even if none actually exist in the parametric variance–covariance matrix $\boldsymbol{\Sigma}$ — and will find covariance outliers to fall in love with in any event. The data reported in Table 1 suggest that FCM systematically finds sample covariance effects to exploit that SIM* and SIM, for reasons explained below, do not see.

4.3.2. *The constant correlation model*

CCM goes one step further than FCM by replacing all pairwise sample correlation coefficients with a single weighted average sample correlation coefficient obtained by replacing the parametric coefficients, $\rho_{ij} = \sigma_{ij}/\sigma_i\sigma_j$, $i \neq j$, on the right-hand side of Equation (12) with sample correlation coefficients, r_{ij}. The sampling distribution of a single correlation coefficient can be known, however, as it will be seen below, only when the parameter ρ_{ij} is known. It follows, therefore, that the sampling distribution of a weighted average sample correlation coefficient can be known only when all the parametric coefficients, ρ_{ij}, are known. For this reason alone, it would be impossible to describe the sampling distribution of this weighted average; but to make matters worse, such a distribution probably does not exist.

The sampling distribution of a correlation coefficient resembles a normal distribution in the large sample case, but must differ from normal because the random variable is bounded by -1 and $+1$. As ρ_{ij} approaches -1 or $+1$, on the other hand, the variance of the sampling distribution of a sample correlation coefficient, r_{ij}, must approach 0. Conversely, as the underlying parameter ρ_{ij} departs from -1 or $+1$, and heads toward 0, the variance of the sampling distribution of r_{ij} must increase. Substituting sample correlation terms, r_{ij}, into Equation (12), therefore, which yields:

$$\boldsymbol{r} = \left[\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}\right\} - n\right] / (n^2 - n) \tag{15}$$

represents the equal pooling on $(n^2 - n)/2$ continuous random variables, any two of which can have the same variance only with probability 0.

There is no basis in statistics to justify this pooling operation, and no mathematically tractable way to explain the statistical properties of the outcome. It can be reasoned, however, that as the size of a correlation matrix grows from a 790 by 790 for the first subsample period, to 1,467 by 1,467, and finally to 2,212 by 2,212, the homoscedasticity assumption that would be necessary to justify such a pooling operation must be subject to greater and greater violation. It can also be reasoned that, in any subsample period, as CCM attempts to diversify by bringing more and more stocks into solution, the cumulative effect of this error of statistical logic must increase.

Contrasting the relative sizes of the CCM and FCM solutions sets shown in Table 1, it can be seen that the data would be perfectly consistent with the preposition that CCM's statistical delusions are affected by both population size and the relative size of a solution set. It is impossible to prove that this will always be the case, but it seems reasonable that it should be. It also seems reasonable to conclude, at this point, that no justification can be found for the statistical operations employed by CCM, or for the model's application.

4.3.3. *Sharpe's single index model*

SIM lies at the opposite extreme. Sharpe, for reasons that had nothing to do with estimation risk or any other statistical nicety, sought to diagonalize the variance–covariance matrix — which he accomplished by brute force. With all off-diagonal elements of the sample variance–covariance matrix forced to 0 — by way of contrast with both FCM and SIM* — SIM could not find any covariance relationships to exploit even if $\boldsymbol{\Sigma}$ was rich with them. It is little wonder, therefore, that, in Table 1, FCM and SIM* are systematically able to find diversification opportunities to exploit that SIM does not find.

SIM's sample portfolio mean:

$$E(R_\mathrm{p}) = \sum_{i=1}^{n} x_i a_i + \left(\sum_{i=1}^{n} x_i b_i\right) E(R_\mathrm{m}) \tag{16}$$

and variance:

$$\mathrm{Var}(R_{\mathrm{p}}) = \sum_{i=1}^{n} x_i^2 \mathrm{Var}(\mathrm{e}_i) + \left(\sum_{i=1}^{n} x_i b_i\right)^2 \mathrm{Var}(R_{\mathrm{m}}) \tag{17}$$

are obtained by replacing parametric variable definitions in Equations (6) and (8), respectively, with sample theory definitions, where a_i and b_i are normally distributed estimated regression coefficients, and the residual and systematic variances, $\mathrm{Var}(\mathrm{e}_i)$ and $\mathrm{Var}(R_{\mathrm{m}})$, respectively, are distributed as χ^2/df. As noted under Equation (7), however, consistent with SIM's restrictive distributional assumption that $\sigma(\epsilon_i, \epsilon_j) = 0$, $\forall_{i \neq j}$, all covariance terms, $\sum_{i=1}^{n} \sum_{j=1}^{n} b_i b_j \mathrm{Var}(R_{\mathrm{m}})$, $i \neq j$, in Equation (17) are set to 0.

By forcing the off-diagonal elements of the variance–covariance matrix to 0, Sharpe (1963) left SIM with no covariance effects to exploit, and, thus, with no other diversification, risk reducing options but application of the law of large numbers. To this effect, though in a somewhat different context, Blume (1971) shows that as the number of stocks included in a portfolio, n, grows without bounds, its diversifiable risk, $\sum_{i=1}^{n} x_i^2 \mathrm{Var}(\mathrm{e}_i)$, must be asymptotic to 0. Contrasting the number of stocks in SIM solutions with the number of stocks contained in solutions obtained by other models at corresponding points along their respective efficient frontiers, it can be seen from Table 1 that, except for the case where $N_{\mathrm{CCM}} = N_{\mathrm{FCM}} = N_{\mathrm{SIM}^*} = N_{\mathrm{SIM.}} = 1$, and in just one other case where $N_{\mathrm{SIM}^*} > N_{\mathrm{SIM.}}$, $N_{\mathrm{SIM.}}$ is larger N_{CCM}, N_{FCM}, or N_{SIM^*} in any row of the table. In many cases, the differences are striking.

The single index analog. SIM*'s sample portfolio mean operator is the same as SIM's, described in Equation (16), but its portfolio standard deviation estimator:

$$SD(R_{\mathrm{p}}) = \left[\sum_{i=1}^{n} x_i^2 \mathrm{Var}(\mathrm{e}_i) + \left(\sum_{i=1}^{n} x_i b_i\right)^2 \mathrm{Var}(R_{\mathrm{m}}) + \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} x_i x_j b_i b_j \mathrm{Var}(R_{\mathrm{m}})\right]^{\frac{1}{2}} \tag{18}$$

is obtained by substituting sample variable definitions in Equation (10) rather than in Equation (8). The residual and systematic variance terms, $\mathrm{Var}(e_i)$ and

$\mathrm{Var}(R_\mathrm{m})$, respectively, are, however, distributed as χ^2/df as in Equation (17). Where the ith element, $i = 1, n$, on the main diagonal of a sample variance–covariance matrix $\boldsymbol{C}$ is represented by $\mathrm{Var}(\mathrm{e}_i)$, element $i = n + 1$ on the main diagonal is represented by $\mathrm{Var}(R_\mathrm{m})$, and any off-diagonal element $i \neq j$ is represented by $\mathrm{Cov}(R_i, R_j) = b_i b_j \mathrm{Var}(R_\mathrm{m})$, it follows from explanations provided by Kotz and Johnson (1988, p. 142) that $\boldsymbol{C} \sim \boldsymbol{W}_\mathrm{m}(\cdot|\boldsymbol{\Sigma})$. If $\boldsymbol{C}$ is distributed as Wishart, what is SIM*'s advantage relative to FCM?

In sample-based applications, FCM searches for a set of optimal security assignments, $\{X_\mathrm{p}|\lambda\}$, by constrained optimization of an objective function:

$$\Phi = \mathrm{Var}(R_\mathrm{p}) - \lambda E(R_\mathrm{p}) \tag{19}$$

obtained by substituting sample variable definitions for the parametric representations in Equation (1). Thus, in each iteration of the model, FCM operates directly on the terms of a sample variance–covariance matrix. In sample space applications, by contrast, SIM* searches to find a set of security weight assignments to maximize the sample-based equivalent of Equation (9):

$$\delta_\mathrm{p} = (E(R_\mathrm{p}) - R)/SD(R_\mathrm{p}) \tag{20}$$

To understand SIM*'s fortuitous advantage, therefore, it is necessary to review the method by which this is accomplished.

SIM* ranks $i = 1, \ldots, n$ securities according to the operator:

$$Z_i = [b_i/\mathrm{Var}(\mathrm{e}_i)][[E(R_i - \boldsymbol{R}] - \phi] + E(R_i) \tag{21}$$

where

$$\phi = \mathrm{Var}(R_\mathrm{m}) \left\{ \left\langle \sum_{j=1}^{k} [[E(R_j) - \boldsymbol{R}]/\mathrm{Var}(\mathrm{e}_j)] b_j \right\rangle \Big/ \left\langle 1 + \mathrm{Var}(R_\mathrm{m}) \sum_{j=1}^{k} (b_j^2/\mathrm{Var}(\mathrm{e}_j)) \right\rangle \right\} \tag{22}$$

and where k is the set of stocks with positive Z_i's, and b_i is an ordinary least-squares regression coefficient. Where $b_i > 0$, securities are ranked according to the excess return to risk ratio, $[E(R_i) - \boldsymbol{R}]/b_i$, in descending order to determine an ordering, and then Equation (22) is used in an iterative fashion, varying k from 1 to n, to determine the portfolio weights. The iterative process continues until Z_j, computed by Equation (21), turns negative. The negative beta stocks are then considered employing a similar logic, but ranked ascending order. Finally, zero beta stocks are considered — where $E(R_i) > \boldsymbol{R}$, a zero beta stock is admitted, and Equation (22) is again employed to calculate the portfolio weight.

It is important to note that, in the context of regression, there is no requirement that the residuals taken over $i = 1, \ldots, n$ individual regressions be cross-sectionally uncorrelated. Sharpe (1963) imposed this distributional constraint by assumption in order to force diagonalization of a variance–covariance matrix. Elton *et al.* (1976), by contrast, paid lip service to Sharp's distributional assumption, perhaps in order to justify CCM as a substitute for FCM, but, as SIM*'s security-ranking procedure does not operate directly on the variance–covariance matrix, and as matrix inversion is not an issue, they had no interest in the variance–covariance matrix and placed no constraint on its distributional form.

While SIM* is left free to exploit the risk-reducing implications of whatever cross-sectional effects may be suggested by the data but assumed away by SIM, it does not suffer the same tendency toward selection bias as FCM because the security-ranking procedure does not operate directly on any random variable that is distributed as Wishart. Table 1 shows that $N_{\text{FCM}} < N_{\text{SIM}^*} < N_{\text{SIM}}$ for any target monthly rate of return less than 3.5% for which there are three feasible solutions. These observations are perfectly consistent with these agreements.

5. Conclusion

The literature is, unfortunately, no closer to a solution to the stochastic optimization problem today than it was more than a half century ago when Harry M. Markowitz (1952) published his Nobel award-winning seminal article. Optimization models in general, and the portfolio selection models in particular, are deterministic in the sense that the decision variables (i.e., parameters, prices, coefficients, or whatever one wishes to call them) are treated as known constants. Arrow (1971) refers to this deterministic decision-making framework as conditions of risk — which he contrasts with conditions of generalized uncertainty which applies when the variables of decision are not know.

The full covariance model was derived under Bayesian assumptions that Markowitz explains in each of his writings, outlines in great detail in his

(1991) Nobel award lecture, and fervently believes in. Sharpe (1963) invoked the same Bayesian nuances when he derived his single index model, citing Markowitz and the authority. Others who have invoked the same Bayesian nuances in the literature may or may not have been less knowledgeable than Markowitz about Bayesian logic, and/or less convinced of the efficacy of "personalized" probability approaches.

When a deterministic portfolio selection model is applied under conditions of generalized uncertainty — that is, where prior distributions are fashioned and/or revised based on statistical information obtained by sampling — it makes no difference if the application is conducted by a distinguished and committed Bayesian such as Markowitz, or an equally committed member of the Neyman and Pierson school of thought. The deleterious effects of sampling error and selection bias in stochastic applications of deterministic models are totally unrelated to the academic convictions and training of the user, and are unavoidable. The sample theory issues raised in this chapter, therefore, apply to Bayesian and objectivist alike, and the empirical results (previously overlooked) reported in Table 1 carry the same implication regardless of one's probabilistic views.

As no mathematically tractable solution to a stochastic optimization problem has yet been devised, or, to the present knowledge, is in the offing, these, now aging, deterministic portfolio selection models are likely to remain in use for some time to come. It behooves us, therefore, even at this late date, to gain a better understanding of their operating characteristics when applied under conditions of generalized uncertainty — which, in the final analysis, is the only game in town.

It seems quite amazing, in this regard, that SIM* should for so long have been so poorly understood. While SIM* may accommodate SIM's restrictive distributional assumptions, as Elton *et al.* (1976) emphasized — perhaps in order to leave room for a Constant Correlation Model which is otherwise devoid of redeeming qualities — the model is also consistent with more general assumptions, as is perfectly clear from Equation (10).

Elton *et al.*'s (1973, p. 1341) declared purpose when deriving SIM* was to overcome "the difficulty of educating portfolio managers to relate to risk return tradeoffs expressed in terms of covariances," which, accordingly, are not directly addressed by the model's security-ranking solution algorithm. Paradoxically, SIM*s primary advantage in stochastic applications is due precisely to this fortuitous omission: SIM* does not operate directly on any random variable whose sampling distribution is distributed as Wishart, and thus is less susceptible than FCM to deleterious effects of a Wishart's, well-documented, chaotic distributional behavior. On the other hand, SIM* does operate on statistical inputs generated by sampling distributions which, while better behaved than any member of the Wishart family of distributions, are nevertheless subject to estimation error and occasional outliers. While less susceptible to the deleterious effects of sampling error, therefore, there is no logical reason to suppose that the model is immune.

The author and his co-authors were preoccupied with *ex post* issues in his 1999 article, and therefore overlooked the empirical implications of *ex ante* results that are summarized here in Table 1. Oddly enough, we also overlooked an *ex post* empirical result that was literally starring us in the face (see Frankfurter *et al.* 1999, Table 3.1, p. 360): in each sample period featured in their study, for any monthly target rate of return equal to or less than 2.5% (roughly equivalent of an annualized target return of 30%), the data clearly show that SIM*'s *ex post* performance satisfies *ex ante* target rate of return objectives better, and more consistently, than CCM, FCM, or SIM. This point may have been overlooked because, then as now, SIM* would seem to be such an unlikely hero.

