Risk Management in Banking

FOURTH EDITION

JOËL BESSIS

WILEY

"For bank managers, risk managers, and students of financial risk management in banking, this indispensable guide provides a comprehensive coverage of all related topics, from balance sheet management to market and credit models. The text groups all that they need to know to understand the techniques, the practices and the main models, to navigate by themselves in the ever-evolving and highly technical literature on risks, and towards further specialisations as needed. There is so much one can talk about, that many authors have a narrow and specialized approach, which does not help readers to appreciate the full scope of the field. On the other hand, this book stays focused on risk management while addressing all angles of the field. The book is an extremely valuable contribution to the knowledge of risk management."

Christian Jimenez, Regional Director, PRMIA Paris

"This comprehensive volume is ideal for finance professionals who aspire to deepen their knowledge of risk management in the banking sector. In an ever ever-changing environment of financial services, this entirely revised edition provides the keys to the sophistication and the technicalities of risk management techniques and models. With a combination of intellectual rigor and pragmatic application, the text integrates concisely a wide body of work, avoiding the narrower approaches of specialists. Overall, Joël Bessis offers a balanced, extensive yet relevant coverage of the far-reaching expertise needed to control and supervise risks in financial institutions."

Elie Heriard-Dubreuil, Senior Director, Global Supranationals, Sovereign Ratings, Standard & Poor's Rating Services

"Understanding how banking firms operate and how risk models are designed and implemented has now become central in modern finance. This book provides a concise overview of these topics and combines analytical rigor with relevance and practices inspired by the academic and professional experiences of the author. A must-read for all students and practitioners who need to have a practical knowledge of how risk management is conducted and will evolve in banking."

Christophe Perignon, Associate Professor of Finance, HEC Paris

"The author's balanced profile, combining academic background with the experience of professional life, shows up in the 'how to' approach for implementing models, techniques and processes, accessible to non-specialists, in the real world. A truly fantastic book and an enduring and worthy part of the financial markets literature."

Professor Moorad Choudhry, Department of Mathematical Sciences, Brunel University, and former Treasurer, Corporate Banking Division, Royal Bank of Scotland

"In a context where banking firms face new challenges of risk management and risk regulations, which have a direct influence of how banks develop, it is more important than ever that all professionals, bank managers and risk managers alike, have a comprehensive view of the diversity of risk management contributions, from asset-liability management to banks systems and risk models. This concise and applied coverage of risk management in banks will enable professionals to effectively to master the technicalities of the field and form educated judgments on what risk managers and risk engineers do, which impacts their own roles."

Patrick Legland, Global Head of Research and Member of Global Capital Markets Executive Committee, Société Générale

Risk Management in Banking

FOURTH EDITION

Joël Bessis

WILEY

This edition first published 2015

Second edition published 2007, third edition published 2011 by John Wiley & Sons, Ltd

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Library of Congress Cataloging-in-Publication Data

Bessis, Joël.
[Gestion des risques et gestion actif-passif des banques. English]
Risk management in banking / Joël Bessis. — Fourth edition.
pages cm
Includes bibliographical references and index.
ISBN 978-1-118-66021-8 (paperback)
1. Bank management. 2. Risk management. 3. Asset-liability management. I. Title.

HG1615.B45713 2015
332.1068'1—dc23 2015010257

A catalogue record for this book is available from the British Library.

ISBN 978-1-118-66021-8 (paperback) ISBN 978-1-118-66018-8 (ebk)
ISBN 978-1-118-66019-5 (ebk)

Cover Design: Wiley
Building detail ©Shutterstock/K images; Background ©iStock.com/svariophoto

Set in 10/12pt TimesNewRomanMTStd by Thomson Digital, Noida, India
Printed in Great Britain by TJ International Ltd, Padstow, Cornwall, UK

Contents

Foreword

It is a truism that while every financial crisis is different, errors made in risk management in banking are timeless. I remember well reading an article from 1994, published by the Federal Reserve Bank of Minneapolis, that highlighted mistakes made by both large and small banks in the US banking crisis of 1980–81. Every single one of the recommendations made by the authors of that paper would have been relevant and applicable to banks that crashed in 2008–09. An ineffective risk management framework, coupled with an aggressive asset origination policy, will always combine to bring badly-run banks down the next time there is an economic downturn. Sound principles of risk management are vital at all times, throughout the cycle. In essence, they are timeless.

This book is timeless. I have been familiar with it since it was first published, and have been its biggest fan ever since. It is great to see it being issued now in its 4th edition. It is one of those rare books that combines the rigor of a sound, balanced academic approach, essential if one is to operate in finance without emotion and with logic, with the accessibility and real-world relevance that is an imperative for the practitioner. It is a genuine "handbook", one can read it and apply its principles right away in just about every type of banking institution in the world, and that bank would be better off as a result.

Every single chapter in the book is worthy of study. I am very enthusiastic about the chapters on ALM gap and hedging. The author places everything in context, and ties in market risk and banking book risk, together with credit risk – a rare, combined approach that plays to my own strong belief about how risk management in banks should be governed by the Asset-Liability Committee (ALCO). Balance sheet risk needs one oversight body that operates with board authority, and as the balance sheet is impacted by ALM, market and credit risk together, it makes sense to view these from the ALCO table.

As a young man I used to play the bass guitar. Being asked to write this Foreword is a bit like being asked by Paul McCartney to play bass on his next album, it is that much of a privilege! Professor Bessis has made a fantastic and most worthwhile contribution to the financial economics literature with this book, right from its first edition, and I am lucky to have had a copy on the desk with me ever since it was first published. I do hope that this exciting and interesting new edition makes balance sheet risk in banking something that is more mainstream at the board level, and furthermore spurs readers on to their own research and investigation – if they follow the application and dedication evident in this work, they will not be going far wrong.

Professor Moorad Choudhry
Department of Mathematical Sciences
Brunel University
Former Treasurer, Corporate Banking Division,
Royal Bank of Scotland
November 2014

Preface

Risk management in banks became, and remained, a hot topic after the financial crisis. Addressing risk management in this context is challenging given that the magnitude of the crisis suggests that risk management was inefficient, that risk models were inadequate and that regulations failed to meet their goal of avoiding a major crisis. Indeed, it is ironic that the crisis started when the new Basel 2 regulations were enforced.

Risk management has made considerable progress, however, as the practices became more sophisticated and as the regulations put pressure on enhancing the resilience of banking firms. It has become a core management field in banking with a large concentration of resources dedicated to better identify, assess and control risks.

The book addresses risk management in three main core sections dedicated, respectively, to asset-liability management, market risk and credit risk. It has been largely inspired from the observation of gaps in the knowledge of the field of risk management in banks.

In business schools and other graduate programs, students are comfortable with corporate finance and capital markets, but much less so with the finance of financial firms. The financial management of banks has not much to do with the corporate finance of commercial and industrial firms. Still many would like to better understand the inside mechanisms of banks and many aim at developing themselves in banking careers. These students of finance do not need standard finance, but rather be acquainted with the specifics of the financial management of banks and the technicalities of risk management. This book is designed to address these needs.

Many professionals in banks perceive themselves as specialists of their own fields and feel that they need more background, conceptual and practical, on the expanding core area of risks. Furthermore, the usage of risk models remained in the hands of a relatively small group of "quants". Experts are embedded in banks, but being embedded does not imply that expertise is shared. The financial literature is broad, specialized and often highly technical in the field of risks. For those professionals of finance who are not model specialists, navigating through the variety of contributions is a challenge. This text is designed to provide a balanced background in risk techniques. The main risk models are introduced through a number of examples that should shed some light where more theoretical texts cannot help.

The volume of literature on market risk and credit risk has grown considerably, but less so in the field of asset-liability management, of which coverage is relatively limited, notably for non-specialists of banks. However, asset-liability management is a core function in banking. It concentrates the financial issues of banks and the attention of regulators who impose new rules on the balance sheet structure of banks. The text provides the minimum background on the area that all students or managers interested in banking should be acquainted with.

In short, this text is designed to address all that is needed to know for students and practitioners to be comfortable with the field and able to navigate further in related areas by themselves, but not more.

This edition has been streamlined compared to previous editions, with a focus entirely on financial issues, and technical developments have been reduced to the minimum for making the text self-contained. Many of my former colleagues and professionals with whom I have had the chance of working in the risk departments of banks have contributed to this text as they shared their experience. All participants in risk management seminars have also helped by raising many excellent and challenging questions, which allowed to refine the approach of the book. They all deserve many thanks for the enrichments that they inspired to this text.

Joël Bessis
Professor of Finance at HEC Paris

About the Author

Joël Bessis is Professor of Finance at HEC Paris, the leading French business school, where he conducts training in risk management throughout Europe, the US and Asia. Over the course of his career Joël has developed a dual expertise – as an academic and as a practitioner, holding permanent consulting assignments in corporations and, later, in banks. Joël worked for over 15 years in risk management departments of financial institutions – as a consultant to the risk departments of several banking institutions in Europe, including Banque Paribas and the European Bank for Development (EIB). Joël took a leave of absence from HEC Paris between 2000 and 2007 where he held positions as Director of Research at Fitch, Head of Risk Analytics and Model Validation at the Risk Department of IXIS, a Paris-based investment bank, and at the Groupe Caisse d'Epargne, a major financial institution in France. Joël graduated as an engineer from École Centrale in Paris, before earning a Master's in Business Administration from Columbia University in New York, and a PhD in Finance from the Université Paris-Dauphine. As an academic, Joël has published various papers and books in the fields of corporate finance, industrial economics and financial markets.

1

Risks and Risk Management

For risk managers and regulators of banks, risk refers to the uncertainty of outcomes and to the negative consequences that it may have on a firm, and both aim at enhancing the resiliency of firms to adverse situations. As a result of their efforts, risks became better identified, assessed and monitored, risk practices improved and risk models became more widespread. Today, risk management has become a core central function for financial firms, banks, funds and insurance companies.

This introductory chapter presents the definitions of financial risks in banking and introduces typical organizations of the risk management function in banks, defining who should be accountable for risk controlling and processes.

Contents

1.1 UNCERTAINTY, RISK AND EXPOSURE TO RISK

Risk has been defined in various ways across time. Some definitions focus on the probability of an event, others refer to the uncertainty of outcomes, positive or negative, and others to risks as the subset of uncertainty that can be quantified.[1]

1 Knight, F. H. (1921), Risk, Uncertainty and Profit, New York.

Risk in finance is defined as the randomness of the return of investments, including both positive and negative outcomes. Under this view, a greater expected return is associated with a greater variability of outcomes.

In the financial industry, the view of risk is different. Risk is defined by the uncertainty that has adverse consequences on earnings or wealth, or the uncertainty associated with negative outcomes only. This view is that of regulators and risk managers. Regulations aim at enhancing the resiliency of financial firms and of the financial system in stressed conditions. Risk managers see their role as being accountable for identifying, assessing and controlling the likelihood and the consequences of adverse events for the firm.

Under this view, risk is seen as the potential of loss resulting from the interaction with uncertainty. The interaction arises from the exposure of financial firms to such randomness. Exposure is the extent to which a business could be affected by certain factors that may have a negative impact on earnings. For example, exposure to foreign exchange rate is the size of revenues in foreign currency; exposure to interest rates can be measured by the size of debt indexed on market rates.

The uncertainty cannot be eliminated but the exposure to uncertainty can be changed. Examples are numerous. A firm having revenues in foreign currency can borrow in the same foreign currency to minimize the earning impact of foreign exchange rate fluctuations. A firm lending floating rate can reduce the fluctuations of net interest income, the interest revenue minus interest cost, etc., by borrowing floating rate.

Exposures can be long or short. Being long is the conventional practice for investing in assets or portfolios. The holder of an asset is long and the risk is that the asset value declines. A short position can be seen as the mirror image of long positions and gains when asset values move down. In investing, a short position is the sale of a borrowed asset, such as a stock, which is later bought back for returning the assets to the lender of the security. In the event of a downside movement, the borrower of the stock buys back the stock at a lower price, hence makes a gain.

Hedging risks can be achieved by taking inverse exposures to long positions. Holding a stock is a long position, which takes a loss if the equities decline. A short position is symmetrical. When a party has both a long and a short position in the same stock, the gains and losses exactly offset. Hence, a perfectly hedged position is subject to uncertainty, but is not exposed to risk.

Hedging can be achieved with cash instruments, but is commonly done with derivatives. Derivatives are instruments, the value of which derives from other underlying assets. For example, the above firm willing to hedge its long exposure in foreign currency could enter into a contract, setting today the future exchange rate for converting foreign revenues in the home currency. This is easier than trying to borrow in the foreign currency. Because of their flexibility, derivatives are extensively used.

1.2 BROAD CLASSES OF FINANCIAL RISK

Financial risks are defined according to the sources of uncertainty. The broad classes of financial risks are credit risk, market risk, liquidity risk and interest rate risk, divided into subclasses relative to the specific events that trigger losses.

1.2.1 Credit Risk

Credit risk is the risk of losses due to borrowers' default or deterioration of credit standing. Default risk is the risk that borrowers fail to comply with their debt obligations. Default triggers a total or partial loss of the amount lent to the counterparty.

Credit risk also refers to the deterioration of the credit standing of a borrower, which does not imply default, but involves a higher likelihood of default. The book value of a loan does not change when the credit quality of the borrower declines, but its economic value is lower because the likelihood of default increases. For a traded debt, an adverse migration triggers a decline of its quoted price.

Recovery risk refers to the uncertain value of recoveries under default. Recoveries depend on the seniority of debt, on any guarantee attached to the transaction and on the workout efforts of the lender. The loss after workout efforts is the loss given default.

Counterparty credit risk exists when both parties of a transaction are potentially exposed to a loss when the other party defaults. A swap contract exchanging fixed for floating interest flows between two parties is a typical example. The party who receives more than it pays is at risk with the other party. The exposure might shift from one party to the other, and its size varies, as a result of the movements of interest rates. Counterparty credit risk exists when exposures are market driven.

1.2.2 Market Risk

Market risk is the risk of losses due to adverse market movements depressing the values of the positions held by market players. The market parameters fluctuating randomly are called "risk factors": they include all interest rates, equity indexes or foreign exchange rates.

Market risk depends on the period required to sell the assets as the magnitude of market movements tends to be wider over longer periods. The liquidation period is lower for instruments easily traded in active markets, and longer for exotic instruments that are traded on a bilateral basis (over the counter). Market risk is a price risk for traded instruments. Instruments that are not traded on organized markets are marked-to-market because their gains or losses are accounted for as variations of value whether or not materialized by a sale.

1.2.3 Liquidity Risk

Liquidity risk is broadly defined as the risk of not being able to raise cash when needed. Banking firms raise cash by borrowing or by selling financial assets in the market.

Funding liquidity refers to borrowing for raising cash. Funding liquidity risk materializes when borrowers are unable to borrow, or to do so at normal conditions. Asset liquidity refers to cash raised from the sale of assets in the market as an alternate source of funds, for example in market disruptions. Asset liquidity also refers to the risk that prices move against the buyer or seller as a result of its own trades when the market cannot absorb the transactions at the current price. Asset liquidity risk also arises when too many players do similar trades. For example, banks raising cash from liquidation of assets in the adverse conditions of the 2008 crisis faced substantial losses from the deep discounts in their trades.

Extreme lack of liquidity results in failure. Such extreme conditions are often the outcome of other risks, such as major markets or credit losses. These unexpected losses raise doubts with respect to the credit standing of the organization, making lenders refrain from further lending to the troubled institution. Massive withdrawals of funds by the public, or the closing of credit lines by other institutions, are potential outcomes of such situations. To that extent, liquidity risk is often a consequence of other risks.

1.2.4 Interest Rate Risk

The interest rate risk is the risk of declines of net interest income, or interest revenues minus interest cost, due to the movements of interest rates. Most of the loans and receivables of the balance sheet of banks, and term or saving deposits, generate revenues and costs that are interest rate driven.

Any party who lends or borrows is subject to interest rate risk. Borrowers and lenders at floating rates have interest costs or revenues indexed to short-term market rates. Fixed-rate loans and debts are also subject to interest rate risk. Fixed-rate lenders could lend at higher than their fixed rate if rates increase and fixed-rate borrowers could benefit from lower interest rates when rates decline. Both are exposed to interest rate fluctuations because of their opportunity costs arising from market movements.

1.2.5 Foreign Exchange Risk

Foreign exchange risk is the risk of incurring losses due to fluctuations of exchange rates. The variations of earnings result from the indexation of revenues and charges to exchange rates, or from the changes of the values of assets and liabilities denominated in foreign currencies (translation risk).

1.2.6 Solvency Risk

Solvency risk is the risk of being unable to absorb losses with the available capital. According to the principle of "capital adequacy" promoted by regulators, a minimum capital base is required to absorb unexpected losses potentially arising from the current risks of the firm. Solvency issues arise when the unexpected losses exceed the capital level, as it did during the 2008 financial crisis for several firms. This capital buffer sets the default probability of the bank, the probability that potential losses exceed the capital base.

1.2.7 Operational Risk

Operational risks are those of malfunctions of the information system, of reporting systems, of internal risk monitoring rules and of procedures designed to take corrective actions on a timely basis. The regulators define operational risk as "the risk of direct or indirect loss resulting from inadequate or failed internal processes, people and systems or from external events".[2] The focus on operational risk developed when regulators imposed that the operational risks should be assigned a capital charge.

1.3 BUSINESS LINES IN BANKING

There is a wide variety of business lines in the banking industry, with different management practices and different sources of risks. This section provides a brief overview of the diversity of activities conducted in banking.

Retail banking tends to be mass oriented and "industrial" because of the large number of transactions. Retail Financial Services (RFS) covers all lending activities to individuals, from credit card and consumer loans, to mortgages. RFS also extends to very small enterprises, such as those of physicians or home services. Lending decisions are based on a combination of automated systems and management monitoring. Statistical techniques are relevant for assessing credit risk.

Standard corporate lending transactions include overnight loans, short-term loans (less than one year), revolving facilities, term loans, committed lines of credit or large commercial and industrial loans. Such transactions are under the responsibility of credit officers and their reporting lines. For the large corporate businesses, relationship banking prevails when the relationship is stable, based on mutual knowledge. Credit analysts are industry specialists who monitor the credit standing of clients. They provide the individual credit assessments of obligors, based on expert judgment, for making lending decisions.

Investment banking is the domain of large transactions customized to the needs of large corporate and financial institutions. It also includes trading activities, under the generic name of "Corporate and Investment Banking" (CIB).

Large corporations demand a variety of services and products, for example from lending facilities and hedging instruments or issuance of securities. A number of very different activities are under the umbrella of the CIB pole. The financing of financial institutions, banks, insurance companies and brokers is organized as separate groups, distinct from those dedicated to commercial and industrial firms. Mergers and acquisitions form another business line.

All activities of specialized, or "structured", finance are also conducted by dedicated units within CIB. The scope of specialized finance includes such activities as project finance, asset financing (ships or aircrafts), commodities finance, commercial real estate and exports. The risk analysis differs radically from the assessment of a corporate borrower. In general, the primary source of repayment is the cash flows generated by the asset(s), from its operations or from the sale of the asset(s). Structuring refers to the assembling of financial products and derivatives, plus contractual clauses ("covenants") in order to make the risk manageable. Securitization is one of the fields of specialized finance: it consists of selling pools of loans, which are normally held in the balance sheet of banks, into the capital markets.

Trading involves traditional proprietary trading and trading for third parties. In proprietary trading, the bank is trading for itself, taking and unfolding positions to make gains. Trading is also client oriented. "Sales" designate trades conducted when the bank acts on behalf of their clients. The "sell side" is the bank, selling products to end-users. The "buy side" designates the clients, corporations and asset managers who buy the products, for example for hedging

---

2 The definition is from the Basel 2 document (2006), [21].

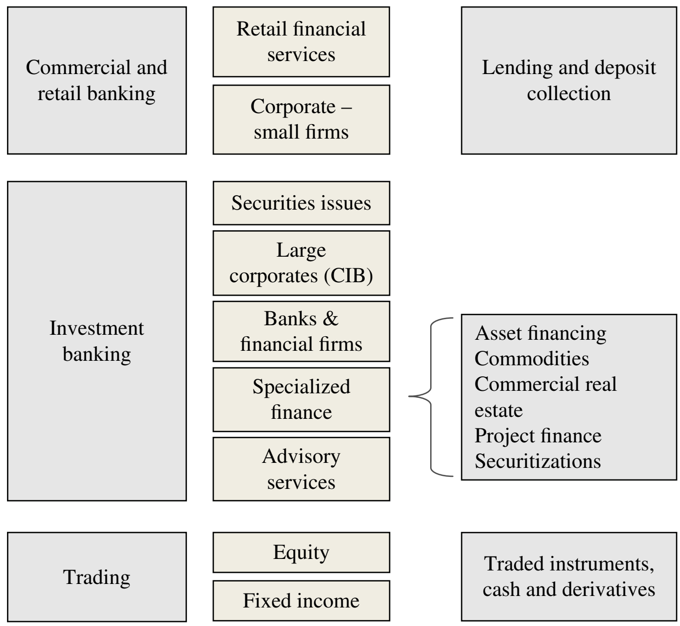


FIGURE 1.1 Business lines in banking

purposes. Traders and lending officers are not allowed to share information, as inside information on a corporate client could inspire trades based on undisclosed information. Banks are also exposed to market risk from their investment portfolio, which is not held for trading but with an objective of long-term performance.

Other activities do not generate directly traditional financial risks. For example, private banking, or asset management, is the activity of wealth management for third parties. Advisory services refer to consulting services offered by banks to corporations considering potential acquisitions, for example, which do not necessarily imply cash outlays. Risks are primarily legal and operational.

Figure 1.1 maps the banking activities grouped into main poles.

1.4 BANKING REGULATIONS AND ACCOUNTING STANDARDS

Banking activities are subject to a wide body of rules. Risks are subject to the regulations rules. Valuation of assets and liabilities and profit and loss are subject to accounting standards.

Risk regulations differ for the banking book and the trading book. The banking book refers to the transactions belonging to the core business of commercial banks, lending and deposit collection. It includes all assets and liabilities that are not actively traded by the institution, and generally held until they mature. The trading book groups capital market transactions, and is exposed to market risk. Positions held for trading are held over a short-term horizon, with the intention of benefiting from expected price movements. The trading book includes proprietary positions, and positions arising from client servicing and market making.

Risk regulations relate directly to the management of the balance sheet, to market risk and credit risk and are detailed in the corresponding sections of this text.

The accounting standards segregate instruments into four classes differing by valuation and treatment of profits and losses:

- Financial assets at fair value through profit and loss;
- Loans and receivables;
- Held-to-maturity investments;
- Available-for-sale financial assets.

The financial assets at fair value include all instruments acquired to take advantage of price fluctuations and are managed with the intention of making short-term profits, the performance of which is evaluated on a fair value basis. Derivatives are held for trading unless they are considered as hedges. The assets and liabilities of the trading book are under this category.

Fair value focuses on the price at which an asset can be sold: it is the amount at which an asset could be exchanged between parties, knowledgeable and willing to exchange. Valuation depends on whether markets are active or not. Active markets are those where the volume of transactions provide clear prices. For other instruments, prices can be derived from other traded instruments in active markets, or valuation is model based.

Accordingly, market instruments fall in either one of three categories: level 1 when quoted prices are available; level 2 when there are market prices for similar instruments; and level 3 for model valuation for instruments that are not extensively traded but have a value derived from models, or mark-to-model, such as options traded over the counter. Model valuation is recognized as fair value in the absence of an active market.

Loans and receivables are instruments with contractual payments and are not quoted in active markets. These assets are held in the banking book. In the banking book, income is determined according to accrual accounting rules of revenues and costs.

Held-to-maturity instruments are financial assets with contractual payments for which the management intention is not trading. Investment portfolios of banks' group financial assets, such as bonds, in which banks invest for the long term with no trading intent, are in this category. All other assets are available for sale.

Liabilities are either at fair value through profit and loss or other liabilities. The liabilities at fair value are held for trading, or designated as such, and the performance is based on fair value. The other liabilities include the normal financing of the bank, debt issued in markets or wholesale debt, which arises from lending and borrowing from other banks or financial institutions.

For the trading book, valuation is based on mark-to-market, or mark-to-model for illiquid instruments. The performance is evaluated on the basis of fair value: profit and loss (P&L) is measured as the variations of value between two dates.

1.5 RISK MANAGEMENT

Risk management requires that the risks of a financial institution be identified, assessed and controlled. Enterprise risk management addresses a combination of credit risk, market risk, interest rate risk, liquidity risk and operational risk. Sound risk practices define who should be accountable for these risks and how the risk processes should be implemented.

1.5.1 Motivations

There are strong reasons motivating the sound assessment and management of risks in decision-making processes, other than compliance with risk regulations.

Risk and return are two sides of the same coin. It is always easy to lend, and to obtain attractive revenues from risky borrowers. The price to pay is a higher risk than the prudent bank and higher potential losses. The prudent bank limits risks by restricting business volume and screening out risky borrowers. It saves potential losses but might suffer from lower market shares and lower revenues. However, after a while, careless risk takers find out that higher losses materialize, and could end up with a lower performance than the prudent lender.

Banks that do not differentiate risks of their customers would suffer from adverse economics. Overpricing good risks would discourage good customers. Underpricing bad risks would attract bad customers. Discouraging the relatively good clients and attracting the relatively bad ones would result in adverse selection.

1.5.2 The Risk Processes

Risk processes include the identification, monitoring and control of risks. Risk models serve for measuring and quantifying risk, and provide the inputs for the management processes and decisions. To be effective, they should be implemented within a dedicated organizational framework that should be enterprise-wide.

All risk processes imply that risk policies be properly defined and that the risk appetite of the firm be well defined. Within this framework, the common process for controlling risks is based on risk limits and risk delegations. Limits impose upper bounds to the potential loss of transactions, or of portfolios of transactions. Delegations serve for decentralizing the risk decisions, within limits.

Limits aim at avoiding that adverse events, affecting a transaction or a portfolio of transactions, impair the credit standing of the firm. Banks need to segment their activities into meaningful portfolios, for example by business unit, product or type of clients. Limits of exposure are set for each segment and down to transactions, forming a hierarchy of limits and sublimits. For credit risk, limits are set by segment, then by counterparty and then by individual transaction. For market risk, limits can be set for specific books of trades, then desks and then trades.

Delegations are authorizations to act and take risks on behalf of the organization. Delegations decentralize and simplify the risk process by allowing local managers to make decisions without referring to the upper levels of the organization, within the scope of their delegations. For example, they simplify the risk process for transactions that are small enough to be dealt with by local procedures.

1.5.2.1 Credit Risk Limits and Delegations

Any limit system requires one or several measures of risk used for determining whether or not a transaction, or a portfolio of transactions, complies with limits. Various risk metrics are used for setting limits for credit risk. The amount at risk, or exposure, is a simple measure of the amount that could be lost in the event of a default of the borrower. Other metrics capture other dimensions of credit risk. For example, trades might be allowed only for eligible borrowers based on their credit quality. Or limits can apply to regulatory capital for credit risk, which combines various components of credit risk, exposure, loss after recoveries and credit quality.

Credit limit systems are based on common criteria, for example:

- Diversify the commitments across various dimensions such as customers, industries and regions.
- Avoid lending to any borrower an amount that would increase its debt beyond its borrowing capacity. The equity of the borrower sets up some reasonable limit to its debt given acceptable levels of debt/equity ratios and repayment ability.
- Set up a maximum risk level, for example defined by the credit standing of borrowers, above which lending is prohibited.
- Ensure a minimum diversification across counterparties and avoid concentrations of risk on a single borrower, an industry or a region.

For being comprehensive and consistent, the limit system has to be bank-wide. Global limit systems aggregate all risks on any single counterparty, no matter which business unit initiates the risk, across all transactions with the bank. Global limits are broken down into sublimits. Sublimits might exist even at the level of a single client. The total usage of sublimits should not exceed the global limit assigned to each counterparty or portfolio of transactions. Limit systems allow sublimits to sum up to more than the global limit because not all sublimits are fully used simultaneously, but the aggregated risk should comply with the global limit. For example, multiple currency facilities are convenient for clients because they allow raising funds in several currencies and as needed, but a client should not use more than its global limit. Utilizations are bounded by either sublimits or global limits, whichever is hit first.

Any excess limit has to be corrected by not entering into a new transaction or mitigating the risk with guarantees. Some limits might be hit while others are not. Banks' systems address the issue with excess limit reports showing which transaction hits which limit.

Credit approval processes vary across banks and across types of transaction. In retail banking, the process relies on procedures that need to accommodate a large volume of transactions. Time-consuming processes are not applicable for the retail portfolio because of the high number of transactions. Instead credit scoring mechanisms and delegations are used, within the guidelines of the credit policy. In normal circumstances, the credit officer in charge of the clients of a branch is authorized to make decisions as long as they comply with the guidelines.

For large transactions, the process involves credit committees. Credit committees bring together the business line, the risk managers and the general management. The business line proposes new transactions, together with a risk analysis, and the committee reviews the deal. Committees need to reach a minimal agreement between members before authorizing a credit decision by examining in detail significant credit applications. The committee makes a yes/no decision, or might issue recommendations for altering the proposed transaction until it complies with risk standards. Collateral, third-party guarantees or contractual clauses, mitigate the risk. The alternate process is through "signatures" whereby the transaction proposal is circulated and approval requires agreement between all credit officers. Whether signatures or committees are used for approval, risk officers remain accountable for the risk decisions and decisions are recorded, eventually with comments and recommendations of participating executives.

1.5.2.2 Market Risk and Trading Activities

For market risk, a common risk metric is the sensitivity of a position or of a portfolio of positions. Sensitivities measure the variations of values due to standard shocks on market parameters such as interest rates, foreign exchange rates or equity indexes. There is a variety of sensitivities depending on the type of products and on the risk factors that influence their values. Other risk metrics involve the capital charge for market risk, which embeds other elements of market risk, such as the market volatility and an assessment of the likelihood of losses of various magnitudes.

As the gains and losses in trading are market driven, a risk tolerance has to be defined for the business lines, the desks and the traders. A risk tolerance is an assessment of the maximum loss, for the business line or for desks, considered as acceptable, but which should not be exceeded.

The policies for trading should be comprehensively documented. Limits depend on expectations about market conditions, as formulated in market committees. Market instability might require tighter limits because the chances of large fluctuations are higher. A daily market committee formalizes the current market conditions. Trading desks operate within their limits. Limits are set for the various desks consistently with aggregate market risk. Traders comply with limits by hedging their risks, or unwinding their positions, eventually at a loss.

1.5.3 Risk Management Organization and Roles

As risk regulations developed and standard practices for risk management spread across the industry, some common views on the organization of the risk management process emerged.

1.5.3.1 The Risk Department and the "Three Lines of Defense" Model

The three lines of defense model is a convenient scheme used for structuring the roles, responsibilities and accountabilities with respect to decision making, risk controlling and for achieving an effective risk governance bank-wide. It illustrates how controls, processes and methods are aligned throughout large organizations. The three lines of defense are:

- The lines of business.
- The central risk function.
- The corporate audit and compliance functions.

The business lines, or front office, make up the first line of defense and are responsible for identifying, measuring and managing all risks within their scope of business. Business lines have the primary responsibility for day-to-day risk management. As the management of the business line is close to the changing nature of risks, it is best able to take actions to manage and mitigate those risks. Lines of business prepare periodic self-assessment reports to identify the status of risk issues, including mitigation plans, if appropriate.

These reports roll up to the executive management and to a central risk department, which enforces the risk discipline. Standard practices impose that the risk management should be centralized and that a "clean break" exists between risk-taking business lines and risk supervising units. The risk department ensures an assessment and a control of risks independent of the business lines. The department is responsible for the guidance and implementation of risk policies, for monitoring their proper execution complying with documented risk processes. It defines, with the top management, the risk policy of the bank. The chief risk officer reports to the senior executive committee, who ultimately provides the risk department with the power of enforcing risk policies.

Given their roles, the perceptions of the same risk reality by the business lines and the risk department might differ. This difference in perspectives is what adds value to the enterprise as a whole and to the risk management process. However, the effectiveness of the risk process can be questioned when there are compelling business reasons to proceed with a transaction. Enforcing the power of a credit committee requires some arbitration process when conflicts arise. The arbitrage between conflicting parties is handled by more senior levels when the process is not conclusive. Moving up in the hierarchy of the bank guarantees that a conclusion will be reached and that the credit proposal is thoroughly examined at each stage.

The existence of a risk department does not suffice to enforce sound risk practices. Both the first line and the second line are accountable for risk assessment and control. Making the risk department the unique function accountable for risks would relieve the business lines from their risk responsibilities. A centralized risk control unit would be overloaded by the number of risk issues raised by the front offices. In large banks, risk managers are "embedded" within the business lines, but report both to the business lines and to the central risk department. They provide the local risk control within the "first line of defense".

The third line of defense is that of internal and external auditors who report independently to the senior committee representing the enterprise's stakeholders. The internal auditors' role is to provide an independent review of the effectiveness and compliance to risk policies of the risk processes. Corporate audit activities are designed to provide reasonable assurance that significant financial, managerial and operating information is materially complete, accurate and reliable; and that employees' actions comply with corporate policies, standards, procedures and applicable laws and regulations. The auditors have the capacity to make recommendations and to supervise their execution.