References

Arrow, KJ (1971). *Essays in the Theory of Risk Bearing*. Chicago: Markham Publishing Company.

Cohen, KJ and JA Pogue (1967). An empirical evaluation of alternative portfolio selection models. *Journal of Business*, 40, 166–193.

Blume, ME (1971). On the assessment of risk. *Journal of Finance*, 26, 1–10.

Brown, S (1979). Estimation risk and optimal portfolio choice: The Sharpe index model. In *Estimation Risk and Optimal Portfolio Choice*, Bawa, VS, SJ Brown and RW Klien (eds.), pp. 145–171. Amsterdam: North-Holland.

Burgess, RC and R Bey (1988). Optimal portfolios: Markowitz full covariance versus simple selection rules. *Journal of Financial Research*, 11, 153–163.

Elton, EJ and M Gruber (1973). Estimating the dependence structure of share prices — Implications for portfolio selection. *Journal of Finance*, 28, 1203–1232.

Elton, EJ, MJ Gruber and MW Padberg (1976). Simple criteria for optimal portfolio selection. *Journal of Finance*, 31, 1341–1357.

Frankfurter, GM and CG Lamoureux (1989). Estimation and selection bias in mean-variance portfolio selection. *The Journal of Financial Research*, 12, 173–181.

Frankfurter, GM, HE Phillips and G Faulk (1999). The *ex post* performance of four portfolio selection algorithms. *Review of Quantitative Finance and Accounting*, 13, 347–366.

Frankfurter, GM, HE Phillips and JP Seagle (1974). Bias in estimating portfolio alpha and beta scores. *Review and Economics and Statistics*, 56, 412–414.

Johnson, RA and DW Wichern (1988). *Applied Multivariate Statistical Analysis*. Englewood Cliffs, NY: Prentice Hall.

Kalymon, BA (1971). Estimation risk in the portfolio selection model. *Journal of Financial and Quantitative Analysis*, 6, 559–582.

Kotz, N and L Johnson (Eds.) (1988). *Encyclopedia of Statistical Sciences*, Vol. 9. New York: Wiley.

McCall, JJ (1971). Probabilistic macroeconomics. *The Bell Journal of Economics and Management Science*, 2, 403–433.

Markowitz, HM (1952). Portfolio selection. *Journal of Finance*, 7, 77–91.

Markowitz, HM (1959). *Portfolio Selection*. New Haven, CT: Yale University Press.

Markowitz, HM (1991). Foundations of portfolio theory. *Journal of Finance*, 46, 469–477.

Phillips, HE (1993). Portfolio optimization algorithms, simplified criteria, and security selection: A contrast and evaluation. *Review of Quantitative Finance and Accounting*, 3, 91–97.

Raiffa, H and R Schlaifer (1961). *Applied Statistical Decision Theory*. Boston, MA: Graduate School of Business, Harvard University.

Sharpe, WF (1963). A simplified model of portfolio analysis. *Management Science*, 9, 277–293.

Tadjudin, S and DA Landgrebe (1999). Covariance estimation with limited training samples. *IEEE Transactions on Geoscience and Remote Sensing*, 37, 2113–2118.

This page intentionally left blank

Chapter 11

Corporate Capital Structure and Firm Value: A Panel Data Evidence from Australia's Dividend Imputation Tax System

Abu Taher Mollik
University of South Australia, Australia

This chapter examines the empirical effects of financial leverage (FL) (corporate capital structure) on the market value of a selection of firms listed on the Australian Stock Exchange, developing a direct value-leverage model. Employing a least square dummy variable method to a pooled time-series and cross-sectional data set, the results suggest that the value of a firm rises significantly with FL. More specifically, there is a statistically significant positive effect of total, total interest bearing, and long-term FL on the market value of a firm, suggesting that leverage matters even under the Australia's (full) dividend imputation tax system.

Keywords: Capital structure; firm value; panel data; imputation tax system.

1. Introduction

Do corporate financing decisions affect firm value? How much do they add and what factors contribute to this effect? Considerable research effort, both theoretical and empirical, has been devoted toward finding sensible answers to these questions since the work of Modigliani and Miller (1958, 1963). Although opinion is not unanimous on which factors are most important or how and whether they contribute to firm value (Shyam-Sunder and Myers, 1998; Fama and French, 1998), the extant literature suggests two sources of effects: tax and nontax sources.[1] When the tax deductibility of interest within the corporation creates a clear gain to leverage (Modigliani and Miller, 1963), personal taxes act to reverse this effect (Miller, 1977). However, the empirical evidence supports only a partial reduction of debt tax shield benefit by personal tax penalties as Graham (2000), for example, reveals that the capitalized tax benefit of debt in the US is about 10% of firm value and that the personal tax penalty reduces this benefit by nearly two-thirds before the tax Reform Act of

[1] See Haugen and Senbet (1986) for review of tax-literature and Harris and Raviv (1991) for nontax literature.

1986 and by slightly less than half after tax reform. Thus, the debt tax shield benefit and the ultimate tax bias to debt prevails under the US classical tax system where a firm's profits, distributed as dividends, are ultimately taxed twice whereas profits distributed as interest is taxed once at a higher rate than personal tax on dividends.

A number of countries, including Australia, now adopt an alternative dividend imputation tax system (DITS) with a view to removing the tax bias to debt. Australia introduced an alternative DITS in 1987 in which a full imputation credit is allowed to eligible shareholders (except certain classes of investors) for taxes paid at the corporate level. Although interest payments remain a tax-deductible corporate expense, this system integrates corporate and personal taxes, and both interest and dividends from Australian operating activities are ultimately taxed once at the investor's marginal tax rate. The level of Australian corporate tax ultimately becomes irrelevant to a certain class of investors. This can lower the tax-reducing benefit of interest deductions at the firm level and provide a tax incentive to reduce the magnitude of debt in the firm's capital structure. Under the Australian (full) imputation tax system, Benge (1997), Peirson *et al.* (1998), and Twite (2001) show that, to the extent tax biases exist, they favor equity rather than debt finance.

On the nontax dimension, the financial distress, agency, corporate control, information, or market timing attributes can impact on the value effect of debt in both directions. The existence of bankruptcy costs is a value-reducing (statutory) event when a levered firm is unable to pay back creditors. Holders of debt bear the *ex post* costs of bankruptcy, passing on the *ex ante* costs of bankruptcy to equity holders in the form of higher interest rates, which lower the valuation of the firm. Thus, bankruptcy costs have a negative effect on firm value, trading-off, any tax advantage to debt (e.g., Baxter, 1967; Kraus and Litzenberger, 1973; Scott, 1976; Kim, 1978; Brennan and Schwartz, 1978). In Jensen and Meckling (1976), debt affects firm value through agency costs. Debt mitigates the manager–shareholder agency conflict and reduces the agency costs of equity by raising the manager's share of ownership in the firm. Debt also use up "free" cash available to managers to engage in their personal pursuits (Jensen, 1986). But debt has monitoring and covenant costs due to "asset substitution effects" creating incentive to invest in value-decreasing projects where the equity holders may benefit from "going for broke," i.e., investing in very risky projects even if they are value-decreasing, but ultimately bear the cost *ex ante*.[2] In a related study, Parrino and Weisbach (1999) empirically estimate that the agency costs of debt are too small to offset the tax benefits. Leverage can also create underinvestment costs as discussed in Stulz (1990) and Peyer and Shivdasani (2001). Existence of several nondebt tax shelters (i.e., investment tax credits, depreciation, and depletion allowances) may limit the tax-advantage of debt as they may reduce utilization of available interest tax credit (DeAngelo and Masulis, 1980; Dammon and Senbet, 1988). There are, however, other theories of capital structure in different settings such as pecking order, corporate control, signaling, information, market timing, and so on[3] that reflect on both positive and/or negative effects of leverage on firm value. On balance, whether leverage has any significant effect on firm value remains an empirical issue.

Empirical evidence on the impacts of corporate leverage (capital structure choice) on firm value is conflicting and inconclusive. A selected few event studies of exchange offers, swaps, stock repurchases, and seasoned equity offerings produce the only direct evidence for the leverage effect on firm value. For example, Masulis (1980) finds that debt-for-equity exchanges generally increase stock prices, while equity-for-debt swaps lower the stock prices. Similar to Masulis (1980), Engel *et al.* (1999) find firms derive substantial net tax benefits when they swap tax-deductible trust preferred stock for nondeductible regular preferred stock. But there are reasons to question whether leverage explain these event study results. First, it is well known that new equity issues lower stock prices (Masulis and Korwar, 1986), but equity repurchases raise stock prices (Vermaelen, 1981). Further evidence along the line that increases in debt that do not involve reductions in equity produce weak stock price response (Eckbo, 1986) reinforces the conclusion that the information effects of changes in equity, rather than the effects of changes in debt, explain Masulis's strong results on exchange offers. Second, these analyses suffer from the sampling bias to only those companies, which change their capital structure during a particular period, as companies which do not change their capital structure (or do not change their capital structure in the way the particular study considers the change) during that particular time period are left out, raising questions about any sustained leverage effect on firm value. Third, although the results may well reflect the value of the

[2]See Jensen and Meckling (1976) for formal derivation.

[3]Harris and Raviv (1991) provide a detailed discussion of these theories.

debt tax shield, the agency, corporate control, information, or market timing attributes of debt-for-equity recapitalizations may contribute to the empirical results. That is, debt may have value effect along nontax dimensions, which these types of studies are unable to illuminate. Thus, the event study results are inconclusive and/or cannot be generalized for all levered firms.

Other empirical research that measure debt tax benefits, for example, Mackie-Mason (1990), Scholes and Wolfson (1992), Trezevant (1992), Francis and Leachman (1994), and Graham (1996, 1999, 2000) provide evidence that the marginal tax rate varies considerably across the companies and the high marginal-tax rate companies borrow more than those with low tax rates, indicating a debt-tax benefit for firms. This evidence provides support for the value-increasing effect of leverage. Graham's (2000) findings in the tax benefit calculation by using firm-level financial statement data for a large sample of Compustat firms also provides a similar indirect support for the value enhancing effect of debt.

In estimating the direct market value effect of debt tax shield, Fama and French (1998) use cross-sectional regressions of firm value on interest expenses (which proxies for debt) and various controls for profitability such as earnings, investment, R & D, and dividend. Fama and French (1998) find a strong negative relationship between leverage and firm value. As it is very hard to interpret their unexpected results even in terms of Miller's (1977) hypothesis that leverage has no net tax benefits (because personal taxes on interest offset the corporate tax savings), they admit that their regressions fail to measure how (or whether) the tax effects of financing decisions affect firm value, concluding that: "The relations between financing decisions and value we observe are unidentified mixes of tax effects and factors that affect profitability." (p. 821) and that "imperfect controls for profitability probably drive the negative relations between debt and value and prevent the regressions from saying anything about the tax benefits of debt" (p. 839).

Unlike Fama and French (1998), Kemsley and Nissim (2002) recognize the fact that the value of a firm's operation is unobservable and could be correlated with debt along nontax dimensions, and use a complex reverse cross-sectional approach by specifying future operating profitability as a function of firm value, debt, and controls for firm-level capitalization rates, to mitigate the effects of the correlation along nontax dimensions. They find a positive, significant value of debt tax shield, net of personal tax disadvantage of debt. They also find the estimated value of the net debt tax shield increases in statutory corporate tax rates over time, and it increases in estimated marginal tax rates across firms. However, their tests are specifically designed to distinguish between Modigliani and Miller's (1963) corporate tax benefit and Miller's (1977) personal tax disadvantage and leverage irrelevance, and therefore, the estimates cannot be used to illuminate the trade-off between tax and nontax factors.

Ruland and Zhou (2005) provide a strong support for the hypothesis that the values of diversified firms increase with leverage. However, specialized firms are not found to have this leverage impact in their study.

In Australia, Twite (2001) examines the changes in corporate capital structure around the introduction of a DITS and finds that dividend imputation provides an incentive for firms to reduce the level of debt financing utilized where this incentive varies across firms depending on their effective corporate tax rate. While this finding provides an indirect support to the theoretical hypothesis of positive tax effect of leverage on firm value (assuming that the firms made a value maximizing decisions), Twite's (2001) study was not designed to examine and report the direct value effect of debt tax shield, and to the best of the author's knowledge, no other study has, as yet, examined the value effect of leverage under an imputation tax system, especially in Australia.

The present study examines the effect of corporate capital structure on firm value in Australia. Specifically, it develops a multivariate econometric model of value–leverage relationship, by specifying market value of a firm as a function of debt, dividend payout, size, growth, expected tax-adjusted earnings, and risk (measured by market beta), and estimates the overall effect (both tax and nontax effects) of capital structure choices on the market value of selected firms in Australia subsequent to the introduction of DITS, over the period from 1988 through 1997. This model is an improvement over Fama and French (1998) in that it uses three different theoretically close measures of debt instead of interest to proxy for leverage and some additional controls for firm size, growth, and risk. Also, this model uses expected tax-adjusted earnings instead of overall earnings to control for profitability. Compared to Kemsley and Nissim's (2002) complex reverse model, this forward model is very simple to estimate with high explanatory power (the Ad-$R^2 = 0.912$). This model, however, does not disaggregate the leverage effect into tax and nontax dimensions. A least square dummy variable (LSDV) method was employed to a pooled time-series and cross-sectional data set. The results suggest that the value of a firm rises significantly with financial leverage (FL), indicating an overall positive effect of leverage. More specifically, the study reveals a statistically significant positive effect of total, total interest bearing, and long-term FL on the market value of a firm. The result supports the Modigliani and Miller (1963) debt tax shield hypothesis, and is consistent with the value effect of leverage in Ruland and Zhou (2005), Kemsley and Nissim's (2002) and the market evidence for a significant positive value effect of leverage in Masulis (1980) and Engel *et al.* (1999). The finding is also consistent with the tax benefit calculations in Graham (2000) and the findings in Mackie-Mason (1990), Scholes and Wolfson (1992), Trezevant (1992), Francis and Leachman (1994), and Graham (1996, 1999) in a sense that debt affects firm value due to taxes. The result also provides support for the recent research finding (Welch, 2004; Mehrota *et al.*, 2005; Leary and Roberts, 2005; Flannery and Rangan, 2006) that companies actively adjust their capital structures in order to maximize their values.