1.5.3.2 The Asset and Liability Management Department

The ALM – asset-liability management – department is in charge of managing the funding and the balance sheet of the bank, and of controlling liquidity and interest rate risks. The function of ALM is the finance function of banks and is often located within the finance department. The scope of ALM extends mainly to the banking portfolio, and less so to trading activities because they rely primarily on short-term financing. For controlling the liquidity risk and the interest rate risk, the ALM sets up limits to future funding requirements and manages the debt of the bank. The interest rate risk is measured by the volatility of target variables such as the net interest income of the bank, using interest rate derivatives.

The ALM committee meets at least monthly, or when needed in adverse conditions. It groups the senior management, the chief finance officer, the head of the ALM team and the executives in charge of business development and commercial policies. The senior management is involved because ALM policies have a strategic influence on the bank's financing profitability. ALM policies also have strong and direct interactions with the commercial policy. The bank exposure to interest rate risk and liquidity risk depends on the product mix in the banking book. ALM policies have also a direct effect on the pricing to clients, as it should absorb the cost of financing the banking book. Furthermore, the ALM unit is in charge of internal prices of funds, the cost of funds charged to lending units and the financial compensation of deposit collection by branches.

1.5.3.3 Enterprise-wide Risk Management (ERM)

Bank-wide management implies that metrics of income and risk at the global bank level be related to similar metrics at the business unit, book and transaction levels.

Policies set global limits and profit objectives at the enterprise level, which are allocated to business units. This top-down process requires that aggregate profit and limits be allocated at lower levels of the hierarchy in a consistent manner. The monitoring and the reporting of risks and performance is bottom-up oriented, starting from transactions, and ending up with aggregated risks and income. Both processes require a sound bank-wide allocation of earnings and of risks.

As funds are transferred to lending activities and from deposits collected, the earnings of business lines depend on internal, or transfer, prices. The transfer pricing system serves to allocate earnings across business lines and transactions and is required for reconciling aggregated earnings with the earnings of business lines, and down to the transaction level.

A similar system should be implemented for allocating a share of the bank's risk to business units. Global limit systems define the hierarchy of limits and sublimits within the organization. But limit systems are distinct from measures of risk.

A key factor for risk aggregation is risk diversification. Because of diversification, risks do not add up arithmetically. Loosely speaking, the sum of individual risks is less than the arithmetic summation of risks. This well-known property of risks being subadditive is the source of the challenging problem of risk allocation. For risks to be aggregated bottom-up, and allocated top-down, a risk allocation mechanism is required. In general, the risk allocation issue is addressed by allocating the capital of the bank to portfolios and transactions and it involves an assessment of diversification effects.

Finally, earnings across transactions or portfolios are not comparable because they are in general exposed to different levels of risk. Performances need to be risk adjusted for being comparable across activities and comparable with the risk-adjusted profitability of the bank. The issue is resolved once earnings and risks are properly allocated, by adjusting earnings with the cost of risk based on the cost of capital backing the transactions.

This shows that three building blocks should be designed and assembled for addressing bank-wide risk management:

- Fund transfer pricing systems;
- Risk and capital allocation systems;
- Risk-adjusted performance measures.

These are the necessary components of risk systems for aligning the measures of earnings and risks,[3] and the related management incentives, across all business lines of large organizations.

3 The fund transfer pricing system is addressed in Chapter 10, the risk allocation issue is discussed in Chapter 27 and risk-adjusted performance is discussed in Chapters 10 and 28.

2

Banking Regulations Overview

The capital adequacy principle is the foundation of regulations aimed at making banks more resilient. Capital adequacy refers to the minimum level of capital for absorbing the potential losses from the current banks' books. Ensuring a proper level of capital fostered the emergence of sound risk management practices and imposed risk models designed for quantifying the potential losses of a bank arising from its current risks.

This chapter is a brief history and overview of the successive risk regulations introduced since 1988, up to the 2008 crisis and to the new regulations introduced as a response to the current crisis. The detailed regulations are presented in subsequent chapters (3, 17 and 26), within the sections dedicated to asset-liability management (ALM), market risk and credit risk.

Contents

2.1 REGULATION PRINCIPLES

The primary purpose of risk regulations is to prevent systemic risk, or the risk of collapse of the entire system due to interconnections between financial firms. However, regulators face dilemmas when attempting to control risks.

Providing more freedom to financial firms has been a long-standing argument for avoiding too many regulations. But relying on codes of conduct, rather than rules, would imply relying on self-discipline, or "self-regulation", which would not inspire trust in the system.

The financial system is subject to moral hazard. Moral hazard is a situation in which a party is more likely to take risks because the costs that could result will not be borne by the party taking the risk. It results in a tendency to be more willing to take a risk, knowing that the potential costs or burdens of taking such risk will be borne, in whole or in part, by others.

Any insurance mechanism potentially generates moral hazard. One the oldest regulations is deposit insurance. Under deposit insurance, depositors are guaranteed the value of their holdings in banks, subject to a cap, that varies across jurisdictions. The regulation provides some safety to depositors, but it does not have much effect on the risk-taking behavior of banks, as depositors are not able to impose a discipline on banks. The protection of bank depositors is an insurance potentially generating moral hazard in the absence of prior penalty for taking risks.

The limited liability of shareholders is another source of the lack of self-discipline. A significant increase of risk can potentially lead to a risk-maximizing behavior. When risks are already high, shareholders have not much to lose and they might prefer to make riskier bets that increase the chances of failing. When banks face serious difficulties, the barriers that limit risks disappear.

The "too big to fail" issue is a potential source of moral hazard. It refers to a situation where no large institution can be allowed to fail for fear of contagion to many others. Many big firms are lending to large financial firms, and would incur large losses if these fail. The domino effect refers to the cascading effect of the failure of large institutions, triggering the failure of others, which, in turn, triggers another wave of failures. Because of fear of such domino effect, the regulators might not allow large institutions to fail, generating moral hazard for the major financial firms. The issue arises from the "interconnectedness" of large financial institutions.

The financial crisis demonstrated that such issues were not hypothetical. In the United States, after assisting in the bail-out of some large financial firms, the financial authorities let Lehman Brothers down in 2008, which is considered as the critical event in the development of the financial crisis. Perhaps the authorities wished to demonstrate that no one single firm was too big to fail. But it ended up in a situation where systemic risk materialized.

Regulations aiming at resolving issues occurring in the event of failure, such as deposit insurance, help to ensure some trust in the system. The resolution plans for an orderly dismantling of large firms, promoted by regulators following the crisis, are another example. But they are after-the-fact rules that do not prevent banks from taking too much risk.

The core concept of risk regulations is the "capital adequacy" principle, which imposes a capital base commensurate with risks to which each bank is exposed. Instead of "dos and don'ts", banks need to have enough capital to make their risks sustainable.

If capital is high enough to absorb large losses, the banks would be safe. The size of the capital base depends on how much risk banks are taking. Capital-based rules raise implementation issues, as the capital charges of banks turn out to be a quantified assessment of their risks. Guidelines for regulations are defined by a group of regulators meeting in Basel at the Bank of International Settlement (BIS), hence the name of "Basel" Accords for the successive rounds of regulations since the initial Basel 1 Accord.

2.2 CAPITAL ADEQUACY

Under the capital adequacy principle, capital is the last "line of defense" for avoiding failure in stressed conditions. The solvency of banks cannot be impaired unless the firm incurs losses in excess of capital. The higher is the capital buffer against losses, the higher is the protection. The capital adequacy principle is a preemptive protection against failure. Capital-based regulations impose that the losses from risks be quantified. The quantification of risk evolved from the simple rules of the 1988 Accord for credit risk, up to more elaborated and complex rules of the current Accords.

The first implementation of capital-based regulations was enforced in 1988 for credit risk with the well-known Cooke ratio, initiated with the Basel 1 Accord.[1] The first Accord focused on credit risk. The Cooke ratio sets the minimum required capital as a fixed percentage of assets weighted according to their credit quality. The capital base included any debt subordinated to other commitments by the bank. Equity represented at least 50% of the total capital base for credit risk, also called the "tier 1" of capital or "core capital". The available capital puts a limit to risk taking, which, in turn, limits the ability to develop business. Under a deficiency of capital, the constraint requires raising new equity, or liquidation of assets, or taking risk-mitigating actions.

The original Cooke ratio of the 1988 Accord stipulates that the capital base should be at least 8% of weighted assets. Risk-weighted assets (RWA) are calculated as the product of the size of loans with risk weights. The risk weights serve for differentiating the capital load according to the credit quality of borrowers. The calculation of the capital, which is still implemented today, is:

$$\text{Capital} = 8\% \times \text{Risk weight} \times \text{Asset size}$$

The 8% is the capital adequacy ratio, which is evolving with regulations and getting closer to around 10% as new regulations are gradually enforced. The regulators' 8% capital adequacy ratio can be interpreted as a view that banks could not lose more than 8% of their total risk-weighted portfolio of loans for credit risk, thanks to risk diversification. With this value of the ratio, the debt-equity ratio is: $92/8 = 11.5$.

The original Basel 1 Accord was designed for keeping calculations simple and allowing an easy implementation. For example, a loan of value 1000 with a risk weight of 100% has a capital charge of 80; if the loan is a mortgage, backed by property, it would have a capital charge of: $50\% \times 8\% \times 1000 = 40$.

The weight scale started from zero, for commitments with sovereign counterparties within the OECD, at the time when Basel 1 was implemented, and up to 100% for non-public businesses. Other weights were: 20% for banks and municipalities within OECD countries, and 50% for residential mortgage-backed loans. Some off-balance sheet commitments, the commitments without any outlay of cash, were weighted 50%, in conjunction with these risk weights.

Today, the same general concepts prevail, using a capital ratio that is a percentage of risk-weighted assets and risk weights being far more risk sensitive.

Regulations do not imply that the true risk of a portfolio is exactly measured by the capital charges. They determine capital charges for portfolios representative of the industry as a whole, not of the specifics of the portfolios of individual banks. Regulators recommended that banks develop their own estimates of credit risk through models. Economic capital refers to better measures of the specific risk of the banks' portfolios.

1 Basel Committee on Banking Supervision (1988), International convergence of capital measurement and capital standards, [17].

The 1988 Accord was followed by capital regulations on market risk in 1996, amended in 1997. The extension to market risk was a major step in 1996/97 as it allowed banks to use models for assessing capital charge for market risk. Since traded assets can be liquidated over short periods, the relevant losses are due to market movements over the same horizon. Capital for market risk should provide a protection against the loss of value that could occur over the liquidation horizon. The regulation promoted the value-at-risk concept. The value-at-risk, or VaR, is the potential future loss for a given portfolio and a given horizon, which is not exceeded in more than a small fraction of outcomes, which is the confidence level. The basic idea is the same, defining the minimum amount of the capital charge, as a function of risks. The risks are assessed either through rules defining capital charge by transaction or VaR-based risk models for market risk. Once VaR-based capital charges were authorized, they became widespread in the financial industry.

The Basel 2 Accord of January 2007 considerably enhanced the credit risk regulations. The schedule of successive Accords, from Basel 1 to Basel 2, is summarized as follows:

1988	Basel 1
1996–1997	Market Risk Amendment allowing usage of internal models for market risk[2]
1999–2007	Consultative packages on the Basel 2 Accord
2007	Basel 2: Qualification of banks for Basel 2 (Europe)
Dec. 2007	Implementation of Basel 2 capital calculations

The approaches of Basel 2 for credit and the update for market risk were published in June 2006 in “International Convergence of Capital Measurement and Capital Standards – A Revised Framework, Comprehensive Version” [21].

The goals of the new Accord were:

- To promote stronger management practices;
- To promote more risk-sensitive capital requirements through a greater use of banks’ own assessment of the credit standing of the borrowers;
- To provide a range of approaches for determining the capital charges for credit risk by allowing banks and supervisors to select the options that were most appropriate for their operations.

The Accord also introduced new capital requirements for operational risks.

For making the capital charge risk sensitive, the Accord provides incentives to use the “internal ratings-based” (IRB) approach, using as inputs for the risk weights the internal ratings, or credit risk assessments, of banks. When not applicable, the banks can rely instead on a “standardized approach” where the risk weights are regulatorily defined.

2.3 SOME LESSONS OF THE FINANCIAL CRISIS

The new waves of regulations, known under Basel 2.5 and Basel 3, are inspired by the lessons of the crisis, which are briefly summarized hereafter. The crisis raised a number of issues, with respect to liquidity, fair value accounting or solvency, as they interact and result in contagion

2 Basel Committee on Banking Supervision (1996, updated 2005), Amendment to the capital accord to incorporate market risk, [19].

and pro-cyclicality. Contagion refers to the waves of failures triggered from individual failures to the system as a whole. Pro-cyclicality refers to the mechanisms that amplify the cycles of the financial system.

2.3.1 Liquidity

The crisis was characterized by the liquidity crunch that plagued the financial system in 2008.[3] A lack of liquidity can emerge from the risk aversion of lenders in stressed times. Players refrain from providing liquidity in a context of failures as no one knows who is next to fail. The "who is going to be next to lose" issue makes potential lenders reduce their exposures to others, for fear that they would suffer unexpected losses of undetermined magnitudes.

Other mechanisms contributed to the liquidity squeeze. The overreliance on short-term funds, a characteristic of the system at the time of the crisis, exacerbates the effect of a liquidity crunch. Once liquidity dried up, all liquidity commitments of banks, whereby banks commit to lend within limits to borrowers, were triggered and translated in a great deal of "involuntary lending". Financial players had to comply with their commitments precisely when they had insufficient liquidity for themselves. Involuntary lending was a source of liquidity for some, but it made liquidity even scarcer as financial firms started to hoard liquidity.

The liquidity crunch lasted even after massive injections of liquidity by central banks through various programs of purchases of financial assets from banks. Presumably, banks and financial firms were "hoarding liquidity" as a protection against a lack of liquidity instead of using it for extending credit. Monetary authorities could not prevent the credit crunch that followed.

2.3.2 Fair Value

Fair value is pro-cyclical as it extends the markdowns to all assets accounted for at fair value, traded or not.[4] Many assets lost value in inactive and illiquid markets. As a consequence of the magnitude of the downturn, fair values appeared disconnected from the fundamental values of assets. The category of assets subject to model valuation extended to assets that, in normal circumstances, would have been fairly valued from prices. Model prices were subject to a negative perception, since many assets lost perceived value as the confidence in models evaporated.

2.3.3 Solvency

When markets move down, fair value rules trigger markdowns of portfolios, even if there is no intention to sell them, which translates into losses and erodes the capital base of banks. Fair

3 There are numerous papers on the liquidity crunch of 2008. See, for example, Brunnermeier, M. K. (2009), Deciphering the liquidity and credit crunch 2007–2008, [39], and Brunnermeier, M. K., Pedersen, L. H. (2009), Market liquidity and funding liquidity, [40].

4 See, for example, Laux, C., Leuz, C. (2010), Did fair-value accounting contribute to the financial crisis?, [89].

value rules in severe conditions made losses unavoidable. Moreover, leveraged banking firms tend to reduce their debt through fire sales of assets. When liquidity relies on fire sales of assets under adverse conditions, markdowns bite the capital base. Under stressed conditions, solvency and liquidity become intertwined. Whether illiquidity or solvency is the initial cause does not matter. Once the mechanism triggered, it operates both ways.

2.3.4 Pro-cyclicality

For borrowing, financial firms pledge their portfolios of securities to lenders. Such collateral-based financing is subject to loan-to-value ratios, whereby the value of pledged assets should be higher than the debt obligation. In a market downturn, complying with ratios imposes that additional collateral be posted for protecting lenders, or, alternatively, that debt be reduced. In a liquidity and credit crunch, cash is raised from the sales of assets for paying back the debt. Fire sales of assets for reducing debt and bringing back asset value in line with loan-to-value ratios add to the market turmoil.[5]

The mechanism is pro-cyclical. Fire sales of assets create a downward pressure on prices, which triggers a new round of collateral calls. This new round results in additional sales of assets and starts another cycle of market decline, and so on. An adverse feedback loop between asset prices and system liquidity develops as a result of such interactions. The mechanism is strongly pro-cyclical: if asset values move down, sales of assets amplify the downturn. In a highly leveraged system, the adverse dynamics develop until the system deleverages itself.

Unregulated funds tend to highly leverage their portfolios for enhancing the return to investors. In a favorable environment, asset values are up and extending collaterized credit to funds is easy. In a stressed environment, the deleveraging of funds puts pressure on the entire system.

Regulations are also pro-cyclical because they impose capital buffers that tend to increase in adverse conditions, while simultaneously the losses shrink the capital base. The process results in credit contraction precisely when credit is needed most by firms chasing funds in illiquid markets, and contributes to the contraction of the whole system.

2.3.5 Securitizations and Contagion of Credit Risk

Securitizations refer to the sales of pools of banking book assets to investors in the capital markets by issuing bonds backed by these assets. Securitizations were a key technique in the “originate and distribute” business model of banks, whereby the banks finance their loans in the markets and, simultaneously, free their capital from backing the risk of sold loans. Prior to the crisis, banks off-loaded and distributed massive amounts of their credit risk into the capital markets.[6] The so-called “toxic assets”, such as subprime loans, were believed to have found their way into the pools sold in the markets and the risk was perceived as disseminated throughout the whole system.

5 On so-called “liquidity spirals”, see Brunnermeier, M. K. (2009), Deciphering the liquidity and credit crunch 2007–2008, [39].

6 See Longstaff, F. A. (2010), The subprime credit crisis and contagion in financial markets, [94].

Rating agencies recognized that they underestimated the risk of asset-backed bonds, many with the highest quality grade, and a wave of downgrades followed. Downgrades command a higher cost of funds, and a higher required return, which translate in a loss of value of the downgraded assets. Investors in asset-backed bonds of securitizations, originally of a high quality, incurred massive losses. The trust in the securitization mechanism disappeared, and with it a major source of funds for the banking system.

2.3.6 Rating Agencies and Credit Enhancers

The frequency of downgrades by rating agencies increased abruptly by the end of 2007, as rating agencies seemed not to have anticipated the effect of the crisis and tried to catch up with bad news. Lagged downgrades were concentrated in time, instead of gradually measuring the actual credit standing of issues.

All entities were hit by rating downgrades. Among those are insurance companies, or monolines, acting as "credit enhancers". These firms enhance the credit quality of assets by providing insurance against credit loss. But the "wrapped" instrument quality is as good as the quality of the insurer. Downgrades of credit enhancers have a leverage effect as any instrument "wrapped" in a guarantee by credit enhancers is also downgraded.

Because monolines extended so many guarantees to assets, they were highly exposed to the risk that erupted in a short period of time. It was not long before credit enhancers were downgraded. AIG, the biggest insurance company in the world, extended credit insurance by trading credit derivatives, and collapsed when lenders required the firm to post collateral against its numerous commitments.

2.4 THE RESPONSES OF REGULATORS TO THE FINANCIAL CRISIS

Following the 2008 financial crisis, the Basel regulators introduced a number of measures to make banks more resilient. A number of significant updates to the regulatory framework have been introduced, reshaping the regulations, after the Basel 2 Accord, into new Basel (2.5 or 3) rules. The main publications include: the global review of the regulatory framework was first published in December 2010, revised in June 2011, "Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems" [24], a "Revision to the Basel 2 Market Risk Framework" is dated February 2011 [25]; the Consultative Document "Fundamental Review of the Trading Book: A Revised Market Risk Framework" [27] was submitted to the industry in October 2013 and covers the regulations for both credit risk and market risk, as of this date.

Figure 2.1 maps the approaches under Basel 2 and the sequential sets of new regulations. The shaded boxes refer to the Basel extensions beyond Basel 2.

The Basel 2 blocks refer to the credit risk treatment for credit capital charges, with the risk-weighted assets according to banks' internal ratings. The current wave of regulations aims at reinforcing both the quality and the quantity of capital. The fraction of equity capital in total capital is reinforced and the capital ratio increases. The subsequent publications imposed new capital requirements, with a series of additions to the Basel 2 capital. The Basel 2 rules are expanded in Chapter 26 dedicated to credit regulations.

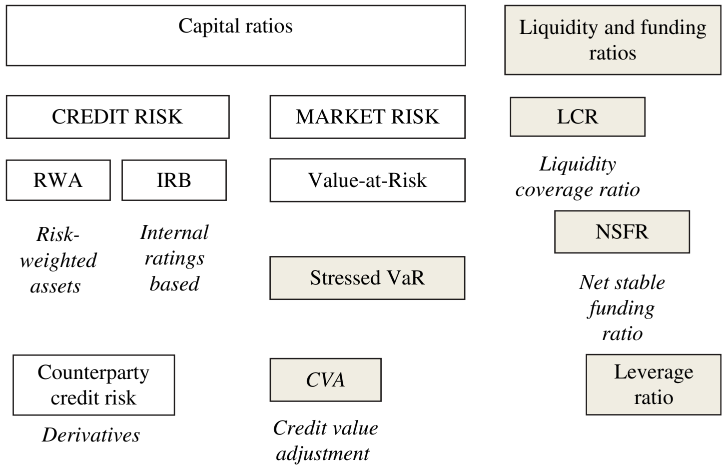


FIGURE 2.1 Overview of Basel regulations

The market risk approaches include the VaR-based capital plus the standardized approach of market risk for firms who do not comply with requirements of internal models. A stressed VaR was introduced as an additional capital charge in 2011. An alternate VaR measure is currently proposed for market risk. All market risk approaches are presented in Chapter 17 in the market risk section.

The treatment of counterparty credit risk has been enhanced with the credit-value adjustment (CVA) that measures the impact of deteriorating credit standing on the value of derivative instruments. The CVA adjustment is introduced in Chapter 22 on counterparty credit risk.

The blocks referring to liquidity and funding ratios cover three ratios gradually introduced by Basel 3. The Liquidity Coverage Ratio (LCR) imposes a minimum level of liquid assets for facing market disruptions to banks' funding. The Net Stable Funding Ratio imposes a minimum level of long-term funding, depending on banks' assets. The leverage ratio caps the size of the balance sheet and of certain off-balance sheet commitments as a function of the capital base. The three ratios are presented in Chapter 3 on balance sheet compliance, where the calculations and the consequences for the balance sheet of banks are detailed through a simplified example.

3

Balance Sheet Management and Regulations

The purpose of asset-liability management (ALM) is to manage assets and liabilities in conjunction with, rather than independently of, the bank so as to finance the bank and control the liquidity risk and the interest rate risk. In the aftermath of the 2008 crisis, regulators recognized that banks were not resilient enough to sustain periods of liquidity and funding stress, and imposed new rules. Under the new Basel regulations, new ratios constrain the balance sheet structure. Conventional ALM techniques apply once the bank complies with the rules. Many current issues relate to the implications of the new Basel 3 on the structure of the balance sheet and on profitability.

This chapter introduces the new regulatory ratios. It shows how compliance with the set of the new regulatory ratios has a direct effect on how banks manage their balance sheet. These impacts are detailed using an example of a typical banking book.

Contents

3.1 THE NEW REGULATORY RATIOS

The 2011 document "A global regulatory framework for more resilient banks and banking systems", dated December 2010, revised June 2011 [24], requires that the capital adequacy ratio be enhanced and introduced two new ratios: the Liquidity Coverage Ratio (LCR) and the

Net Stable Funding Ratio (NSFR). The reform is gradual, with full changes being enforced by 2018. A new leverage ratio is also introduced, aimed at preventing an excess buildup of credit in expansion phases.

3.1.1 Capital Adequacy

Regulatory capital is divided into Tier 1 and Tier 2. Tier 1 capital includes equity and retained earnings. Tier 2 capital is made of subordinated debts. Tier 1 capital is available when the bank is solvent and operates as a going concern. If a bank's losses exceed its equity base, it should pay back all creditors. Tier 2 capital is relevant when the bank is no longer a going concern, under the "gone-concern" view.

Under the new regulation mentioned above, Common Equity Tier 1 must be at least 4.5% of risk-weighted assets at all times. Tier 1 capital must be at least 6.0% of risk-weighted assets. Total capital (Tier 1 plus Tier 2) must be at least 8.0% of risk-weighted assets at all times. An additional equity capital conservation buffer is required to ensure that banks build up capital buffers outside periods of stress, which can be drawn down when losses are incurred. A countercyclical buffer can be required for protecting the banking sector from excess aggregate credit growth by raising the cost of credit.

The cumulative effect of these requirements is that the ratio of core capital to risk-weighted assets would reach 10%, or more, by the time the new regulations would be fully enforced.

The Basel 3 document also addresses the systemic risk implications of large banks, with additional capital buffers, and a greater supervisory discipline over banks designated as "systemic" because of their interconnectedness with the rest of the financial system.[1]

3.1.2 The Liquidity Coverage Ratio (LCR)

The goal of the LCR is to improve the short-term resilience of a bank's liquidity risk profile by ensuring that it has sufficient high-quality liquid assets to survive an acute stress scenario lasting for one month.

The LCR imposes that the liquidation value of eligible short-term assets be higher or equal to the "net stressed outflows" measured over a period of 30 days. The stress scenario might include: a significant downgrade of the institution's public credit rating; a partial loss of deposits; a loss of unsecured wholesale funding; a significant increase in secured funding haircuts; and increases in collateral calls on contractual and non-contractual off-balance sheet exposures, including committed credit and liquidity facilities.

The net cash flows result from the runoffs of assets and liabilities under adverse conditions. They are measured from factors, or percentage runoffs, applied to assets and liabilities. On the asset side, the factors are fractions of assets expected as cash inflows. Term loans have low factors, while traded assets of good quality, which can be sold easily, have higher factors.

1 See "Addressing systemic risk and interconnectedness", paragraph 32 of the 2010 Basel 3 document [24]. Systemic risk is addressed in the literature: see Arnold, B., Borio, C., Ellis, L., Moshirian, F. (2012), Systemic risk, macroprudential policy frameworks, monitoring financial systems and the evolution of capital adequacy [14]; or Nijskens, R., Wagner, W. (2011), Credit risk transfer activities and systemic risk: How banks became less risky individually but posed greater risks to the financial system at the same time [111].

The stressed outflows are calculated from runoff factors applied to liabilities, which measure the expected outlays of cash for various resources. Short-term wholesale debt has a runoff rate of 100%. Other resources, such as deposits, are more stable but they are nevertheless expected to face withdrawals under stressed conditions. Long-term resources, such as capital and issued bonds, have no runoff over the short term.

3.1.3 The Net Stable Funding Ratio (NSFR)

The objective of the NSFR is to promote the resilience over a longer time horizon than the LCR by creating additional incentives for a bank to fund its activities with stable sources of financing. The NSFR aims to limit overreliance on short-term wholesale funding during times of buoyant market liquidity and encourages better assessment of liquidity risk across all on- and off-balance sheet items. The NSFR has a time horizon of one year and imposes that the resources that regulators see as non-volatile over one year be at least equal to the amount of assets that are seen to stay in place under the same horizon.

The ratio is implemented by comparing the available stable funds (ASF) to the required stable funds (RSF). The stable funds are the financing that is expected to stay in place for an extended period of at least one year, excluding any volatile debt. The required stable funds represent the amount of assets that supervisors believe should be supported with stable funding.

According to the NSFR, the ratio of ASF to RSF should be above one. For measuring required and available stable funds, percentages called factors, are used for weighting assets and liabilities. The RSF are derived from assets, by applying RSF factors to existing assets.

The RSF factors measure how easy it is to turn assets into cash. They range from 0%, for those assets that do not require stable funding, to 100% for those assets that should be supported entirely by long-term funds. Good quality market instruments, such as investment grade bonds, do not require 100% of stable funds because they can be sold or financed by pledging them for borrowing. On the other hand, some loans with maturity longer than one year would require close to 100% stable funding.

The ASF factors measure how stable the resources are and are in the range of 0 to 100%. Factors in the upper range are applied to stable resources, such as equity and bonds issued by the bank. Factors in the lower range are used for resources that are considered as less stable, or volatile, such as the fluctuating fraction of deposits and short-term interbank debt.

3.1.4 The Leverage Ratio

The leverage ratio is intended to prevent the excessive buildup of exposures during expansion period. Banks can build up leverage in expansion, even when they maintain a strong capital base. The purpose of the ratio is to limit the expansion of the balance sheet in growth periods, and, consequently, limit the deleveraging of the balance sheet under recession periods.[2]

The leverage ratio imposes that core capital be at least 3% of the balance sheet size, plus some off-balance sheet commitments such as uncancellable banking commitments. The ratio is implemented over a test period extending until 2017.

2 Sources on leverage and liquidity include Acharya, V. V., Viswanathan, S. (2011), Leverage, moral hazard, and liquidity, [2], and Adrian, T., Shin, H-S. (2010), Liquidity and leverage, [4].

3.2 COMPLIANCE OF A COMMERCIAL BALANCE SHEET: EXAMPLE

The combination of the capital adequacy ratio, the LCR, the NSFR and the leverage ratio results in four constraints on the balance sheet of a bank. An example of a typical balance sheet is used below to show how compliance can potentially reshape a typical balance sheet and impact the profitability of a bank.

The combination of four ratios makes the management of the balance sheet a constrained exercise. The mix of assets and liabilities of the banking book is business driven. In a typical commercial bank, the bank holds corporate loans and loans to individuals arising from consumer lending and mortgage loans, plus an investment portfolio made of financial assets managed under a buy and hold policy. On the liabilities side, commercial resources are the deposits.

Assuming that the mix of commercial loans and deposits is given, the bank should manage its portfolio of liquid assets and its financing in order to comply with regulatory constraints.

The following example relies on a simplified balance sheet of a hypothetical commercial bank (Table 3.1). The commercial activities include mortgages plus corporate and retail loans. The bank also has an investment portfolio, which is assumed to be made of investment-grade

TABLE 3.1 A sample balance sheet

€ billion	*Natexia*
ASSETS	
Mortgage loans	20.04
General lending	30.07
Investment portfolio	16.79
Money market assets	1.20
Balancing treasury securities	*0.00*
Earning assets	**68.10**
Other assets	8.00
TOTAL ASSETS	**76.10**
LIABILITIES AND EQUITY	
Deposits	37.10
Wholesale debt	19.30
Issued bonds	5.00
Balancing interbank debt	*0.00*
Other liabilities	8.00
Total liabilities	69.40
Subordinated debt	3.35
Equity	**3.35**
Total equity and liabilities	**76.10**

bonds. On the resource side, the bank has long-term resources combining capital and subordinated debt, plus bonds issued in the capital markets. The bank's main resources are retail deposits and wholesale debt. The purpose of the example is to determine whether the initial balance sheet is Basel 3 compliant and what are the required changes to make it compliant.

3.2.1 Capital Base

The starting point of calculations should be the capital base since it sets a cap on the maximum size of risk-weighted assets and of the (unweighted) balance sheet size.[3] The capital ratio sets the maximum size of risk-weighted assets. Using a capital ratio of 10%, in line with Basel 3 expected requirements, the maximum value of risk-weighted assets is 3.35/10% = 33.5.

The leverage ratio sets the maximum size of the balance sheet. Given core capital, this size cannot exceed 3.35/3% = 111.7.

3.2.2 Risk Weights, Asset Mix and Capital

The next step addresses asset size and asset mix between mortgages and lending activities, which requires the risk weights. It is assumed that risk weights for those commercial assets are as in Table 3.2.

The acceptable size of assets and the acceptable size of risk-weighted assets depend on the asset mix. Investments should also be included, but, if the core business is in commercial activities, they need not be considered in a first step.

The regulatory ratios do not allow any commercial asset mix. For example, if the bank were lending only through mortgages, with a low risk weight, the maximum size of the mortgage portfolio is reached when they have a risk-weighted value of 33.5. This size would be 33.5/20% = 167.5. Lending through mortgages only would result in a size exceeding the maximum value from the leverage ratio, given existing core capital of 3.35. On the other hand, if the bank activities were concentrated on general lending, given a risk weight of 100%, the size of the loan portfolio would be capped at 33.5.

If the asset mix is kept as it is currently, 40% mortgages and 60% lending, the average risk weight is 68%.[4] The maximum size of the balance sheet, given the maximum risk-weighted assets of 33.5 is: 33.5/68% = 49.3. This size is within the limit set by the leverage ratio. Note,

TABLE 3.2 Natexia risk weights for mortgages and general lending

Natexia	*Risk weights*
Mortgages	20%
General lending	100%
Investment portfolio	20%
Balancing treasuries	0%

3 In this simplified example, off-balance sheet commitments are ignored.

4 The average risk weight is $20\% \times 40\% + 100\% \times 60\% = 68\%$.

however, that the new total size of assets is now lower than the current size of the loan portfolio, 50.1. The reason is the higher capital ratio of 10%.[5] This illustrates the simple fact that a higher capital requirement implies a cap on the volume of credit.

3.2.3 Compliance with NSFR

For checking compliance with the NSFR, ASF and RSF factors are required. If loans are entirely long term, a sensible RSF factor is 100%. In this example, it is assumed that investments do not require any stable funding. For ASF factors, capital and issued debt are factored 100% stable resources. A fraction of the deposit base can be short term, due to seasonal and short-term volatility, for example 10%. Finally, wholesale debt is volatile and is assigned an ASF factor of 0%. The calculation of the NSFR is detailed in Table 3.3.

The NSFR is lower than one, which means that there is a deficit of available stable funds. The reason is the large reliance on wholesale debt. For compliance, some long-term debt should replace some of the wholesale debt. Adding the investment portfolio to the calculation would increase the required stable funds, which would further constrain the size of the loan portfolio given the existing financing.

Assuming that the current financing mix is maintained, in a first step, the size of assets has to be downsized again with the 0.92 scaling factor, in order to align RSF with ASF. Given that mortgages and loans have an ASF factor of 100%, their new size becomes 92% of the original size, or 45.09.