The remainder of the chapter is organized as follows: Section 2 discusses the theoretical basis and develops the value-leverage models; Section 3 describes the empirical models and estimation methods; Section 4 presents the regression results; and Section 5 provides summary and conclusions.

2. The Theoretical Framework and the Model

2.1. *Theoretical framework*

Modigliani and Miller (1963) show that when corporate tax laws permit the deductibility of interest payments, the market value of a firm is an increasing function of leverage.[4] The equilibrium market value of a levered firm is given by:[5]

$$V_{\mathrm{L}} = \overline{X}(1 - \tau_{\mathrm{c}})/\rho + \tau_{\mathrm{c}} D_{\mathrm{L}} \tag{1}$$

where $\overline{X}$ equals expected earnings before interest and taxes, τ_{c} is corporate income tax rate, ρ is the expected cost of capital or capitalization factor to an all-equity firm on after tax basis, $\overline{X}(1 - \tau_{\mathrm{c}})/\rho = V_{\mathrm{u}}$, value of the firm if

[4]While he derived leverage irrelevance with no taxes (Modigliani and Miller, 1958).
[5]The formal derivation can be found in most of the finance textbooks, including Brealey and Myers (2006).

all-equity-financed, and $\tau_c D_L$ is the present value of the interest tax-shield, the tax advantage of debt.

It is given that $\overline{X}$, V_L increase with leverage, because interest is a tax-exempt expense. While this relationship successfully introduces the potential effects of corporate taxes into capital structure theory, providing a supply side explanation of the existence of debt, it leads to an extreme *corner solution* as the firm value is maximized when 100% debt finance is used. In reality, few such firms exist probably because of the uncertainty of interest tax-savings, and the existence of personal taxes (Miller, 1977) and nondebt tax shields (DeAngelo and Masulis, 1980) putting a limit to the tax advantage of debt.

Taking into account Miller's (1977)[6] demand side effect incorporating personal income tax along with corporation income tax, the gain from leverage, G_L, for stockholders in a firm holding real assets can be shown to be given by $G_L = D_L(1 - [(1 - \tau_c)(1 - \tau_{pe})/(1 - \tau_{pd})])$. Therefore, the market value of a levered firm incorporating the effects of both corporate and personal taxes can be expressed as follows:

$$V_L = V_U + D_L(1 - [(1 - \tau_c)(1 - \tau_{pe})/(1 - \tau_{pd})]) \qquad (2)$$

where τ_{pe} is the marginal tax rates applicable to a firm's equity holders and τ_{pd} is the marginal tax rates applicable to a firm's debt holders.

An important implication of Equation (2) is that the tax gain from leverage is now lower than $\tau_c D_L$, because the higher tax liabilities on interest at the personal level offsets the interest tax-shield benefit from leverage at the corporate level.

Miller suggests tax-induced preferences for investors' debt/equity choices (Miller's "clientele effect") with (i) $(1 - \tau_c)(1 - \tau_{pe}) > (1 - \tau_{pd})$ indicating a preference for equity, (ii) $(1 - \tau_c)(1 - \tau_{pe}) < (1 - \tau_{pd})$ indicating a preference for debt and, (iii) $(1 - \tau_c)(1 - \tau_{pe}) = (1 - \tau_{pd})$, investors being indifferent between debt and equity; but firms will supply both the securities, because the tax advantage of debt vanishes completely at this point and, in equilibrium, the equality of (iii) holds for all firms. Hence, there is an aggregate optimum debt level for firms as a whole, which depends on the corporate tax rate and the funds available to individual investors in the various tax brackets. No single

[6]Although Miller's (1977) theory was developed long after the bankruptcy costs theory, the former was discussed before the bankruptcy theory to maintain continuity with the irrelevance hypothesis since Miller (1977) also obtains leverage irrelevancy at the firm level.

firm can influence that. Therefore, for an individual firm, capital structure does not matter.[7]

2.1.1. *Bankruptcy cost*

Up until now the premise of the present analysis was limited by the assumption of a perfectly competitive frictionless capital market with common information and certainty of (investment) outcomes. The existence of bankruptcy costs is capable of changing the conclusions of classical models because it is a value-reducing (statutory) event when a levered firm is unable to pay back creditors. Holders of debt bear the *ex post* costs of bankruptcy, passing on the *ex ante* costs of bankruptcy to equity holders in the form of higher interest rates, which lower the valuation of the firm. Thus, bankruptcy costs have a negative effect on firm value, trading-off any tax advantage to debt. Therefore, a number of authors (e.g., Baxter, 1967; Kraus and Litzenberger, 1973; Scott, 1976; Kim, 1978; Brennan and Schwartz, 1978) have suggested that the expected cost of bankruptcy/financial distress is the factor missing in the Modigliani and Miller (1958, 1963) models,[8] and that balancing the tax advantages of debt against the expected cost of bankruptcy raises the possibility of an optimal interior capital structure for a firm. The fundamental format for each of those models can be summarized as follows:

$$V_{\mathrm{L}} = \overline{X}(1-\tau_{\mathrm{c}})/\rho + \lambda D_{\mathrm{L}} - bD_{\mathrm{L}} \tag{3}$$

where $\lambda = (1 - [(1-\tau_{\mathrm{c}})(1-\tau_{\mathrm{pe}})/(1-\tau_{\mathrm{pd}})])$ and b is the present value of the bankruptcy cost per dollar of debt.[9] The optimal capital structure will occur by maximizing the firm value at the point where the present value of the marginal tax shield (benefit) on interest payments equals the present value of

[7] Arditti *et al.* (1977) obtained the same formula as Miller, but using an arbitrage argument, and proved that this is an equilibrium relationship between the value of the levered and unlevered firms.

[8] Although Modigliani and Miller (1958, 1963) recognized the existence of the probability of bankruptcy, they assumed that there is no bankruptcy (or bankruptcy is costless).

[9] The last term in Equation (3), can be clarified by specifying the probability of bankruptcy as the probability that the firm's cash flow is not sufficient to meet its financial obligation, that is, $\psi[X \leq f(D)]$, where X is the firm's random cash flow and $f(D)$ is its financial obligation. The firm's financial obligation is a function of its use of debt financing. Given the firm's probability distribution of cash flows, X, the increased use of FL will, other things equal, increase the probability of bankruptcy or financial distress.

the marginal expected bankruptcy costs of debt, i.e., the following condition holds[10]:

$$\partial V/\partial D_{\mathrm{L}} = 0; \quad \text{or} \quad \partial(\lambda D_{\mathrm{L}})/\partial D_{\mathrm{L}} = \partial(bD_{\mathrm{L}})/\partial D_{\mathrm{L}} \tag{4}$$

In this trade-off model, the optimal capital structure for a firm can be obtained without resorting to Miller's personal taxes, in which a higher tax advantage of debt, $\tau_{\mathrm{c}} D_{\mathrm{L}}$ is traded-off against the present value of bankruptcy cost of debt, bD_{L}.

The bankruptcy costs model has considerable intuitive appeal because it provides an explanation for the coexistence of debt and equity in the individual firm's capital structure. Its argument, however, rests on how significant these costs are to completely offset the tax benefits of leverage. Myers (1984), while acknowledging the existence of bankruptcy costs, cast doubt on the magnitude of these costs.

2.1.2. *Agency cost*

Jensen and Meckling (1976) used the agency relationship and thus agency costs to explain the existence of optimal capital structure at the firm level. They argue that separation of the firm's control (management) from its ownership may create conflicts of interest between agents and costs to the firm, known as agency costs of equity, because managers may be engaged in value non-maximizing activities such as investing less effort in managing firm resources and/or transferring firm resources for personal benefit, e.g., excess perquisites consumption. In a related paper, Parrino and Weisbach (1999) empirically estimate that the agency costs of debt are too small to offset the tax benefits. However, debt mitigates the manager–shareholder conflict and reduces the agency costs of equity by raising the manager's share of ownership in the firm, as increase in debt, holding the manager's absolute investment in the firm constant, increases the manager's share of equity. Also, debt can reduce agency costs of equity by reducing the amount of "free" cash available to managers to engage in their personal pursuits (Jensen, 1986) since debt commits the firm to pay out cash.

But debt can create "asset substitution effects" by creating the incentive to invest in value-decreasing projects where the equity holders may benefit

[10] $\partial V/\partial D_{\mathrm{L}} = \partial(\lambda D_{\mathrm{L}})/\partial D_{\mathrm{L}} - \partial(bD_{\mathrm{L}})/\partial D_{\mathrm{L}} = 0.$

from "going for broke," i.e., investing in very risky projects even if they are value-decreasing, but ultimately bear the cost *ex ante*.[11] Also, to protect themselves, debt holders monitor the firm (imposing monitoring costs) and impose covenants (covenant costs), both of which can be described as agency costs of debt. Debt can also cause under investment problems, as the manager of a highly geared firm may miss out on valuable investment opportunities, reducing the value of the firm.

Due to the agency costs attached to both debt and equity, an optimal capital structure is obtained in the agency approach by trading-off the agency costs of equity (the benefit of debt) against the agency costs of debt and by minimizing the total agency costs involved in issuing debt and equity. If a dollar of debt reduces the present value of the agency costs of equity by c_{e} and increases the present value of the agency costs incurred by c_{d} (which is assumed to be generally lower than c_{e}), the total benefit of debt used by the firm, D_{L}, would be $(c_{\mathrm{e}} - c_{\mathrm{d}})D_{\mathrm{L}}$. Considering this as an addition to the total value of the firm in excess of the value added by debt-tax shields, the total market value of the firm can be expressed as:

$$\begin{gathered} V_{\mathrm{L}} = \overline{X}(1 - \tau_c)/\rho + (\lambda - b)D_{\mathrm{L}} + (c_{\mathrm{e}} - c_{\mathrm{d}})D_{\mathrm{L}} \quad \text{or} \\ V_{\mathrm{L}} = \overline{X}(1 - \tau_{\mathrm{c}})/\rho + \Phi D_{\mathrm{L}} \end{gathered} \tag{5}$$

where $\Phi = (\lambda + c_{\mathrm{e}}) - (b + c_{\mathrm{d}})$. In equilibrium, $(\lambda + c_{\mathrm{e}}) = (b + c_{\mathrm{d}})$, and an optimal capital structure for individual firms is obtained at the margin.

The above theoretical discussion and simple mathematical exercises show that there may be both tax and nontax benefits associated with debt financing of a firm, which are traded off against the tax and nontax costs associated with debt to obtain the firm's optimal capital structure. So far, the effects of two nontax theories along with the tax theories have been included to derive Equation (5). There are, however, other theories of capital structure in different settings such as nondebt tax shield, pecking order, corporate control, signaling, and so on.[12] Existence of several nondebt tax shelters (i.e., investment tax credits, depreciation, and depletion allowances) may limit the tax-advantage of debt as they may reduce utilization of available interest tax credit (DeAngelo and Masulis, 1980; Dammon and Senbet, 1988). However, the costs and benefits of debt that these theories may suggest can be impounded inΦ, the coefficient of D_{L}, in a similar fashion keeping the form of Equation (5) above

[11] See Jensen and Meckling (1976) for formal derivation.
[12] Harris and Raviv (1991) provide a detailed discussion of these theories.

unchanged. Φ, therefore, includes both the tax and nontax effects of debt, and estimating Φ empirically for Australian firms is the main objective of the present study.

2.2. *Australian imputation tax system and firm valuation*

Australia, like Germany, Italy, and New Zealand, adopted an alternative (full) DITS for Australian resident taxpayers from July 1, 1987 (prior to which a classical tax system existed), integrating corporate and personal taxes where both interest and dividends are ultimately taxed once at the investor's marginal tax rate. Australia's DITS applies to dividends paid by resident Australian companies to resident individual shareholders and a limited class of other shareholders, including Australian pension funds.

In general, for Australian resident taxpayers Australia's DITS works as follows.[13] Corporate tax is paid by firms. Interest paid is tax deductible for companies, but lenders pay tax on interest income at the personal tax rate τ_p and receive $(1 - \tau_p)$ after tax. The tax treatment of dividends depends on whether dividends are "franked" (paid from after tax income) or "unfranked" (paid from before tax income). Unfranked dividends do not include any tax credit, but the same personal tax rate applies. For each dollar of unfranked dividends shareholders pay tax of τ_p and receive $(1 - \tau_p)$ after tax. On franked dividends shareholders can claim full credit ($(\tau_c/(1 - \tau_c))$ per cent of the amount of franked dividend) for the taxes paid at the firm level, so that for each dollar of taxable income, the company pays τ_c in corporate tax and shareholders pay, after allowing for the tax credit τ_c, a further $(\tau_p - \tau_c)$ in personal tax and receive $(1 - \tau_c) - (\tau_p - \tau_c) = (1 - \tau_p)$ after tax. Thus, income distributed as dividends and interest to resident shareholders and bondholders in Australia is effectively taxed once at the personal tax rate. The corporate tax is ultimately eliminated, and therefore, irrelevant. That is, with the corporate tax rate effectively zero, the size of the gain from leverage, G_L (as demonstrated by Miller (1977) and included in Equation (2) in this chapter), will depend on the balance between the personal tax on interest income and that on equity income (dividends and/or capital gains), in this case, as follows:

$$G_L = D_L[1 - \{(1 - \tau_{pe})/(1 - \tau_{pd})\}] \tag{6}$$

[13] This paragraph largely borrows from Benge (1997).

Since, $\tau_{pe} = \tau_{pd} = \tau_p$, under the Australian (full) imputation tax system, $G_L = 0$ is obtained. Therefore, it is likely that the choice of capital structure does not affect firm value, when only the tax benefit to leverage is considered. For example, Mr. A earns \$80,000 as salaried income, \$140 as fully franked dividend after 30% corporate tax payment and \$200 as interest income. Mr. A is a resident Australian tax payer and is eligible to claim 100% imputation credit back for the tax paid at the corporate level. According to the current income tax rates applicable for individuals in Australia (2006–2007), the tax rate applicable for any taxable income between \$75,001 and \$150,000 is 40%. Therefore, the marginal personal tax rate applicable on Mr. A's grossed up dividend and interest income is the same 40%. His grossed up dividend income is \$200 [140/(1 − 0.30)] and total taxable income is \$80,400 (\$80,000 +\$200 +\$200). Mr. A's total tax liability is \$20,010 (\$17,850 on income up to \$75,000 plus \$2,160 (40%) on \$5,400). The tax payable by Mr. A after adjustment for the imputation credit of \$60 (\$200–\$140) for dividend income is \$19,950 (\$20,010–\$60). From this example, it is evident that Mr. A has paid tax at the same marginal rate on both grossed up dividend and interest income. Considering Mr. A's case as the standard, Equation (6) shows that there is no gain from corporate leverage as $\tau_{pe} = \tau_{pd} = 0.40$ or (40%).[14] Therefore, an effective full DITS favors neither debt nor equity. While this result is from a fully franked and unfranked dividend, the outcome is unaffected by whether the dividend is fully franked, unfranked, or partially franked (mixture of fully franked and unfranked) for a certain class of investors who are eligible for full imputation credit. This is referred to as a fully integrated system as investors are effectively taxed at their personal tax rates.