The unweighted existing resources are 68.1. As they exceed the new size of assets, the volume of debt financing needs to be adjusted. In Table 3.4, the wholesale debt is the balancing

TABLE 3.3 The new commercial balance sheet: compliant with NSFR

	RSF factor	*Assets*	*RSF*
Mortgages	100%	19.71	19.71
General lending	100%	29.56	29.56
Investments	20%	0.00	0.00
Required stable funding		***49.26***	**49.26**
	ASF factor	**Liabilities**	**ASF**
Capital	100%	3.35	3.35
Subordinated debt	100%	3.35	3.35
Issued debt	100%	5.00	5.00
Deposits	90%	37.10	33.39
Wholesale debt	0%	19.30	0.00
Available stable funding		***68.1***	**45.09**
NSFR			**0.92**

5 It can easily be checked that the initial balance sheet is compliant with core capital and subordinated debt, and a capital ratio of 8%.

TABLE 3.4 The new compliant balance sheet

ASSETS	
Mortgages	18.04
General lending	27.05
Total assets	**45.09**
LIABILITIES AND EQUITY	
Deposits	37.10
Wholesale debt – balancing	*−3.71*
Issued debt	5.00
Subordinated debt	3.35
Capital	3.35
Total equity and liabilities	**45.09**

item, adjusted for matching the volume of assets and of liabilities. The resources should be reduced by 68.1 – 45.09, or 23.01. This is more than the initial wholesale debt. In other words, the bank now has excess resources from its stable financing that would show up as liquid assets for an amount of 23.01 – 19.30 = 3.71.

The new commercial balance sheet is now compliant with the stable funding ratio. But the adjustment process is not completed since a final step is needed to check compliance with the LCR.

3.2.4 LCR Compliance

The calculation of LCR requires the runoff rates of resources and the liquid factors for assets. If excess resources are invested in liquid assets, the liquid asset cushion is 3.71. The unique runoff factor needed is that of deposits, the current balance of which is 37.10. With a runoff factor of 10%, the corresponding outflow would be 3.71, which matches the excess of resources. The liquid portfolio is high enough. The current balance sheet is now compliant with the LCR.

The rebalancing illustrates the effect of the new constraints. The mechanism was illustrated by assuming existing resources as given. It is obvious that using more long-term debt, or increasing the capital base, would allow the banking book to keep its current size and even to expand it. Similarly, the investment portfolio, which was ignored in this simulation, considering that the core business was commercial banking, could be accommodated by adjusting the capital base and the issued debt. The purpose was only to illustrate that compliance with the new constraints would reshape a balance sheet.

3.3 CREATION OF VALUE

The simulation does not take into account the profitability of the bank and ignores, notably, the higher cost of stable debt versus wholesale debt. The higher cost of stable resources potentially implies that the profitability of existing activities be adversely affected.

A primary criterion for economic performance is the bank's return on capital, the ratio of net income to regulatory capital. There are strong interdependencies between the structure and size of a banking portfolio, its financing, its risk and the final return on capital.

On the asset side, a change of asset volume and of asset mix has an impact on both profit and risk. The size of the banking portfolio has an impact on the volume of revenues and the capital charge. The structure of assets has an impact on margins and on capital, as these depend on margins and risk weights differentiated by asset segment. On the liability side, the structure of debt determines the financing cost of the bank. The final impact on the return on capital combines all elements of revenues and financing costs, and capital charge.

A simulation of return of capital derives from the simulation of the balance sheet. The return on capital is risk adjusted as the capital is a metric of risk. Its determination depends on both risk weights and income. Figure 3.1 shows how the various elements of profitability and risk combine into the return of capital. Some critical parameters are in shaded boxes.

On the risk side, the capital depends on the volume of assets and asset mix. In the example, the new asset volume is 45.09 and the asset mix corresponds to an average risk weight of 68%. The risk-weighted assets are 30.66, and, with a capital ratio of 10%, the capital charge is 3.07.

On the profitability side, the net interest income depends on the margins of the asset segments and on the cost of financing. Presumably, reliance on additional long-term financing, plus the low return of liquid assets, has an adverse effect on the bank's return. It is assumed here that the net interest income margin is 2.5% of the balance of earnings assets, once the financing mix is adjusted for compliance, resulting in a net interest income of 1.13 (45.09 × 2.5%).

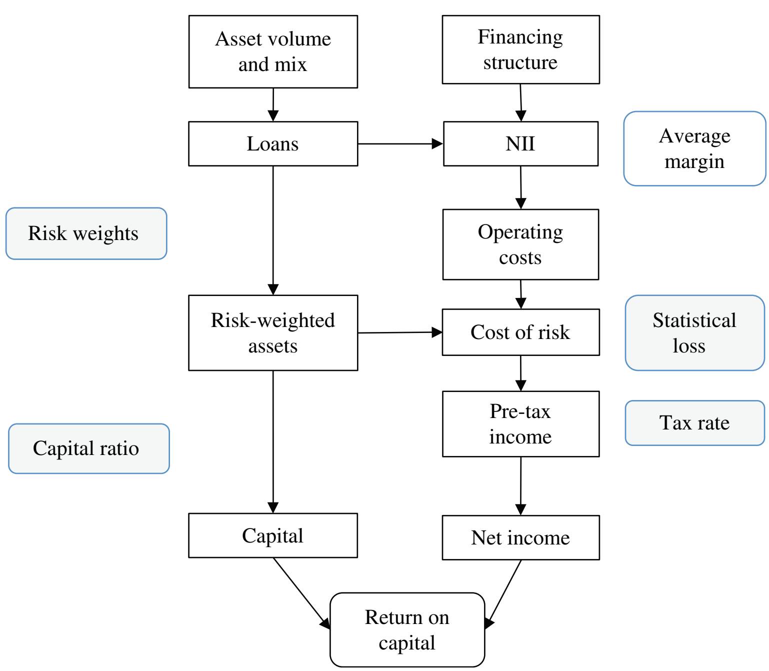


FIGURE 3.1 Return on capital

For determining the profit after tax, operating costs, the provision from loan loss and the tax should be deducted. The loan loss provision depends on the risk of loans. The operating cost depends on the management ratios of the bank.

For example, sensible parameters could be as follows: a cost to income ratio of 50%, a provision for loan losses of 0.5% and a tax rate of 34%. With these values, operating cost is 0.56, loan loss provision is 0.23, taxable income is 0.34 and net income is 0.22. Using these figures, the return on capital is 7.28%. Other scenarios could easily be simulated.

4

Liquidity Management and Liquidity Gaps

The balance sheet ratios of Basel 3 take care of minimum liquidity requirements. For managing liquidity in a non-stressed environment, banks need to make financing and investing decisions depending on their cash positions, current and expected. For financing decisions, they need to assess how much they need, in the case of projected deficits, from which date and until when. Managing the funding requires projecting the existing assets and liabilities for determining any future imbalance. These mismatches are called liquidity gaps.

This chapter presents the conventional gaps for managing liquidity and addresses some of their limitations.

Contents

4.1 LIQUIDITY AND LIQUIDITY RISK

Before the financial crisis, a shortage of liquidity referred to a situation when a bank raises cash at a higher cost than usual conditions, with an extreme shortage leading to failure. The financial crisis of 2008 demonstrated that liquidity disruption could be system-wide, that liquidity could evaporate in a matter of days and that liquidity is highly scenario dependent: “One moment it is there in abundance, the next it is gone.”[1] Accordingly, the cost of liquidity cannot be ignored anymore and has to be fully recognized in the bank’s cost of funds: the cost of funding liquidity is addressed in Chapter 10 dedicated to internal pricing of funds.

4.1.1 Liquidity and Financing

Liquidity is often impaired by adverse events, which are not directly related to liquidity issues. Bank-specific events include adverse earnings or loss announcements, or a downgrade of the firm. System-wide events are market disruptions, as the last financial crisis demonstrated. However, the chances that events materialize in a liquidity crisis depend on how vulnerable the sources of liquidity are at the firm level.

Banks raise cash by borrowing short-term or stable resources, secured or not, or alternatively raise cash from the sale of assets. Borrowing uses a variety of instruments such as interbank debt, wholesale debt or customers’ deposits. Banks also rely on committed lines of credit from other banks, on which they can draw if need be. Funding risk has to do with the availability or unavailability of sources of funds at a reasonable cost of funds. The trust of the providers of liquidity is critical. Wholesale lenders can stop lending almost whenever they decide. Banks can refrain from letting borrowers draw on a credit line when they think that the risk of a legal procedure is lower than the risk of losing the money lent. The reliance on stable sources of funds mitigates funding risk. Stable funds include bond issues, secured borrowings over a horizon beyond one year and the stable fraction of deposits.

An alternate source of cash is the sale of market assets. Most financial firms hold financial assets. All assets are liquid if we extend the time frame. Even loans can be traded or securitized if time permits. High-quality and short-term assets are valuable because they can be easily sold, and without significant discount since they are not sensitive to interest rate variations. Many other assets are tradable but they are less liquid, which implies a longer time frame and higher potential discounts from variations of interest rate or because the volume of sales depresses their prices.

Banks might refrain from holding large balances of high-quality and short-term assets because of their lower earnings. Because such assets are the unique alternate source of funds in the presence of funding disruptions, regulators introduced the liquidity cover ratio for imposing a minimum liquidity buffer to banks.

Cash can also be obtained without selling assets, by pledging them for raising secured debt. The amount of secured borrowing that can be potentially raised is the asset current value minus haircuts[2] required by lenders. Good quality assets, those assets with a good credit standing of issuer and that can be traded, can always be “turned into cash”, either through sales or secured borrowing.

Banks tend to finance long-term assets with short-term resources. The resulting mismatch is a major source of liquidity risk. Mismatches of maturities generate liquidity risk because the

1 “When the river runs dry”, The Economist, February 11th, 2010.

2 The haircut is the gap between the asset value and the smaller volume of borrowings obtained from pledging the asset.

short-term debts need to be rolled over for financing the long-term assets. They also have interest rate risk because the spread between long-term and short-term interest rates fluctuates with market movements.

4.1.2 Liquidity Risk in the Banking Book

All stochastic factors affecting cash flows contribute to the level of liquidity risk of banks. Stochastic cash flows occur because clients hold options to increase or decrease their loans or deposits at the bank.

A significant portion of assets and liabilities within a bank's balance sheets typically consists of so-called "non-maturing accounts", or accounts with an indeterminate maturity. The cash flows to and from these accounts are not deterministic but stochastic, and contribute to liquidity risk.

Such accounts include demand deposits on the liability side and consumer loans or overdrafts on the asset side. On the liability side, banks' clients may freely increase or withdraw their deposits. On the asset side, consumer loans and credit card accounts are rolled over within the bank's authorization limits, at customers' initiative. Clients can increase or repay their drawn amounts on consumer loans, such as credit cards and overdrafts, at any point in time and without penalty. Those loans are the equivalent of deposits on the asset side. Fixed-rate mortgage loans can be prepaid by borrowers who would like to take advantage of declining rates. Stochastic cash flows arise from such rights embedded in banking products.

4.2 LIQUIDITY GAP TIME PROFILES

Liquidity gaps refer to the projected imbalances of sources and uses of funds. Gap reports provide the necessary information for taking funding or investing decisions. Gap management consists of managing the projected mismatches between assets and liabilities.

The liquidity gaps are the differences, at future dates, between the projected balances of assets and liabilities of the banking portfolio. The existing assets and liabilities amortize (run off) gradually over time, and the time profiles of their balances are declining. Projections are "static" when they ignore new loans, new deposits or debts at future dates.

Static liquidity gaps have a time profile depending on how assets and liabilities run off. At a future date, t, the liquidity gap is the algebraic difference between the projected balances of existing assets and liabilities. There are as many gaps as there are time points over the planning horizon. At each future date t:

$$\text{Liquidity gap}\,(t) = \text{Assets}\,(t) - \text{Liabilities}\,(t)$$

Figure 4.1 shows typical situations for the time profiles of assets, liabilities and gaps. The current gap, as of today, is zero as the sizes of assets and liabilities match. Any excess of funds, or any deficit of funds, at this starting date is supposedly fully funded or invested.

As time passes, the existing assets and liabilities amortize. For static gaps, new loans and new debts are ignored, and the outstanding balances of existing assets and liabilities decline. Once all debts are fully repaid, the time profile of liabilities hits the floor of capital. The interest rate flows generated by existing assets and liabilities are not shown in these charts. With the

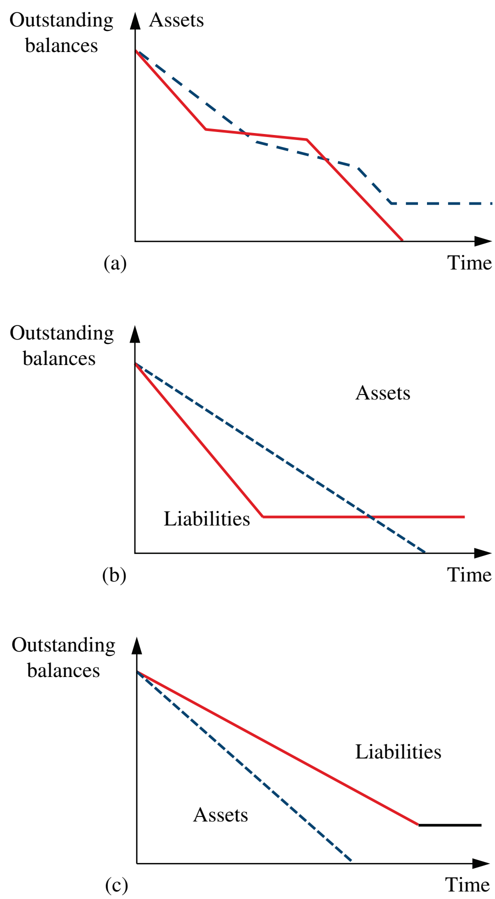


FIGURE 4.1 Three basic situations (a) Near zero liquidity gaps: cash matching (b) Deficits (c) Excess funds

algebraic definition of liquidity gaps, a positive gap between assets and liabilities is equivalent to a deficit, and vice versa.

The time profiles of liquidity gaps can have three basic shapes: assets and liabilities are matched; there are deficits of funds; or there are excesses of funds. Positive gaps mean that the balance sheet generates projected deficits of resources relative to assets. This is the gap profile of asset-driven banks. These banks have to raise funds periodically to match the development of their assets. The opposite case is that of liability-driven banks. As time passes, these firms have excess resources and need to expand new lending or investing activities.

Cash matching implies that the time profiles of amortization of assets and liabilities are – approximately – identical. In this case, there is no significant maturity mismatch and all liquidity gaps are near zero. The amortization of the balance sheet does not generate any deficit or excess of funds, and the repayment schedule of debt replicates the repayment schedule of loans. The balance sheet remains balanced, as time passes, without generating any need or investment of funds. Full matching eliminates liquidity risk.

Matching the runoff profiles of assets and liabilities is the reference case, with liquidity gaps closed. Furthermore, if interest rates have the same nature – fixed or floating – on both sides of the balance sheet, there is no uncertainty on net interest income since all interest revenues and costs are deterministic.

Consider the case of a single loan, for example a bullet loan. If the loan is financed with a debt replicating the time profile of the loan, the liquidity gaps until maturity of the loan are zero. Assume further that both loan and debt are fixed rate. The loan rate will usually be higher than the one of debt, the difference being the spread between the client's rate and the cost of debt. The net interest income (NII), the interest revenue minus interest cost, of the portfolio combining the loan plus the mirroring debt is locked for the duration of the loan.

Under cash matching, a similar conclusion applies to a floating-rate loan backed with a floating-rate debt. Assume that the loan is indexed to the Libor 3 month. Over a year, there are four reset dates, including the date at which the loan is granted. At each reset date, both the loan rate and the debt rate are reset with the same Libor reference, and the NII remains locked from origination.

Alternatively, any mismatch between maturities and interest rate generates both liquidity risk and interest rate risk. A standard position of banks consists of lending for longer maturities than those of liabilities, or being asset driven. Some banks have this position because they prefer to rely on short-term debt for capturing a positive spread between long-term and short-term rates. In such instances, the bank has recurring deficits of funds and needs to roll over the debt, the interest rate of which fluctuates as time passes. In the opposite case of positive gaps, there is a recurring excess funding. There is no liquidity risk, but there is interest rate risk, because the interest rates of future loans or investments from excess funds are unknown viewed from today.

When mismatches result from a deliberate financing policy of the bank, they should be consistent with expectations on interest rates. For instance, in a fixed-rate universe, keeping a balance sheet underfunded (positive liquidity gaps) makes sense if short-term rates are lower than long-term rates, or when betting on declining interest rates so that any delayed financing will be at lower rates. Conversely, being overfunded (negative liquidity gaps) means that funds are raised in advance, which could make sense if higher interest rates are expected.

4.3 TYPES OF LIQUIDITY GAPS

Several liquidity gaps could be considered: marginal gaps, static and dynamic gaps. The usual liquidity gaps are static, derived with a scenario of the runoff of existing assets and liabilities only. When gaps include the new loans or debts, they are called dynamic. Under a dynamic view, the outstanding balances of assets and liabilities often increase with time rather than amortize, as with static gaps, because new loans and new debts pile up over the amortizing loans and debts.

Incremental or marginal gaps are the differential variations between two adjacent time points of assets and liabilities. The cumulated marginal gaps, from today up to a date t, match the gap between the outstanding balances of assets and liabilities as of the same date. Table 4.1 is an example of the time profile of both cumulative gaps and incremental gaps.

In the example in the table, assets amortize slower than liabilities. The inflows from repayments of loans are smaller than the outflows from the amortization of debts. Deficits cumulate from one period to the other, except in period 5.

TABLE 4.1 Time profiles of outstanding assets and liabilities and of liquidity gaps

Dates	*1*	*2*	*3*	*4*	*5*	*6*
Assets	1,000	900	700	650	500	300
Liabilities	1,000	800	500	400	350	100
Gap(1)	0	100	200	250	150	200
Assets amortization		−100	−200	−50	−150	−200
Liabilities amortization		−200	−300	−100	−50	−250
Marginal gap(2)		100	100	50	−100	50
Cumulative gap(3)		100	200	250	150	200

(1) Calculated as the difference between assets and liabilities. A positive gap is a deficit that requires funding. A negative gap is an excess of resources to be invested.
(2) Calculated as the algebraic variation of assets minus the algebraic variation of liabilities between t and $t-1$.
(3) The cumulative gaps are identical to the gaps calculated with the outstanding balances of assets and liabilities.

A positive marginal gap means that the algebraic variation of assets exceeds the algebraic variation of liabilities: it is an inflow. A negative marginal gap means that assets amortize at a slower pace than liabilities at that period: it is an outflow.[3]

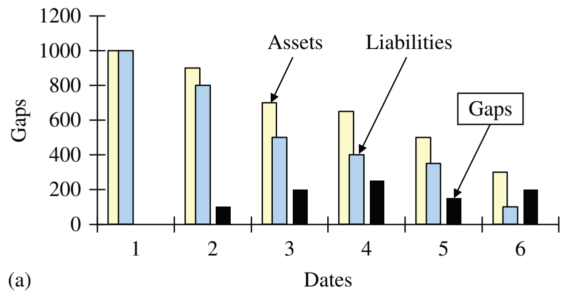
The marginal gaps represent the new funds to be raised over the period, or the new excess funds of the period available for investing (Figure 4.2). For example, the marginal gap for the period between date 2 and date 3 is +100. It is the additional deficit of the period. It is also the amount of funding required if previous deficits up to date 2 have been closed.[4] Marginal gaps represent the new amounts raised, or invested, over a given horizon, if, and only if, the incremental financing or investments between initial date and current date are still in place today. On the other hand, cumulative gaps are cumulated deficits or excesses of funds, assuming no intermediate fundraising or investing took place between the initial date and the date at which the gap is projected.

Table 4.2 shows a typical liquidity gap report. The time buckets are quite large in this sample table. Most gap reports provide gaps over frequent time points, for example monthly for up to one to three years.

The calculation of gaps is simple, but it is conducted over a very large number of lines in the balance sheet. Standard asset-liability management (ALM) information systems provide gap reports, given the details of the amortization of each item of the balance sheet. A sample output of such software, used in a retail bank, is shown in Figure 4.3. The top part of the picture shows how assets amortize through time by monthly steps. The grayed area is the outstanding monthly balance, measured along the Y-axis. The liabilities are shown as another grayed area in the down section of the same graph. The time profile of liquidity gaps is the difference between assets and liabilities. The gaps are the differences between those two areas. The isolated lines show the average interest rates of assets and of liabilities, respectively.

3 For instance, if the amortization of assets is 3, and that of liabilities is 5, the algebraic marginal gap, as defined above, is $-3-(-5)=+2$, and it is an outflow. It is important to keep a clear convention, so that the figures can be easily interpreted.

4 In other words, 100 raised at date 2 is not amortized until a date posterior to date 3.



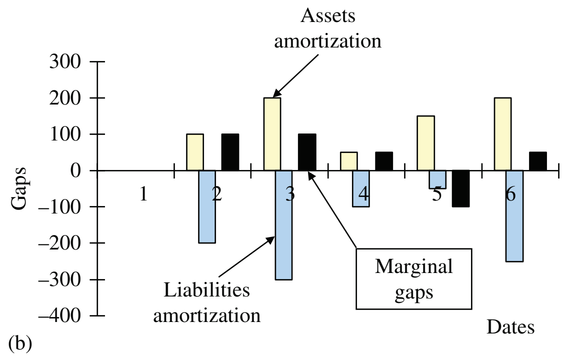


FIGURE 4.2 Time profile of liquidity gaps: cumulative and marginal (a) Cumulative (b) Marginal

Figure 4.4 shows the time profile of liquidity gaps, magnified from Figure 4.3. The gaps are cumulative up to each future date. They decline to zero in the long term as the balances of assets and liabilities amortize completely. In general, the gaps are subject to limits, which cap the amount of financing required. The limit is a horizontal line in the figure that caps the financing needs of the bank. If the limit is breached at some future point, a new financing has to be raised in order to keep the projected deficits below the limit.

Information systems dedicated to ALM embed a number of useful functionalities. For example, outstanding projected balances can be end-of-month or monthly averages. It is possible to drill down in the gap for any date to find out which transactions, or products, contribute to the gap. Similarly, the variations in the gap profile can be traced back to the corresponding transactions.

TABLE 4.2 Sample liquidity gap time profile

In million 31/12/2010	*≤1 month*	*1–3 months*	*3–12 months*	*1–5 years*	*5–10 years*	*>10 years*	*Not defined*	*Total*
Assets	98,044	25,163	32,168	53,836	29,167	29,938	36,271	304,587
Liabilities	133,253	40,492	37,355	39,252	11,487	8,736	34,012	304,587
Liquidity gap	−35,209	−15,329	−5,187	14,584	17,680	21,202	2,259	—

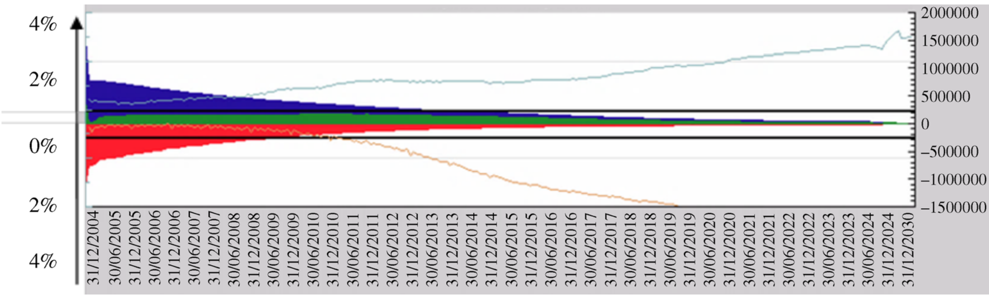


FIGURE 4.3 **Time profiles of assets and liabilities (in thousands of euros)**

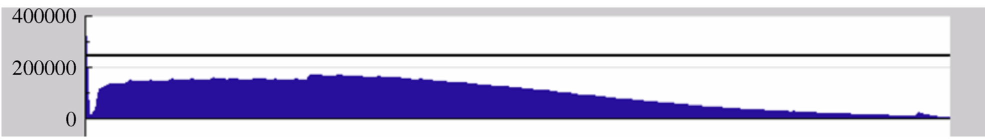


FIGURE 4.4 **Time profile of liquidity gap and limits**

4.4 MANAGING INCREMENTAL GAPS

As in all projections, gaps have limitations. Notably, they are based on assumptions and rules for projecting the runoffs of accounts with indeterminate maturities. Gap reports are not intended to model the behavior of these accounts, but they need to be based on sensible assumptions.

In gap reports, projections are based on discrete scenarios, where runoffs are seen as deterministic. But a significant fraction of assets and liabilities generate stochastic cash flows. Deposits have no maturity, but are sticky; mortgages can be pre-paid or their customer's rate be renegotiated; credit lines of off-balance sheet commitments can be drawn at the client's initiative within limits; consumer and credit card loans tend to roll over. With indeterminate maturities, the runoffs are stochastic rather than deterministic. A general rule is that the liquidity scenarios should be constructed from effective, rather than contractual, runoff schedules.

Frequent updates of reports take care of short-term uncertainty. Between two consecutive gap reports, the effective balance of all accounts, including those with random cash flows, is observed. In Figure 4.5, the gaps shown are seen from date t. At this date, the liquidity gap is closed but the gap at $t + 1$, viewed from t, is open. The picture changes as time passes. If projections are exact and if there are no new debts or loans, the gap at $t + 1$ is identical to the projected gap viewed from the initial date t. Once this next date, $t + 1$, is reached, the positive gap is closed with new funding.

However, in general, the new gap at $t + 1$ differs from its projected value because of random cash flows from assets and liabilities existing at t, and because of new loans and debts. The actual runoff of resources, deterministic or not, is now recorded at $t + 1$. The entire time profiles of assets and liabilities are also updated. Frequent updates allow gap management to be incremental, with periodic adjustments, which allows adjusting to expected and unexpected changes as they occur.

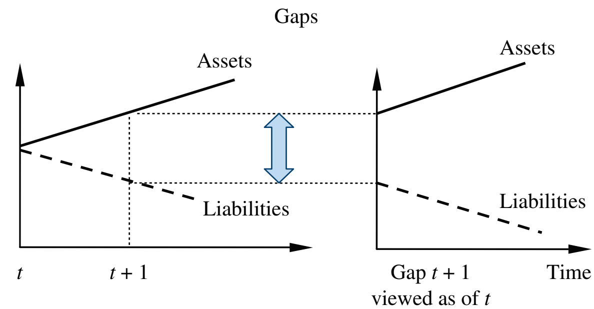


FIGURE 4.5 How gap profiles change

Incremental gap management takes care of the short-term uncertainty. But, in general, gap reports extend over long horizons, eventually up to the longest maturity of accounts in the balance sheet. Gaps over the next two or three years are actively managed.

Non-maturing accounts and mortgage runoffs have generated a considerable debate in the industry on how to manage risks. Discrete scenarios use projections relying on conventions and measures of effective runoffs. For loans, statistics of amortization are available. Effective runoffs are derived from historical data, as the average amortization for pools of loans. Statistics allow more elaborated modeling of the behavior of the accounts, for example for prepayments of mortgages.

On the liability side, demand deposits have a legal zero maturity because they can be instantly withdrawn from the bank. However, a substantial fraction of current deposits is stable over time, or “sticky”. The remaining fraction is volatile, and is considered as short-term debt. The core deposits represent a stable resource, for which an amortization schedule is required for projecting liquidity gaps.

In many instances, rules rely on conservative conventions. For example, an annual constant amortization rate of 5%, or 10%, can be chosen for gap reports. These rates correspond to durations of, respectively, 20 or 10 years. Such conventions generate an additional liquidity gap equal to annual amortization, which, in general, is not in line with reality. However, conventional amortization is a common practice because it is conservative.

The alternate solution is to define an effective runoff for the stable fraction of deposits. Effective runoffs can be inferred from time series of the number of accounts and of deposit balances. Only those accounts that existed at some historical point in time, used as origin, should be monitored, which implies that existing accounts at origin be isolated from new accounts opened later.

The number of existing accounts decreases as time passes, as some accounts are closed, but the behavior of their aggregated balance also depends on the balance of each individual account. If the average unit balance, the amount held per account, were constant, the outstanding balance of deposits would amortize at the same rate as the number of accounts. But if the unit balance increases with the age of account, the aggregated balance of accounts decreases at a slower pace than the number of accounts, and can eventually increase over time.

Such analyses are used to support the projections of deposits. They have implications other than obtaining sensible liquidity gaps. For example, the sensitivity of the economic value of the balance sheet is a target variable for ALM (Chapter 7). It depends, among other factors, on the duration assigned to deposits.

An extended framework is required for a more advanced modeling of non-maturing accounts. Details are provided in Chapter 8, and the modeling of mortgages with the embedded repayment option is detailed in Chapter 9.

4.5 DYNAMIC LIQUIDITY GAPS

For liquidity management purposes, it is a common practice to focus primarily on the existing assets and liabilities and static gaps. The rationale of the usage of static gaps is that there is no need to obtain funds in advance for new transactions, or to manage today resources that are not yet collected. Funding the deficits or investing excesses of funds from new business occurs when they appear in the balance sheet, not earlier. Moreover, the liquidity gaps are continuously updated, and the new loans or debts will appear gradually in the updated static liquidity gap profile.

Dynamic liquidity gaps add to the amortization profiles of existing assets and liabilities the projected balances of new loans and new deposits, at the time of the gap report. Such balance sheet projections are used for budgeting purposes, and also serve to make sure that limits on financing raised at any future periods will not be exceeded.

Total assets and liabilities, existing plus new ones, tend to increase in general, rather than amortize. Gaps for both existing and new assets and liabilities are required to project the total excesses or deficits of funds between today and the future dates. Figure 4.6 shows new transactions combined with the existing assets and liabilities. Note that volumes of new assets and liabilities should be net from the amortization of the new loans and of the new deposits, since the amortization starts from the origination date of these transactions.

4.6 FUNDING LIQUIDITY MANAGEMENT

Controlling liquidity risk implies spreading over time the required amounts of funding, and avoiding unexpected gaps or gaps in excess of what the firm usually raises in the market. Limits serve for making sure that funds raised in future periods will remain within acceptable boundaries. The process requires spreading the debts raised in a sensible way across periods.

Illustrated below are two examples of profiles of existing assets and liabilities with initial deficits. In both cases, the resources amortize quicker than assets and the deficits extend, with

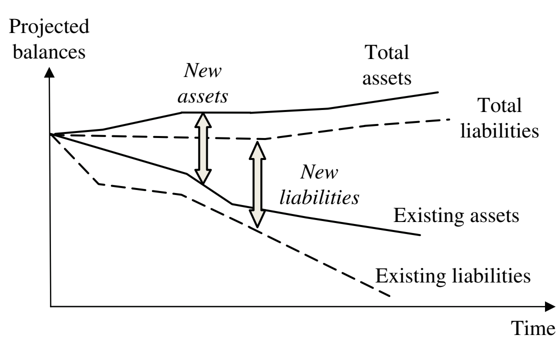


FIGURE 4.6 Gap profiles with existing and new transactions

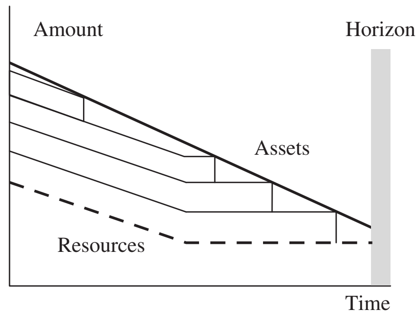


FIGURE 4.7 Closing liquidity gaps

varying amplitude, up to the horizon. The purpose of the examples is to show how gaps can be managed. It is assumed, for example, that the ALM committee wishes to close all gaps. This means that new resources should be contracted in order to raise the time profile of resources up to the one of assets. The issue is to define which new debts are consistent with the new goal.

For defining how much cash should be raised, when and until which date, defining a planning horizon is a prerequisite. New debts are defined by their size and by the initial date when they are contracted and by the final dates when they are repaid. "Layers" of debts starting from the final horizon are combined in order to raise the time profile of liabilities until it reaches the time profile of assets.

Figure 4.7 illustrates a first case. In the left-hand side, the gaps continuously decline until the horizon. Because the sizes of gaps vary with dates, several layers of debt are required to close them. More funds are needed today than at the management horizon, and cash has to be raised today. Layers of debts of varying maturities ensure that the resources profile replicates the asset profile. A first debt serves for closing the final gap, which is smaller than the current gap. Other debts close the intermediate gaps.

A second case is illustrated in Figure 4.8. In this second case, it is not feasible to close all gaps with debts contracted today. The same process does not apply because the gap starts increasing before declining again until the horizon. Closing the final gap with debt raised today would result in excess cash today because the final gap is larger than the initial gap. The maximum amount of debt contracted today, without having excess funds, is the initial gap.

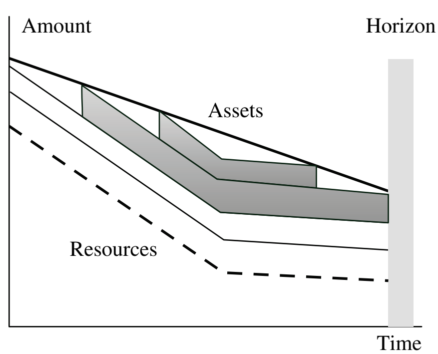


FIGURE 4.8 Closing liquidity gaps

Since this initial gap is open until horizon, the debt contracted should be in place until the final horizon. The final gap is not closed with this financing because the gap increases with time, peaks and then narrows down. It is easy to observe that intermediate liquidity gaps remain open. From a liquidity standpoint, this not an issue as long the required debts for closing those intermediate gaps do not exceed limits.