Corporate tax becomes relevant when the system is not fully integrated. There are a number of factors that detract from its being fully integrated such as the existence of a mixture of investors who face different tax regimes. Excluding certain classes of investors, the Australia's DITS is only partially integrated. Prior to July 1, 2000, imputation credits were not refunded and could not be carried forward to future tax years. Shareholders could only off-set the surplus imputation credits against the current tax liability of their other incomes. This caused the DITS to be only partially integrated, creating several tax-based clienteles for the investors, based on their capacity to utilize

[14]Please visit http://en.wikipedia.org/wiki/Dividend_imputation, for a better understanding of the Australian tax system.

imputation credits over the sampled study periods from 1988 to 1997. Tax-exempt and some very low-income investors could not utilize imputation credits at all because they had no other tax liability, and were effectively taxed (the dividend income) at the company tax rate τ_c. Some low-income investors could utilize imputation credits only partially due to a lower tax liability than imputation credits, and consequently ended up paying tax at a rate higher than their marginal personal tax rate τ_p. Obviously, these investors would benefit by being lenders instead of shareholders under this system, as there would be either no tax or less tax on interest than on dividends. However, this does not necessarily mean that shareholders can gain by investing in companies borrowing from tax-exempt investors, because there are few tax-exempt investors and the interest rate paid by companies will have to be sufficiently large to attract taxable investors to become lenders.

The case of retention of profits is somewhat complex. Retention of both franked and unfranked dividends raises capital gains to shareholders. Realized capital gains and losses are taxed at the resident investor's statutory marginal income tax rate τ_p.[15] Nonresident shareholders are not subject to Australian capital gain taxes on gains and losses in Australian listed shares/equities. This tax rate on capital gains is defined as τ_g to differentiate it from the statutory personal tax rate on dividend τ_p. To the extent that capital gains are taxed upon realization, it can be expected that $\tau_g < \tau_p$, and a tax-induced preference for equity exists in case of capital gains arising from retention of unfranked dividends. But capital gains arising from the retention of franked dividends would be effectively subject to double taxation, once at the corporate level at τ_c and then at the personal level at τ_g. The preferences will depend on the combined effects of the three tax rates (i.e., τ_c, τ_p, and τ_g). Based on these tax rates, for $(1 - \tau_c)(1 - \tau_g) > (1 - \tau_p)$, investors prefer equity; for $(1 - \tau_c)(1 - \tau_g) < (1 - \tau_p)$, investors prefer debt; and for $(1 - \tau_c)(1 - \tau_g) = (1 - \tau_p)$, investors are at best neutral or prefer debt if the time value of money is taken into consideration. However, in the short-run, it is expected that $(1 - \tau_c)(1 - \tau_g) < (1 - \tau_p)$ and investors prefer debt or retention of unfranked dividends. Therefore, if a firm adopts an optimal dividend

[15]Capital losses in excess of capital gains in current year cannot be offset against other income, but can be carried forward to be offset against the future capital gains. The cost base of the asset is inflation indexed for the purpose of calculating the capital gain (but not the capital loss) for assets held for 12 months or more.

policy of distributing franked dividends and retaining unfranked dividends, which Twite (2001) also assumed, to derive his tax-induced preference for equity, then capital gains will be corporate tax free and the value of $1 distributed as franked dividends and capital gains through retained earnings will dominate the value of $1 distributed as interest, since $\tau_g < \tau_p$, creating a clear tax-induced preference for equity. But in practice, it is generally observed that Australian firms pay both franked and unfranked dividends and may retain profits from the franking account. Therefore, it can be expected that the corporate tax rate τ_c will have influence on investors' preference between debt and equity of these firms.

Another important point that most discussions in this regard miss is the fact that the interest tax-shield increases the shareholders' potential capital gains by raising the firm's capacity to retain unfranked dividends under the Australian imputation system. All else equal, a levered firm will have higher unfranked dividends available to be distributed as capital gains via retained-earnings, due to the tax saving on interest. Also, it appears from the above analysis that while all tax-paying investors benefit by investing in firms that distribute profits as capital gains via retention of unfranked dividends, higher tax bracket (higher than corporate tax rate) investors benefit even more as it reduces their additional current tax liability and at the same time increases their future capital gains. Since a growing firm has more opportunity to efficiently employ its retained earnings and maintain its share prices, higher tax bracket investors will maximize their return by investing in the equity of optimally levered growth firms. Thus, firms have incentive to supply both the securities — debt and equities.

Nonresident foreign investors are not eligible for imputation tax credits. They are subject to withholding taxes on unfranked dividends received, but tax exempt foreign funds can usually get an exemption from withholding taxes on unfranked dividends. Franked dividends are not subject to withholding tax for them. However, they might be subject to tax in their home country on both franked and unfranked dividend income. This can be an important factor that may help companies to gain from leverage because of the fact that overseas investors in Australian companies are outside the imputation tax system and are effectively still operating under the classical tax system. Consequently, companies with large overseas ownership are expected to reap tax benefits.

Australian pension funds are taxed at 15%, a lower rate than both the corporate and marginal personal tax rates. Investors end up paying 15% of the pretax earnings of the firm irrespective of the form of distribution as interest or dividends, excepting franked dividend distributed as capital gains through retained earnings in which case the effective tax rate would be higher than 15% due to double taxation. Therefore, for pension funds, debt remains tax-preferred to equity, and all domestic investors have a tax preference for the distribution of franked dividends and the retention of unfranked dividends.[16]

There are also several other factors such as immediate payout of all profits as franked or unfranked dividends, and no time delay between the payment of corporate tax and the benefit of the imputation tax credit to shareholders. The degree to which these factors do not hold contributes to a DITS not being fully integrated. In these circumstances, the analysis under the classical tax system holds, and the corporate rate may exert nontrivial influence on the investors' preference for debt and equity.

To summarize, whether corporate tax and the tax benefit to leverage matters depends on which of the above mentioned taxpayer groups is marginal investor and determines the firm value. It is theoretically likely that for a certain class of investors who are eligible to utilize full imputation credit, the Australian DITS provides no tax benefit to leverage. Other results are possible if profits are retained or investors are tax-exempt or nonresident foreign. Thus, firms have incentive to issue both debt and equity securities, as some investors gain from borrowing at the corporate level. The crucial point is while the magnitude of tax effect under dividend imputation maybe less, the premise of tax-benefit analysis remains the same as it is under the classical tax system. The overall gain from leverage at the corporation level is subject to empirical examination. The purpose of the present study is to examine the overall effect of leverage on firm value in Australia.

2.3. *The value-leverage model*

Based on the above analysis of leverage and valuation theories, Equation (5) can be estimated in the following structural equation form for the value of a levered firm, V_{L}:

$$V_{\mathrm{L}} = a_0 + a_1 \overline{X}(1 - \tau_{\mathrm{c}}) + a_2 D_L + \varepsilon \tag{7}$$

[16]Twite (2001) demonstrates this in detail.

where $a_1 = 1/\rho$ is the inverse of marginal capitalization rate, $a_2 = \Phi$ (in Equation 5), which measures overall (both the tax and nontax) effects of leverage on a firm's value is the main focus of this analysis, and ε is the stochastic or random component of V_L or simply the error term.

There are, of course, many other factors that influence firm valuations in practice. Some of them are so large and systematic that they need to be directly incorporated into the model rather than impounding them in the general disturbance term. Following Miller and Modigliani (1966), Crockett and Friend (1967), Robichek *et al.* (1967), Sarma and Rao (1969), and Rao and Litzenberger (1971), growth, dividend, and size variables are introduced into Equation (7) to control their valuation effects. As the sample incudes firms from different industry groups, a risk variable is also included to control for the risk differences as follows:

$$V_L = a_0 + a_1\overline{X}(1 - \tau_c) + a_2 D_L + a_3 R + a_4 P + a_5 Z + a_6 G + \varepsilon \qquad (8)$$

where $a_1 = 1/\rho$ is the marginal capitalization rate, $a_2 = \Phi$ (in Equation [5]), which measures overall (both the tax and nontax) effects of leverage on a firm's value is the main focus of this analysis, and a_6, a_3, a_4, and a_5, respectively, measure of the effects of growth potential, systematic risk, dividend payouts, and the effects of size or scale on firm value. ε is the stochastic or random component of V_L or simply the error term.

3. Empirical Models and Estimation Method

3.1. *Data and the variables and measures*

This study analyses the effects of FL on firm value of a selection of Australian firms. The variables were analyzed over the time period from 1988 through 1997. The source of all the data is Data Stream. All Australian nonfinancial firms available in Data Stream (at the time of data extraction) were considered for the preliminary analysis. The year-end company accounts and balance sheet items relevant to the indicators were used for the present analysis.

From the total sample, all the firms that did not have a complete and consistent record of the variables included in the analysis over the 10-year period were deleted. Furthermore, since many of the variables are scaled by total assets or average operating income, a small number of observations that included negative values for one of those variables were deleted. These requirements may have biased the sample toward relatively old and larger firms. In total, 45 firms were available for the final analysis.

Three-year moving averages of the sample variables were used, except when mentioned otherwise. This averaging reduces the measurement error due to random year-to-year fluctuations in the variables.

The key variables involved are the market value of the firm, the tax adjusted earnings, $(X^{\tau} - \tau R)$, of the firm, and different debt ratios. In addition, growth, firm size, dividend payout, and beta risk have been used as control variables. The effects of debt–equity choices on the firms' market value are estimated in a panel data framework. Since most of the variables used are proxies for actual variables, it is very important that the constructed variables are defined properly.

3.1.1. *Financial leverage (FL)*

FL or the debt–equity ratio (DR) is one of the most important variables in this study. Previous empirical capital structure studies have used a large variety of DRs as measures of FL. Taking account of the issues and comments made by Bowman (1980), Titman and Wessels (1988), Bradley *et al.* (1984), Long and Malitz (1985), Weiss (1990), and Hamson (1992) about firm value and debt measures, the value of a firm, V (equivalently V_{L}), has been defined as the market value of equity plus book value of long-term debt and total short-term or current liabilities. This definition of the market value of the firm assumes that the face value of debt is fixed while the interest rate is variable. It also assumes noninterest bearing short-term liabilities as a part of the total debt of the firm. Although this is inconsistent with tax-based theories, it is fairly consistent with the agency costs, pecking order, and other information asymmetry theories of corporate capital structure. V scaled by the fixed assets (FA) has been used as the measure of market value of the firm.

Debt has been defined in three different ways. They are: (i) book value of long-term debt plus book value of total current liabilities (D1); (ii) book value of long-term and short-term interest bearing debt only (D2); (iii) book value of only long-term debt (D3). These three measures of debt scaled by the fixed assets (FA) of the firm have been used as the measure of FL or debt–equity ratio (DRs) of the firm as: DR1 (= D1/FA), DR2 (= D2/FA), DR3 (= D3/FA). Leverage measures scaled by the market value of the firm have not been employed to avoid the potential inverse correlation effect with the dependent variable, which is unavoidably the market value of the firm. Following Sarma and Rao (1969), total assets instead of fixed assets were experimented with for deflating the firm value and leverage variables, but meaningful results were found when fixed assets were used.

3.1.2. *Control variables*

The expected tax adjusted earnings ($X^{\tau} - \tau R$), is the key control variable in the value–leverage relationship analysis. This is the expected income to the equity holders of the firm had there been no debt, and capitalizing this expected income (earnings) at the rate of return expected by them (stockholders) would give the total market value of the firm, *ceteris paribus*. $X^{\tau}[= X(1-\tau)+\tau R]$ is the expected earnings after taxes and before interest, as it actually comes onto the market for sale to the various security purchasers.[17] $R = iD$, is the total interest on outstanding debt. Preferred dividends have been excluded, because both the number of sampled firms having preferred stocks outstanding and the amount involved were insignificant and no separate disclosure were found for their preferred dividends. Therefore, to avoid measurement error, those firms were excluded from the present analysis. Firms which had a net loss during the study period were excluded, because this particular variable is undefined in this case. As a result, the number of firms was reduced to 45.

Dividend payment. The firm's dividend or payout policy is represented by the payout ratio calculated as the average cash dividend paid over the current and preceding 2 years divided by the average earnings available for common stockholders over the same period. This measure has an advantage over averaging the yearly payout ratios in that the former procedure prevents years with near-zero income from dominating the average.[18] However, in instances of negative payout ratio, it was arbitrarily defined to be 100% since a payout cannot be negative. Since a negative ratio occurs only four or five

[17] Total profit after-tax and interest would be $(X - R)(1 - \tau)$, X = earnings before interest and taxes as defined in the valuation equations in earlier chapters. Hence, profit after tax, but before interest would be $(X - R)(1 - \tau) + R = X(1 - \tau) + \tau R$.

[18] If the earning in any period is zero or close to zero, the ratio becomes extremely large. Computing an average of the payout ratios over several years does not adequately deal with this problem, because one extreme year will still dominate the average. The result of the present measure equals to a weighted average of yearly payout ratios, where each year's weights are equal to the proportion of each year's earnings to the total earnings over the averaging period.

times out of a possible 315 cases, the empirical results are not sensitive to the particular procedure used to remove this anomaly. This difficulty also arises in all of the risk and other measures, which use the observed profitability as denominator.

Following the course of previous valuation studies (Miller and Modigliani, 1966; Sarma and Rao, 1969; Rao and Litzenberger, 1971), a dividend term with unspecified coefficient was added to the structural Equation (7) to control for the valuation effect of dividend payments and let the sample determine its value.