However, in a fixed-rate universe, such mismatches would generate interest rate risk since cash needs to be raised later at unknown rates. Hedging the interest rate risk would require the usage of derivatives, such as swaps, which allow setting today an interest rate for debts raised later, if needed.

4.7 LIQUIDITY CRISES AND STRESS SCENARIOS

Gaps do not address the situations of liquidity crises. Stress tests and contingency plans are better suited to such extreme situations. The framework for addressing liquidity management, notably under stressed conditions, has been now expanded with new regulations and new principles.

Liquidity crises can be triggered by many adverse events, which can be specific to a firm or system-wide. At the level of the firm, an unexpected release of information might cause usual lenders to cut down their credit lines. Unexpected and sudden losses threaten solvency and disrupt the willingness to lend by other financial players. A downgrade of the bank's rating can also have major effects on both the asset and the liability side. The cost of funding will increase and the bank might become ineligible for further lending by other institutions. The same cause can trigger margin calls, through which the bank is supposed to post more collateral in its debt, or, alternatively, reduce its debt.

Global events include: a system-wide "run on deposits"; a liquidity crunch caused by failures; and the fear of contagion of adverse events across financial institutions causing an increase of risk aversion and reluctance to lend. There are several examples of such crises: the Russian debt crisis, the "Y2K" fear of information systems failure; the 9/11 crisis; and the subprime crisis of 2007 and 2008, which showed that a freeze of the most liquid markets could effectively happen.

Stress tests consist of simulating what could happen in a worst-case event. The possible origins of a crisis and the historical events provide examples of such situations and can serve as a reference for designing the stress scenarios.

Contingent plans serve for assessing the ability to generate cash flows to meet debt commitments and deposit withdrawals, or margin calls on collaterals, under extreme circumstances. In order to provide some practical inputs to such plans, banks have to consider the size of the liquid asset portfolio, balance sheet and funding limits, maximum repos[5] that can be contracted, the diversification of sources of funding and the levels of unsecured funding. Once these are identified, a bank can further diversify sources of funding, increase the base for repo financing, etc. The report on sound liquidity management practices[6] provides a number of criteria for addressing such contingency plans.

5 A repo, or repurchase agreement, consists of lending a security in exchange for cash with agreement to buy it back. It is equivalent to a debt secured by the assets.

6 Basel Committee on Bank Supervision (2008), Principles for sound liquidity risk management and supervision [22].

5

Interest Rate Gaps

Interest rate risk refers to the sensitivity of earnings to shock on interest rates. Interest rate risk arises from the mismatches of maturities or of interest rate sensitivities of assets and liabilities. Interest rate gaps are differences of projected assets and liabilities calculated over subsets of balance sheet items that share the same interest rate reference. By contrast, the liquidity gaps are differences between the projected balances of all assets and liabilities of a financial firm.

The "gap model" relates the net interest income of a bank to the variations of interest rates. Interest rate gaps are appealing because they are extremely simple and easy to use for managing the interest rate exposure.

This chapter addresses interest rate risk and explains how interest rate gaps are used to control the effect of interest rate risk on the net interest income of banks. The limitations of conventional gaps are also discussed.

Contents

5.1 INTEREST RATE RISK

Interest rate risk exists when earnings are sensitive to the movements of interest rates. Loans or investments, term or saving deposits and financial debt generate revenues or costs that are driven by interest rates. Since interest rates are unstable, so are earnings.

All lenders and borrowers are subject to interest rate risk. Variable-rate borrowers and lenders have interest revenues or costs indexed to short-term market rates. Loans and debts that are fixed rate are also subject to interest rate risk. Fixed-rate lenders could lend at higher than their fixed rate if rates increase and face the risk of not lending at such higher rate. Fixed-rate borrowers would be better off by borrowing at lower rates when rates decline. Fixed-rate transactions are exposed to interest rate risk because lenders and borrowers might face a random opportunity cost arising from market movements.

There is no neutral position for interest rate risk, since there are always direct, or opportunity, gains and losses due to interest rate movements. Lenders and borrowers can only choose which type of exposure, variable rate or fixed rate, they want to have. They do so for new loans or debts. For existing loans and debts, they can still change their exposure, according to their perception of interest rates, by using interest rate derivatives.[1] Derivatives do not modify the original contract (debt or loan), but are new contracts.

Banks have exposure to interest rates both as lenders and borrowers. They also have an interest exposure on the future contractual cash flows already embedded in their existing assets and liabilities because these future flows will be financed or lent at interest rates that are unknown as of today. Liquidity risk therefore generates interest rate risk. Both projected and unexpected excesses and deficits of funds will be financed at interest rates unknown today. Implicit or explicit options embedded in banking products are another source of both liquidity and interest rate risk.

5.1.1 Interest Rate Risk, the Term Structure and Mismatch Risk

Mismatch risk refers to the risk of maintaining a gap between maturities and/or interest rate reset dates of assets and liabilities.

Such mismatches have an economic rationale when long-term rates are higher than short-term rates. By lending long while borrowing short, the bank captures the positive spread between long- and short-term rates. The same outcome occurs when lending at long-term rates while borrowing floating rates and matching maturities. The second position eliminates liquidity risk. Since floating rates are indexed to short-term rates, banks can still capture the spread between the loan and short-term rates. However, both situations are exposed to the fluctuations of the spread between long and short rates.

Mismatch risk is not specific to banks. Many financial entities tap liquidity in the short-term market while asset maturity is much longer. The liquidity risk is sustainable as long as there is no market disruption and the interest rate risk is significant. The savings and loans debacle in the US is an example of adverse effects of mismatch risk. The firms were lending at fixed rates and borrowing at floating (short) rates. When interest rates suddenly spiked because of anti-inflationary policies by the Federal Reserve in the early 1980s, they cumulated interest losses.

1 Interest rate derivatives are discussed in Chapter 6.

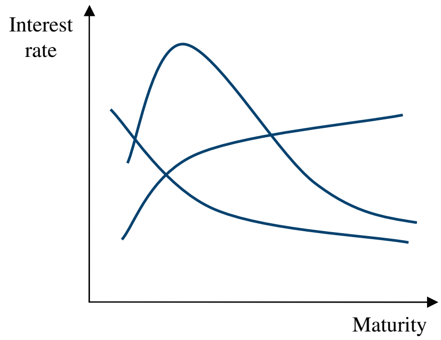


FIGURE 5.1 Typical shapes of term structures of interest rates

A mismatched position is a position on the slope of the yield curve. The yield curve, or the term structure of interest rates, plots the market rates against maturities. It is generally, but not always, upward sloping, meaning that long rates are higher than short rates.

There are many such curves as there are many interest rates. The yield to maturity is the return of a bond held until maturity. It is defined as the discount rate that makes the discounted value of all future flows, interest and principal, equal to the current price of the bond. When there is no intermediate coupon, the yield to maturity becomes the zero-coupon rate applying to the unique final flow. Zero-coupon rates apply to a unique flow to be received at maturity of a zero bond. Such zero rates are usually derived from non-zero rates by replication of zero-coupon bonds from other bonds that are liquid. Floating rates are indexed on short-term rates from overnight to one year. The reference for short-term rates is the Libor (London Interbank Offering Rate) interest rate. The Libor is an average of the rates at which main banks lend and borrow from each other, and is fixed daily.

Interest rates also vary with the credit standing of the borrowers, and there are as many curves for risky rates as there are credit standings, measured by credit ratings. The rates for sovereign borrowers with the best credit standing are considered as risk-free rates. The rates paid by a risky borrower include a spread above risk-free rate. The spread compensates lenders for the default risk of the risky debt, and is, accordingly, called a credit spread. It is related to credit rating, among other factors, which measures the credit standing of a particular debt issue.

The shapes of term structures of interest rates are usually upward sloping, downward sloping, with a bump or are simply approximately flat (Figure 5.1).

With upward sloping curves, the mismatched position of banks earns the positive spread between the interest rate of assets and the interest rate of liabilities. The spread is market driven and does not depend on the commercial margins of the bank, which are the spread between the clients' rates and the market rates. Earning such market spread is attractive for financial firms but not without risks. The risk is that short-term rates rise, or that the term structure becomes flat, erasing any spread, or even becoming inverted with short-term rates higher than long-term rates. Figure 5.2 shows the changes of shares of the US term structure at the inception of the financial crisis. Shortly after, both short-term rates and long-term rates declined to historical lows.

The interest rates along the curve tend to be correlated, but they have no reason to be perfectly correlated. The curves evolve though time and their shapes can change. The figure shows that the US government curve moved from upward sloping to a near flat curve between

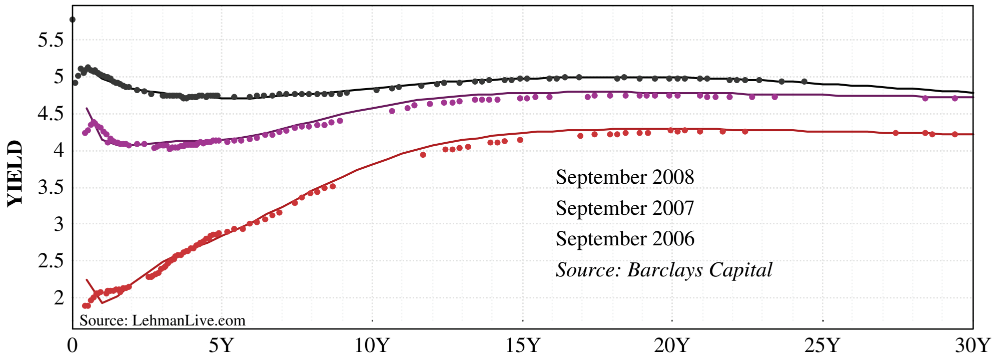


FIGURE 5.2 Bond yields US, 2006–2008 (*Source:* Barclays Capital.)

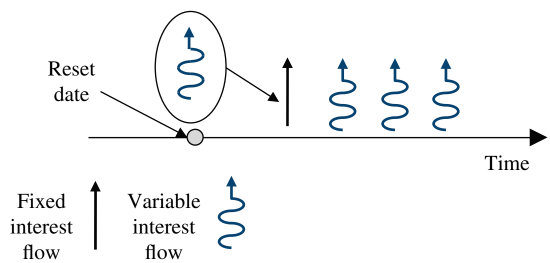


FIGURE 5.3 Fixed and variable interest rates

2006 and 2008. Shocks on interest rates can result in alterations of the shape of the curve. A parallel shock, often used as a first approximation, is a unique shock applied to all interest rates simultaneously.

Interest rates can be fixed or variable. Floating rates are periodically reset based on a market index, usually a Libor. Libor rates apply to various maturities up to one year. For example, the Libor 1 month or the Libor 1 year are reset, respectively, each month or each year from initial date.

Between two consecutive reset dates, a floating rate is "fixed" until the next reset date. In Figure 5.3, straight lines stand for fixed rates and curved lines represent variable, unknown, interest rates. Reset dates can occur anywhere within a period according to the reset period. The 1-month Libor changes only at reset dates. Over the period elapsing from the beginning of month and up to the next reset date, the rate is fixed.

Some rates, used as references by banks, do not refer to Libor rates, but are variable. The prime rate, or a rate based on the average of daily rates over a month, is an example. The prime rate is the interest rate used by banks for their best customers, and changes by steps rather than continuously.

5.2 INTEREST RATE GAPS

Interest rate gaps are algebraic differences between assets and liabilities of which interest rate shares a common reference. Static interest rate gaps are calculated from the current and projected balances of existing assets and liabilities, and form a time profile over the horizon of projections. Interest rate gaps are calculated for fixed-rate or for variable-rate assets and liabilities.

The fixed interest rate gap at a given date is the difference between fixed-rate assets and fixed-rate liabilities. The variable interest rate gap at a given date is the difference between variable-rate, or interest rate sensitive, assets and liabilities. Since there are many floating-rate references, there are potentially as many floating-rate gaps.

Like liquidity gaps, static gaps are derived from the projected runoffs of existing assets and liabilities. Static gaps are used for interest rate risk management because there is usually no need to hedge today transactions not yet in the balance sheet.

Floating-rate interest gaps can be defined for all floating-rate references (1-month Libor, 1-year Libor, etc.). These floating-rate gaps are not fungible: they cannot be aggregated unless assuming a parallel shift of all rates. The calculation of variable-rate gaps requires a proper mapping of the interest rates of the items in the balance sheet to selected reference rates. On the other hand, there is only one fixed-rate gap for a given horizon.

Gaps are calculated at various time points until some long-term horizon. The time bands between these points in time should not be long because the calculation of gaps does not show at which exact date, within each time band, a variable rate is reset. For example, periodical gaps can be calculated on a monthly basis. The gap profile can extend up to the longest maturity of the portfolio, but gaps are more useful up to shorter horizons, such as one, two or three years. By convention, gaps are calculated as differences between assets and liabilities, either fixed-rate or interest-rate sensitive (IRS) over a future period defined by its begin and end dates (t and $t+1$).

$$\text{Fixed interest rate gap}\,(t, t+1) = \text{Fixed rate assets}\,(t) - \text{Fixed rate liabilities}\,(t)$$

$$\text{Variable interest rate gap}\,(t, t+1) = \text{Interest rate sensitive assets}\,(t) - \text{Interest rate sensitive liabilities}\,(t)$$

5.3 CALCULATIONS OF INTEREST RATE GAP

The calculation of interest rate gap is straightforward once proper interest rate references have been defined and assigned to all items in the balance sheet. Fixed-rate gaps are simpler to report because there is no need to define all interest rate references. Non-interest bearing assets and liabilities should be excluded from the calculation of interest rate gaps, but they influence these gaps through their liquidity gap, as the calculation in paragraph 3.1 illustrates.

5.3.1 Calculation of Interest Rate Gaps

Figure 5.4 shows a point in time balance sheet, as of t, at a future date. All rates that are known as of this date are "fixed" and all other items are considered as interest rate sensitive and aggregated. The interest rate gaps (fixed or variable) depend on the gap between fixed assets and equity.

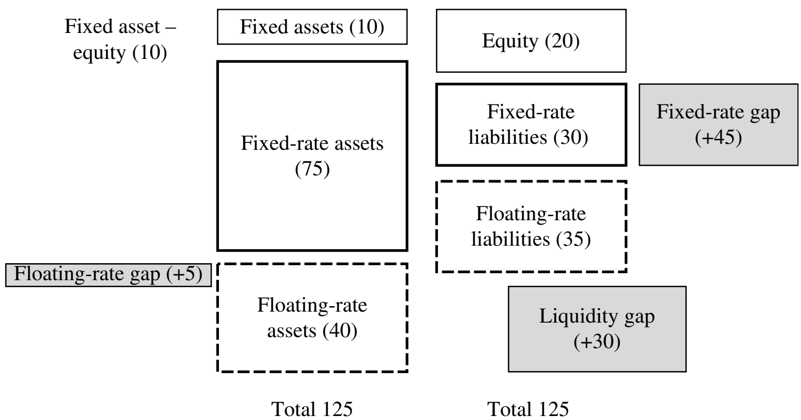


FIGURE 5.4 Interest rate gaps

If there was no liquidity gap, the fixed-rate gap and the floating-rate gap[2] would be identical in absolute values. But, in general, there is a liquidity gap for future dates. A projected liquidity gap generates an interest rate gap because excess funds will be invested, or deficits will be funded, at future dates at unknown interest rates. A projected deficit of funds is equivalent to an interest-sensitive liability. A projected excess of funds is equivalent to an interest-sensitive asset.

In the example of Figure 5.4, the floating-rate gaps differ, before and after the deficit, by the liquidity gap, equivalent here to a variable-rate liability:

$$\text{Floating-rate gap before liquidity gap} = 40 - 35 = +5$$

$$\text{Floating-rate gap after liquidity gap} = +5 - 40 = -35$$

The variable-rate gap can be calculated without considering the liquidity gaps or inclusive of the liquidity gap. Since the liquidity gap is a floating-rate item, it makes sense to include it in the calculation. In this case, the floating-rate gap is the floating-rate gap before the liquidity gap minus the liquidity gap:

$$\text{Floating-rate gap post funding} = \text{Floating-rate gap before funding} - \text{Liquidity gap} = -35 = +5 - 40$$

The calculation makes sense when all interest rate-sensitive items are aggregated. However, if interest rate gaps should be broken down by type of floating rates (Libor 1 month, Libor 1 quarter, etc.), the calculation cannot be done unless the interest rate references of the open liquidity gap are known.

5.3.2 Mapping Interest Rates to Selected Risk Factors

There are many floating-rate references. In addition, any future funding can be indexed to floating interest rates or to interest rates for longer maturities. The interest rates of the balance

2 This is the gap combining all interest rate-sensitive assets and liabilities at a point in time.

sheet items, and those of projected financing or investments, have to be mapped to selected market rates.

The mapping exercise is simple with existing assets and liabilities when the contractual rates are market rates, such as Libor. If only a subset of market rates is selected, some interest rates will not coincide with the market rates used as reference. The process creates basis risk, the residual risk due to differences between selected and actual rates.

The same happens with customers' rates, as these might not fluctuate mechanically with market rates. One solution is to relate actual rates, by product segment, to the selected reference rates and to use sensitivities for calculating "standardized gaps". Statistical techniques provide the sensitivities. The average rate of return of a subportfolio, for a product family for instance, is the ratio of interest revenues (or costs) to the total outstanding balance. It is feasible to construct times series of such average rates over many periods. A statistical fit to observed data provides the relationship between the average rate of the portfolio and the selected rates. A linear relation is such as:

$$\text{Rate} = \beta_0 + \beta_1 \text{ Index}_1 + \varepsilon$$

The coefficient β_1 is the sensitivity of the loan portfolio rate with respect to the interest rate index_1. The standardized gap weights the assets and liabilities by their sensitivities to the selected reference rates. For example, if the return of a segment of loans has a sensitivity of 0.8 with respect to the short-term market rate, it will be weighted with 0.8 in the gap calculation.

5.3.3 Sample Gap Reports

The simplified report (Table 5.1) is representative of what banks disclose in their annual reports.

Internal information systems provide far more details as illustrated by Figure 5.5. The graph uses the average balances of existing assets and liabilities over each time band. The horizontal axis shows dates, and the calculation is done on a monthly basis. The assets and liabilities shown are fixed rate. The top section shows assets and the bottom section shows the liabilities, the difference being the fixed-rate gap. The graph also shows the projected average rates of fixed-rate assets and liabilities.

TABLE 5.1 Sample interest rate gap report

In millions

31/12/2010	*<=1 month*	*1–3 months*	*3–12 months*	*1–5 years*	*5–10 years*	*>10 years*	*Not defined*	*Total*
Assets	110,942	55,572	120,418	117,677	48,374	16,969	26,557	496,509
Liabilities	120,989	63,858	117,070	110,542	41,619	13,495	28,936	496,509
Interest-sensitive gap	−10,047	−8,286	3,348	7,135	6,755	3,474	−2,379	-

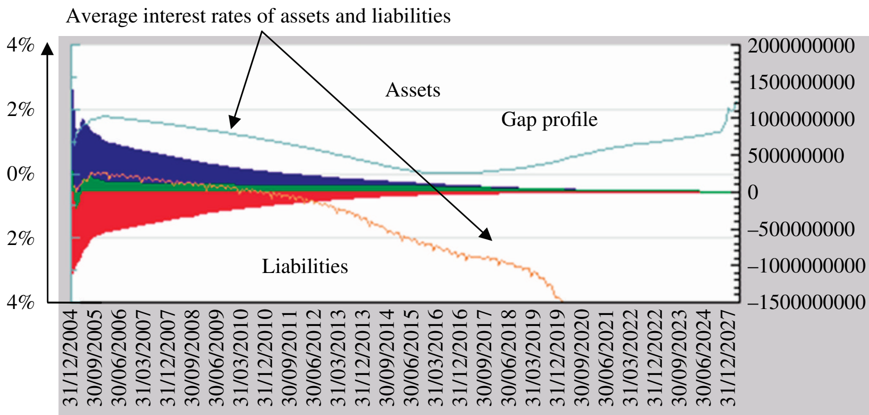


FIGURE 5.5 Assets, liabilities and interest rate gap time profiles

Internal models provide capabilities for drilling down into the gaps and finding out which subsets of transactions contribute to gaps. Subsets of transactions can be defined by product type, such as consumer loans, mortgage loans, or based on any other dimension.

5.4 THE GAP MODEL

Interest rate gaps measure the sensitivity of the net interest income (NII) to a shift of interest rates. When the variable-rate gap is positive, the volume of assets that are interest rate sensitive is larger than the volume of interest rate-sensitive liabilities. With a positive and parallel shock of interest rates, the interest revenue increases more than the interest costs and the NII increases.

If IRSA and IRSL represent the interest rate-sensitive assets and liabilities and Δi is the variation of the floating rates, the change of NII due to the change in interest rate Δi is:

$$\Delta\text{NII} = (\text{IRSA} - \text{IRSL})\,\Delta i$$

If, for example, the gap is +200, the variation of the NII is 2 when the rate changes by 1%. The basic formula relating the variable-rate gap to NII is:

$$\Delta\text{NII} = (\text{IRSA} - \text{IRSL})\,\Delta i = (\text{Interest rate gap})\,\Delta i$$

When the interest rate-sensitive gap is positive, the NII increases with interest rates, and conversely when the variable-rate gap is negative. When the variable-rate gap is zero, the NII is insensitive to changes of interest rates. It is said to be “immune” to variations of rates. The sensitivity of NII to a given parallel shock on interest rates is its variation, and is identical to the variable-rate gap.

Using a single variable-rate gap assumes parallel shifts of all interest rates. Regulations impose testing the effect of parallel shifts on NII, using shocks of 1% or 2%. When referring to other than parallel shocks on rates, such as shocks on the slope of the curve, several variable-rate gaps should be used. The variable-rate gaps by interest rate reference provide the sensitivities of the NII to each selected reference.

A gap nets the sizes of assets and liabilities. A gap derived from sizes as of the beginning of the month would assume that the sizes of assets and liabilities do not change within a month. An alternate common practice uses the average of daily sizes of assets and liabilities over a month for smoothing out the variations of sizes occurring within periods. The practice makes sense since interest revenues and costs are based on daily sizes.

A gap does not measure accurately the sensitivity of earnings because reset dates can be anywhere within the period over which the gap is calculated. If a change of interest rates occurs at the beginning of a period, the interest accrued over the entire period changes. If the change occurs at the end of a period, there is almost no effect on net interest revenue. If the reset dates are close to the end of the month, the change of interest income is overestimated, and it is underestimated if the reset dates are closer to the beginning of the month. The Appendix in section 5.8 shows an exact calculation of the NII when the gap is zero but reset dates differ. The example shows how the NII remains sensitive to a variation of interest rates even though the gap is zero.

The gap model is popular because it is a very simple way to measure approximately the sensitivity of the NII to interest rate movements. Implementing gap management for controlling the sensitivity of the NII is simply achieved by controlling the size of gaps using derivatives such as interest rate swaps.[3]

5.5 NET INTEREST INCOME AND INTEREST RATE GAPS

Additional assumptions on commercial margins are required for modeling the behavior of net interest income using gaps. Table 5.2 provides a simplified balance sheet, projected at a future date 1, one year from now, with balances being averages over a future period starting in one year. The purpose of the example is to calculate the NII and to measure directly how sensitive it is to a parallel shock on rates.

TABLE 5.2 Balance sheet projections and gap calculations

Dates	*1*
Banking portfolio	
Fixed-rate assets	19
Interest-sensitive assets	17
Total assets	36
Fixed-rate resources	15
Interest-sensitive resources	9
Total liabilities	24
Liquidity gap	+12
Variable interest rate gap	+8
Variable-rate gap after funding	−4

3 Interest rate swaps are introduced in Chapter 6.

The liquidity gap shows a deficit of 12 (Table 5.2). The variable-rate gap before closing the financing gap is +8. But the deficit of 12 is similar to a variable-rate debt if its rate is not locked in advance. Therefore, the post-funding variable interest rate gap is −4.

In this example, the NII, when the liquidity gap is ignored, increases with a positive shock on interest rates because the gap calculated for loans and deposits is positive (+8). However, after including the liquidity gap, the gap becomes negative (−4) and the reverse occurs if liquidity is raised at the future date. Let us assume that all variable-rate items, including the financing gap, are dependent on the same market rate. There is a shock on this single interest rate, and it is assumed that this reference rate moves from 8% to 11%. Gaps are used to model the behavior of the NII in the event of such a shock.

This is feasible only if the behavior of customers' rates as a function of the selected reference rate is modeled. The customers' rates depend on the bank's commercial margins. The commercial margins are the spreads between customers' rates and some internal references indexed to market rates.[4] The margins are defined here as differences between the customers' rates and market rates expressed in percentages of the balances of assets and liabilities. For lending, the margins are positive because the bank lends above market rates. Margins are negative for deposits because banks collect money at a cost below market rates.

For example, if commercial margins are 3% for assets and –3% for liabilities, the average customer rate for loans is 3% above the market rate, and the customer rate for deposits is, on average, 3% lower than the market rate. If the current market rate is 8%, those rates are 11% and 5%. Using customers' rates, the initial NII from commercial activities only is:

$$36 \times 11\% - 24 \times 5\% = 3.96 - 1.20 = 2.76$$

The bank's earnings differ from this commercial margin because of the cost of funding. For financing the liquidity gap, the same interest rate is used. The NII is the commercial earnings minus the cost of funding a deficit of 12, which is, initially: $8\% \times 12 = 0.96$. The bank's NII is therefore: $2.76 - 0.96 = 1.80$.

The gap models correctly the sensitivity of the NII only if commercial margins are constant. For fixed-rate items, the customers' rates remain unchanged when the market rate moves. The direct calculations of the NII, when the interest rate shifts from 8% to 11%, are shown in Table 5.3.

The NII, after financing the deficit of 12, decreases by $1.80 - 1.68 = -0.12$. This is because the funding cost is indexed to the market rate and increases by $12 \times 3\% = 0.36$. The negative variation, −0.12, results from the variations of the commercial earnings minus the additional cost of funds when rates are up. The calculation is consistent with the negative interest rate gap after financing, which is −4. This gap, multiplied by 3%, results in the −0.12 change in NII.

The commercial earnings are calculated on the commercial portfolio only, loans and deposits, ignoring the financing. The commercial earrings increase from 2.76 to 3, a variation of +0.24. This is consistent with the interest rate gap of the commercial portfolio, which is +8, and a variation of interest rate of +3%, since: $8 \times 3\% = 0.24$.

4 Commercial margins are calculated relative to internal references indexed on market rates, as detailed in Chapter 10. Margins are differences between customers' rates and the internal rates used for charging loans or for compensating deposit collections. It is shown in this chapter that these internal rates are indexed to market rates.

TABLE 5.3 Net interest income and interest rates

	Volume	*Initial rate*	*Revenues/costs*	*Final rate*	*Revenues/costs*
Fixed-rate assets	19	11%	2.09	11%	2.09
Interest-sensitive assets	17	11%	1.87	14%	2.38
Revenues			3.96		4.47
Fixed-rate resources	15	5%	0.75	5%	0.75
Interest-sensitive resources	9	5%	0.45	8%	0.72
Costs			1.20		1.47
Commercial NII			2.76		3.00
Liquidity gap	12	8%	0.96	11%	1.32
NII			1.80		1.68

5.6 GAP MANAGEMENT AND HEDGING

When the gaps are open, the bank is sensitive to variations of interest rates. Depending on the interest rate policy, gaps can be kept open or closed. Keeping gaps open means that the bank is betting on expected variations of interest rates. When the variable-rate gap is positive, an increase of interest rates is favorable. If the gap is negative, a decline of interest rates has a positive impact on earnings.

Gap management consists of controlling the size and the sign of interest rate gaps. Adjusting the gaps can be done on balance sheet, at least for future loans and debts. It is more convenient to use interest rate derivatives, as discussed in Chapter 6. Static gaps are relevant for risk management purposes if there is no need to hedge in advance future gaps. If so, managing the risk of new transactions can be differed until they enter the balance sheet. When new transactions enter the balance sheet, the static gaps are updated and hedges are adjusted. In other cases, it might be desirable to control interest rate risk for existing and new transactions, if the projections for the new portfolio are reliable. In this case, dynamic gaps should be used.

Often, balance sheet projections are used for budgeting purposes primarily because the budget depends on the new business that accrues over the next years. Dynamic gaps factor in the new transactions and model the behavior of the balance sheet according to commercial projections. Note, however, that future transactions have an unknown rate as of today, whether they have a fixed or a variable rate. Future fixed-rate loans will earn an interest that depends on future market rates, and should be considered as variable-rate assets as of today. New liabilities also carry an unknown rate as of today. Therefore, new transactions affect the magnitude of the variable-rate gaps, not that of the fixed-rate gap.

5.7 LIMITATIONS OF INTEREST RATE GAPS

Gap reports are appealing since they provide a measure of bank exposure, on how it behaves through time, and have an immediate application for hedging, or managing the gap with derivatives. But they embed a number of assumptions and conventions.

5.7.1 Customers' Rates

The gap model is linear, meaning that the sensitivity to a shock on interest rates is constant. One underlying assumption is that the spread between customers' rates and market rates is constant. To some extent, using standardized gaps based on sensitivities of customers' rates to market rates takes care of the issue.

But banks have the option to adjust customers' rates for smoothing out the variations of market rates. For example, lending rates smooth out upwards movements and deposit rates tend to lag when market rates decline. In addition, the sensitivity of clients' rates to market rates can be asymmetric for ups and downs (downside movements are slower than upside movements).[5] This implies that the spreads between customers' rates and market are not constant. The customers' rates become a function of current and lagged market rates, with different sensitivities for up and down movements. The relationship between earnings and market rates is not symmetric as basic gaps suggest, and it involves lagged market rates.

5.7.2 Regulated Rates

The gap methodology can be adapted to rates that differ from market rates. An example is the rate of regulated saving deposits in European countries. This rate is a function of the inflation rate and a short-term rate. This means that the sensitivity to each item is less than one. For example, assume that an amount of 100 is indexed to the average of the short-term rate and the inflation rate. This implies that each component is weighted 50% in the calculation of accrued interest. The variation of interest becomes: Δ interest accrued $= 100 \times (\frac{1}{2}\Delta$ Short-term rate $+ \frac{1}{2}\Delta$ Inflation). This is the same as having 50 fully indexed to the short-term rate and 50 fully indexed to inflation and is equivalent to using weighted gaps with respect to each index.

Note that regulations might create exposure to inflation rate. Inflation indexed hedges (such as inflation swaps) might be used, which are usually derived from those government bonds which rate is indexed to inflation. But the technique does not fully resolve the issue, because regulated rates might also move by steps, for example, only when the composite index moves by at least 0.5%. Stepwise functions are not consistent with the gap model.

5.7.3 Embedded Options

Banking products have embedded options. Fixed-rate mortgages can be renegotiated when interest rates decline. Variable-rate mortgages have caps on the interest rate paid by the borrower. Deposits can be withdrawn at any time, and be transferred into interest earning accounts when interest rates rise.

Embedded options are a challenge for liquidity gaps because they result in stochastic cash flows, eventually dependent on interest rates. For interest rate risk, they are a major issue as well. Embedded options are always adverse to the bank: fixed-rate loans are renegotiated when rates are down, resulting in lower revenues for the bank; deposits contract when interest rates are up, which is also when the bank will have to face higher interest costs. Embedded options

5 The models of customers' rates are briefly discussed in Chapter 8.

therefore create dependencies between the balances of mortgages or of deposits and the level of interest rates.

The effect of embedded options, whether explicit or implicit, is that the nature of interest rate assigned to a loan – fixed or variable – can change with the level of market rates. For a variable-rate loan with a cap, the rate becomes fixed when hitting the cap. When renegotiating the rate of a fixed-rate loan, the interest rate is initially fixed, but it becomes variable at the time of the renegotiation.

Such dependencies make volumes dependent on the level of interest rates and gaps cannot be considered constant anymore when rates vary: the basic gap model collapses. This does not mean that gaps cannot be used. It implies that gaps be adjusted as options are being exercised, using some models of how balances shift with interest rates. It also implies that the adverse effect of options should be modeled separately since it is ignored in the standard gap models. Extensions on modeling the prepayment risk of mortgages, on modeling volumes and on the optional interest rate risk of banks are provided in Chapters 8 and 9.

5.8 APPENDIX: GAPS AND INTEREST RATE SENSITIVITY

The example of Figure 5.6 considers the negative gap generated by a positive and a negative flow occurring at different dates, within a same period. The negative gap suggests that the NII should increase when the interest rate decreases, for example from 10% to 8%. However, this ignores the reinvestment of the positive intermediate flow of date 90, at the lower rate and for 270 days. On the other hand, the negative flow is similar to a debt that costs less over the remaining 180 days. The NII is interest rate sensitive if interest revenues and costs are calculated using the accurate dates of flows.

The interest revenues and costs are:

$$\text{Inflow at day 90: } 1,000 \times (1.10^{270/360} - 1.08^{270/360}) = -14.70$$

$$\text{Outflow at day 180: } -1,536 \times (1.10^{180/360} - 1.08^{180/360}) = +14.70$$

They match exactly because interest revenues and costs are proportional to the size of flows and to the residual period of reinvestment or funding. The first flow is smaller than the second

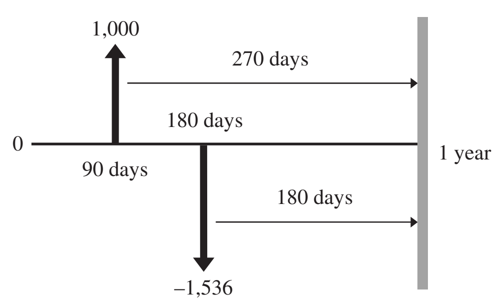


FIGURE 5.6 Negative gap and fixed NII

flow but generates interest revenues over a longer period. The values are such that the maturity differential compensates exactly the size differential.

The example shows how gaps can lead to errors. In the example, the gap model suggests plugging another flow of 536 to hedge the NII. Actually, doing so would put the margin at risk. The exact condition under which the margin is insensitive to interest rates is relatively easy to derive for any set of flows.

6

Hedging and Gap Management

Under conventional gap management, interest rate gaps measure the sensitivity of the net interest income (NII). Gap management consists of controlling the size and the sign of interest rate gaps over the management horizon, depending on the views of interest rates of the asset-liability management (ALM). Banks can decide to keep open positions depending on their views on interest rates. For managing gaps, banks rely on standard interest rate derivatives, such as interest rate swaps.

This chapter introduces the standard interest rate derivatives, explains how they are used for managing gaps and discusses the hedging policies of banks.

Contents

6.1 THE TRADE-OFFS OF GAP MANAGEMENT

There are various ways to finance a loan depending on which trade-offs between risk and return the bank is willing to achieve. Suppose that a client from a bank borrows for 10 years at a fixed rate of 5%. The bank has to raise the funds for the amount of the loan. The Treasury can operate in various ways, which differ in terms of the spread between the customer's rate and the cost of financing, and in terms of liquidity and interest rate risk. In general, maximizing the expected profit implies higher risk.

For maximizing the spread between the loan rate and the cost of debt, a short-term financing can be contracted at a floating rate, for example the Euribor 3 month, assumed to be 1%. The spread is 5% − 1% = 4%. The Treasury is now facing liquidity risk, because the financing has to be rolled over the duration of the loan, and interest rate risk, because the short-term rate can fluctuate. Presumably such financing would not be allowed by the guidelines of the finance department due to the high liquidity risk.

The Treasury can eliminate the liquidity risk by borrowing for 10 years at a floating rate. With this financing, there is no liquidity risk, but there is a mismatch between the fixed interest rate of the loan and the floating rate of the debt. The cost of the debt will be higher than for a short-term debt by the amount of a liquidity premium, for example 40 basis points, and the spread would narrow to 3.6%. Many financial firms can do so within their guidelines and limits. The NII remains fully exposed to the variations of the short-term rate but the Treasury still benefits from a large spread between the long-term rate of the loan and the floating rate.

However, the Treasury has now an additional risk due to the fluctuation of the liquidity premium. The liquidity premium is not correlated with the level of interest rates, and rather depends on the preferences for short-term commitments of financial players. Liquidity premiums are higher in period of stress, when they can become substantial.

For eliminating liquidity risk and interest rate risk, the Treasury could fund the loan with a debt fixed rate with same maturity. The spread would be much lower when long-term rates are above short-term rates. If the fixed-rate debt costs 4%, the spread drops down to 1%. This situation is the reference because it eliminates both liquidity and interest rate risk from the transaction. On the other hand, the Treasury sacrifices a major fraction of the spread.

Similar options exist at the level of the balance sheet, except that the mismatches of maturities and interest rates are managed at an aggregated level. This example shows that banks can decide to keep open positions to maximize their returns, within the limits imposed on liquidity and interest rate risks set by the ALM committee. The gaps are managed on-balance sheet when raising funds (in the case of deficits) and off-balance sheet using interest rate derivatives.

6.2 MANAGING INTEREST RATE GAPS WITH INTEREST RATE DERIVATIVES

The generic instruments for interest rate risk management include:

- Interest rate swaps (IRS), contracted immediately or effective later (forward swap);
- Caps or floors, which provide the buyer with a guaranteed maximum interest rate, or a guaranteed minimum interest rate;
- Option on interest rate swaps (known as a swap option or swaption), which is a right to enter into a swap at a later date, but not an obligation;
- Or do nothing.

A short presentation of these standard instruments follows.

IRS exchange a floating rate against a fixed rate, and vice versa, or a specific floating rate (1-month Libor, for example) against another one (Libor 1 year, for example). A swap has a

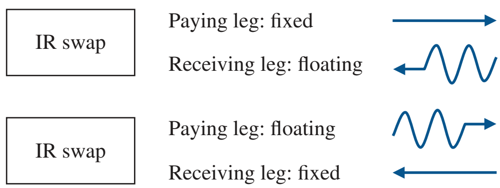


FIGURE 6.1 Interest rate swaps

"receiving" leg and a "paying" leg (Figure 6.1). Receiver swaps designate swaps that receive the fixed rate and pay floating. The "notional" of the swap is the reference amount serving for calculating the amount of interest payments received and paid. The swap fixed rate is set such that the value at inception is zero. The swap fixed rates for various maturities form the "swap curve".

Interest rate swaps are extremely convenient for all financial players: borrowers, lenders and banks. Corporations use IRSs for managing the interest rate risk of their debts. For example, a fixed-rate borrower expecting that short-term rates will decline would prefer to pay a floating rate rather than the fixed rate. By entering into a swap, the borrower receives the fixed swap rate and pays the floating rate. The fixed-rate payments from the loan are partially offset by the fixed rate received from the swap. The offset is not complete because the original fixed rate of the loan is generally different from the fixed swap rate at the time when the swap contract is effective. The borrower using a swap pays the algebraic difference between the fixed rate of the loan and the fixed rate received from the swap plus the floating rate. Similarly, a swap paying the fixed rate and receiving the floating would be appealing for a floating-rate borrower who fears that its interest cost rises.

The cost of entering into a swap is minimal. Instead of receiving the fixed rate, a receiver swap would receive a rate slightly lower than the swap rate. The difference depends on who contracts the swaps and compensates the seller. Interest rate swaps are extremely convenient because of their low cost and because they can be reversed. Suppose that the borrower made a mistake when entering into a receiver swap, if the floating rate goes up instead of down as expected. It is always possible to reverse the swap and end up paying a fixed rate with an acceptable cost.

Caps are optional instruments, setting an upper bound on the interest rate of a floating-rate borrower. The guaranteed rate of the cap is the strike. The cost of purchasing the option is the premium. When the interest rate is below the guaranteed rate, the borrower pays the floating rate and benefits from any decline. If the floating rate exceeds the guaranteed rate, the seller of the cap compensates the difference. The cap provides the best of all worlds, low rates and high rates. Because of this benefit, the cap is usually expensive. Before entering into a cap, the borrower should check that the expected payoff of the cap will exceed its price, the premium paid. A floor guarantees a minimum rate, and is adapted to lenders fearing lower rates, rather than borrowers.

Such options are expensive. Because of their high premium, players can rely on collars, which combine a purchased cap and a sold floor at a guaranteed rate lower than the maximum rate of the cap (Figures 6.2 and 6.3). The buyer of the collar pays no less than the minimum guaranteed rate and no more than the maximum guaranteed rate, and pays the prevailing rate when it is within the guaranteed range. The benefit of the collar is that it has a lower premium than a straight cap because the premium received from selling the floor offsets the premium

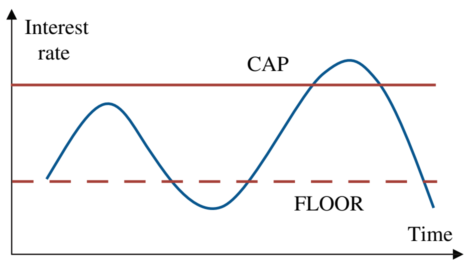


FIGURE 6.2 Cap, floor and collar

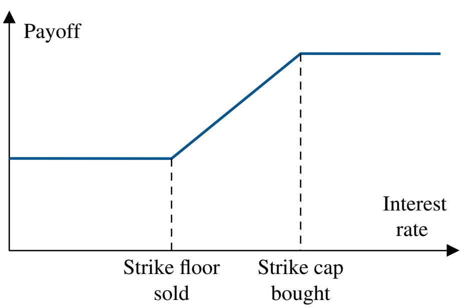


FIGURE 6.3 Payoff of a purchased cap and a sold floor

paid for the purchased cap. Banks quote zero-premium collars, obtained by adjusting the upper and lower rates so that both premiums offset exactly.

6.3 MANAGING INTEREST RATE GAPS

The usage of swap for gap management is straightforward. In the example of the balance sheet in Chapter 5 (Table 6.1), the liquidity gap is +12 and the variable-rate gap of the commercial balance sheet is +8. It is assumed that the gaps are calculated for a future date and a future period, for example in one year from now and the gap is an average over the year following this date. The financing of the liquidity gap is a variable-rate item, as it will be contracted in the future. If the liquidity gap is indexed on the same market rates as other variable-rate items, the floating-rate gap, including the liquidity gap, is −4.

If the bank keeps the gap open, it would benefit from a decline of interest rates. Suppose the bank does not wish to bet on such a decline and that it decides to close the gap, making its NII immune to interest rate changes.

The liquidity gap will be closed when the funds are raised in one year, but no cash is needed today. For managing the interest rate gap, the appropriate instrument is a swap. The bank has too many variable-rate liabilities relative to variable-rate assets, as it pays more variable-rate interest flows than it receives from its assets. For reducing the outflows from variable-rate interest payments, it should contract a swap receiving the floating rate and paying the fixed. The notional of the swap should be equal to the size of the variable-rate gap, or 4. Combined with the swaps, the balance sheet has now a zero interest rate gap, the desired result.

TABLE 6.1 Liquidity and interest rate gaps

Banking portfolio	
Fixed-rate assets	19
Interest-sensitive assets	17
Total assets	36
Fixed-rate resources	15
Interest-sensitive resources	9
Total liabilities	24
Gaps	
Liquidity gap	+12
Variable interest rate gap	+8
Variable interest rate gap after funding	−4

In this particular example, the open liquidity gap is +12. A zero interest rate gap occurs when a fraction, of size 4, of this future funding is fixed rate, with a rate set today. The rest of the funding can remain variable. The appropriate swap is a forward swap starting at the date of the projected balance sheet and with duration equal to the period over which the projected gaps are constant.

6.4 SETTING LIMITS TO GAPS

Limits on gaps can be defined in order to cap the variations of the NII. Under the gap model, the variations of earnings are proportional to the size of gaps and to the variations of interest rate. For setting a limit on adverse deviations, it is sufficient to set a limit on the size of the gap. The variation of NII is a linear function of the interest rate variation, the coefficient being the gap:

$$\Delta\text{NII} = \text{Variable-rate gap} \times \Delta\text{Interest rate}$$

With a given shock of 100 bps on interest rates, the variation of NII is 1% times the gap. Limits cap such variation to some maximum percentage of the earnings of the current and next years, for example. This result can be achieved by capping the size of the open interest rate gap with derivatives.

An alternate way of setting limits is to set a value of earnings-at-risk for the period. Standard shocks do not capture the volatility of interest rates. Earnings-at-risk is a measure of the potential adverse deviations of earnings that can be exceeded only in a small fraction of possible outcome, the small fraction being the confidence level.[1] The variation of NII due to an interest rate shock Δi is proportional to the gap: $\Delta\text{NII} = \text{Gap}\,\Delta i$.

With a fixed gap, the distribution of the variations of the NII results from the distribution of the interest rate shocks. The maximum deviation of the earnings at a preset confidence level

1 Earnings-at-risk are also discussed in Chapter 16 on interest rate simulations.

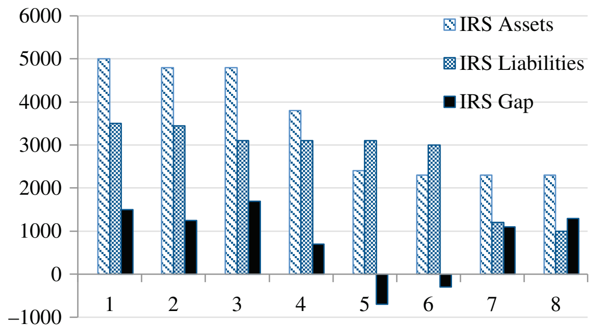


FIGURE 6.4 Time profile of variable rate gaps

results directly from the maximum deviation of the interest rate at the same confidence level. If the gap is positive, adverse deviations occur when interest rates are down, and the opposite occurs with negative gaps.

The exact derivation should use a sensible distribution of interest rate shocks.[2] Suppose that the maximum downside variation of interest rates at the confidence level α is $\Delta i(\alpha)$. This lower bound corresponds to an adverse variation of the NII equal to Gap $\times$ $\Delta i(\alpha)$ if the gap is positive. A limit to this deviation implies a maximum size for the gap.

The usage of limits implies that hedging programs be defined. A hedging program caps the sizes of gaps over several periods. The size of gaps varies from one period to the next, and the hedges have varying amounts over the management horizon. Figure 6.4 provides an example of a series of interest rate-sensitive (IRS) gaps over several periods. There is a limit on interest rate gaps applying across all periods up to horizon, assumed constant and equal to 1,000. Figure 6.4 shows the time profiles of assets and liabilities and Figure 6.5 shows the limit and the series of periodical variable-rate gaps. Any floating-rate gap above the upper limit is an excess gap. A lower negative limit is ignored here since all negative floating-rate gaps are lower, in absolute value, than 1000.

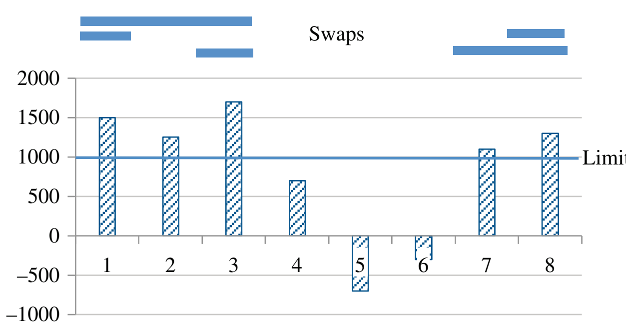


FIGURE 6.5 Hedging program complying with limits

2 With a normal distribution, used as a proxy when interest rates are not close to zero, the one-sided deviation not exceeded in more than 1% of all outcomes is 2.33 times volatility.

The excess gaps are the differences between floating-rate assets and floating-rate liabilities. Each excess gap is hedged with a swap receiving the fixed rate and paying the variable. The hedges are defined starting from the end of a planning horizon and moving backward for defining the size and the dates of the hedging swaps. These hedges offset all excess gaps over the horizon.

There are three excess gaps over the first periods, 500, 250 and 700. A first swap hedges the excess gap of 250, which lasts over all three first periods. A second swap offsets the remaining excess gap of 250 (or $500 - 250$) for the first period only. A third one offsets the gap of period 3 in excess of 250, the notional of the first swap, or $700 - 250 = 450$. Other swaps are required for hedging the excess gaps of later periods. The five swaps are represented by horizontal lines on top of the graph.

Limits are often tighter for the near future, and wider as the horizon extends. As views on interest rates are less reliable for longer horizons, tight limits over longer horizons might result in overhedging.

6.5 HEDGING THE VARIATIONS OF THE TERM STRUCTURE OF INTEREST RATES (CASE STUDY)

In some particular instances it might be desirable to rely on forward contracts on interest rates. This would be relevant for an institution having recurring excesses of funds while the interest rates are declining. The issue for such a net lender is to decide what should be done for the projected excesses of funds. Doing nothing would imply that the investment of future excesses of funds will provide lower and lower returns.

This example is inspired by the behavior of the European term structures of interest rates in the years preceding the euro event. The national yield curves were upward sloping a few years before the euro. The slope of the yield curve was steep, and declined when approaching the euro, moving down until it became almost flat when reaching the date of the creation of the euro. Such movement was interpreted as a convergence of rates of various European countries towards the lowest rates prevailing, which were the German rates.

In this case, forward contracts on interest rates were the appropriate instrument. A forward contract sets the rate today for a future period. A forward rate is an interest rate viewed from today (t) applicable between two future dates t_1 and t_2. There are as many forward rates as there are couples of futures dates, seen from the current date (t) and forward rates are indexed by the three dates, as $F(t, t_1, t_2)$. Forward rates are used to lock in future lending rates for any future period, for lenders who fear that future rates fall below the forward rates.

All forward rates can be derived from the current term structure of interest rates. This can be shown by finding the today transactions that can replicate such contracts. Assume that a lender expects a future inflow $X = 100$ one year from now and that the spot rates for one and two years are, respectively, 4% and 5%. The lender wants to lock in a rate today for a forward loan starting at date 1 and for one year until date 2. A forward loan can be replicated by borrowing today for one year and lending two years the same amount, as shown in Figure 6.6.

Given the expected inflow of 100, the exact amount that can be borrowed today at 4% with a repayment of 100 in one year is: $100/(1 + 5\%) = 95.2381$. Lending this amount for two years at 5% yields: $96.2381 \times (1 + 6\%)^2 = 107.0095$. These are the proceeds after lending 100 in

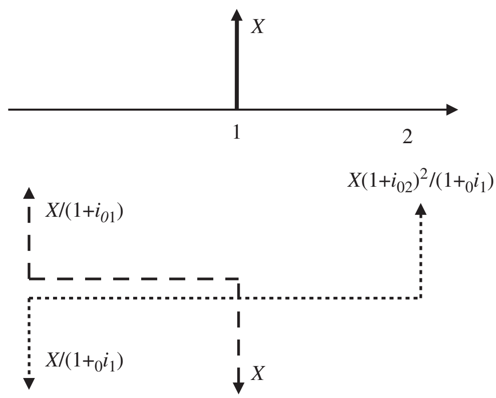


FIGURE 6.6 Calculation of the (12 × 24) forward rate

one year and for one year, and the corresponding rate is the forward rate, 7.0095%, for one year and starting in one year.

The figure shows the exact calculation of this forward rate. The two subscripts of interest rates are the initial (0) and the final dates (1 or 2) to which spot interest rates apply. The one-year forward rate applying one year from now is the 12 × 24 forward, 12 and 24 designating the number of months for start and end dates.

The forward contract does not imply any cash flow today, but allows lending the expected inflow of date 1 until date 2. Also, the forward rate is higher than the two spot rates. This is a general rule: forward rates are above spot rates when the term structure of rates is upward sloping and below if it is downward sloping.

In the context of the pre-euro event, all yield curves were upward sloping, and forward rates were above the spot rates, which kept declining as time passed. When the slopes were steep, there was an opportunity to benefit from forward rates way above the spot rates.[3] Since rates were declining, the chances of reversal of the trend seemed remote.

One option was to wait and do nothing. A possible rationale for waiting was that rates were close to low historical rates and that they should start rising again. Another option was to hedge gradually, progressively entering into swaps receiving the fixed rate, as the decline of interest rates continued. Early swaps would have captured the higher rates of the beginning of the period. But later swaps would have captured very low fixed rates, making them useless if rates were expected to move up again. In such a case, the value of waiting is negative and late hedges are generally useless. The forward rates provided an opportunity to earn higher returns than current and future rates, if contracted early enough.

When interest rates reached the bottom line, in late 1999 and early 2000, the yield curve became lower and flatter. The spread between forward and spot rates narrowed. The "window of opportunity" for investing in forward rates higher than subsequent rates closed. Taking such "directional" positions was possible because the trend of interest rates persisted for several years in a row.

3 In the above example, one could lend forward at a rate of 7%, well above the two spot rates for one and two years.

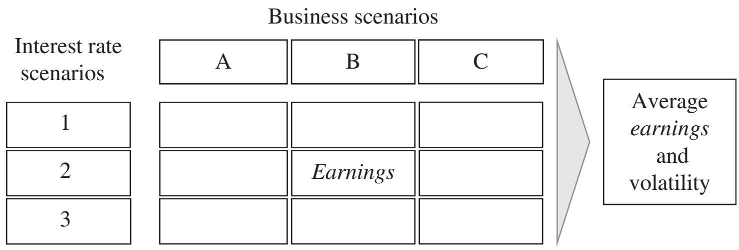


FIGURE 6.7 Cross-tabulation of business scenarios and interest rate scenarios

6.6 HEDGING BUSINESS RISK AND INTEREST RATE RISK

Controlling interest rate risk with the gap model is simple for deterministic scenarios. But projected gaps are not deterministic and there is more than a single scenario for which hedging solutions should be defined. Furthermore, the projected volumes might be partially dependent on interest rates because of embedded options in the balance sheet, or, for example, due to uncertainty in the future business volumes.

Several scenarios imply several gap profiles. With multiple scenarios, volume and interest rate uncertainties can be addressed using combinations of interest rates and volume scenarios. One technique for handling the uncertainty of volumes and rates is to cross-tabulate these scenarios and form a matrix. In Figure 6.7, the lines represent interest rate scenarios and the columns represent volume scenarios, assuming a finite number of such scenarios. For each combination of volumes and interest rates, the earnings can be calculated. There is one value of earnings for each cell of the matrix. The matrix can be summarized by the average earnings and the dispersion of earnings across cells. This is a convenient way to summarize the current position of the bank in terms of expected earnings and the potential volatility of earnings across scenarios.

Since the interest income depends upon the financing and hedging solutions, there are as many matrixes as there are hedging solutions, as illustrated in Figure 6.8. Since the gaps vary across scenarios, there is no hedging solution that closes the gaps across scenarios. Instead, various hedging solutions can be tested, by recalculating all earnings for all cells of the matrix, once hedges are in place. If various hedges are tested, there are as many matrices as hedging

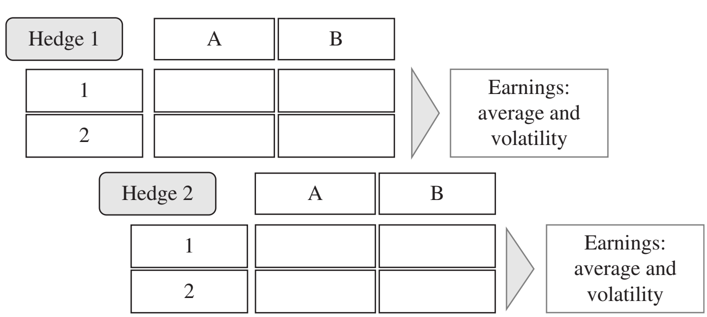


FIGURE 6.8 Cross-tabulation for various hedging scenarios

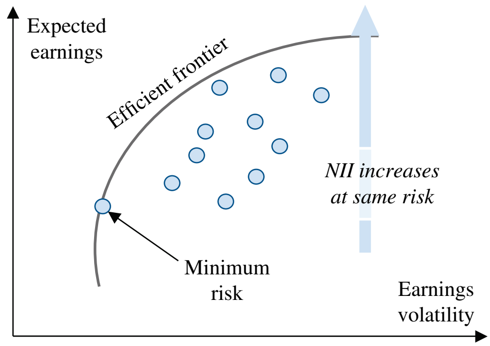


FIGURE 6.9 Risk–return combinations

solutions. For handling these tables, each matrix is summarized by the average earnings and their standard deviations across the cells of each matrix. By changing hedges, a series of expected earnings and earnings volatility is generated.

Expected earnings and earnings volatility can be plotted, with each dot summarizing the entire matrix of earnings for a given hedge (Figure 6.9).

The Sharpe ratio, the ratio of the expected earnings to the earnings volatility, can be used as a risk-adjusted measure of profitability.[4] The best hedges are those that maximize average earnings given the volatility, or those that minimize volatility given the level of earnings, and are "efficient". There is no hedge that makes earnings immune to interest rate changes, due to the uncertainty on the sizes of gaps. But there can be inefficient solutions. Efficient solutions for various levels of risk can then be isolated.

4 The Sharpe ratio is a measure of the risk-adjusted performance of a portfolio. It also serves when modeling the credit risk of the portfolio.

7

Economic Value of the Banking Book

Standard gap management has many drawbacks. Gaps are not adapted to long horizons and measure the risk over a limited period, whether or not banks have very long commitments. They are unable to account for the optional risk embedded in banking products. Also they are only approximations of the actual sensitivity of the net interest income (NII) to shocks on interest rates over a period.

An alternate approach is to assess the sensitivity of the value of the balance sheet of a bank to shocks of interest rates. If a bank lends above market rates and collects resources below market rates, its banking book has a positive value. Its risk can be measured by the sensitivity of this value to shocks on interest rates. One benefit of the approach is that it measures the risk of the banking book over the entire horizon of existing facilities.

This chapter explains how economic value (EV) relates to the net interest incomes of the bank and how its sensitivity can be addressed.

Contents

7.1 ECONOMIC VALUE AND ITS SENSITIVITY

The banking portfolio is long on loans and short on financing, and its value is the net value of assets and liabilities. The valuation of a banking portfolio faces many conceptual and practical challenges. Loans and deposits are not traded, there is no price for these assets, and valuation collapses to a mark-to-market exercise. Both banking loans and deposits embed options, which are given away to the customers, and valuation should be net of the value of these options. Loans are also exposed to the credit risk of borrowers, which has an impact on their values. Given the stochastic cash flows of banking assets and liabilities, the valuation exercise is not simple. An extended framework is proposed in Chapter 8, which can, conceptually, address these difficulties.

The focus of this chapter is not value but is rather the sensitivity of this value to shocks on interest rates. The credit risk adjustment for loans is of lesser interest in this context. The mark-to-market value of a credit risky loan can be obtained by discounting its contractual cash flows at a risky rate, adding to the risk-free rate the credit spread matching its credit quality.

In practice, for the sole purpose of risk measurement, the sensitivity of the value is of interest, more than the value itself. Under this view, there is no need to adjust the value of loans for credit risk, since credit risk is kept constant when assessing the effect of shocks on interest rates. Ignoring credit risk means that discounting is conducted with risk-free rates.

The underlying assumption is that the sensitivity of the value non-adjusted for credit risk is an acceptable proxy of the sensitivity of the fair values of credit risky assets. The assumption makes sense if the credit spreads are small and independent of shocks on interest rates. Using a risk-free curve also means that the entire spread between customers' rates and risk-free rate, inclusive of credit spreads, is earned by the bank, which is true for non-defaulted loans.

The approximation has huge benefits since a single curve of interest rates can be used for all assets and debts instead of differentiating the rates with the risk of individual items. In what follows, the EV designates the value of the balance sheet using a reference curve that is relevant for banks.

The reference interest rate curve for a bank combines the Libor rates, under one year, and the swap curve for longer maturities. The Libor rates are interest rates that apply to well-known banks. Beyond one year, the reference rates are those of swaps exchanging the Libor against a fixed rate. The swap curve is made of the fixed rates of swap for various maturities. An interest rate swap is a contract for exchanging interest payments at a fixed rate, the swap rate, with floating rate payments, at an interbank rate, the Libor. Swap fixed rates are used as reference rates, and considered as near risk-free because they do not include the credit spread of market debts, such as bonds, which compensate investors for the credit risk of debts. Instead, swap rates are unfunded to the extent that they do not require any exchange of principal between parties. However, they embed the small spread above Treasury rates of participating banks. In order to earn the swap rate, a bank would need to lend at the interbank Libor rate and enter into a swap. Such interbank lending and borrowing transactions carry the credit spreads of highly rated banks.

The present value (PV) of all cash flows discounts all cash flows from existing assets and liabilities of the banking book. They include capital flows, such as principal reimbursements from loans, and interest flows. All amortization cash flows are summarized in liquidity gaps. Therefore the PV of all flows is the PV for all liquidity gaps plus all interest flows. The PV of the stream of future flows F_t from is:

$$V = \frac{\sum_t F_t}{(1 + r_t)^t}$$

In the formula, the discount rate is the risk-free rate or the Libor and swap rates, following the rationale of the above conventions. The EV is the net value of a portfolio long in assets and short in liabilities. It can be positive or negative. The change in EV depends entirely upon the sensitivities of these two portfolios.

This approach ignores the optionality of banking products. Optionality reduces the value of the banking book as it is always adverse to the bank since mortgages are renegotiated when interest rates are down and deposits are withdrawn when interest rates are up, requiring costlier financing for lost deposits. The adjustment for optionality is addressed in dedicated chapters (Chapters 8 and 9). At this stage, the purpose is only to make explicit the interpretation of the sensitivity of EV.

7.2 ECONOMIC VALUE AND NET INTEREST INCOME

Intuitively, the EV should be a measure of long-term performance. If the value of assets is above face value, it means that their return is above the market rates, a normal situation if loans generate a spread above market rates. If the deposits pay lower than market rates, their value is below their face value. EV is positive since assets are valued above par and liabilities are below par. The value should increase with the magnitude of the spreads of loans and deposits relative to market rates. It can be seen in an example that the EV, calculated with above conventions, is directly related to the stream of periodic NIIs.

7.2.1 Sample Bank Balance Sheet

A sample balance sheet is described in Table 7.1. The current market rate is 4%. For simplicity, a flat interest rate curve is used, which is not restrictive. The bank has a one-year debt with a spread of 1% above the market rate, or 6%. The debt rate is fixed for one period and is reset after, when it is rolled over for matching the longer maturity of assets. The loans have a maturity of three years and a contractual fixed rate of 7%, or 3% above the market rate. The NII is immune to interest rate changes for the first year only, and it is reset at a new rate for the second and third years.

If there is a shock on interest rates after the first period, the interest cost of the debt is reset, but the fixed rate of loans does not change. The EV discounts all cash flows until the final horizon. At the end of the third period, both loans and debt are repaid. Table 7.2 shows the balances of assets and liabilities at end of period, and the corresponding cash flows.

TABLE 7.1 Example of a simplified balance sheet

	Assets	*Liabilities*
Amount	1,000	900
Fixed rate	7%	5%
Maturity	3 years	1 year

TABLE 7.2 Stream of principal flows generated by assets and liabilities

Outstanding balances	*0*	*1*	*2*	*3*
Assets	1,000	1,000	1,000	0
Liabilities	900	900	900	0
Cash flows				
Asset flows		0	0	1,000
Liability flows		0	0	−900
Net capital flows		0	0	100

TABLE 7.3 The interest flows

Interest revenues and costs at t *(using balances as of* t − 1*)*			
Revenues	70.0	70.0	70.0
Costs	45.0	45.0	45.0
NII	25.0	25.0	25.0

Table 7.3 shows the calculation of the net interest flows. If there is no change of interest rates, both interest income and interest costs are constant from one period to the next. They are calculated from the balances at the beginning of each period.

The calculation of EV adds up the present values of the principal flows and of the interest flows. The discount factors are calculated with the market rate, 4%. The present values of all-in flows, flows combining interest and capital flows, is the EV, 158.3 (Table 7.4). It sums up the present values of interest flows and of principal flows.

The value of net interest flows is calculated separately (Table 7.5).

It can be seen that the difference with the EV is the PV of the final equity flow positioned at the last date (Table 7.6).

These calculations show that the EV sums up the present values of interest flows and of the final capital flow:

$$\text{EV} = \text{Present value of NII} + \text{Present value (equity as of date 3)} = 69.4 + 88.9 = 158.3$$

TABLE 7.4 Calculation of EV

Discount factors – final rate		*0.952*	*0.907*	*0.864*
All-in flows		25.0	25.0	125.0
PV (all-in flows)		24.0	23.1	111.1
EV: ΣPV (all-in flows)	158.28			

TABLE 7.5 Value of interest flows

Interest revenues and costs at t (using balances of assets and liabilities as of t − 1)				
Revenues	–	70.0	70.0	70.0
Costs	–	45.0	45.0	45.0
NII	–	25.0	25.0	25.0
PV(NII)		24.0	23.1	22.2
ΣPV(NII)	69.38			

TABLE 7.6 Value of capital flows

Dates		*1*	*2*	*3*
Discount factors – final rate		*0.952*	*0.907*	*0.864*
PV (capital flow)		0.0	0.0	88.9
ΣPV (capital flow)	88.90			

The implication is that controlling the sensitivity of the EV is equivalent to controlling the sensitivity of the entire stream of earnings up to the final horizon, plus the value of equity positioned at the last date.

When there is a positive shock on interest rates, the EV declines because the NIIs are lower after the first period. This can be checked by applying a shock on interest rates that makes the NII negative. For example, when the market rate moves up from 4% to 8%, the net incomes of periods 2 and 3 are negative. The principal flows are unchanged, and the interest cost of the debt is updated when the debt is rolled over after the first period. The new net interest flows are shown in Table 7.7.

The EV is lower because it sums up the PV of lower NII up to horizon, plus the PV of a constant term (Table 7.8). The discount factors use the new rate, 8%.

The example shows how EV relates to the stream of NIIs in a simple situation. In a more general case, assets and liabilities amortize and the final horizon can be the longest maturity or shorter.

TABLE 7.7 Interest flows after a shock on interest rates

Interest revenues and costs at t (using balances of assets and liabilities as of t − 1)				
		1	*2*	*3*
Revenues	–	70.0	70.0	70.0
Costs	–	45.0	90.0	90.0
NII	–	25.0	−20.0	−20.0

TABLE 7.8 The new EV after a shock on interest rates[1]

		1	2	3
Discount factors – final rate		*0.926*	*0.857*	*0.794*
All-in flows		25.0	−20.0	80.0
PV (all-in flows)		23.1	−17.1	63.5
EV = ΣPV (all-in flows)	69.51			

[1] The last flow is 80, the sum of equity and of the negative income, −20.

In the general case, when assets and liabilities runoff, the EV includes the value of the stream of NII up to the horizon, plus the PV of the liquidity gaps up to the selected horizon. If liquidity gaps are not affected by shocks on interest rates, the sensitivity of EV relates directly to that of the stream of NIIs.