Size of the firm. The valuation functions have been assumed as linear homogenous functions of the independent variables. This homogeneity implies, among other things, that a given proportionate change in the values of all the independent variables leads to an equal proportionate change in the market value of the firm. The previous studies (Gordon, 1962, for example), however, suggest that the true market capitalization rate for the expected earnings of large firms may possibly differ systematically from that of small firms in the same industry, implying a nonlinear relationship between size and market value of firms. Therefore, following the suggestion of Crockett and Friend (1967), Robichek *et al.* (1967), and Rao and Litzenberger (1971), a separate independent variable is included into the valuation model for an explicit accounting of the size effect.

In time series analysis, the general growth of the economy can be reflected in these variables. Therefore, following Gatward and Sharpe (1996), the natural logarithm of the ratio of total assets to the M3 definition of the Australian money supply was used to remove the effects of general growth in the economy over time from this variable.[19] While the functional relationship with the dependent variable should get priority, the log transformation of the variables more closely conforms to the distribution properties of symmetry and normality. Also the cross-section coefficient of variation is greatly reduced with the log transformation.

[19]Since the present analysis involves pooled time series and cross-sectional data, it is necessary to scale total assets by a stock variable, which reflects the general growth in the economy over time. The author has experimented with the natural logarithm of total asset as well as with total value of the firm as measured in the V1 definition used in measuring the debt ratios, but this measure was found to contribute more in terms of adjusted R^2 value in the estimated regressions.

Growth. Considering the Miller and Modigliani (1966) explanation of the corporate earnings growth models, Equation (5) can be modified to express the current market value of the firm including the effects of its potential future growth opportunity, the net present value of the future earnings growth. Although Miller and Modigliani have applied his valuation model in cross-sectional study, the same analogy applies to the dynamic or panel data analysis. Growth has been defined in this study as growth in total assets. Specifically, 3-year moving averages of the yearly growth rates of total assets over the period under study were computed. Since the theory suggests growth opportunities of firms are negatively related to leverage, the indicator of growth is expected to have the same negative relationship.

Systematic or beta risk. Risk has a major impact on the value of a firm.[20] Increased risk lowers the market value of the firm's stock and reduced risk increases the value of its stock *ceteris paribus*. Therefore, Modigliani and Miller (1958, 1963) assumed homogeneity of risk classes. The market model (CAPM) beta has been used to incorporate the effects of risk in determining firm value.[21]

Empirical estimates of the familiar market model beta, b_i, have been used as a surrogate measure for the systematic or theoretical beta risk of the firms. Betas were estimated using monthly returns over a moving 36 months period ending in June for each of the 45 firms in least-square regressions of the following form: $R_{it} = a_i + b_i R_{mt} + e_{it}$ $i = 1, \ldots, 45$; $t = 1, \ldots, 36$; where R_{it} and R_{mt} are *ex post* returns for security i and the market, respectively; e_{it} is the disturbance term in the equation and b_i is the estimate of the theoretical beta for security i. The use of moving 3-year (36-month) estimates considers the beta nonstationarity over the years. Moreover, this estimation does not involve the portfolio beta.

[20] See for example, the fundamental common stock valuation model, the dividend valuation model, that specifies the market value V of a share of stock i as: $V = \text{P0} = \text{D0}(1+g)/(\text{Ri}-g)$, where assumed constant growth in dividend is g, P0 is the price today, right after the receipt of the cash dividend, D0, and Ri is the required rate of return on the stock which is based on the risk-free rate of return and the risk premium appropriate to the stock in question.

[21] Most text books on finance and investment provide a discussion on CAPM and systematic risk.

3.2. *The Statistical Estimation*

Considering the panel nature of data the LSDV or fixed effect model was used for the analysis, as suggested by the Hausman (1978) specification test statistic (H-value is reported in the last row of each result table).[22] Specifically, to evaluate the impact of the debt–equity choice on the value of the firm, Equation (8) has been estimated by using the computer package LIMDEP 7.0, as follows[23]:

$$\begin{aligned}(V/\mathrm{FA})_{it} &= a_1((X^{\tau} - \tau R)/\mathrm{FA})_{it} + a_2\mathrm{DR}_{it} + a_3\beta_{it} + a_4\mathrm{DP}_{it} \\ &\quad + a_5 \ln(\mathrm{TAM3})_{it} + a_6\mathrm{GR}_{it} + \varepsilon_{it} \end{aligned} \tag{9}$$

where $i = 1, \ldots, 45$, the number of firms; $t = 1, \ldots, 7$, the number of periods; V is the total market value of the firm as defined earlier and $(X^{\tau} - \tau R)$ is the expected tax adjusted income available for the equity holders of the firms had there been no debt; GR, ln(TAM3) and DP are respectively the growth rate of total assets measuring the *growth* opportunities of the firms; natural logarithm of the ratio of total assets to M3, the definition of Australian money supply, representing size of the firm, SZ and dividend paid scaled by the income available for the ordinary shareholders, measuring the dividend payments to ordinary shareholder; $\beta =$ the estimates of the familiar market model beta, b_i represents the systematic risk of the firms; FA $=$ net fixed assets of the firms; DR $=$ the debt–equity ratio (as defined earlier), measures the FL used by the firms; and ε is error terms.

Since all of the DRs are based on fixed assets, firm value V and the expected tax adjusted income have been scaled by fixed assets. It should be noted that firms which incurred net losses during any of the study period have been excluded from this analysis, because their value, V, is not defined based on the variable $(X^{\tau} - \tau R)$ for a net loss. As a result, the number of firms was reduced to 45. This may introduce survivorship bias to the estimated coefficient of leverage, making it more positive than the population parameter as profitable firms are expected to have more market value and therefore less leverage than their unprofitable counterparts. Also, profitable firms will use less leverage for future investment according to the pecking order theory. However, free

---

[22] See Hsiao (1986) for the benefits of using panel data for econometric estimation, over the conventional cross-sectional or time-series data approaches.

[23] The individual firm-specific intercept term is suppressed, since the fixed effects estimator uses only the variation within an individual unit's set of observations (i.e., within group estimation).

cash flow theory suggests that profitable firms should use more debt to reduce agency costs. If the free cash flow and the pecking order theories reasonably describe the capital structure decisions of Australian firms, the survivorship bias was expected to be at its minimum and should not have statistically significant effect on the coefficient estimates.

The parameter of interest is the coefficient estimate for the leverage variable, a_2, which measures the overall (both the tax and the nontax) effects of leverage on firm value. A significant positive coefficient is, therefore, the prediction of both the tax-based and the information asymmetry theories under imperfect market conditions. However, the magnitude of the tax-effect is expected to be less (hence could be insignificant) under the imputation tax system than that under the classical tax system. The regression coefficient for the tax adjusted income variable, a_1, may be roughly interpreted as the inverse of the cost of capital to an all-equity firm. The sign and magnitude of the coefficient estimate for the DP (dividend payment) variable, a_4, are unspecified. However, an insignificant a_4 (or equivalently a zero coefficient) would support the Miller and Modigliani (1961) theory that dividend policy is a matter of indifference in an efficient capital market. Since firm size is negatively related to transaction and bankruptcy costs according to the information asymmetry theories, a positive size effect is expected. The coefficient for the β (risk) variable should be negative, while the coefficient for the GR (growth) variable should be positive, in general.

The efficient and unbiased estimates of the coefficients of Equation (9) will depend on how well the resulting equations conform to the assumptions of the classical regression model, specifically the assumptions of serial independence of the disturbance terms, homoskedasticity, and normality. Other likely problems include multicollinearity, spurious correlation, and under/over-specification of the equations. Also, using variables in ratio forms, instead of levels, reduces the likelihood of heteroskedasticity. Considering the panel nature of data, both heteroskedasticity and autocorrelation (serial dependence) corrected estimates of the coefficients have been obtained (where necessary) employing the techniques available in “LIMDEP 7.0”.

Multicollinearity has not been a problem in this study. If the explanatory variables are highly intercorrelated (collinear), it becomes difficult to disentangle the separate effects of each of the explanatory variables on the dependent variable. Although there is no single appropriate method to detect multicollinearity, two counts of tests suggest no such multicollinearity problem in the estimated model. The correlation matrix of explanatory

Table 1. Correlation matrix of the explanatory variables.

Variables	$(X^{\tau} - \tau R)$/FA	GR (growth)	DP (dividend paid)	Risk (beta)	lnTAM3 (size)	TIBD/FA	TD/FA	LD/FA
$(X^{\tau} - \tau R)$/FA	1							
GR (growth)	0.277	1						
DP	−0.302	−0.198	1					
Risk (beta)	−0.076	−0.121	0.132	1				
lnTAM3 (size)	−0.315	−0.042	0.169	0.348	1			
TIBD/FA	0.332	0.189	−0.233	0.032	0.179	1		
TD/FA	0.654	0.32	−0.272	−0.056	−0.115	0.774	1	
LD/FA	0.273	0.207	−0.253	0.023	0.283	0.905	0.661	1

Table 2. Multicollinearity test by auxiliary regression method.

Dependent variables	Goodness of fit, R^2s	
	R^2	Adj-R^2
V1/FA	0.93	0.91
$(X^\tau - \tau R)$/FA	0.81	0.76
GR (growth)	0.61	0.51
Risk (beta)	0.44	0.31
DP (dividend paid)	0.66	0.57
lnTAM3 (size)	0.89	0.88

variables in Table 1 shows no such variable to be highly correlated with the others that can cause multicollinearity problem. "Klien's rule of thumb" also suggests[24] that multicollinearity was not a troublesome problem in the model as all of the R^2s of the auxiliary regressions are less than the $R^2 (= 0.93)$ of the main regression model in Table 2.

Simultaneity between leverage and firm value may be another problem affecting the estimated results. This study examines the effect of leverage on firm value in a structural equation model considering leverage as exogenously determined. Leverage may have value effect due both tax savings and reduction of agency costs.[25] However, it is possible that leverage and firm value are endogenously determined and the estimation of the structural equation suffers from simultaneity bias. This could happen if firms with larger market values have higher growth opportunities requiring greater external financing for further investment in positive NPV projects. Assuming that pecking order theory reasonably describes firms' financing decisions, firms with higher values should be observed to have higher leverage. Thus, if the pecking order and other capital structure theories make equal contributions in explaining the relationship between firm value and leverage, the empirical estimates of the ordinary least square (OLS) coefficients may suffer from simultaneity bias. If there is simultaneity, the methods of two stage least square (2SLS) and instrumental variables give estimators that are consistent and efficient,[26] as does OLS under no simultaneity. But applying 2SLS

[24]See Gujarati (1995).

[25]There are other factors identified by a number of capital structure theories as well.

[26]Palia (2001) and Ruland and Zhou used 2SLS to correct for the simultaneity bias in their models.

and instrumental variables when there is no simultaneity will yield estimators that are consistent but not efficient. Therefore, it is required to check for the simultaneity problem before one discard OLS in favor of the alternatives. To find out which is the case in the model of the present study, Hausman's specification error test was used to the following simultaneous equation system:

$$\begin{aligned}(V/\mathrm{FA})_{it} &= a_1((X^\tau - \tau R)/\mathrm{FA})_{it} + a_2\mathrm{DR}_{it} + a_3\beta_{it} + a_4\mathrm{DP}_{it} \\ &\quad + a_5 \ln(\mathrm{TAM3})_{it} + a_6\mathrm{GR}_{it} + \varepsilon_{it} \qquad (10)\end{aligned}$$

$$\begin{aligned}(\mathrm{DR})_{it} &= b_1(V/\mathrm{FA})_{it} + b_2\beta_{it} + b_3\mathrm{TAX}_{it} + b_4\mathrm{DP}_{it} + b_5 \ln(\mathrm{TAM3})_{it} \\ &\quad + b_6\mathrm{GR}_{it} + v_{it} \qquad (11)\end{aligned}$$

In this system, the two endogenous variables are V/FA (firm value) and DR (leverage). It is assumed that all other variables are exogenous. Technically, the simultaneity problem arises because some of the regressors are endogenous (in this case DR and V/FA in Equations [10] and [11], respectively) and are therefore likely to be correlated with the error terms in the respective equations. This is a violation of one of the assumptions of the classical regression model. As the study regresses value on leverage, the Hausman test for simultaneity between firm value and leverage involves regressing DR (leverage) on all the exogenous variables in the system first in Equation (12) as follows:

$$\begin{aligned}(\mathrm{DR})_{it} &= b_1((X^\tau - \tau R)/\mathrm{FA})_{it} + b_2\beta_{it} + b_3\mathrm{TAX}_{it} + b_4\mathrm{DP}_{it} \\ &\quad + b_5 \ln(\mathrm{TAM3})_{it} + b_6\mathrm{GR}_{it} + w_{it} \qquad (12)\end{aligned}$$

Save the predicted value of DR and w (disturbance term) from estimated regression, then regress V/FA on the predicted values $\hat{\mathrm{DR}}$ and $\hat{w}$ in the following equation:

$$(V/\mathrm{FA})_{it} = a_1\hat{DR}_{it} + a_2\hat{w}_{it} + u_{it} \qquad (13)$$

Now testing the simultaneity in the system is to perform a t-test on the coefficient of $\hat{w}_{it}, \alpha_2$. If it is significant, do not reject the hypothesis of simultaneity and apply 2SLS and instrumental variables methods for consistent estimators of Equation (10), the main model used in the study; otherwise reject the hypothesis of simultaneity and apply OLS. Equation (13) has been estimated to obtain the following results:

$$(V/\mathrm{FA})_{it} = 0.0043\hat{DR}_{it} \quad t = (4.61) + 0.47\hat{w}_{it}\ (t = 1.46), \quad R^2 = 0.81$$

The coefficients of $\hat{w}_{it}$ in all three equations for three types of leverage definitions are found more or less similar and insignificant, suggesting no simultaneity problem in Equation (10).

Sample outliers were identified and removed using primarily the residual method, which is usually suggested for multiple regressions (Maddala, 1993). The OLS regression is relatively robust to moderate departures from normality for large sample sizes (Schmidt, 1976), and given that the other corrections have been made and the reasonably large sample size, the results should not be affected by the nonnormality of the disturbances distribution (Maddala, 1993), if there is any.

Taking into consideration that with panel data it is possible to obtain consistent estimates of the parameters of interest even in the face of correlated omitted variable effects and the fact that the variables/attributes used are those suggested by the extant theories of corporate finance, the estimated models should not suffer from any under/over-specification problems.

4. Regression Results

The main objective of this study was to examine the effects of debt–equity choices on firm value. To estimate these effects, multiple regression analysis was performed based on Equation (8), using the LSDV method separately for each of the three categories of debt mentioned in the previous section. Results of all of these estimates are presented in Table 3.