By definition, the EV discounts the "all-in" cash flows, capital plus interest. We can split the two types of cash flows by subperiod, t to $t+1$:

$$\text{“All-in” flows } (t, t+1) = \text{Capital flows } (t, t+1) + \text{Interest flows } (t, t+1)$$

The net capital flows over the periods are the amortizations of assets and liabilities, which are identical to the variations of the liquidity gaps, and equal to the marginal liquidity gaps of each period. The net interest flows are identical to the NIIs. The EV is the PV of all NIIs and of the marginal liquidity gaps:

$$\text{EV} = \text{PV (Future interest margins)} + \text{PV}(\Delta\text{Liquidity gaps})$$

Using the same balance sheet as before, assets and liabilities are now allowed to amortize progressively. The initial amounts of debts and loans are the same, but the outstanding balances now decline from the initial date (Table 7.9). Liquidity gaps are positive until end of horizon, at date 3. The loans are fixed rate (7%) until the horizon. The debt has a fixed rate for the first period (5%), but is rolled over the next periods at a new interest rate if there is an interest rate shock.

TABLE 7.9 Balance sheet amortization up to a fixed horizon

Dates		*1*	*2*	*3*
Assets	1,000	900	800	600
Liabilities	900	800	600	500
Liquidity gaps		100	200	100
Asset flows		100	100	200
Liability flows		−100	−200	−100
Net capital flows		0	−100	100

TABLE 7.10 Interest revenues and costs

Dates	*1*	*2*	*3*
Interest revenues and costs at t *(using balances as of* t – *1)*			
Revenues	70.0	63.0	56.0
Costs	45.0	40.0	30.0
NII	25.0	25.0	26.0

The interest revenues and costs of period t are calculated from the outstanding balances of assets and liabilities as of $t-1$ using the rates applicable to loans and debts (Table 7.10). The NII changes as time passes because both balances of assets and liabilities decline over time.

The PV of all cash flows sums up the PV of all NIIs and of the net capital flows up to the horizon (Table 7.11). The EV is 64.9, the sum of the PV of net capital flows (–3.6) and of the PV of the interest incomes over all periods (68.4) up to the horizon. Conversely, the PV of NIIs is the EV minus the PV of capital flows over any horizon:

$$\text{PV (NIIs)} = \text{EV} - \text{PV (Capital flows)} = \text{PV (NIIs)} = 68.4 - (-3.6) = 64.9$$

These identities hold for any horizon and any amortization profile for assets and liabilities. When extending the horizon to the longest maturity date of all assets and debts, the present value of the entire stream of NIIs is the economic value minus the discounted value of equity.

EV is a target variable of asset-liability management (ALM) because it summarizes in a single value the PV of the entire stream of NIIs up to any cut-off horizon, given the time profile of liquidity gaps.

TABLE 7.11 Present values of all-in flows, capital and interest flows

Discount factors – initial rate		0.962	0.925	0.889
PV (capital flows)		0.0	–92.5	88.9
ΣPV (capital flows)	–3.56			
Discount factors – initial rate		0.962	0.925	0.889
PV (NII)		24.0	21.3	23.1
ΣPV (NII)	68.42			
Discount factors – initial rate		0.962	0.925	0.889
All-in flows		25.0	–77.0	126.0
PV (all-In flows)		24.0	–71.2	112.0
ΣPV (all-In flows)	64.86			

For managing interest rate risk, gaps serve for controlling the sensitivity of the NII over the near term. But closing the interest rate gaps does not imply that EV is immune to interest rate shocks. For controlling the sensitivity of the entire stream of NII of all existing assets and liabilities, it is necessary to control the sensitivity of EV. The conclusion is that it makes sense to use both periodical NII and EV as target variables for ALM. Gap management is dedicated to the near term while controlling the behavior of EV takes care of the long term.

7.3 THE SENSITIVITY OF ECONOMIC VALUE TO INTEREST RATES

The sensitivity of the value of a bank's balance sheet to shocks on interest rates is a long-term target variable for interest rate risk management. This interest rate sensitivity depends on the relative sensitivities of assets and of liabilities.

7.3.1 Sensitivity of Fixed Income Assets or Debts

The value of fixed-rate assets is the PV of all contractual cash flows at the market rates. There is an inverse relationship between value and interest rates (Figure 7.1). The relation means that a fixed-rate loan has more value for the lender when it pays more than the market. The value of a fixed-rate debt also goes up if interest rates decline, which means that the debt value is higher for the borrower if it costs more than the market. The value of a deposit paying a zero rate behaves as a zero-coupon bond. If the market rates go up, it has less value for the depositor because of the larger opportunity cost of not investing the money in markets. Simultaneously, the bank benefits from a debt cheaper than market debt and more so when market rates are up.

For floating-rate assets, the value of cash flows is constant. For a single period, the final cash flow is the sum of principal (N) and interest (i), usually a Libor-based rate. The today value discounts a final flow using the same rate, resulting in a constant value equal to face value $N : N(1+i)/(1+i) = N$. A floating-rate note with maturity of several periods can be

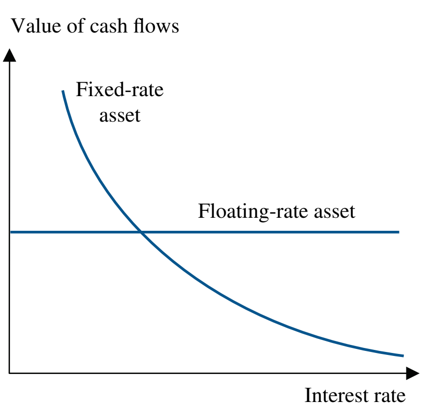


FIGURE 7.1 Market value

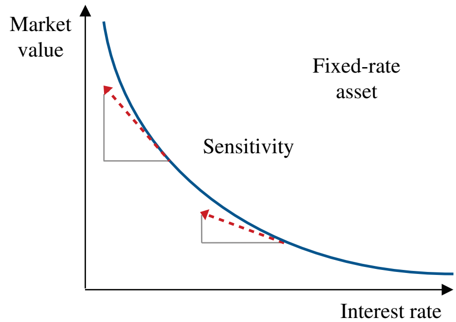


FIGURE 7.2 Sensitivity and interest rates

replicated by a series of single-period notes, with constant value, and is also constant. Floating-rate assets are therefore insensitive to shocks on the floating rate.[1]

The inverse relation between value and rates for a fixed-rate asset is non-linear. The effect of a decrease of rates, from 9% to 8%, for example, is larger than the change generated by a decline of lower rates, such as from 4% to 3%. These changes measure the sensitivity to shocks on rates, which is represented graphically by the slope of the line tangent to the curve at the initial rate (Figure 7.2).

The sensitivity changes along the curve because of its convexity. For small changes of interest rates, the sensitivity is approximately constant. But its variations are not negligible with significant shocks on interest rates. The sensitivity is a local measure, meaning that it depends on current conditions. The convexity is a measure of the curvature in the relationship between debt values and rates that measures how the duration changes with the interest rate. It results from the non-linear functions that represent discount factors.

The formula for the asset value with contractual fixed cash flows is:

$$V = \sum_{t=1,T} \frac{F_t}{(1 + r_t)^t}$$

The F_t are the contractual cash flows and r_t is the discount rate applicable at date t. A similar formula applies with a constant discount rate when this rate is the yield to maturity r. The sensitivity of a bond value with respect to this discount rate relates the change of value to a shock on rates:

$$\Delta V = -[D/(1 + r)]V\,\Delta r$$

The sensitivity is $D/(1 + r)$, V is the initial value and Δr represents the shock on the yield. The coefficient $D/(1 + r)$ is the modified duration. The duration is the first derivative of value with respect to a shock applied to all rates used in discount factors. The calculation shows that the duration is the weighted average of the dates of cash flows, with weights equal to the ratio of

1 Ignoring a constant spread over the discount rate that may apply.

discounted cash flows to the value.[2] For floating-rate assets, the value is constant and the duration is zero.

The duration has a number of important properties.[3] The duration of a portfolio is the average of the durations of assets weighted by their market values. This property allows deriving the duration of the balance sheet as a weighted average of the durations of the bank's accounts. The duration is a "local" measure because it changes if the interest rate changes. The duration drifts when time passes: it increases with maturity, but less than proportionately. If the duration is two years today, it will be longer than one year after one year, because it declines at a slower pace than residual maturity. The implication is that any condition applying to durations at a point in time does not hold permanently, when interest rates change and time drifts.

7.3.2 Sensitivity of Economic Value and Duration Gaps

The EV is the difference of values of assets and liabilities. Its sensitivity depends on their sensitivities and on how convex the relations are between value and rates.

For neutralizing the sensitivity of the EV, the variations of the values of assets and of liabilities should be identical. This condition implies a relationship between the values and the durations of assets and liabilities. With a parallel shift of the yield curve equal to Δr, the sensitivity of EV is:

$$\Delta \mathrm{EV}/\Delta r = \Delta(V_A - V_L)/\Delta r$$

V_A and V_L are the market values of assets and liabilities. The changes in these values result from their durations, D_A and D_L. The sensitivity of EV is the sensitivity of a portfolio long with assets and short with liabilities. Its duration is a linear function of these, using market value weights:

$$\Delta \mathrm{EV} = \frac{[-D_A V_A + D_L V_L]}{(1+r)} \Delta r$$

The duration gap is the difference of durations of assets and liabilities weighted by the market values of assets and liabilities.

$$\text{Duration gap} = -D_A V_A + D_L V_L$$

The duration gap is the equivalent of the interest rate gap, used for controlling the variations of the NII, for controlling the sensitivity of the EV.

The immunization condition, stating that value is insensitive to variations of rates, because the assets and liabilities vary by the same amount, is:

$$V_A D_A = V_L D_L$$

2 Consider a common shock Δr applying to all rates r_t, which become $r_t + \Delta r$. The derivative with respect to Δr is the duration: Duration $= \sum_{t=1,N} t\,F_t/(1+r_t)^t / \sum_{t=1,N} F_t/(1+r_t)^t$.

3 The book by Bierwag (1987, [30]) is entirely dedicated to duration definitions and properties.

The duration gap is in value, not in years. For durations in years, the same immunization condition implies that the ratio of the durations of assets and liabilities should be equal to the ratio of the market values of liabilities and assets:

$$\frac{D_A}{D_L} = \frac{V_L}{V_A}$$

The percentage sensitivity of EV to a shock on interest rates is the ratio of the duration gap, in value, to EV, and is the also the duration of EV. The duration of EV is expressed in years rather than in value:

$$\frac{\Delta \mathrm{EV}}{\mathrm{EV}} = \frac{[-D_A V_A + D_L V_L]}{(1+r)} \frac{\Delta r}{\mathrm{EV}}$$

Sometimes, this duration of EV is called "duration of equity". For instance, using book values of assets (A) and liabilities (L) as a proxy for their values, $V_L = 92\% \times V_A$ when equity is 8% of assets (weighted according to regulatory ratios). EV represents 8% of the balance sheet. If $D_L = 1$ and $D_A = 2$, the weighted duration gap, in percentage of assets, is around:

$$-2 \times 100\% + 1 \times 92\% = -108\%$$

The duration in years is obtained by dividing by the value, and is: $-1.08/8\% = -13.5$. This duration means that the percentage change of the EV for an interest rate shock of 1% is 13.5%.

The duration gap is weighted by market values. The unweighted duration gap has a different meaning. It is zero when durations of assets and liabilities are equal ($D_A = D_L$). When durations match, any change of rates generates the same percentage change for assets and for debts. Since the change in the debt to asset ratio is the difference of those two percentage changes, it becomes zero.[4] If the debt to asset ratio is constant, the debt to equity ratio of a bank is also constant. Hence, matching durations means that the market value leverage is immune to changes of interest rates.[5]

The sensitivity of EV should be compared to capital. It should be managed in such a way that a shock on interest rates does not impair the capital base excessively. This is a possible rationale for setting limits on the sensitivity of EV.

Managing the duration gap implies adjusting the values of assets and liabilities and their duration. To a certain extent, this can be done on-balance sheet. But the characteristics of the banking book are driven by the commercial policy and by the demand of the bank's customers. For managing the duration gap, more flexibility is required.

---

4 The change in the asset to debt ratio is: $\Delta(V_A/V_L)/\Delta i = \left(-1/V_L^2\right)(V_L \Delta V_A - V_A \Delta V_L)/\Delta i = (-1/V_L^2)\left[\frac{-V_L V_A D_A}{1+i} + \frac{V_A V_L D_L}{1+i}\right]$. This can be simplified by factoring out common terms and becomes a function of: $(D_A - D_L)$. The sensitivity of the ratio is zero when $D_A = D_L$. This condition implies that the ratio of assets to liability is constant, or, equivalently, that the leverage ratio of debt to equity is also constant.

5 Such a condition would apply, for example, to a fund that uses a constant leverage in terms of market values. Typically, a fund has assets and debt complying with the constraint that asset value remains above that of debt, the asset to value ratio. Maintaining such ratio would imply equal duration of assets and of debt.

The alternate method is to use derivatives, or off-balance sheet management. Derivatives modify duration to the extent that they modify the interest flows. A fixed-rate asset combined with a swap paying fixed and receiving floating behaves like a floating-rate asset.[6] Future contracts[7] have durations identical to the underlying asset because they track its value. Forward contracts are similar contracts, which are traded over the counter, unlike futures traded in organized exchanges. Both have a similar behavior when there are interest rate shocks. For example, a forward loan would lock in an interest rate for a future period. It is interest sensitive since it gains value if interest rates decline and loses value if they move up.

7.3.3 Behavior of Value with Shocks on Interest Rates

EV is, in general, sensitive to interest rate variations. The shape of the relation depends upon the relative durations of the assets and the liabilities. When the duration gap is equal to zero, the EV is immune to variations of interest rates as long as they are not too important. A duration mismatch (weighted by asset and liability values) between assets and liabilities makes the EV sensitive to rates. The standard mismatch position of banks, a mismatch between maturities of assets and liabilities, also implies an open duration gap.

Figure 7.3 shows the sensitivity of the EV with a mismatch of durations. In this example, there is a value of interest rates such that EV is zero. If interest rates deviate from this position, the EV can be positive or negative.

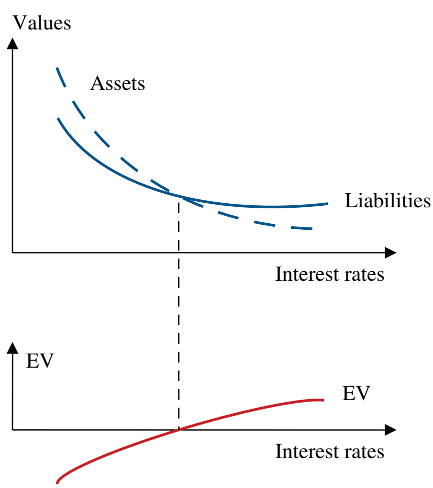


FIGURE 7.3 EV and duration mismatch

6 There would remain, in general, a fixed-rate interest flow, equal to the difference between the fixed rate received by the asset minus the fixed rate paid by the swap, and they are not equal in general.

7 A futures is a contract setting today the price of an asset for a future date. These contracts are traded on organized markets. The value of a future contract tracks the price of the underlying asset from which it differs by the net carry, the net cost of holding the underlying until delivery of the asset.

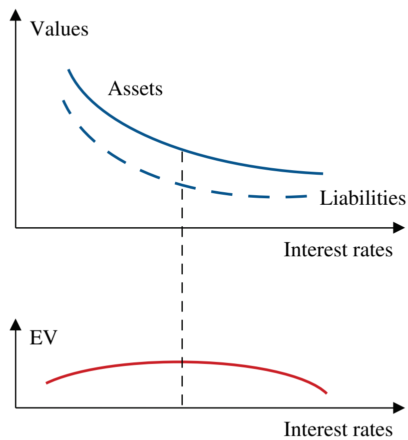


FIGURE 7.4 EV and duration match

Figure 7.4 illustrates a situation when the weighted durations of assets and liabilities match. The convexities[8] of assets and liabilities are not too important, so that small variations of interest rates only slightly affect the EV, and the EV is immune to small variations of interest rates, in the neighborhood of current interest rates. In the example, the profiles of assets and liabilities have their usual shape, with an upward-looking convexity. The EV is positive because the value of assets exceeds that of liabilities, and the slopes of the curves are identical at the initial interest rate because their durations match.

In general, the convexities of assets and liabilities differ and they generate a sensitivity of EV for significant variations of rates. Convexity can theoretically be beneficial for the bank. For example, Figure 7.5 shows that EV increases with up and down variations of interest rates. This happens when the curvature of assets is more important than the convexity of liabilities.

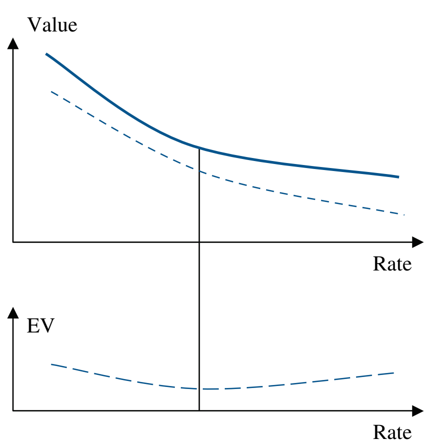


FIGURE 7.5 Convexities and EV sensitivity to interest rate

8 Convexities are the curvatures of the shapes representing the relation between values and interest rates.

The difference of the convexities of assets and liabilities is the “convexity gap”. The convexity can be shown to increase with the time dispersion of the cash flows. This effect can be illustrated by comparing two assets, a zero-coupon bond and a portfolio of two zero-coupon bonds with the same duration (see Appendix below).

Unfortunately, unlike the favorable situation when the convexity has a positive impact on value, it is shown in Chapter 8 that convexity is always adverse to the bank due to the embedded optionality in banking products.

7.4 APPENDIX: CONVEXITY

Convexity, given duration, increases with the dispersion of cash flows through time. This can be seen by comparing a single bond with a portfolio of bonds with same duration as the single bond. The zero-coupon bond has duration equal to maturity. The portfolio of two zero-coupon bonds is constructed such that its duration matches that of the single zero-coupon bond. By changing both the nominal and the maturity of each zero-coupon bond of the portfolio of two bonds, the portfolio duration and the portfolio value can be made equal to those of the single bond.[9] Figure 7.6 shows the portfolio (continuous line) and the single bond (dashed line).

When the interest rate is 6%, they have exactly the same duration. But the portfolio line, of which cash flows have different maturities, is more convex.

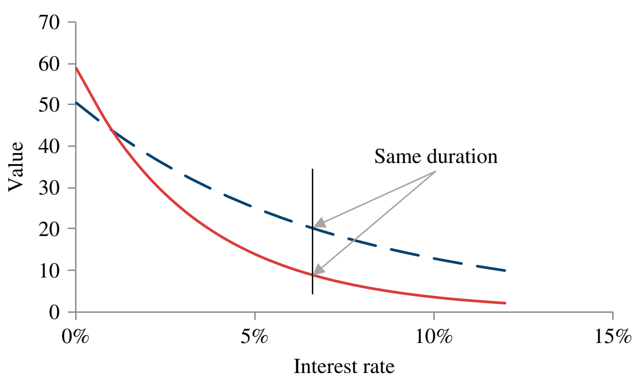


FIGURE 7.6 Convexity and dispersion of cash flows across time

9 We have a system of two equations: one for equating the value-weighted duration, and one for equating the sum of value of the portfolio to that of the single bond. The portfolio used in the example is made of two zero-coupon bonds with, respectively, values 35.71 and 8.10, rates 1% and 4% and the same maturity, 30 years. The same duration zero-coupon bond has a maturity of 15 years, a 5% rate and a value 43.81, equal to the value of the portfolio of the two zero bonds.

8

Convexity Risk in Banking

The optionality of banking products is always adverse to the bank. Fixed-rate loans are renegotiated when interest rates are down with an adverse effect on revenues. Deposits are withdrawn when interest rates are up making the cost of financing higher. Furthermore, such events are dependent on interest rates.

The consequences are numerous. Cash flows are stochastic, which triggers liquidity risk. Standard gaps or duration gaps underestimate the interest rate sensitivity of net income and economic value. The valuation of the balance sheet is adversely affected by the value of the options given away to clients. Finally, the embedded options in banking products make the assessment of risk and of value far more complex than with the basic tools.

This chapter has two main purposes. The first is to explain the nature and consequences of optionality in banking. The second is to explain how an extended framework can address the conceptual and practical issues raised by optionality. The next chapter shows an example of how the principles outlined here can be implemented in the case of the prepayment option of mortgages.

Contents

8.1 CONVEXITY RISK AND ECONOMIC VALUE

Borrowers can renegotiate the rates of their loans and depositors can withdraw or add freely their funds to their accounts. These rights are the embedded options in banking products. Such optionality is adverse to the bank.

Fixed-rate mortgages are renegotiated when interest rates move down, while the fixed-rate debt backing the mortgages stays in place in the balance sheet. Variable-rate mortgages often embed a cap that is triggered in rising market rates, while the variable-rate debt backing the mortgages is not capped. In the absence of any hedge, the net interest income is down in both cases. On the liability side, deposits are withdrawn and funds transferred to interest earning accounts when interest rates rise. If the resources remain with the bank, they cost more. If they are withdrawn, new and costlier financing has to be substituted to the withdrawn deposits.

The options embedded in loans and deposits benefit the client and entail a loss for the bank. These options are held by the clients and given away by the bank. Their value should be deducted from the value of the balance sheet. The value of options depends on interest rates. It materializes when interest rates vary significantly on either side, up or down. When they move up, the incentive to withdraw funds from deposits gains value. When they move down, the benefit from renegotiation of fixed-rate loans gains value. The bank loses the value of these options when interest rates have significant swings and the value of its balance sheet behaves as an inverted U curve, declining when rates move on either side.

This impact of options on value has a graphical representation. In the absence of optionality, both assets and liabilities are inverse and monotonous functions of interest rates. The effect of optionality is to inverse the convexity of these curves.

Fixed-rate loans behave as if they were variable rate when interest rates are moving down and a renegotiation or a prepayment occurs. Their duration moves down instead of up. The upward section of the curve, showing how their value behaves when rates are up, flattens out when rates have gone up and the value becomes insensitive to the rise of rates (Figure 8.1). Prepaying the loans has the same effect since the funds made available by the client can only be reinvested at the new rates.

Variable-rate mortgages behave as if they were fixed rate when rising market rates hit the caps embedded in the contracts, and their values start declining instead of staying flat. In both cases, the convexity of loans is inverted.

Deposits would behave as a zero-coupon bond if they were not subject to free withdrawal or additions. In a low interest rate environment, there is little incentive to withdraw or add funds. In a high interest rate environment, there is an incentive to move the funds away from deposit accounts into interest earning assets. The funds moved to floating-rate assets become insensitive to variations of rates, the curve becomes horizontal instead of declining and the value of the liability becomes higher than the value of non-callable funds.

Figure 8.1 shows the effect of optionality on value. Non-callable loans and deposits follow the usual monotonous inverse relationship with market rates. When rates are down, the value of callable loans flattens out on the left-hand side. When rates are up, the value of deposits flattens out on the right-hand side. The gaps with the profiles of non-callable loans or deposits represent the value of the options, which increases as the deviations of market rates extend to the left or to the right. The net effect is an inverted U curve for economic value.

Duration gaps capture the first-order effect of a small change of interest rates on value. In the mid-range of the graph, the value is positive and durations of assets and liabilities have the

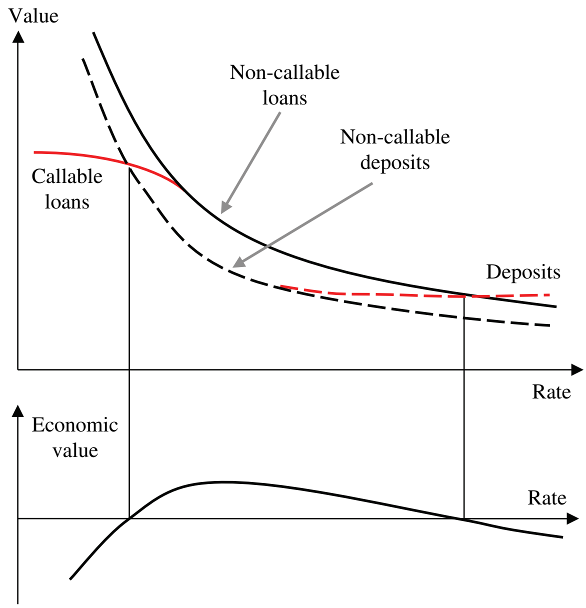


FIGURE 8.1 "Scissors effect" on economic value due to embedded options

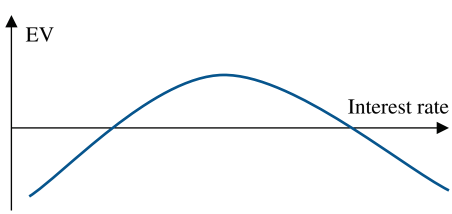


FIGURE 8.2 Effect of convexity on economic value

same signs. When moving away on the left-hand side, the duration of assets starts decreasing with rates. On the right-hand side, the duration of deposits rises to zero instead of declining (Figure 8.2). Convexity measures how the duration changes when rates vary. When the curves of assets and liabilities cross, the economic value is zero, and it becomes negative beyond, on either side. This is the general case for commercial banks.

The effect of optionality on economic value has a direct translation in terms of net interest income. Adverse deviations of economic values imply adverse deviations of net interest incomes.

The convexity is due to options, of which payoff has a kink.[1] When options are out-of-the-money, their payoff is zero, when they are in-the-money, their payoff and their value start

1 The payoff of an option is near-zero if out-of-the-money and becomes a function of the underlying asset value if in-the-money. Accordingly, the payoff from immediate exercise has a kink, as the sensitivity changes from zero to a positive number. The relationship between the option value and the underlying is convex.

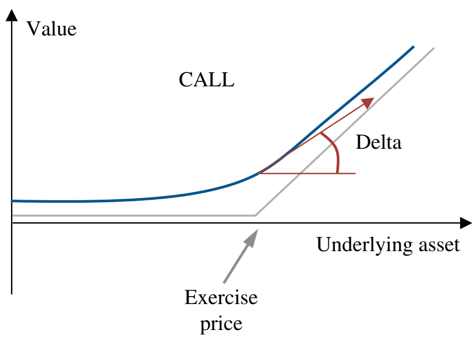


FIGURE 8.3 The value of a call option as a function of the underlying

moving like the underlying asset (Figure 8.3). Their convexity is positive, which means that the payoff and the value increase more proportionally than the underlying asset. The figure shows the profile of a call option on an underlying asset.

The call option is the right, but not the obligation, to buy the asset at a preset price, the exercise price. This right has little or no value when the value of the asset is lower than the exercise price, and starts gaining value as the price of the asset moves above the exercise price, hence the kink of the curve. The sensitivity of the option, its delta, is its rate of change for a unit change of the underlying. The delta of the call option is positive and moves from zero to a value of 1 when the option is deep in-the-money.

Options embedded in banking products are held by clients, and hence “sold” by the bank. Their profiles are inverted and downside, relative to a purchased option, hence the negative convexity of callable loans.

A positive convexity has value in volatile environments, as it is shown in the Appendix in section 8.3. Sold options have a negative convexity, the value of which is negative. A bank is short on convexity, meaning that the value declines when volatility is up. The drift of sensitivities, the deltas, shows up in the change of the slopes of the curves representing assets and liabilities. The changes of deltas are measured by “gamma”, the second derivative of the curves. A bank is “short gamma”: this is the opposite of the favorable situation, being long in volatility, when value increases with volatility, as it does for a purchased call option.

The options embedded in callable loans have a value that can, theoretically, be priced to the client, if competition allows such mark-up. Hedging the optional risk can be achieved by purchasing options, of which cost can be offset by pricing the options. For example, the risk for callable loans is that the interest revenues decline with rates. The appropriate hedging option is a floor, an option that guarantees a minimum return for a lender. The pricing of the call option and its hedging are detailed in Chapter 9.

8.2 AN EXTENDED FRAMEWORK FOR STOCHASTIC CASH FLOWS: VALUATION

The optionality of accounts with an indeterminate maturity makes the volumes and the cash flows of these accounts stochastic. A fraction of the variance is independent of rates, but another fraction depends on interest rates. With stochastic cash flows, standard gaps underestimate the sensitivity of earnings and valuation should include the adverse convexity effects of embedded options. The standard gaps or the duration gaps make sense when the volume of fixed or variable-rate accounts is independent of interest rates, but these simple models fail to properly address the sensitivities to interest rates if gaps vary with interest rates.

The sensitivity and valuation of the balance sheet have to be adjusted for optionality. Such issues require a significant extension of the modeling and conceptual framework. An appropriate extended framework requires several building blocks for addressing the variety and complexity of issues raised by stochastic cash flows and dependency on interest rates.

The purpose of this section is to detail these issues, which have generated much debate in the industry. The implementation varies across banks, which rely on a variety of tools to reach satisfactory conclusions on how to deal with this complexity. Examples of how the principles enunciated hereafter can be turned into practical applications follow in Chapter 9, which addresses the optionality of callable loans, a major issue in interest rate risk management of the banking portfolio.

Valuation cannot follow the basic processes previously addressed. Deriving a value of the balance sheet with deterministic scenarios is misleading because it ignores the value of options given away. The discounting of stochastic cash flows requires reverting to the principles of valuation when both interest rates and cash flows are random. The so-called risk-neutral valuation is required. Risk-neutral valuation requires discounting at risk-free rates across all interest rate scenarios and deriving value as the average of the present values generated for each scenario. An interest rate model is required to generate a comprehensive set of interest rate scenarios covering the variety of outcomes. Such extended framework should allow addressing convexity risk, valuing the options given away to clients, and obtaining an option-adjusted sensitivity of value.

The behavior of accounts depends on time, on portfolio characteristics and on the clients' behavior with respect to the movements of customers' rates and market rates. Customers' rates depend on market rates, and the volume of accounts depends on both market interest rates and on customers' interest rates. Interest rate scenarios are inputs. A comprehensive framework for modeling the stochastic behavior of accounts should therefore include three building blocks:

- An interest rate model;
- A customers' rate model;
- A volume model.

Interest rate scenarios are inputs to the customers' rate model and the volume model. Such interest rate scenarios have to defined, or they have to be modeled when the purpose is to describe a wide spectrum of outcomes, or for valuing the options in embedded banking products. By adjusting customers' rates, the bank makes its loans or deposits more or less attractive for clients, and can smooth out the variations of market rates. The purpose of the customers' rate function is to define how clients' rates behave as a function of market rates. The volume of accounts, and how they run off, is a function of time and depends on how customers behave given market rates and customers' rates. Both are inputs for explaining the variations of runoffs as time passes.

8.2.1 Customers' Rate Models

Customers' rates depend on interest rates, but banks have the option to adjust clients' rates for smoothing the reactions of customers and the resulting variations of volumes.

The differential between customers' rates and benchmark market rates has a potential influence on the volumes of mortgage prepayments, or of deposits inflows and outflows. Customers' rates do not adjust mechanically to market rates, unless they are floating rates. They do adjust for new transactions, fixed or floating rate, as banks attempt to stabilize volumes or adjust to competition. The speed of adjustment is usually not instantaneous and reduces the variations of market rates. Consequently, customers' rates become a function of both current and lagged market rates.

It has also been observed that the speed of adjustment also depends on whether rates are rising or falling or whether they are above or below benchmark rates, those rates that match some long-term margin.[2] When market rates increase, the customers' rates tend to be stickier than when interest rates decline.

A customers' rate function would use as inputs the current and past market rates. For capturing the asymmetry of variations, coefficients can be differentiated for up movements and down movements.

8.2.2 Volume Models

In general, the volume of accounts depends on some trend, a random noise, plus a function of the customers' and the market rates. For example, the runoffs of existing fixed-rate mortgages depend on contractual amortization, on factors independent of interest rates accounting for geographical mobility or other random elements, plus the differential between the customers' fixed rates and market rates.

Deposit volumes follow some trend and have seasonality and random fluctuations, which can be, or not, explained by market and customers' rates. Various functional forms have been proposed in the literature. The form of the trend can be derived from time series models. Seasonal fluctuations are easily captured with dummy variables. The dummy variables for months are indicator functions, a dummy variable taking the value 1 or 0 depending on the month. The variation of a short-term market index might have an impact on volumes, or not, depending on the context.

For example, the overall balance of existing accounts often increases. But the same growth can be obtained from various combinations of the variations of the number of accounts and of the average balance of each individual account. The unit balance – by account – can depend on the age of deposits. The implication is that the aggregated volume of deposits, and how they run off with time, requires isolating various effects. Deposits are also sensitive to many factors other than interest rates, such as the bank's commercial policy or regulations. Some banks propose a teaser rate for attracting new customers. Regulations include tax benefits of some saving accounts, plus a cap on the balance of eligible accounts. Regulated saving accounts also earn a regulated rate, such as a combination of the short-term rate and the inflation, which changes by steps rather than smoothly.

Such effects should be controlled for determining whether the volume of existing deposits is sensitive to interest rates. For testing a dependency of deposits on interest rates, the function should include market rates and allow lags for capturing the stickiness of customers' rates.[3]

---

2 Jarrow, R. A., Van Deventer, D. R. (1998), The arbitrage-free valuation and hedging of demand deposits and credit card loans, [82].

3 See the chapter "Modeling Non-maturing Products", in Matz, L., Neu, P. (2006) Liquidity Risk Measurement and Management: A Practitioner's Guide to Global Best Practices, [97].

8.2.3 Interest Rate Models

A small number of discrete scenarios can be used for interest rates. But, for generating a wide spectrum of interest rate scenarios, interest rates need to be modeled.[4]

The interest rates are modeled as a stochastic process. A process defines the interest rate variation over an infinitesimally small period with two basic components: the drift, which is the deterministic trend factor; and a random noise, which is the diffusion process, of which volatility matches that of the modeled interest rate.[5]

The choice of the model depends on the purpose. For example, interest rates can be used as inputs in volume models and are then used for generating cash flows. A simple technique consists of using principal component analysis, presented in Chapter 16 on market risk.

For the purpose of valuation, a calibration to market rates has to be performed. The calibration ensures that actual market prices of Treasury bills and bonds appear to the model neither too high nor too low. All rates have to be adjusted in order to achieve the fit between interest rates and market prices.

Valuation requires discounting of cash flows using risk-free rates as if in a risk-neutral world. In a risk-neutral world, investors behave as if there was no aversion to risk. Valuation is conducted by discounting cash flows at risk-free rates and averaging the values across all interest rate scenarios.

Several functional forms for interest rate processes have been used. In the example of prepayment option used in Chapter 9, a binomial model for the short-term rate is used.[6] Interest rate scenarios are time paths of interest rates, up to a horizon divided in short time periods. The calculation of the interest rate at each interval depends on the previous rate. Each sequence of short-term rates is an interest rate scenario. The scenario is used to make the cash flows from prepayments dependent on how the interest rates evolve across successive time periods.

8.2.4 A Comprehensive Framework

The three above modules are interacting. The interest rate scenarios are inputs to the customers' rate model and to volume models. Volume models use market rates and customers' rates. Models are behavioral in that volumes depend on customers' initiative. The interaction can be summarized in Figure 8.4.

This framework can be used for modeling cash flows that are contingent on interest rates. In the case of mortgages discussed in Chapter 9, a short-term rate model provides multiple time paths for the short rate until the date of each cash flow. Cash flows are contingent on interest rate scenarios: they occur at random times and have random values depending on interest rate. They are known only for a given interest rate scenario. The interest rate model is then used for generating contingent (i.e. optional) cash flows.

Once cash flows are modeled as a function of interest rates, the valuation mechanism follows. Discounting uses each interest rate scenario for calculating the present value of future

4 A short review of interest rest rates is in Hull [78]. Models of interest rates are reviewed in Brigo, D., Mercurio, F. (2007), Interest Rate Models – Theory and Practice: With Smile, Inflation and Credit, [36].

5 Some common processes are presented in Chapter 11.

6 The model is derived from Black, F., Derman, E., Toy, W. (1990), A one-factor model of interest rates and its application to Treasury bond options, [31].

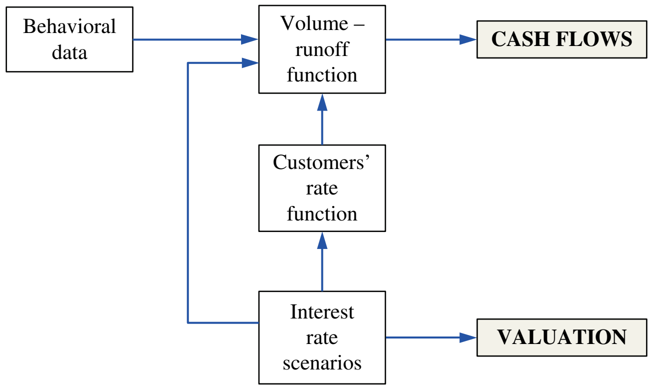


FIGURE 8.4 A comprehensive framework for valuation (*Source:* Adapted from L. Matz and P. Neu, [97].)

cash flows. A cash flow dependent on interest rates is defined for each scenario. Its present value is known for this interest rate scenario. With many interest rate scenarios, there are also several present values of this single future cash flow. The today value of the random cash flow is the expectation of all such values. The valuation process averages the present values of all random cash flows across all possible outcomes.

In general valuation uses either risk-free rates or risky rates for discounting. For traded assets, the contractual cash flows can be discounted with risky rates, which add an additional spread to the risk-free rate. The spread is observable as the difference between risky yields and risk-free rates. The alternate approach averages the present values of random cash flows, across all possible outcomes, after discounting with risk-free rates. In the first case, the risk is accounted for in the discount rate. In the second approach, the risk is embedded in the probability distribution of cash flows.

The valuation of the banking book follows this second approach by averaging the value of assets and liabilities across all outcomes. The probabilities of the outcomes are called risk neutral. They differ from real-world probabilities because they embed the risk aversion of market players. These risk-neutral probabilities are those that, if applied to the cash flows of a credit risky asset discounted at risk-free rates, would provide the exact price of the asset in markets. As the risk is now embedded in the various outcomes for the cash flows, the relevant discount rate should be risk-free.[7]

Hence, for valuation purpose, discounting is conducted with scenarios of risk-free rates and the uncertainty is taken care of by the multiple scenarios from which interest rate-dependent cash flows derive.

The valuation mechanism, in the presence of options, is illustrated in the case of prepayment of mortgages in Chapter 9. It addresses many issues raised by convexity risk:

- Adjusting the economic value of the balance sheet for embedded options;
- Adjusting the sensitivity of the economic value to shock on interest rates in the presence of options;
- Pricing the embedded options to clients, once they are valued.

7 Risk-neutral valuation is a standard topic in finance texts addressing the valuation of market assets. See, for example, Hull, J. (2014), Options, Futures and Other Derivatives, [78].

Discounting deterministic earnings and cash flows provides the economic value of the balance sheet in the absence of optionality. The omission ignores convexity risk and the short "gamma" position of banks. The economic value netted from the value of options given away to clients is closer to the true value. The sensitivity, like value, should be option adjusted. When optionality is accounted for, the sensitivity of value to shocks on interest rates, adjusted for convexity, can be derived from two valuation simulations, with and without a shock on rates. The sensitivity is the variation of value and is now adjusted for convexity effects. Finally, assigning a value to embedded options allows pricing them to clients.

8.3 APPENDIX: THE VALUE OF CONVEXITY

Convexity is relevant because it provides an additional return when the volatility of interest rates becomes significant. This notion is illustrated in Figure 8.5.

It is assumed that the interest rate can have two values with equal probability, at equal distance from the current value. Without convexity, the expected value of an asset would be on the straight line tangent to the curve at the initial point matching the average value of interest rate.

A convex asset has higher values when the interest rate shifts up or down from its initial value. Its average value is also on the straight line joining the two points representing the values for these two interest rates. But the line is above the tangent and the expected value of the asset is higher than the asset value at the initial interest rate. The higher the volatility of interest rates, the higher the value of convexity.

A higher convexity of assets increases the value of the balance sheet and, conversely, a lower convexity of liabilities makes their value lower.

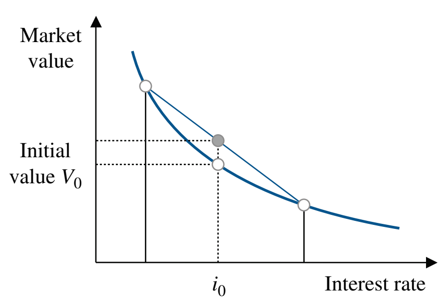


FIGURE 8.5 Value of convexity

9

Convexity Risk: The Case of Mortgages

The renegotiation of fixed-rate mortgages is a well-known case of convexity risk. Callable loans raise many issues: How should the risk be managed? How can the risk be measured? How can the risk be valued and priced to customers?

The first section of this chapter addresses the management issue for the mortgage portfolio. From a management standpoint, the runoff rate of mortgages is a function of the differential between the customers' and market rates for new mortgages. The management of liquidity and interest rate risk depends on the views, or scenarios, for interest rates. Gap management needs to be adapted accordingly, with frequent rebalancing of financing and hedging as a function of current and future conditions on interest rates.

The second part of this chapter addresses the valuation issue. It relies on the simple binomial model of short-term interest rates. Using a simplified example, it is shown how contingent cash flows are modeled from interest rate scenarios, how the value of the prepayment option is derived, and how it can be priced to clients.

Contents

9.1 THE CONVEXITY RISK OF MORTGAGES

Once mortgages are originated, they have a lifecycle with several phases. The exposure to interest rate risk starts from the initiation phase, when the client signs the contract. Over the initial delay between signature and lending, the agreement on the contractual rate is binding for the bank, thereby generating interest rate risk. The regular amortization period starts once the loan is extended and funded by the bank.

Convexity risk might emerge at this stage. For fixed-rate mortgages, a fraction of prepayments is due to unexpected events and another fraction is triggered by declining interest rates. The bank might be forced to accept a decline in the client's rate, while the loan is already financed with fixed-rate debt contracted at the time of the cash outlay, the effect of which is a decline of bank's earnings.

Floating-rate mortgages might include a cap on the floating rate paid by the borrower. As long as the cap is not in-the-money, the bank's margin is not at risk if the loan is financed with a floating-rate debt. But the loan interest revenue is often capped while the bank's floating-rate debt is not, resulting in a potential loss of margin in periods of rising rates.

The gain from prepayments for the client increases with the positive differential between the contractual rate of the customers' and the current market rate, and with the residual maturity of the loan. The magnitude of the risk can be monitored, at a point in time, by breaking down the mortgage portfolio according to clients' rate and to residual maturity. For clients' rates above market rates, the risk is material, and more so when the gap with market rate and the residual maturity are larger. The size of subportfolios for which renegotiation is profitable for the clients measures the potential exposure to prepayment risk.

The area in white of Table 9.1 shows which pools of loans have rates close to current rates.[1] The grayed area shows pools for which the prepayment option is in-the-money. Such breakdown provides, at best, a view of potential exposure, not an estimate of the balance of the mortgage portfolio that could be prepaid. First, prepayments also depend on other random events, such as geographical mobility, unemployment or changes of civil status, for example, which are unrelated to interest rates. Second, exercising the right to renegotiate or prepay the loan depends on the payoff for the client, and on clients' behavior, with some lags in adapting to the interest rate environment. The expected payoff of the prepayment option depends on which interest rate scenarios are anticipated and on any penalty from renegotiation. The valuation of the option, in the subsequent section 2, shows how such payoffs can be modeled.

9.1.1 Mortgages: Runoff Function

A runoff function models the behavior of prepayment rates[2] and how they react to interest rates. Several runoff functions have been considered, differing on how simple they are. The simplest runoff function could be estimated as a stable prepayment rate for accounts beyond some threshold age. Prepayments do not occur too early when a loan is contracted, or too late when principal is low, since the customers' gains drop when getting closer to maturity. The seasoning of a portfolio refers to the age required for a newly issued mortgage to attain a stabilized

1 Table 9.1 uses a common rate, 4.5%, for breaking down the portfolio around current rates, but such a cutoff rate depends on maturity.

2 The prepayment rate is expressed as a percentage of the portfolio.

TABLE 9.1 Sample breakdown of a fixed-rate mortgage portfolio

	Duration (years)				
Client rate (%)	*0–5*	*6–10*	*11–15*	*16–20*	*Total*
2.00	1	0	0	1	3
2.50	93	0	1	3	98
3.00	253	12	5	6	276
3.50	126	44	37	33	241
4.00	154	92	63	136	444
4.50	*102*	*167*	*149*	*194*	*612*
5.00	99	106	208	129	542
5.50	87	67	79	263	496
5.75	66	29	39	55	189
6.00	53	19	24	46	143
6.50	63	18	28	47	156
7.00	42	11	9	15	77
7.50	11	1	2	4	18
8.00	4	0	0	1	7
8.50	1	0	0	0	2
9.00	1	1	0	1	3
Total	1,157	568	646	936	3,307

benchmark rate. But constant prepayment rates, beyond some value of the age of accounts, are a crude approximation. Some clients always prepay their loans regardless of the level of interest rates. Others will prepay when the payoff is significant. Finally, some never prepay. In the European case, prepayment rates are in the range of 10% to 30% or more. Because prepayment rates are significant, the impact on earnings cannot be ignored.

A runoff function can be modeled as an "S curve". There is a minimum prepayment rate for a period and a maximum prepayment rate. In between, the runoff rate is an increasing function of the algebraic differential between customers' rate and market rate. The shape of the curve is illustrated in Figure 9.1. At the initial interest rate, the differential is zero, and the average runoff rate is, for example, 20%. Such effective amortization of mortgages is inferred from historical statistics for projecting the future balances. The S function shows how the prepayment rate deviates as a function of the differential of initial and current rates. In the figure, the lowest prepayment rate on the curve is 10%, the median value is 20% and the maximum value is 30%.

If the S curve is determined, it becomes a volume model that can be implemented for managing the risk. The curve is used to assess the departures from the average prepayment rate dependent on interest rates. The issue for the bank is to address the deviations from the current runoff rate dependent on market rates. Such deviations generate both liquidity and interest rate risk. Liquidity risk arises because the amortization will move away from the initial effective amortization schedule. The interest rate risk results from the volume of the mortgage portfolio, the rate of which is renegotiated or prepaid.

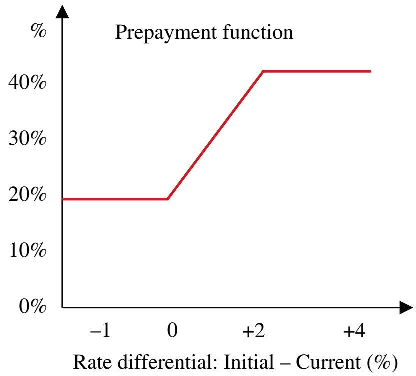


FIGURE 9.1 Simplified runoff function for fixed-rate mortgages

At this point, an incremental risk management can be initiated. The initial hedges in place are adjusted as time passes. If interest rates move down, the amortization accelerates, and the new funding of the bank might be reduced. Any hedge in place for interest rate risk should be adjusted. If the S curve is relatively stable, it can be used for anticipating the next adjustments with forward-looking interest rate scenarios.

This continuous adjustment is similar to a delta hedge.[3] In delta hedging, hedges are used, given a point in time exposure to interest rates, and adjusted with the movements of market rates. For example, closing a gap with a swap is such that the notional of the swap matches the size of the gap. When interest rates move, a new mismatch appears because the gap is sensitive to interest rates due to the shift in amortization. As the exposure changes, following the S curve, the hedge is adjusted so as to close the new incremental gap. When using incremental hedges for mortgages, the rebalancing of the hedge mimics the behavior of embedded options.

9.1.2 Optional Hedges

Optional hedges can be used to offset convexity risk of both fixed-rate mortgages and floating-rate mortgages. Floors hedge the risk of fixed-rate loans and caps hedge the risk of floating-rate loans when they embed a contractual cap. The standard options on interest rates are “caps” and “floors”. A cap guarantees a maximum rate to the buyer. A floor guarantees a minimum rate to the buyer. Lenders are interested by floors because they are willing to be protected against declining interest revenues.

Banks borrow fixed rate for financing fixed-rate loans. They would consider floors if they fear substantial prepayments. Borrowers paying a floating rate are interested in caps because they guarantee a maximum rate. For floating-rate loans embedding a cap, banks borrow floating rate if they wish to lock in the spread between the lending and borrowing rates. They would

3 Delta hedging occurs when a position is hedged by another instrument that has the same sensitivity as the hedged instrument. Well-known examples involve options on stocks and shares. For example, the value of a call option varies as delta times the value of the underlying stock. Such sensitivity can be offset by holding an opposite position in stock, with value being delta times the stock value. As the underlying asset price changes, the delta changes and the proportion of stock held against the option has to be continually adjusted.

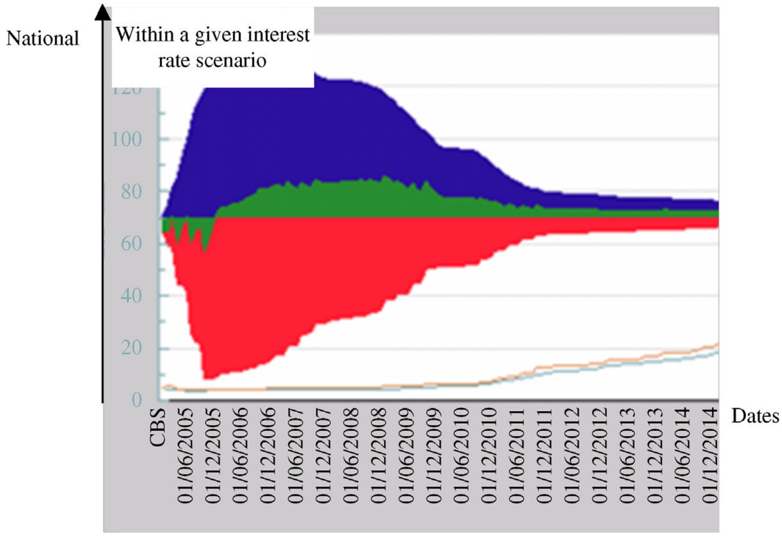


FIGURE 9.2 Sample optional gap report

consider caps if they fear that the caps of loans will become in-the-money because of higher short-term rates.

Because the cost of such options is high, banks would enter into such contracts when they are out-of-the-money, when their cost is still low. They would thus gain protection against wide fluctuations of volatile market rates. The policy is a protection against the downside risk of economic value. Banks are short volatility and the hedge from purchased caps and floors is a long position on volatility.

It is not simple to determine how effective such hedges are. Consider the case of hedging capped variable-rate loans. The effectiveness depends on interest rate scenarios and on the levels of caps embedded on loans. Suppose that optional hedges against the rise of the debt cost are in place. An "optional gap" shows the size of capped loans relative to the notional of hedging caps acquired by the bank. In this case, the gap profile shown in Figure 9.2 is a "gap of caps".

Such optional gap is not sufficient to assess the sensitivity of the bank's earnings to a change of interest rates, because a change of interest rates has an effect on interest revenues and costs only when the options are in-the-money. The sensitivity of the portfolio to variations of interest rates depends on which options, those in the loan portfolio and those purchased by the bank, are in-the-money or not, and by how much.

An "optional gap report" shows the net effect, given an interest rate scenario, over the management horizon. The net payoff equals the gain from hedging caps minus the losses from caps held by clients. Figure 9.3 illustrates the effect of options for a given interest rate scenario. The bars show the interest revenue given up on the top section of the graph and the interest payoff from bank's caps on the bottom section. Positive numbers are losses and negative numbers are gains in this figure. The effect on net interest income is the difference, shown as the intermediate time profile. In the example, the savings from purchased caps less than offset the revenue lost from loans.

An optional gap report is not an optimization tool. It provides information for contracting optional hedges. For example, in the interest rate scenario considered, the figure shows how much can be lost, and when, from capped loans. But the figure shows only a differential payoff,

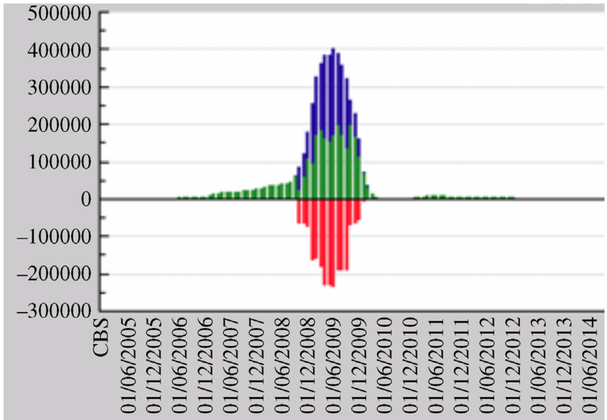


FIGURE 9.3 Interest revenue from loans given up and interest income from hedging caps

not the initial cost of purchasing the caps. The cost of purchased options should not exceed this net payoff.[4] The report is an aid for deciding how much hedge, and over which horizons, should be purchased by the bank in a given interest rate scenario.

The valuation of options sold and options purchased serves for showing the impact on the economic value of the balance sheet. The next section addresses the valuation of embedded prepayment options in fixed-rate loans.

9.2 THE VALUATION OF THE PREPAYMENT OPTION AND ITS PRICING

The valuation of options is important for measuring the value given away to customers, or for determining how much should be charged to them for the benefits of holding their rights. Valuation requires an interest rate model and modeling the cash flows from prepayments of fixed-rate mortgages. The interest rate model used to illustrate the process is a binomial model of the short-term rates. Given the volatility of the market rates, the model allows generating a large number of interest rate time paths. Each time path of interest rates is an interest rate scenario. For each interest rate path, the benefits from exercising the prepayment option or not are valued for all time points.

The model is first used as a cash flow generator, the cash flows from prepayments being contingent on interest rates. For each of the possible scenarios, the valuation process then discounts these contingent cash flows using the simulated values of the short-term rates along each time path. Valuation is the average of all discounted values for all time paths of rates. The value of the callable loan is that of the non-callable loan minus the value of the option.

9.2.1 Payoff of Prepayment

For a fixed-rate loan, the option payoff is the present value of the differential cash savings before and after exercising the renegotiation option. The payoff increases with the differential

4 This is meant for a given interest rate scenario and if ignoring the time value of money.

between the initial and the new interest rates and the residual maturity of the loan. The payoff at a point in time is lower than the value of the prepayment option because renegotiating the interest rate at a later date might be more beneficial than today. This is the time value of the option.

The renegotiation usually generates a cost to the borrower. For instance, a prepayment penalty of 3% of the outstanding balance might be imposed on the borrower. But this cost can be lower than the expected gain from exercise.

The expected value of payoffs, at all points in time until maturity, is the value of the option. Calculating the immediate payoff is a necessary step for valuing the option. Because the option is held by the clients and given away by the bank, the value of the loan is equal to the value of a standard loan (non-callable), minus the value of the option.

The time path of interest rates is relevant for options. If the rates decline substantially at the early stage of a loan, they trigger an early prepayment. If the decline occurs latter, the prepayment will also occur later. The implication is that prepayments do not relate only to the current interest level.

9.2.2 Payoff of Prepayment with Immediate Exercise

Consider a fixed-rate amortizing loan, repaid with constant annuities including interest and capital (Table 9.2). The customer renews the loan at a new fixed rate when exercising the renegotiation option. Table 9.3 shows the amortizing schedule of a loan used as an example throughout the rest of the chapter.

The present value of all annuities at the loan rate, 10%, is equal to the amount borrowed, or 1,000.

Some conventions are used for dating the cash flows and discounting. The date 0 is the date of origination of the loan. Subsequent dates refer to end of period: date 1 is the end of first year, and so on. The borrower pays the annuity immediately after each date. Hence, the outstanding debt at a date t includes the forthcoming undiscounted annuity due for the elapsed period. The value at the terminal date, 5, is equal to the last annuity due, 263.80. Table 9.4 shows the

TABLE 9.2 Characteristics of original loan

Original loan	1,000
Original maturity (years)	5
Original fixed rate	10.00%
Original annuity	263.80

TABLE 9.3 Repayment schedule of the original loan

Dates	*1*	*2*	*3*	*4*	*5*
Original annuity	−263.80	−263.80	−263.80	−263.80	−263.80
Principal repayment	163.80	180.18	198.19	218.01	239.82
Outstanding amount (end of period)	836.20	656.03	457.83	239.82	0.00

TABLE 9.4 Repayment schedule of loan

Dates	*0*	*1*	*2*	*3*	*4*	*5*
Loan	1,000					
Annuities		263.80	263.80	263.80	263.80	263.80
Principal repayment		163.80	180.18	198.19	218.01	239.82
Balance after payment		836.20	656.03	457.83	239.82	0.00

TABLE 9.5 Calculation of payoff of renegotiation

New loan	*675.71*		
New loan dates	*3*	*4*	*5*
Annuity new loan	−257.48	−257.48	−257.48
Principal repayment	210.18	224.89	240.63
Outstanding principal	465.53	240.63	0.00
Cash savings: new annuity – old annuity	6.32	6.32	6.32
V(New debt) – *V*(Old debt)	16.58		

original repayment schedule of the loan. The constant annuity is 263.80 with an original rate of 10%. The present value of all annuities at a discount rate of 10% is exactly 1,000.[5]

The borrower gains from a prepayment if the value of the cash savings, at the new rate, relative to the original loan exceeds the penalty paid. For example, if the interest rate declines to 7% in date 2, the borrower will assess the gain from prepayment. If the borrower prepays, the principal of the new loan should include the penalty for prepayment. With a 3% penalty, the principal of the new loan is $656.03 \times (1 + 3\%) = 675.71$. If its maturity is three years, the residual maturity of the original loan, the new annuity is 257.48.

The cash savings for the borrower are the differences between the old and new annuities, or $263.80 - 257.48 = 6.32$. The present value of this differential gain for the borrower, at the new rate, is 16.58, the value gained by the borrower who prepays (Table 9.5).

This value, as of the date of renegotiation, T, is the difference between the values of the old and the new debts, calculated at the new loan rate:

$$\begin{aligned}&\text{Value}\,(T,\ \text{New rate})\,\text{Original loan} - \text{Value}\,(T,\ \text{New rate})\,\text{New loan}\\&\quad = \text{Value}\,(T,\ \text{New rate})\,\text{Original loan} - \text{Outstanding balance}\,(1 + \text{penalty})\end{aligned}$$

The right-hand term states that the gain from the renegotiation is equal to the value of the original loan calculated at the new fixed rate minus the outstanding principal plus penalty. This is the payoff of a call option on the original loan with strike equal to the outstanding balance as of the prepayment date plus the penalty.

5 The value 1,000 discounts all annuities of 263.8, starting at end of period 1, so that: $1000 = \sum_{t=1,5} 263.8/(1 + i)^t$

But the value of the original loan at all intermediate dates between origination and maturity is random because the interest rate is. The gain from immediate exercise is lower than the value of the option to prepay because the option allows waiting in the hope of further declines of interest rates. The difference is the time value of the option.

9.2.3 Modeling Prepayments

The value of an option combines its liquidation value, from the payoff under immediate exercise, plus the value of waiting further for larger payoffs. The valuation process should consider all future outcomes of interest rates for finding out whether the option should be exercised and when.

The interest rate scenarios are generated for risk-free rates, since valuation is conducted in a risk-neutral world. For each scenario, the option value is obtained by discounting cash flows using the interest rates of the particular scenario. The interest rate scenarios serve for generating the payoffs of immediate exercise at each time up to the loan maturity. Prepayment might occur at any intermediate time, somewhere after origination and before maturity. With intermediate exercise allowed, the prepayment option is American.[6] At a given point in time, exercise can be optimal or not for the borrower depending on whether the immediate gain is higher or lower than expected future gains. The first step in the valuation process is to generate interest rate scenarios. The simple binomial model[7] of interest rates is used for generating scenarios of the time path of short-term rates.

9.2.4 The "Binomial" Model of Short-term Interest Rates

The horizon is broken down into small and equal intervals. Each "step" is a small interval of time between two consecutive dates, t and $t + 1$.[8] Given a value of the short-term rate at date t, there are only two possible values at $t + 1$. The rate can only move up or down by a fixed amount. The magnitudes of these movements are assigned the percentage changes, u and d, for up and down movements over each interval. If i_t is the interest rate at date t, the up and down movements are $i_{t+1} = ui_t$ and di_t. It is convenient to choose $d = 1/u$ because, with this choice, an up movement from an initial value followed by a down movement ends up with the initial value.[9] This minimizes the number of the possible values of the interest rate.

The sequence of consecutive values of the interest rate makes up a binomial tree, which looks like Figure 9.4.

6 American options can be exercised before maturity. European options can be exercised only at maturity of the contract.

7 The binomial model of short-term rates is presented in numerous texts. See, for example, Black, F., Derman, E., Toy, W. (1990), A one-factor model of interest rates and its application to Treasury bond options, [31]. A technical presentation is in: Hull, J., White, A. (1993), One-factor interest-rate models and the valuation of interest-rate derivative securities, [79].

8 The literature uses a small interval Δt, which tends towards zero.

9 The tree of possible values is said to be "recombining".

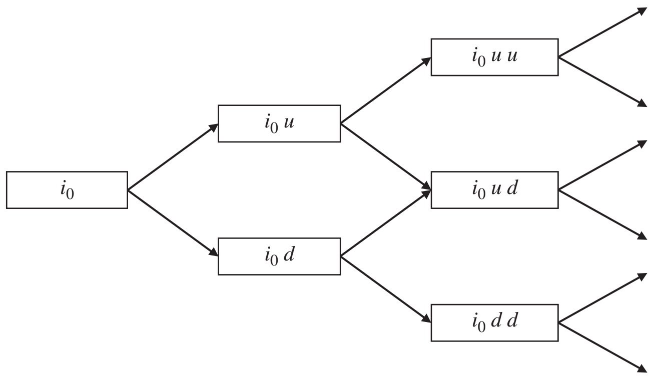


FIGURE 9.4 The binomial tree of rates

Probabilities for the up and down movements have to be defined, for example 50% for each.[10] Under this choice, the volatility of the short-term rate has to match the market volatility per unit of time, or σ. The values of u and d matching the volatility of short-term rates are derived from the formulas: $u = \exp(\sigma\sqrt{\Delta t})$ and $d = \exp(-\sigma\sqrt{\Delta t})$.[11]

In the example used in this section, the annual volatility is 20%, and the steps correspond to a one-year period.[12] The numerical values selected for u and d are:

$$u = \exp(20\%) = 1.2214 \quad \text{and} \quad d = 1/u = \exp(-20\%) = 0.8187$$

The interest rate is known at the initial step, since it is the current rate. Uncertainty appears only from the second step and further. The binomial tree of rates is shown in Table 9.6.

Once all interest rate values are found, the value of an asset depending on the interest rate can be derived from the tree.

TABLE 9.6 Binomial tree of interest rates

Date	*0*	*1*	*2*	*3*	*4*	*5*
1	10.00%	12.21%	14.92%	18.22%	22.26%	27.18%
2		8.19%	10.00%	12.21%	14.92%	18.22%
3			6.70%	8.19%	10.00%	12.21%
4				5.49%	6.70%	8.19%
5					4.49%	5.49%

10 When these probabilities are assigned, the magnitudes of the up and down movements are assigned from the volatility of the short-term rate.

11 With such up and down movements, it is easy to check that the volatility of rates at each node is equal to σ.

12 Shorter durations of each step improve accuracy.

9.2.5 Valuation of the Loan with all Interest Rate Scenarios

Consider the valuation of an asset, such as a bond, which depends on the interest rate. In a deterministic world, there would be a unique time path for interest rate, and the value would be the present value of the cash flows generated, using the sequence of interest rates. But with stochastic rates, there are numerous interest rate paths diverging from each other. Each sequence of rates is a realization of uncertain paths of rates over time. For each realization, there is a discounted value of the cash flows generated from the asset. There are as many discounted values as there are time paths of interest rates. The value of the asset is the average of these values, assuming that all time paths of rates are equally probable.[13]

It is shown in the Appendix in section 9.3, which uses the example of a simple two-period zero-coupon bond, that such a discounting process can start from the end, moving backwards to the initial step of the tree. In the general case of options with maturity extending over several periods, the unknown is the initial value of the bonds that depends on which outcomes for interest rates occur later. By contrast, at maturity, the value of an option is known. With the callable loan, the value of the call option at maturity (date 5) is zero because there is no benefit in prepaying a loan that is already maturing.

The discounting process is first used to find the value of the loan if non-callable at each node of the above tree because this is the underlying of the call option. The mark-to-market value of the original loan varies with interest rates and is the present value of all subsequent annuities, given the sequence of short-term rates.

Before proceeding, an additional technical adjustment of the tree of interest rates is required. Using an original short-term rate of 10% results in a value of the debt at date 0 higher than 1,000. The initial value is an average of values along all the time paths of interest rates but these values are not a linear function of rates. In order to find exactly the value of 1,000 when moving along the "tree" of rates, it is necessary to adjust all rates by deducting a constant drift of −0.171%[14] (Table 9.7). Instead of starting at 10% we start at 10% − 0.171% = 9.83%. The same drift applies to all rates of the tree.

TABLE 9.7 Binomial tree of rates after calibration

Dates	*0*	*1*	*2*	*3*	*4*	*5*
Drift	−0.171%	−0.171%	−0.171%	−0.171%	−0.171%	−0.171%
Calibrated rates	9.83%	12.04%	14.75%	18.05%	22.08%	27.01%
		8.02%	9.83%	12.04%	14.75%	18.05%
			6.53%	8.02%	9.83%	12.04%
				5.32%	6.53%	8.02%
					4.32%	5.32%
						3.51%

13 The maximum number of time paths is 2^n where n is the number of periods, since, at each date, there are twice as many nodes than at the preceding date.

14 The determination of this constant drift is empirical, given the initial tree of interest rates.

TABLE 9.8 Mark-to-market value of a loan(1)

Dates	*0*	*1*	*2*	*3*	*4*	*5*
Flows	0.00	263.80	263.80	263.80	263.80	263.80
	1,000.00	1,061.72	867.37	676.15	479.88	263.80
		1,134.86	920.66	709.02	493.69	263.80
			961.11	733.82	503.99	263.80
				751.89	511.42	263.80
					516.67	263.80
						263.80

(1) Values at each node include the annuity paid immediately after each date.

This adjustment is the calibration that would be required for matching the value of listed assets to the value calculated with the simulated interest rates. In order to link observed prices and calculated prices, the rates of each node of the tree need an adjustment. Calibration to observed prices is the process that guarantees that calculated prices are identical to observed prices. This is illustrated in the Appendix in section 9.4.