The last three rows of Table 3 reports the diagnostic test statistics of the estimated regressions such as the Adj-R^2, the Akaike information criterion (AIC), the Hausman H-value, and its probability, respectively. On the basis of the H-values, the models are estimated using the LSDV method. The Adj-R^2 values show that the estimated models are able to explain about 91% (Adj-R^2 ranges from 0.905 to 0.912) of the variations in firm value (although they include individual-specific effects). The AIC is like the Adj-R^2 or minimum variance ($\hat{\sigma}^2$) type criteria, the one with the lowest AIC is chosen.

Table 3 clearly shows that firm value is an increasing function of FL as all of the estimated coefficients for the FL variable are positive. The coefficients are significant for the leverage variables for total, total interest bearing

Table 3. Effects of capital structure choices on total value (V1) of the firms (dependent variable = V1/FA).

Indep. variables ↓	Estimated regressions models		
	1 (TD)	2 (TIBD)	3 (LD)
$(X\tau - \tau R)$/FA	12.86 (16.58)*	13.79 (18.39)*	13.83 (18.49)*
GR (growth)	−0.68 (−1.75)***	−0.62 (−1.59)	−0.60 (−1.55)
Risk (beta)	−0.07 (−1.41)	−0.06 (−1.22)	−0.05 (−1.11)
DP (dividend paid)	0.88 (3.87)*	0.95 (4.20)*	0.99 (4.40)*
lnTAM3 (size)	0.52 (3.17)*	0.56 (3.42)*	0.53 (3.27)*
TD/FA	0.50 (2.74)*		
TIBD/FA		0.39 (2.28)**	
LD/FA			0.48 (2.49)*
Adj-R^2	0.91	0.91	0.91
Ak. Info. Crt.	1.17	1.18	1.18
H-value (Hausman)	34.91 (pv = 0.000)	44.39 (pv = 0.000)	42.59 (pv = 0.000)

Figures in the brackets are t-ratios.
*Significant at 1% level.
**Significant at 5% level.
***Significant at 10% level.

and long-term debts. The results are consistent with the predictions of both Modigliani and Miller's (1963) hypothesis that, with corporate taxes, firm values increase with leverage and the other capital structure theories based on information asymmetry that predict a positive effect of leverage on firm value. Although it is expected that the tax benefit of debt would be less under the Australian full imputation tax system than it would be under the classical tax system, it might still have significant positive effect on firm value. This study, however, does not disentangle the tax effect of leverage from its non-tax effects. Therefore, the results of this study only suggest that corporate leverage has significant positive effect on firm value in Australia (since all the coefficient estimates for the leverage variable in Table 3 are significant at 1–5% level), which could be due to tax or nontax effects or a combined effect of both.

The plausible tax benefit explanation for the result is that, investors prefer leverage on corporate accounts due to tax savings and the financial markets efficiently capture their expectations in Australia. It could be the marginal investors of the sampled firms are nonresident foreign investors who are still under the classical tax regimes, and therefore, value the tax savings.

Australian pension fund, resident Australian tax-exempt and high-tax investors with personal tax rates greater than the corporate tax rate also value tax savings at the corporate level. As discussed earlier, high-tax investors prefer firms with optimal leverage as debt tax shields enhance the potential for the unfranked dividend to be retained and distributed later as capital gain to avoid further tax liabilities (more current tax payments). On the nontax dimension, leverage has an overall positive effect on firm value.

The coefficient estimates for the growth variable are consistently negative in sign and significant in one out of three regressions. A positive coefficient for the growth variable could be attributed to the expected rates of return on new investments above the industry's "normal" cost of capital. The coefficients for the risk variable are also consistently negative as expected, but insignificant in all of the regressions. The insignificance of the coefficients for the risk variable might be a reflection of the sample characteristic that the firms with positive profits have been used in this analysis. Firms with positive profits/earnings should be less risky, *ceteris paribus*, and relatively highly valued in the market.

The regression coefficients for the "dividend paid", DP, variable are consistently positive and significant at the 1% level in all of the regressions. The strong significant positive effect of the DP variable thus suggests a consistency with the traditional thesis that investors have a preference for current dividends. However, the result should be used with caution, because the DP variable was used as a control variable and cannot be claimed to have provided a rigorous test of the effect of dividend policy on valuation.

The regression coefficient for the tax adjusted income/earnings $(X^\tau - \tau R)$ variable varies approximately from 13 to 14 depending on the specification of the regression equations. The difference, however, is meagre. Theoretically, the coefficient is equal to the inverse of the cost of capital or marginal capitalization factor, $1/\rho$, to an all equity firm. In the present models, it roughly estimates the expected marginal capitalization factor for firms, had the firms used no debt. Therefore, the estimated expected cost of equity capital of the sampled firms, ρ, is approximately 0.077 $(1/\rho = 13)$ or 7.7% at the maximum. Although no other recent study was found to compare this with, the estimate seems to be reasonable.

The coefficients for the firm size variable are consistently positive and statistically significant at the 1% level in all of the estimated three regressions, suggesting that larger firms have higher value per unit of investment.

5. Summary and Conclusions

The study provides an empirical analysis of corporate capital structure by examining its effects on the market value of a selection of firms listed at the ASX, adding new empirical evidence to the capital structure literature under the Australian full-imputation tax system.

The results reveal that the value of a firm rises significantly with FL. More specifically, there is a statistically significant positive effect of total, total interest bearing, and long-term FL on the market value of a firm in Australia. These results are obtained when the overall effect (of both tax and nontax factors of leverage) is considered. Thus, the findings of the study suggest that financial policy or corporate leverage matters. The existence of a firm level optimal capital structure is justified on the ground that it has a statistically significant positive impact on the firm's market value.

Although it was not the objective to examine the impact of dividend policy, it has been included as a control variable in the value-leverage relationship model. The findings imply that companies that pay more current dividends issue more debt, and are highly valued in the market, probably because investors prefer current dividends and pay higher prices for the shares of those companies.

The results also imply that by expanding its size or equivalently increasing its volume of investment a firm can increase its market value per unit of investment, *ceteris paribus*. A strongly significant positive size effect was found on the market value of firms. Thus, the result supports the mergers and acquisitions activities of corporations to maximize the market value of a firm.

Due to the nonavailability of theoretically consistent actual data, using proxy variables is a common practice in finance research. This study is no exception to that. So the usual caveats for using proxies apply. The results of the study are specific to the nonfinancial firms listed to the ASX, as the sample excludes financial firms. The data requirements to fit the models necessitated the use of a small number of 45 firms having positive profits during the 10-year study period. Although these are unlikely to constrain the statistical analysis and affect the results, as there were enough data points (450) to analyze, generalizing the results for all firms (other than the sampled firms) becomes limited due to the nature of empirical model used.

Empirical evidence by further such research in other countries will validate this result. Being constrained by the model requirements, this study analyses a small number of firms with positive earnings. Further extension may be directed toward developing a model to include all types of firms in a large sample size to confirm the findings. It is beyond the scope of the present study to examine the reasons for the revealed positive effect of leverage. An immediate extension of this study might, therefore, look into the particular reasons for the effects of leverage, separating the tax and nontax effects. Similar study under different tax systems would provide further support for the findings.

Acknowledgment

The author is indebted to anonymous reviewers of the paper of 13th PBFEAB conference and Dr. Chris Jones, Senior Lecturer, School of Economics, ANU and Professor K. P. Kalirajan, Professor, GRIPS/FASID, National Graduate Institute for Policy Studies, Japan, for their valuable critical comments and suggestions.

References

Arditti, F, H Levy and M Sarnat (1977). Taxes, capital structure and the cost of capital: Some extensions. *Quarterly Review of Economics and Business*, 17, 89–95.

Baxter, D (1967). Leverage, risk of ruin and the cost of capital. *Journal of Finance*, 22(4), 395–403.

Benge, M (1997). Taxes, corporate financial policy and investment decisions in Australia. *The Economic Record*, 73(220), 1–15.

Bowman, J (1980). The importance of a market value measurement of debt in assessing leverage. *Journal of Accounting Research*, 18, 242–252.

Bradley, M, GA Jarrell and EH Kim (1984). On the existence of an optimal capital structure: Theory and evidence. *Journal of Finance*, 39(3), 857–878.

Brealey, RA and SC Myers (2006). *Principles of Corporate Finance*. New York, NY: McGraw-Hill Companies, Inc.

Brennan, M and E Schwartz (1978). Corporate income taxes, valuation, and the problem of optimal capital structure. *Journal of Business* 51, 103-114.

Crockett, J and L Friend (1967). Some estimates of the cost of capital to the electric utility industry, 1954-1957: Comment. *American Economic Review*, 57(5), 1258–1267.

Dammon, R and L Senbet (1988). The effect of taxes and depreciation on corporate investment and financial leverage. *Journal of Finance*, 43, 357–371.

DeAngelo, H and R Masulis (1980). Optimal capital structure under corporate and personal taxation. *Journal of Financial Economics*, 8, 3–29.

Eckbo, BE (1986). Valuation effects and corporate debt offerings. *Journal of Financial Economics*, 15, 119–152.

Engel, E, E Merle and M Edward (1999). Debt equity hybrid securities. *Journal of Accounting Research*, 37, 249–273.

Fama, E and K French (1998). Taxes, financing costs, and firm value. *Journal of Finance*, 53, 819–843.

Flannery, MJ and KP Rangan (2006). Partial adjustment toward target capital structures. *Journal of Financial Economics*, 79, 469–506.

Francis, BB and LL Leachman (1994). A time series approach to exploring aggregate optimal capital structure: Cointegration analysis. *Applied Financial Economics*, 4, 41–54.

Gatward, P and IG Sharpe (1996). Capital structure dynamics with interrelated adjustment: Australian evidence. *Australian Journal of Management*, 21(2), 89–112.

Gordon, MJ (1962). *The Investment, Financing, and Valuation of the Corporation*. Homewood, IL: Irwin.

Graham, JR (2000). How big are the tax effects of debt? *Journal of Finance*, 55(5), 1901–1941.

Graham, JR (1999). Do personal taxes affect corporate financing decisions? *Journal of Public Economics*, 73, 147–185.

Graham, JR (1996). Debt and the marginal tax rate. *Journal of Financial Economics* 41, 41–73.

Gujarati, DN (1995). *Basic Econometrics*, 3rd Ed. New York: McGraw-Hill Inc.

Hamson, DF (1992). An empirical examination of corporate capital structure in Australia and the USA. Unpublished, Department of Commerce, University of Queensland, Australia.

Harris, M and A Raviv (1991). The theory of capital structure. *Journal of Finance*, 46, 297–355.

Haugen, RA and LW Senbet (1986). Corporate finance and taxes: A review. *Financial Management*, 15, 5–21.

Hausman, J (1978). Specification test in econometrics. *Econometrica*, 46, 1251–1271.

Hsiao, C (1986). *Analysis of Panel Data*. Cambridge: Cambridge University Press.

Jensen, MC and W Meckling (1976). Theory of the firm: Managerial behaviour, agency costs, and ownership structure. *Journal of Financial Economics*, 3, 305–360.

Jensen, MC (1986). Agency costs of free cash flow, corporate finance and takeovers. *American Economic Review*, 76, 323–329.

Kemsley, D and D Nissim (2002). Valuation of the debt tax shield. *Journal of Finance*, 57, 2045–2073.

Kim, EH (1978). A mean-variance theory of optimal capital structure and corporate debt capacity. *Journal of Finance*, 33, 45–64.

Kraus, A and R Litzenberger (1973). A state-preference model of optimal capital structure. *Journal of Finance*, 28, 911–921.

Leary, MT and MR Roberts (2005). Do firms rebalance their capital structure? *Journal of Finance*, 60, 2575–2619.

Long, M and I Malitz (1985). Investment patterns and financial leverage. In *Corporate Capital Structure and Financial Intermediaries*, B Friedman (ed.) Chicago: The University of Chicago Press.

Mackie-Mason, JK (1990). Do taxes affect corporate financing decisions. *Journal of Finance*, 45, 1471–1493.

Maddala, GS (1993). *The Econometrics of Panel Data*, Vol. I, pp. xii-xxi. Cambridge, UK: University Press.

Masulis, RW (1980). The effect of capital structure changes on security prices: A study of exchange offers. *Journal of Financial Economics*, 8, 139–177.

Masulis, RW and AW Korwar (1986). Seasoned equity offerings: An empirical investigation. *Journal of Financial Economics*, 15, 91–118.

Mehrota, V, W Mikkelson and M Partch (2005). Do managers have capital structure targets? Evidence from corporate spinoffs. *Journal of Applied Corporate Finance*, 7, 18–25.

Miller, MH and F Modigliani (1961). Dividend policy, growth and the valuation of shares. *Journal of Business*, 34, 411–433.

Miller, MH and F Modigliani (1966). Some estimates of the cost of capital to the electric utility industry, 1954–57. *American Economic Review*, 56(3), 333–391.

Miller, MH (1977). Debt and taxes. *Journal of Finance*, 32, 261–273.

Modigliani, F and MH Miller (1958). The cost of capital, corporation finance and the theory of investment. *American Economic Review*, 48, 261–297.

Modigliani, F and MH Miller (1963). Corporation income taxes and the cost of capital: A correction. *American Economic Review*, 53, 433–443.

Myers, SC (1984). The capital structure puzzle. *Journal of Finance*, 39, 575–592.

Palia, D (2001). The endogeneity of managerial compensation in firm valuation: A solution. *Review of Financial Studies*, 14(3), 735–764.

Parrino, R and M Weisbach (1999). Measuring investment distortions arising from stockholders-bondholders conflicts. *Journal of Financial Economics*, 52, 3–42.

Peirson, G, R Brown, S Easton and P Howard (1998). *Peirson and Bird's Business Finance*, 7th ed., pp. 489–490. Roseville, Australia: McGraw-Hill Book Company Australia Pty Limited.

Peyer, UC and A Shivdasani (2001). Leverage and internal capital markets: Evidence from leverage recapitalizations. *Journal of Financial Economics*, 59, 477–515.

Rao, CU and RH Litzenberger (1971). Leverage and the cost of capital in a less developed capital market: Comment. *Journal of Finance*, 26(3), 777–782.

Robicheck, AA, JG McDonald and RC Higgins (1967). Some estimates of the cost of capital to the electric utility industry, 1954–57: Comment. *American Economic Review*, 57(5), 1278–1288.

Ruland, W and P Zhou (2005). Debt, diversification, and valuation. *Review of Quantitative Finance and Accounting*, 25, 277–291.

Sarma, LVLN and KSH Rao (1969). Leverage and the value of the firm. *Journal of Finance*, 24(4), 673–677.

Schmidt, R (1976). Determinants of corporate debt ratios in Germany. *European Finance Association 1975 Proceedings*. North Holland, Amsterdam.

Scholes, M and GP Wolfson (1992). *Taxes and Business Strategy: A Planning Approach*. Englewood Cliffs, NJ: Prentice Hall.

Scott, J Jr (1976). A theory of optimal capital structure. *Bell Journal of Economics and Management Science*, 7, 33–53.

Shyam-Sunder, L and S Myers (1998). Testing static trade-off against pecking-order models of capital structure. *Journal of Financial Economics*, 51, 219–244.

Stulz, RM (1990). Managerial discretion and optimal financing policies. *Journal of Financial Economics*, 26, 3–27.

Titman, S and R Wessels (1988). The determinants of capital structure choice. *Journal of Finance*, 43, 1–19.

Trezevant, R (1992). Debt financing and tax status: Test of the substitution effect and the tax exhaustion hypothesis using firms responses to the Economic Recovery Tax Act of 1981. *Journal of Finance*, 45, 1557–1568.

Twite, G (2001). Capital structure choices and taxes: Evidence from the Australian dividend imputation tax system. *International Review of Finance*, 2(4), 217–234.

Vermaelen, T (1981). Common stock repurchases and market signaling: An empirical study. *Journal of Financial Economics*, 9, 139–183.

Welch, I, (2004). Capital structure and stock returns. *Journal of Political Economy*, 112, 106–131.

Weiss, LA (1990). Bankruptcy resolution: Direct costs and violation of priority claims. *Journal of Financial Economics*, 27, 285–314.

This page intentionally left blank

Chapter 12

The Momentum and Mean Reversion of Nikkei Index Futures: A Markov Chain Analysis

Ke Peng
University of Bradford, UK

Shiyun Wang
Southwestern University of Finance and Economics, P. R. China

This chapter finds that the intraday Nikkei futures returns exhibit different patterns of momentum or mean reversion when changing observation intervals. Using a Markov chains methodology, a significant return momentum was found at 1-min observation interval. However, a significant return mean reversion was found at 10-min observation interval. This switching pattern of momentum to mean reversion is robust to intraday seasonality. Further, the sources that contribute to the high-frequency momentum and mean reversion are explored and it is concluded that large limit orders and the bid-ask effect can play the role.

Keywords: Intraday patterns; momentum; mean reversion; Markov chains.

1. Introduction

In the high-frequency literature, the high-frequency successive returns are normally characterized as negatively first-order auto-correlated. This mean reversion pattern has been supported by large volumes of empirical work. For example, Miller *et al.* (1994), using S&P 500 index futures, report a negative autocorrelation (−0.029) for the 15-min price changes. Ederington and Lee (1995) document 87%, 82%, and 61% price reversals for interest rate futures, Eurodollar, and DM futures, respectively, at tick-by-tick intervals. The high-frequency mean reversion is believed to be mainly due to the bid-ask "bounce", as "the price movement attributable to new information may be small relative to the size of the bid-ask spread" (Miller *et al.*, 1994).

However, in this chapter, using Nikkei index futures intraday data, it is found that the pattern of high-frequency serial correlation behavior largely depends on what observation intervals are used. While the mean reversion pattern is found when using 10-min observation interval, this is not the case when 1-min returns are used, where a continuation pattern is documented.

This switching pattern of serial correlation along the observation intervals is similar to what has been documented in the low-frequency literature when daily or monthly intervals are used.[1]

In the low-frequency literature many studies have been carried out to explain the documented disparities of the univariate time-series properties of security returns. Some researchers attribute the disparities to the methodologies employed (e.g., Richardson and Smith, 1994), whereas others attribute those differences to the underreaction and overreaction hypotheses at different time horizons (see, e.g., Hong and Stein, 1999). While most of the previous theories have provided sound explanations for some aspects of the low-frequency security return properties, few studies have investigated the switching pattern of successive returns in high-frequency when the overreaction/underreaction can hardly play a key role in ultra-short time. This chapter is an effort to explore the serial behavior of successive returns at different observation intervals at high frequency.

This chapter is organized as follows. Section 2 describes the data and methodologies. Section 3 gives the empirical results and discussion. Section 4 offers conclusions.

2. Data and Methodology

2.1. *Data*

The data cover January 1, 1997 to December 31, 2000. The primary data set consists of a time-stamped transaction record including transaction, bid, and ask prices. The Nikkei 225 index futures contracts are traded following the March, June, September, and December cycles. The nearest contracts are used as they are almost always the most heavily traded. The data are provided by Singapore Exchange Limited (SGX). Using the data set, 1-, 5-, 10-, and 20-min interval subsequences of the futures trading prices and bid-ask quotes

[1]It is well known that stock prices exhibit different patterns of momentum, mean reversion, or even random walks along different return horizons. For example, momentum (persistence) is more often documented for medium horizons (3–12 months, see, e.g., Jegadeesh, 1990), whereas contrarian (mean reversion, or reversals) is more documented for short (daily, weekly, and monthly, see, e.g., Lo and MacKinlay, 1990) or for long horizons (3–5 years, see, e.g., French and Fama, 1988).

are constructed. The data were broken into two subsamples, 1997–1998 and 1999–2000, to check whether there are substantial changes at different times.

Table 1 gives the descriptive statistics of Nikkei 225 index futures for 1- and 10-min interval data.

The results for the two subsamples are quite similar. At 1-min interval, only about 5% out of the intervals have no trades, and about one-third of the intervals have four to six trades, with an average of five trades per minute. At 10-min interval, the average number of trades per 10-min is about 50. Part 2 gives descriptive statistics of successive price changes and log price changes. As expected, the average 1- and 10-min price and log price changes are very small with the log price changes close to 0. The successive price changes follow a roughly symmetric but not normal distribution with the kurtosis much higher than 3 (leptokurtic).

2.2. *Methodology*

The traditional serial correlation tests utilize the restrictive assumptions of normality distribution and linearity property of successive returns that are easily violated. As stated by McQueen and Thorley (1991), "the variance ratio and regression tests assume linearity, yet the alternatives of fads or rational speculative bubbles suggest the possibility of nonlinear patterns in returns". Richardson and Smith (1994) also argue that the importance of the normality assumption (of variance-ratio statistics) makes the results difficult to interpret outside that context. A Markov chain methodology has been employed to overcome those limitations by the traditional tests. Markov chain methodology has advantages in the studies of price behavior because it assumes neither linearity nor normality of returns.[2]

The transition probability matrix was constructed by defining a two-state Markov chain $\{I_t = 1, 0\}$ as follows,

$$I_t = \begin{cases} 1, & \text{if}, \quad r_t > \bar{r} \\ 0, & \text{if}, \quad r_t \le \bar{r} \end{cases} \tag{1}$$

[2]Some other authors have employed the Markov chain methodologies in the study of successive price behavior. For example, McQueen and Thorley (1991), by letting one state represent high returns and the other represent low returns, allege that low (high) returns tend to follow runs of high (low) returns in the post-war period.

Table 1. Descriptive statistics of Nikkei 225 index futures during 1997–2000.

Part 1. Number of trades per 1 and 10 min

Number of trades		0	1–3	4–6	7–9	>10	Average trades per min
Panel A: Number of trades per 1 min							
1997–1998	1-min intervals	6676	40565	55696	36970	14888	
	%	4.3%	26.2%	36.0%	23.9%	9.6%	5.321
1999–2000	1-min intervals	8153	46032	54326	32724	13510	
	%	5.3%	29.8%	35.1%	21.2%	8.7%	5.016

Panel B: Number of trades per 10 min

Number of trades		0	1–30	31–60	61–90	>90	Average trades per 10 min
1997–1998	10-min intervals	1	3207	7158	4918	880	
	%	0.0%	19.8%	44.3%	30.4%	5.4%	50.95
1999–2000	10-min intervals	1	3862	7315	4253	733	
	%	0.0%	23.9%	45.3%	26.3%	4.5%	48.02

Part 2. Price changes per 1 and 10 min

		Mean	Std	Min	Max	Skew	Kurtosis	Count
Panel C: Price changes per 1 min								
1997–1998	Price changes	−0.035	12.4	−150	200	0.001	8.59	
	(Log)	0	0.001	−0.009	0.012	0.003	9.31	147706
1999–2000	Price changes	−0.032	9.89	−245	130	0.015	9.42	
	(Log)	0.000	0.001	−0.013	0.007	0.013	8.49	146177
Panel D: Price changes per 10-min								
1997–1998	Price changes	−0.478	40.798	−380.000	615.000	0.052	10.543	
	(Log)	0.000	0.002	−0.023	0.038	0.077	11.143	143278
1999–2000	Price changes	−0.342	31.724	−495	495	0.007	7.569	
	(Log)	0.000	0.002	−0.034	0.034	0.005	7.669	141749

1-min intervals is the total number of 1-min observation intervals within which there were trades taking place, as shown in the first row; % is the percentage of time in the whole period; Price changes is the price changes during the specific observation intervals; Log is the logarithm price changes within the specific observation intervals.

where $\{r_t\}$ is the intraday continuous compounded return series and $\overline{r}$ is the mean of intraday returns. In the present analysis, it is important to set states in a way that each state has roughly the same chance to appear within the same trading day, otherwise the comparison among different transition probabilities will tend to favor certain states. The use of the mean intraday return as the cutoff-point is because $\{r_t\}$ follows an approximately symmetric distribution with the mean around 0 (see Table 1).

Following Wang *et al.* (1999), a Markov chain long-run probability distribution is used to examine the transitions between successive returns. If prices follow the random walk, then the Markov chain long-run transition probability should be without structure, that is, the transition probabilities should be equal for all prior information.[3]

A second-order Markov chain is constructed below[4]

The transition probability matrix

Previous states		Current states	
		1	0
1	1	$P_{11,1}$	$P_{11,0}$
1	0	$P_{10,1}$	$P_{10,0}$
0	1	$P_{01,1}$	$P_{01,0}$
0	0	$P_{00,1}$	$P_{00,0}$

where $P_{xy,z}$ stands for the probability of obtaining a state z, conditional on two prior state of x, y successively. For the successive three states, $x \to y \to z$, there are two transitions involved, that is, transition $x \to y$ and transition $y \to z$. In the long-run probability test, $P_{i,j}$ is defined as the probability of a transition i $(x \to y)$ followed by a transition j $(y \to z)$. The long-run probability of a transition $y \to z$ can be obtained by,

$$\pi_j = \sum_{i=1}^{T} \pi_i P_{i,j} \tag{2}$$

$$\sum_{j=1}^{T} \pi_j = 1 \tag{3}$$

[3]Strictly speaking, the test of no pattern between successive returns is equivalent to the test of the Martingale model of which the random walk model can be viewed as a special case.

[4]The second-order Markov chains have been employed by various authors such as Neftci (1984), and McQueen and Thorley (1991). The second-order Markov chains have been used for the derivation of the long-run transition probabilities.

where π_j is the long-run probability of a transition $j(y \to z)$. j can be a transition of 11, 10, 01, and 00. T is the total number of transitions ($T = 4$). Under the null hypothesis of random walk, $\pi_{11} = \pi_{01} = \pi_{10} = \pi_{00} = 1/4$. The test statistic is constructed as,

$$x^2 = N \sum_{j=1}^{T} \frac{(\pi_j - 1/4)^2}{1/4} \tag{4}$$

where N is the total number of observations. Under the null hypothesis of no structure in the transition probabilities, x^2 follows an asymptotic chi-square distribution with $(T - 1)$ degrees of freedom.

3. Empirical Results

3.1. *Mean reversion or continuation?*

To compare the serial behavior of successive returns at different observation intervals, the 1-, 5-, 10-, 20-min interval returns were obtained and their first-order autocorrelations are reported.

The results in Table 2, Panel A show that the first-order autocorrelation changes along time intervals. For 1997–1998, there is significant persistence of successive returns if returns are measured using intervals of less than 5 min. However, this pattern of persistence changes to a significant mean reversion if returns are calculated at longer than 5-min intervals. The results using the data of 1999–2000 confirm the changing pattern of serial behavior of returns.

Table 2. First-order autocorrelation at different observation intervals.

Periods	1-min	5-min	10-min	20-min
Panel A: Log returns				
1997–1998	0.019*	0.020*	−0.016*	−0.055*
1999–2000	−0.009*	−0.005	−0.055*	−0.055*
1997–2000	0.009*	0.011*	−0.031*	−0.055*
Panel B: Arrivals of buy/sell orders				
1997–1998	0.026*	0.006*	−0.001	−0.003
1999–2000	0.024*	0.007*	−0.003	−0.000
1997–2000	0.024*	0.006*	−0.002	−0.002

*Indicates rejection of zero autocorrelation at 1% significant level.

At 1-min interval, there is a significant price reversion. However, at 5-min interval, this reversion becomes insignificant, and at intervals longer than 10-min, the mean reversion becomes significant again. The results using the whole period of 1997–2000 are similar to that of 1997–1998 — a momentum at shorter observation intervals (1-, 5-min) and a continuation at longer observation intervals (10-, 20-min). Table 3 gives the results by Markov chain long-run probability tests.

The long-run probabilities predict the transition patterns between successive returns contingent on the available information. At 1-min interval, it is observed that from both subsamples and the whole sample, that the probability of continuation ($\pi_{11} + \pi_{00}$) is significantly higher than that of reversion ($\pi_{10} + \pi_{01}$), where π_{10} is the long-run probability of a transition from state 1 (higher return) to state 0 (lower return). Alternatively, π_{10} can be viewed as the proportion in the whole time period when the transition takes the pattern of $1 \rightarrow 0$ (higher return transit to lower return). For the two subsamples, the

Table 3. Markov chain long-run transition probability distribution between successive returns.

		π_{11}	π_{10}	π_{01}	π_{00}	χ_1^2 stat	$\pi_{11} + \pi_{00}$	$\pi_{10} + \pi_{01}$	χ_2^2 stat
1-min interval	1997–1998	0.270	0.237	0.237	0.257	475.6*	0.527	0.473	423.8*
	1999–2000	0.268	0.236	0.236	0.260	455.3*	0.528	0.472	436.5*
	1997–2000	0.269	0.236	0.236	0.258	926.7*	0.527	0.473	860.3*
5-min interval	1997–1998	0.259	0.248	0.248	0.245	12.7*	0.504	0.496	1.8
	1999–2000	0.253	0.249	0.249	0.249	1.400	0.502	0.498	0.6
	1997–2000	0.256	0.248	0.248	0.247	11.000	0.503	0.497	2.3
10-min interval	1997–1998	0.255	0.255	0.255	0.236	15.9*	0.491	0.509	5.1
	1999–2000	0.240	0.262	0.262	0.237	31.7*	0.477	0.523	31.5*
	1997–2000	0.247	0.258	0.258	0.236	38.1*	0.483	0.517	31.0*
20-min interval	1997–1998	0.251	0.257	0.257	0.236	7.900	0.487	0.513	4.8
	1999–2000	0.247	0.259	0.259	0.235	10.400	0.482	0.518	8.6*
	1997–2000	0.249	0.258	0.258	0.236	17.9*	0.485	0.515	13.2*

States {1, 0} is defined as whether the current return is {higher, lower} than the average. π_{ij} is the long-run probability of state i followed by state j. The χ^2 statistic serves to test the null hypothesis of no pattern between successive returns. χ_1^2 stat is the chi-square statistic to test the null hypothesis H_0: $\pi_{11} = \pi_{10} = \pi_{01} = \pi_{00}$, and χ_2^2 stat is the chi-square statistic to test the null hypothesis H_0: $(\pi_{11} + \pi_{00}) = (\pi_{10} + \pi_{01})$.

*Indicates rejection of the null hypothesis at 5% significant level.

proportion of continuations is 52.7% and 52.8%, and the proportion of reversions is 47.3% and 47.2%, respectively. The chi-square tests strongly reject the null hypothesis of equal proportion for all transitions. It is concluded, based on 1-min interval returns, that there is a continuation pattern between successive returns.

However, at 5-min internal, there is no clear pattern between successive returns. The null hypotheses of equal probability of continuations and reversions cannot be rejected at 1% significant level for all samples. For example, during 1997–2000, there is roughly a half chance (50.3%) for continuations and a half chance (49.7%) for reversions. At 10-min interval, the pattern changes again. For both subsamples and the whole sample, there is a significant tendency for price reversion ($\pi_{10} + \pi_{01} > \pi_{11} + \pi_{00}$), which is in contrast to that at 1-min interval. At 20-min interval, the pattern of mean reversion continues.

3.2. *Rejection or acceptance of random walks?*

In the previous section, it has been shown that changing the observation intervals will change the inference to the pattern of the successive returns. When the return horizon increases, a price reversion, a continuation, or even a random walk may be obtained at different times. As shown in Table 3, the random walk model of no pattern between successive returns which is rejected for both 1- and 10-min intervals cannot, however, be rejected at 5-min interval. We conclude that the choice of return intervals plays a role in the test of the random walk model of successive returns.

3.3. *Sources of the mean reversion and continuation of intraday index futures returns*

Roll (1984) shows that under the efficient market hypothesis, the observed price changes will be first-order negatively auto-correlated, due to the so-called bid-ask bounce, even the "true" price changes are serially uncorrelated. This pattern of negative autocorrelation is unconditional on the observation interval used. Following this theory, it is suggested that the negative first-order autocorrelation in observed price changes is more likely for extremely short intervals where the price movement attributable to new information may be small relative to the size of the bid-ask spread (see, e.g., Miller *et al.*, 1994). However, the bid-ask bounce theory fails to explain the intraday results. Although a reversion pattern was found at 10-min interval, there is a significant continuation pattern at 1-min interval.

Limit orders can play a role in the high-frequency return behavior. Under the random walk with barriers theory, if there are limit orders, when the price falls the buy orders become effective, and when the price rises the sell orders become effective, resulting in an observed negative autocorrelation between consecutive price changes (see Granger and Morgenstern, 1970, p. 153). On the other hand, limit orders can also lead to a positive correlation between consecutive returns at higher frequency observations. This is because that when there are limit orders, a large order may move along the limit order book, that is, buy orders follow buys and sells follow sells, resulting in a positive autocorrelation of returns. As stated by Goodhart and O'Hara (1997) "such effects (positive autocorrelation) would, one would expect, be more prominent the higher the frequency of the data."

Following Hasbrouck and Ho (1987), a buy/sell indicator series was constructed to investigate the arrivals of buy/sell orders. A "buy" was classified for any order that transacted at a price greater than the midpoint of the current bid and ask prices, and the indicator was assigned a value of $+1$, whereas a "sell" was any transaction at a price lower than the midpoint and the indicator was assigned a value of -1. If the transaction occurred at the midpoint, the indicator was assigned a value of 0. The first-order autocorrelations of the indicator series at different intervals are presented in Panel B of Table 2. At 1-min interval, orders are positively auto-correlated for both the subsamples and the whole sample. For example, the first-order autocorrelation during 1997–2000 is significantly positive, 0.24, consistent to the positive autocorrelation of returns at 1-min interval. At 10-min interval, the first-order autocorrelations for both subsamples and the whole samples are insignificantly negative, -0.001, -0.003, and -0.002, respectively. Thus, the negative first-order autocorrelation of returns at 10-min interval can be due to the combined effect of the arrivals of the orders and the bid-ask bounce.

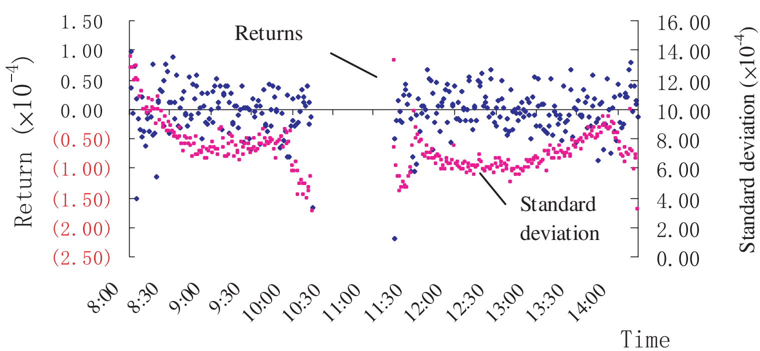
3.4. *Seasonality*

The intraday returns are often found to follow a U-shaped pattern, reaching a high near market open and rising again near market close (see, e.g., Harris, 1986; Jain and Joh, 1988). If there is a U-shaped pattern in the present data sample, the return transition patterns in the opening and closing hours should be significantly different from those in the other hours, as successive transitions should happen more frequently between higher returns in the opening and closing hours than do in the other time. Therefore, the "bouncing" pattern of mean reversion or continuation may be different during different hours of the day.

Figure 1 depicts the intraday minute-by-minute average returns and the volatilities for the two subsamples. Although the intraday volatilities clearly follow a U-shaped pattern, this is not the case for the intraday returns, which fluctuate around 0 at almost equal frequency.

Table 4 reports the hourly long-run transition probability distribution between successive returns. As the results for the two subsamples and the whole sample are very similar, only the results were reported for the whole

A. 1997–1998



B. 1999–2000

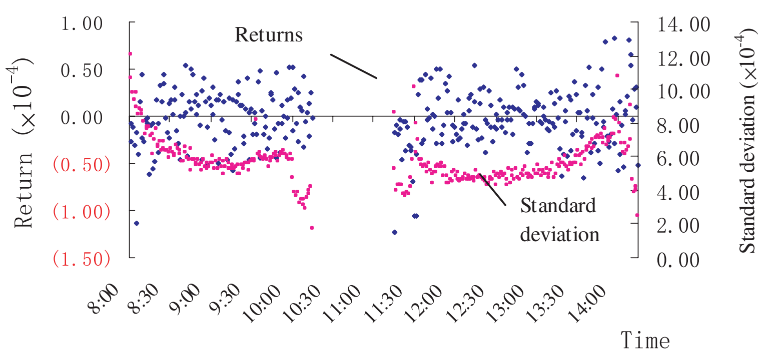


Figure 1. Intraday minute-by-minute average returns and standard deviations.

Table 4. Hourly long-run transition probability distribution between successive returns, 1997–2000.

	Hour	π_{11}	π_{10}	π_{01}	π_{00}	χ_1^2 stat	$\pi_{11}+\pi_{00}$	$\pi_{10}+\pi_{01}$	χ_2^2 stat
1-min interval	08:00–09:00 h	0.270	0.237	0.237	0.257	175.2*	0.526	0.474	155.9*
	09:00–10:00 h	0.265	0.239	0.239	0.257	115.8*	0.522	0.478	107.3*
	10:00–10:15 h	0.273	0.228	0.228	0.271	86.7*	0.545	0.455	86.6*
	11:15–12:00 h	0.270	0.233	0.233	0.264	176.2*	0.534	0.466	173.6*
	12:00–13:00 h	0.271	0.235	0.235	0.258	203.6*	0.529	0.471	187.0*
	13:00–14:00 h	0.268	0.236	0.236	0.260	172.3*	0.527	0.473	165.3*
	14:00–14:15 h	0.272	0.244	0.244	0.239	34.0*	0.511	0.489	6.2*
10-min interval	08:00–09:00 h	0.255	0.262	0.262	0.222	16.9*	0.477	0.523	8.4*
	09:00–10:00 h	0.242	0.253	0.253	0.252	1.9	0.493	0.507	0.9
	10:00–10:15 h	0.177	0.261	0.261	0.302	31.8*	0.479	0.521	1.7
	11:15–12:00 h	0.255	0.251	0.250	0.244	0.7	0.499	0.501	0.1
	12:00–13:00 h	0.261	0.258	0.258	0.223	19.7*	0.484	0.516	5.0*
	13:00–14:00 h	0.238	0.266	0.266	0.230	20.1*	0.468	0.532	19.5*
	14:00–14:15 h	0.260	0.268	0.268	0.204	11.3*	0.464	0.536	5.1*

States $\{1, 0\}$ is defined as whether the current return is {higher, lower} than the average. π_{ij} is the long-run probability of state i followed by state j. The χ^2 statistic serves to test the null hypothesis of no pattern between successive returns. χ_1^2 stat is the chi-square statistic to test the null hypothesis $H_0 : \pi_{11} = \pi_{10} = \pi_{01} = \pi_{00}$, and χ_2^2 stat is the chi-square statistic to test the null hypothesis $H_0 : (\pi_{11} + \pi_{00}) = (\pi_{10} + \pi_{01})$.

*Indicates rejection of the null hypothesis at 5% significant level.

sample. At 1-min interval, significant continuations of returns ($\pi_{11} + \pi_{00} > \pi_{10} + \pi_{01}$) are observed in all different hours from opening to closing. However at 10-min interval, the mean reversion pattern ($\pi_{10} + \pi_{01} > \pi_{11} + \pi_{00}$) exists in all hours, although this pattern is not significant at 5% level in some hours. We conclude that the switching patterns of a continuation at 1-min interval and a mean reversion at 10-min interval are robust to the intraday seasonality.

4. Summary and Conclusions

In this chapter, it is presented that the choice of return intervals plays an important role in the pattern of successive returns. Using a Markov chains methodology that does not require any presumption of the distribution of returns, a continuation pattern of successive return at 1-min interval, a random walk at 5-min interval, and a price reversion at 10-min interval were documented. This bouncing pattern from continuation to mean reversion of successive returns is robust to intraday seasonality. Further, the arrivals of the orders are examined and found that at 1-min interval, orders are positively first-order auto-correlated, that is, buy orders tend to be followed by buys and sells by sells. At 10-min interval, orders do not show any significant autocorrelation. The present findings are consistent with the limit order theory that at very high frequent observations, a large order may move along the limit order book, resulting in order continuations, and subsequently positive return autocorrelations. At longer observation intervals such as 10-min interval, the effect of large order decays and the random walk with barriers theory and the bid-ask effect can explain the negative autocorrelation of successive returns.

Acknowledgment

The author is grateful for comments and advice from Gishan Dissanaike, an anonymous referee, and participants at various seminars and conferences.

References

Ederington, LH and JH Lee (1995). The short-run dynamics of the price adjustment to new information. *Journal of Financial and Quantitative Analysis*, 30(1), 117–134.

French, KR and EF Fama Sr (1988). Permanent and temporary components of stock prices. *Journal of Political Economy*, 96(2), 246–273.

Goodhart, CAE and M O'Hara (1997). High frequency data in financial markets: Issues and applications. *Journal of Empirical Finance*, 4(2–3), 73–114.

Granger, CWJ and O Morgenstern (1970). *Predictability of Stock Market Prices*. Lexington, MA: Heath Lexington.

Harris, L (1986). A transaction data study of weekly and intradaily patterns in stock returns. *Journal of Financial Economics*, 16(1), 99–117.

Hasbrouck, J and TSY Ho (1987). Order arrival, quote behavior, and the return-generating process. *Journal of Finance*, 42(4), 1035–1048.

Hong, H and JC Stein (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance*, 54(6), 2143–2184.

Jain, PC and G-H Joh (1988). The dependence between hourly prices and trading volume. *Journal of Financial and Quantitative Analysis*, 23(3), 269–283.

Jegadeesh, N (1990). Evidence of predictable behavior of security returns. *Journal of Finance*, 55(3), 881–898.

Lo, AW and AC MacKinlay (1990). When are contrarian profits due to stock market overreaction? *Review of Financial Studies*, 3(2), 175–205.

McQueen, G and S Thorley (1991). Are stock returns predictable? A test using Markov chains. *Journal of Finance*, 46(1), 239–263.

Miller, MH, J Muthuswamy and RE Whaley (1994). Mean reversion of Standard & Poor's 500 index basis changes: Arbitrage-induced or statistical illusion? *Journal of Finance*, 49(2), 479–513.

Neftci, SN (1984). Are economic time series asymmetric over the business cycle? *Journal of Political Economy*, 92(2), 307–328.

Richardson, MP and T Smith (1994). A unified approach to testing for serial correlation in stock returns. *Journal of Business*, 67(3), 371–399.

Roll, R (1984). A simple implicit measure of the bid/ask spread in an efficient market. *Journal of Finance*, 39(4), 1127–1139.

Wang, S, KG Lim and CW Chang (1999). A new methodology for studying intraday dynamics of Nikkei index futures using Markov chains. *Journal of International Financial Markets, Institutions and Money*, 9(3), 247–265.

This page intentionally left blank

Index