Using this tree of short-term rates, and going backward to initial date, the value of the loan is exactly 1,000. Table 9.8 shows the values of the non-callable loan at each node of the tree. The convention used in these calculations is that the constant annuity of 263.80 is paid immediately after each date. As a consequence, the terminal values of the loan at date 5 are all equal to this constant annuity.[15] At any intermediate date, the value of the debt discounts the future annuities along each path of rates.

The table is constructed starting from the end and moving backwards. The final values, before payment of annuity, are all equal to the annuity since there is no discounting. At the previous period 4, the upper interest rate applicable is 22.08% and the two possible final flows are 263.80 with equal probability 0.5. The value of the loan at this node is obtained from the following calculation:

$$479.88 = 263.80 + \frac{1}{2}\frac{(263.80 + 263.80)}{(1 + 22.08\%)}$$

All values at date 4 are obtained from a similar calculation. At date 3, the same discounting process is used. The upper value of the column is calculated with the 18.05% value of the interest rate as:

$$676.15 = 263.80 + \frac{1}{2}\frac{(479.88 + 493.69)}{(1 + 18.05\%)}$$

The value of the original loan at each node is obtained by taking the average of the discounted values, at the prevailing rate, of the loan at the next two nodes and adding the undiscounted annuity 263.80. The same recursive calculation provides the initial loan value

15 The value of debt at date 6 is then zero, as it should be.

1,000. The next step consists of deriving the payoffs of the callable option contingent on interest rates.

9.2.6 The Payoffs from the Prepayment Option

The immediate payoff from prepayment is the difference between the current value of the original loan and the strike price of the prepayment option at the same date. This strike price is equal to the outstanding principal of the original loan at the current date, plus the penalty for prepayment. The time profile of the exercise price is shown in Table 9.9.

The payoff under immediate exercise is the difference between the value of the original loan and the strike price at the same date. The strike is the value of the old loan, and it varies across dates according to the realization of interest rates.

A sample calculation is detailed for date 1, when the payoff is 9.77 (second column of Table 9.10). The value of the loan at the lower node of date 1, before payment of the annuity, is 1,134.86. After payment of annuity, the value of the loan is 871.06. This value is compared to the strike price at date 1, which is equal to the outstanding principal balance of the original loan plus the 3% penalty: $836.20(1 + 3\%) = 861.29$. The difference is the gain from immediate prepayment at date 2, or $871.06 - 861.29 = 9.77$. The next values along the tree are calculated using the same technique in Table 9.10.

However, a positive immediate payoff does not necessarily trigger exercise, since immediate renegotiation may be less profitable than a deferred exercise. At any date, when the option is in-the-money, a decision has to be made. Exercising the option depends upon whether later exercises are more profitable or not. Therefore, an American option[16] has two values at each

TABLE 9.9 Outstanding capital due and penalty for early repayment

Dates	*0*	*1*	*2*	*3*	*4*
Outstanding capital	1,000.00	836.20	656.03	457.83	239.82
Capital (1 + 3%)	1,030.00	861.29	675.71	471.57	247.01

TABLE 9.10 Payoff of options under immediate exercise[(1)]

Dates	*0*	*1*	*2*	*3*	*4*	*5*
	0.00	0.00	0.00	0.00	0.00	0.00
		9.77	0.00	0.00	0.00	0.00
			21.60	0.00	0.00	0.00
				16.52	0.61	0.00
					5.86	0.00
						0.00

(1) Payoff = Loan value − Principal (1 + 3%).

16 An American option can be exercised at any date until maturity.

TABLE 9.11 Expected payoffs from the prepayment option(1)

Dates	*0*	*1*	*2*	*3*	*4*	*5*
	4.61	0.06	0.00	0.00	0.00	0.00
		10.06	0.13	0.00	0.00	0.00
			21.60	0.28	0.00	0.00
				16.52	0.61	0.00
					5.86	0.00
						0.00

(1) Option value = Max (Exercise value at *t*; expected exercise values of $t+1$ discounted to *t*; 0).

node (date *t*) of the tree: a value under no exercise and the immediate exercise value. If exercise is deferred to later periods, the value is the average of the two possible values at the next date $t+1$. If exercised, the value is the immediate payoff. The sensible rule is to use the maximum of those two values: the borrower makes the optimal decision by comparing the immediate gain with the expected gains of the next period. The process starts from the gain at the final date 5, when exercise is useless and payoff is zero. Moving backward along the tree, all optimum gains from exercise are determined.

The calculation is illustrated for the lower node of date 1 in Table 9.11. The immediate payoff is 9.77 from Table 9.10. The expected value of deferred gains at the next period is the average of their present values, using the 8.02% discount rate. This is:

$$\frac{1}{2}\frac{(21.60+0.13)}{(1+8.02\%)}=10.06$$

Since the value of deferred payoffs is higher than 9.77, the optimum choice is to defer exercise.

The payoffs are calculated in Table 9.11 using the same rule from the recursive process starting from the end.[17]

These gains represent the value of the prepayment option at each point in time. Moving back to the initial date, the present value of all expected gains is the current value of the prepayment option. The value of the option is 4.61, or, for a loan of 1,000, 0.461% of the amount of the loan. The value of the loan for the bank is net of the value of the prepayment option. It is: $1,000-4.61=995.39$.

For having a sensitivity of the value of the callable loan, a constant shock is applied to all rates of all time paths, and a new value obtained. The departure from the initial value measures the sensitivity of the option-adjusted value, and is the option-adjusted sensitivity to interest rates.

9.2.7 Pricing the Option: the Option-adjusted Spread

Once the value of the option is known, it can be priced by the bank to the clients. Pricing requires converting the today value of the option into a percentage mark-up over the customers' rate that would apply to non-callable loans. The mark-up reflects the difference of values

17 The numbers 21.60 and 0.13 are the optimum deferred payoffs.

TABLE 9.12 Summary of characteristics of the loan

A. Loan without prepayment option

Debt	*Maturity*	*Rate*	*Annuity*
1,000	5	10.00%	263.80

B. Loan with prepayment option

Current rate	*Value of the option*	*Value of loan without option*	*Value of debt with option*(1)	*OAS*
10.00%	4.61	1000	995.39	0.29%

(1) The difference between the loan without options and the option value is rounded.

between a fixed-rate loan that does not allow prepayment and a callable loan, the same loan combined with the prepayment option. Viewed from the bank, the callable loan has a lower value, for the bank, than the straight loan, the difference being the option value. Viewed by the borrower, the liability has a lower value than a similar non-callable debt.

The percentage mark-up is the “option-adjusted spread” (OAS). The OAS is the spread added to the loan rate making the value of the callable loan identical to the value of a straight loan.[18] To bring the value of the straight loan in line with the value of the callable loan, it is necessary to increase the discount rates applied to the future cash flows of the callable debt. The additional spread is the OAS. For pricing the option, the OAS is the mark-up that should be included in the customer rate, if competition allows.

In the example discussed, the differential value compared to a straight non-callable loan is 4.61. The bank can price the value of the option given away to the client by adding a mark-up over customers’ rate applicable to a non-callable loan. The OAS is an additional drift applied to all rates bringing back the value of the callable loan to 1,000. For increasing the value of the loan, a negative drift is needed for decreasing all discounting rates (Table 9.12). The numerical value of the drift in the example is −0.29%.

The 0.29% drift is the OAS spread that should be added to the loan rate for pricing the option to the borrower. This value would apply if the loan runoff were contractual, but the effective runoff is shorter. The spread should therefore be adjusted upward to account for the fact that its collected only over the effective schedule. If not, the spread collected over the effective runoff will not fully compensate the bank. The methodology used to determine the OAS is similar to that of calibration. However, the goal differs. The calibration serves for linking the model to the market. The OAS serves for determining the cost of the option embedded in a loan and the required mark-up for pricing it to the borrower.

9.3 APPENDIX 1: VALUATION OF A BOND USING AN INTEREST RATE TREE

This appendix uses the example of a simple zero-coupon bond generating a single terminal flow of 100. The value of the bond discounts this cash flow over the first step at the known rate of

18 See Hayre, Lakhbir S. (1990), Understanding option-adjusted spreads and their use, [75].

TABLE 9.13 Values of the bond at various dates

Dates	*0*	*1*	*2*
	82.574	89.590	100
Values		92.073	100
			100

10%. Uncertain rates appear after step 1. The present values as of end of step 1 use the unknown rates at period 2. There are only two possible sequences of rates in this example, hence two possible values of the same asset as of date 0:

$$\frac{100}{(1+10\%)(1+11.62\%)} = 81.445$$

$$\frac{100}{(1+10\%)(1+8.61\%)} = 83.702$$

The average value is $(81.445 + 83.702)/2 = 82.574$, which is the expected value given the uncertainty on interest rates. The two-period interest rate is a geometric average of the two short-term rates, the first one being known. Table 9.13 shows the values at steps 1 and 2.

There is another equivalent technique for calculating the same value, which serves later. At date 2, the value of the final flow is 100 for all time paths of rates. However, this flow has two possible values at date 1 corresponding to the two possible values of the rate between dates 1 and 2: $100/(1 + 11.62\%) = 89.590$ and $100/(1 + 8.61\%) = 92.073$. The current value results from discounting the average value at date 1, using the current rate, which gives:

$$\frac{1}{1+10\%}\frac{89.590+92.073}{2} = 82.574$$

The result is identical to that above. In the former case, the value as of date 0 is obtained by discounting twice the final value 100 and averaging after. The second method calculates two values as of date 1 by discounting once, 89.950 and 92.073, and averages them into a single value at 1, or $(89.590 + 92.073)/2 = 90.8315$, and then discounts this single averaged value to present $(90.8315/1 + 10\%) = 82.574$.

Mathematically, the formula is identical. But the second method calculates the current value using a recursive process starting from the end, and moving backward to date 0. The process can be generalized. Starting from final values, values can be derived at the immediately preceding date and so on until the current date.

The only difference with options is that the discounted flows change with rates, since they become the cash payoffs of the option instead of being independent of interest rate scenarios.

9.4 APPENDIX 2: CALIBRATION OF THE BINOMIAL TREE

The value of listed assets, calculated with all the time paths of interest rates, should replicate those observed on the market. The above binomial tree does not as yet include all available information. The volatility sets the upward and the downward moves at each step. This adjustment does not capture the trend of rates, any liquidity premium or credit spread observed in markets. All factors should show up in simulated rates for mimicking the actual prices. The calibration should capture all missing factors in the model.

In the above example, the value of the zero coupon is 82.574. Its actual price might diverge from this. For instance, let us assume that its observed price is 81.5. The simulated rates need an adjustment to reconcile the two values. In this example, the calculated value is too high, and the simulated rates should be higher, so that the calculated price decreases until it reaches the observed price. The volatility should not be changed, since the up and down moves already replicate the market volatility. In the case of the zero coupon maturing at date 2, the initial value of the interest rate is 10% and is fixed. Hence, only rates as of date 1 need an adjustment.

The easiest way to change them is to add a common constant to both rates at date 1. The constant is empirical. Its value is such that it makes identical the calculated and the observed prices. This constant is the "drift". The numerical calculation shows that the required drift to obtain a value of 81.5 is around 1.40%. The calibrated rates at 1 become: $i_{1u} = 11.62\% + 1.40\% = 13.02\%$ and $i_{1d} = 8.61\% + 1.40\% = 10.01\%$.

Once the rates at date 1 are adjusted, the calibration extends to subsequent periods using the prices of listed assets having longer maturity. It is necessary to repeat this process for extending the adjustment of the whole tree. This technique enforces the external consistency with market data.

10

Funds Transfer Pricing Systems

All business units of a financial institution share a common resource: liquidity. The primary functions of funds transfer pricing systems (FTP) are to exchange funds between business units and to determine the profit and loss (P&L) of business lines. But transfer systems have several other critical roles. They serve to ensure that interest rate risk and liquidity risk are withdrawn from commercial lines and transferred to the central books of the asset-liability management (ALM) unit, and they are also used as benchmarks for defining the proper pricing of loans to clients.

This chapter discusses the design of internal prices and explains the mechanism of P&L allocation across business lines, the transfer mechanism of the liquidity and interest rate risks to the ALM portfolio, and the definition of appropriate economic benchmarks for internal prices. It also discusses how risk can be priced to borrowers by adding the cost of credit risk to these internal benchmarks.

Contents

10.1 INTERNAL FUND PRICING SYSTEMS

A major purpose of internal prices is to determine the P&L of the business lines. Transfer prices are internal prices of funds charged to business units or compensating cheap resources such as deposits.

Transfer pricing systems are notably designed for the banking book, for compensating resources collected from depositors and for charging funds used for lending. Internal prices also serve for exchanging funds between units with deficits of funds and units with excesses of funds. As they are used for calculating the P&L of a business line, they perform income allocations across business lines.

Internal prices apply to all business lines using funding liquidity, such as the trading portfolio. The starting point for defining the internal prices of funds is the banks' cost of funds. This cost depends on the mix of financing resources, market funding or resources collected from clients. The relevant financing mix depends on which book is considered. The banking book includes the resources from depositors, which are often cheaper than market debts (bonds or wholesale debt). The trading book often uses dedicated financing sources, for example a financing secured by the assets held, of which cost is lower than unsecured resources.

Liquidity was abundant before the 2008 crisis, and the cost of liquidity was considered negligible. After the liquidity crisis, it became necessary to fully recognize the cost of liquidity and to include it in the internal prices. If the credit spread of the bank were independent of maturity, the cost of liquidity could be measured by the incremental cost of funds over the reference rates of banks as maturity gets longer. The benchmark rates are generally the Libor rates for the short-term and swap rates beyond one year.[1]

The banks' own cost of funds includes the credit spread, which is the compensation for banks' own credit risk, and the liquidity premium, which is the pure cost of liquidity. For measuring the liquidity premium, the premium should be isolated from the spread for credit risk within the overall spread applicable to the banks' cost of funds.

Accordingly, there are various proxy measures of the cost of funding liquidity. One measure is the difference between the funded interest rate and the unfunded interest rate, or, respectively, the bond rate and the swap fixed rate. However, the bond rate includes the credit spread for banks' own risk. Hence, the cost of liquidity can, in theory, be better measured by the difference between the bond spread and the credit default swap (CDS) premium on the same issuer (see Chapter 29). A CDS is a traded instrument providing an insurance against the credit risk of a reference debt. The buyer of a CDS is the buyer of the protection who pays a premium that compensates for the credit risk of the reference debt. Therefore, the CDS premium includes only an issuer's credit spread, whether an issuer's bond yield includes both the spread for credit risk and the liquidity premium. The difference is the "basis" and is a measure of the cost of liquidity, as the credit spread components of both instruments should theoretically cancel out.[2] The incremental cost of funds as maturity extends is another measure of the liquidity premium. Banks must select how they assess their cost of liquidity from these alternate measures.

The cost of liquidity, or liquidity premium, should be included in internal prices charged to users of funds for determining the fair P&L of a business line. If not, the bank would absorb the

1 The overnight indexed swap rate (OIS) tends to be used as a reference, instead of Libor-based swaps, as it is closer to a risk-free rate. An OIS is an interest rate swap where the periodic floating payment is based on a daily return. The reference for a daily compounded rate is an overnight rate, such as the rate for overnight unsecured lending between banks, for example the federal funds rate for US dollars. The fixed rate of OIS is considered less risky than the interbank rate (Libor) because credit risk is limited.

2 Measures of the basis include the asset-swap spread and the z-spread. The asset-swap spread is the drift added on top of a bond yield for its value to equal its observed price. The z-spread is the additional drift added on top of the zero-coupon rates for a bond value to equal its market price. The basis is the difference between the asset-swap spread, or the z-spread, and the CDS premium.

additional cost instead of the business line. The relevant references differ for the banking book and the trading book, as the financing mix differs for these books.

The fair internal price for financing a trading portfolio depends on its effective financing. The cost of funds is lower for secured financing, such as repurchase agreements. A repurchase agreement consists of lending a security in exchange for cash, with the commitment to buy back the security at a predetermined price with cash. This is economically equivalent to a debt pledged by the security exchanged. If a bank assigns properly resources to subportfolios of traded assets, such dedicated costs of funds should be used as internal prices for determining the P&L of the portfolio.

The financing of a traded portfolio depends on the assets. For derivatives, the cost of funds should include a funding value adjustment, or FVA. FVA appears when collateral is pledged for derivative transactions by one party, and when the other party, for example a corporate, does not pledge collateral. The collateral posted by the bank requires a financing, of which cost is usually higher than interest earned on collateral. The negative difference is a differential cost, which should be added on top of the cost of financing the derivative portfolio, and is the FVA. As the maturity of derivatives is often long term, the relevant cost of funds should include the full cost of funding, inclusive of credit spread, liquidity premium and FVA.

For the banking book, the financing mix includes wholesale debt, issued bonds and deposits. The banking book can be seen as a portfolio long on loans and short on deposits from clients. For loans, the liquidity premium is a cost that should be charged to loans. Resources from depositors have a lower interest rate than market rates with similar maturity, and they carry a negative spread relative to market rates, which should serve as a basis for compensating internally these resources. Similarly, it could be that the cost of liquidity from deposits be lower than the market cost of liquidity, and the difference should be allocated to these resources.

The rest of this chapter is dedicated to the banking book, as the exchanges of funds and the corresponding internal prices should have a number of desirable properties. They should allow the calculation of a fair P&L for business lines and they should transfer the liquidity and the interest rate risk to the ALM books.

10.2 FUNDS TRANSFER PRICING IN THE BANKING BOOK

All transfer systems provide an internal allocation of funds, so that deficits and excesses of funds of branches or business lines are netted through a central pool. The central unit finances any deficit, or invests any excess, in the capital markets. But different designs can fulfill this function.

10.2.1 Transfers of Funds

Commercial activities generate uses and resources of cash, which are generally not balanced for each business unit. Cash netting is the simplest transfer system (Figure 10.1). Cash is allocated to units having liquidity deficits, or excesses are purchased where they appear. Net cash transfers compensate excesses and deficits of funds across business units. The central Treasury is the unit that, in the end, raises debt or invests excesses of funds in the market for compensating any aggregated net excess of deficit of cash. In a simple cash netting system, transfer prices apply only to net balances. The system records excesses and deficits and nets

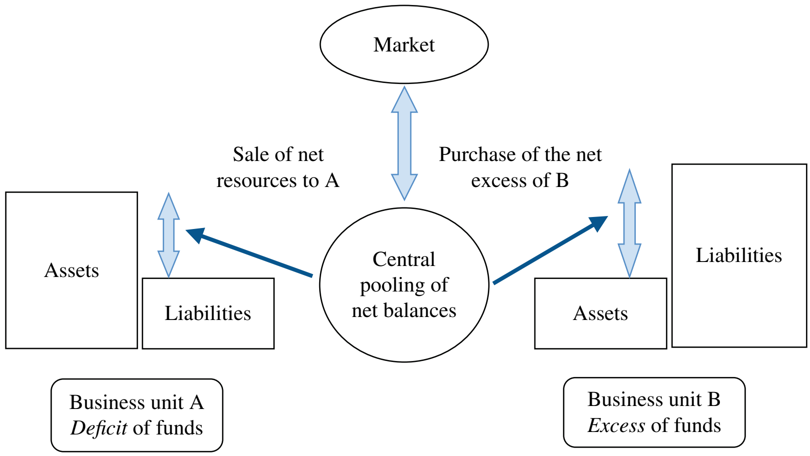


FIGURE 10.1 Transfers of net balances only

them. Some systems use several pools of funds, for instance by grouping them according to maturity.

Since netting occurs before any transfer to the central pool, transfer prices apply only to netted balances. A loan is not exchanged with the central pool unless it is allocated to the net cash balance transiting through the pool. A consequence is that the funds used for other loans would have an indeterminate cost. The same would apply to deposits collected from customers. An excess balance of deposits over uses of funds could be purchased internally by the central pool, but deposits not exchanged would have indeterminate revenue. Cash netting systems do not allow the allocation of earnings, which is a major drawback.

The alternate solution is to exchange the total balances of funds used and of funds collected, without prior local netting of assets and liabilities, as shown in Figure 10.2. The funds are exchanged through a central pool, under the ALM control. Internal exchanges of funds require

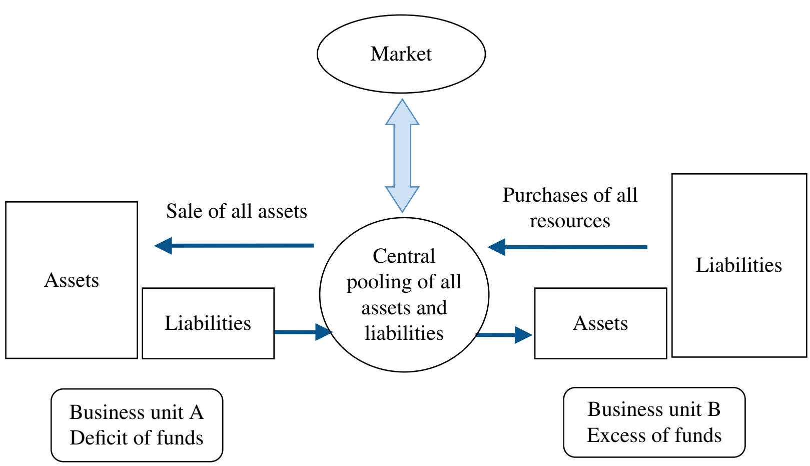


FIGURE 10.2 Central cash pool of all assets and liabilities

a pricing system. The central pool buys funds collected from clients and charges funds used for lending and investing. The earnings of each unit derive from the spread between these internal prices and customers' prices. Such exchanges are analytical since only cash balances need to be physically netted. The whole balances of assets and liabilities transit through the central pool.

This scheme creates an internal market for funds with internal prices. The ALM invests any aggregated excess of cash, or finances any global deficit of cash, in the capital markets. The internal prices are applied to the entire balances of assets and liabilities of each business unit. In such a system, each unit, and each individual transaction, has a net interest income, which is the difference between internal transfer prices and customers' prices. The system effectively allocates earnings to business units and to each individual transaction. What follows refers to this system of analytical transfers.

A transfer pricing system can also be used as a commercial tool since internal prices can serve for differentiating incentives to selected activities, by applying mark-ups or mark-downs over internal benchmarks. But the internal references should be fair and based on sound economic benchmarks in order to avoid conflicts.

10.2.2 The Allocation of Net Interest Income

The spread between customer prices and the internal prices measures the earnings of commercial activities. The ALM unit also makes a profit or a loss from internal exchanges and given the market rates for borrowing and lending in the capital markets. In a comprehensive system, the ALM portfolio is the mirror image of the banking portfolio since the ALM unit buys all liabilities and sells all assets. The sum of the earnings generated by the business units and those generated by the ALM portfolio should equal the actual earnings of the bank, since internal analytical exchanges cancel out.

The reconciliation of analytical earnings of business units and ALM with the accounting earnings of the bank is illustrated with a simple example of a commercial bank holding loans and collecting deposits. For the bank, the accounting earnings sum up revenues from lending, deduct the cost of deposits and add up any revenue or cost from lending and borrowing in the market. For lending, the revenues result from the spread between customer prices and the cost of the internal purchases of resources at transfer prices. For resources collected from depositors, the costs are the interest paid to customers (depositors) and the revenues from selling these resources internally at transfer prices. For the ALM unit, the revenues arise from charging to the loans the cost of funds, and the costs arise from the purchase of resources from customers' deposits. In addition, the ALM has financial costs, or revenues, from borrowing, or investing, the net aggregated cash balances of commercial lines.

The calculations are illustrated with a simplified balance sheet (Table 10.1). In the example, the average customer price for borrowers is 12% and the average customer rate paid to depositors is 6%. The ALM borrows in the market at the current market rate, 9%. The commercial balance sheet generates a deficit of cash, which is financed by the ALM unit. A unique transfer price, 9.20%, applies to funds charged to business units and to resources purchased from business units. This transfer price differs from the market rate.

The direct calculation of the accounting earnings is straightforward since it depends only on customers' rates and the funding cost by the ALM:

$$2,000 \times 12\% - 1,200 \times 6\% - 800 \times 9\% = 96$$

TABLE 10.1 Bank's balance sheet

	Volume	*Rate*
Assets		
Loans	2,000	12.00%
Total	2,000	
Resources		
Deposits	1,200	6.00%
Funding	800	9.00%
Total	2,000	

The other calculations show the earnings from the commercial activities, broken down into lending and collecting resources (Table 10.2). The ALM finances externally the bank's deficit. The allocated net interest income (NII) of each commercial unit results from the customer rates and the transfer price. The commercial margin is $12\% - 9.20\% = 2.80\%$ for loans, and $6\% - 9.20\% = -2.80\%$ for deposits. The negative sign for deposits simply means that the bank pays less to depositors than what ALM pays for the resources.

The loans generate a margin of 56.0, and the deposits generate 38.4. The total commercial NII is 94.4, which is lower than the bank accounting NII of 96. This commercial interest margin adds up the contributions of the lending activity and of collecting resources.

For reconciling the NII of business units with the NII of the bank, the ALM unit should also have an income statement. Revenues are collected from lending to business lines, and costs reflect the internal purchases of resources from business lines. The ALM has revenues and costs symmetrical to those of business units, plus the cost of funding the deficit of 800 in the market. Its NII is 1.6, exactly the difference between the commercial earnings and the total bank's earnings. The sum of the aggregated commercial margins and of the ALM NII is equal to the

TABLE 10.2 Calculation of NII

Market rate	9.00%	
Transfer price	9.20%	
Margin	*Calculation*	*Value*
Direct calculation of NII		
Accounting NII	2,000 × 12% − 1,200 × 6.00% − 800 × 9.00%	96.0
Commercial NII		
Loans	2,000 × (12% − 9.20%)	56.0
Deposits	1,200 × (9.20% − 6.00%)	38.4
Total commercial NII	2,000 × (12% − 9.20%) + 1,200 × (9.20% − 6.00%)	94.4
ALM NII		
ALM NII	2,000 × 9.20% − 1,200 × 9.20% − 800 × 9.00%	1.6
Bank NII	Commercial NII + ALM NII	96.0

bank's NII. The ALM profit and loss statement allows reconciling the sum of analytical profits of commercial activities and of ALM activities with the net income of the bank.

In this example, the ALM net income is positive because it overcharges commercial units with a transfer price of 9.20%, higher than the market rate (9%). If the transfer price changes, the allocation of the bank's NII changes, but the overall earnings of the bank remains identical. The ALM NII is zero when the internal price is identical to the market rate and the bank's NII becomes equal to the commercial NII. This is the reference case. The ALM is neutral to the extent that it generates neither a profit nor a loss. The entire bank's NII is "within" the commercial units and the allocated earnings of commercial activities sum up to the bank's profit.[3]

10.3 ECONOMIC TRANSFER PRICES IN THE BANKING BOOK

Transfer, or internal, prices serve for allocating funds and earnings across business lines. They should be fair, and avoid any conflict generated by discrepancies between financing conditions inside the bank and in the market. For the banking book, they are also used as benchmarks for defining a proper pricing to clients, such that the bank earns its target profit.

10.3.1 Using Market Rates as Benchmarks

Fair transfer prices should be based on market rates in order to eliminate discrepancies with market rates. Distortions from market references create subsidies or penalties that would be considered unfair. On the lending side, charging a cost of funds lower than market rates could end up in customers' prices lower than market rates, and would not compensate the cost of funds if loans are financed at market rates. Charging a higher cost than market rates would also be unfair.

For resources collected as deposits, a fair compensation should be in line with what the business units would earn if they invested outside. If not, collecting cheap resources from clients, at rates are lower than market rates, would not be properly compensated. This would generate disincentives for collecting cheap resources and destroy value.

Some alternate choices to market rates could be considered, but they imply conventions and generate economic inconsistencies. For example, the actual cost of financing could be used. Or systems pooling uses and resources of funds by maturity buckets could be used to properly assign the actual cost of resources to the matching uses of funds. But such views have conceptual flows. There are many sources of financing for banks of which costs differ: demand deposits, term deposits, wholesale debt or financial debt issued in the capital market. Allocating specific sources of financing to lending activities would rely on conventions. If the cost of resources backing loans is subsidized by low-cost deposits, the collection of these cheap resources would not be compensated and the system would transfer the low cost of deposits to the margin allocated to loans. Lending activities would be subsidized by the deposit base and might charge customers' rates lower than market rates. Moreover, such allocation assumes than

3 This is easily checked by replicating the above calculations of allocated earnings when setting the transfer price to 9%.

new deposits match each new dollar of loans, which is unrealistic. Given such inconsistencies and unrealistic assumptions, matching assets with existing resources is not economic.

Such a "horizontal" view of the balance sheet, where commercial loans are backed by designated resources, is inappropriate. The commercial profit should be generated by the whole banking book, with both loans and deposits. The banking book is then financed on the market if it generates a liquidity deficit, or excess funds are invested in the market. In the first case, the relevant financing mix is made of the capital base and the banks' financial debt, and, in the second case, the relevant return is a market return for excess funds.

10.3.2 Transfer Prices for Loans

The relevant financial debt backing the loans is the debt that eliminates liquidity and interest rate risk. The transfer price for loans is the market cost of this debt. For a loan, the funding should replicate exactly the amortization profile with a rate, fixed or floating, matching the nature of the loan interest rate. Such financing is "notional" rather than material, because it serves as a reference without being actually contracted. Its cost is independent of the existing resources, as it should be.

In some cases, the replication is simple and the cost of funds is easily defined. A bullet debt matches a bullet loan. The relevant cost is the market rate corresponding to the maturity of the transaction. For an amortizing loan, the outstanding balance varies over time until maturity. Using the market rate of the same maturity would mean that we use a bullet debt for an amortizing loan, which is not a replication. The funding that replicates the time profile of the loan should be the mirror image of the loan: it is an amortizing debt with exactly the same repayment schedule.

The transfer price is the cost of this debt, which is the cost of a combination of debts of various maturities. In the example in Figure 10.3, a fixed-rate loan of 100 amortizes in two years, the amortizations of capital being 40 and 60. The replicating debt combines two spot bullet debts, of maturities 1 and 2 years, contracted at the market rates,[4] as shown in Figure 10.3. With a floating-

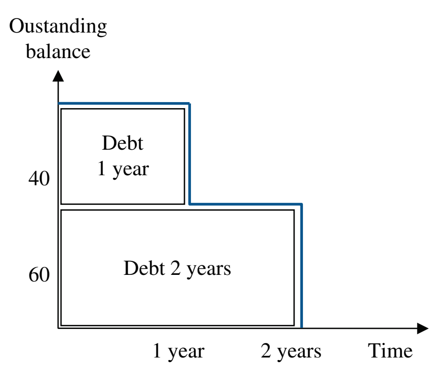


FIGURE 10.3 A two-year amortizing loan

4 There is another solution for backing the loan. For instance, a spot debt for one year could be contracted for the full amount of the loan that is 100 for one year at the spot market rate, followed by another one-year debt, for an amount of 60, starting one year from now. Nevertheless, such funding does not replicate the loan, since a fraction of debt needs renewal for another year at a rate unknown today.

rate loan, the replication would be set up with floating-rate references, including any liquidity premium for long maturities.

The cost of the replicating debt is its yield. For a fixed-rate loan, the mirror debt combines several zero-coupon debts, with volumes and maturities matching the time profile of the amortizing debt. The yield is a function of zero-coupon rates. In the example in Figure 10.3, there are two layers of debt: 60 for 2 years, and 40 for 1 year. The relevant rates are the market rates for these maturities. The transfer price is the average cost of funds of the two debts: it is the discount rate making the present value of the future flows generated by the two debts equal to the amount borrowed.[5]

The future outflows are the capital repayments and interests. The interest flows cumulate those of the one-year and the two-year market rates. If the debts are zero coupon, the interest payments are at maturity. The flows are 40 $(1 + r_1)$ in one year and 60 $(1 + r_2)^2$ in two years. The yield, y, of this composite funding is a discount rate such that:

$$100 = 40(1 + r_1)/(1 + y) + 60(1 + r_2)^2/(1 + y)^2$$

The discount rate is somewhere between the two market rates. The cost of the debt is approximately the weighted average of the market rates for one and two years, using weights combining the size of each debt and its maturity.[6] With rates r_1 and r_2 equal to 8% and 9%, y is approximately 8.75%. In practice, tables of transfer prices provide the yield of the mirroring debt, which is a composite of market rates, given the time profile of loans and the current market rates.

Using the replicating debt has numerous economic benefits. The margin of the loan is immune to interest rate movements. There is no need of conventions for assigning existing resources to usages of funds. There is no transfer of income generated from collecting resources to the income of lending activities.

A replicating debt can be defined for loans with a determinate maturity schedule. However, contractual maturities are not effective maturities, due to prepayments of mortgage loans, for example. Hence, the behavioral, or effective, maturity schedule of loans should be used.

The debt mirroring the loan should also embed the same option as the loan. For a fixed-rate mortgage, the debt should be callable, as the loan is. For a floating-rate loan with a cap, the debt should embed the same cap. The additional cost of the option would be charged to the loan, on top of the cost of a straight debt with no option. This cost is evaluated as the option-adjusted spread of the loan in Chapter 9.

The ALM unit does not need to adopt the funding policy that fully protects the interest margin of the bank. The replicating debt is notional. In some cases, the bank has enough stable funds from deposits and no debt needs to be raised. In other cases, the ALM might wish to maintain an open position on interest rate risk. The cost of the notional debt is used independently from the actual financing policy.

10.3.3 Transfer Prices for Resources

For deposits, applying the same principle necessitates defining a "mirror" asset replicating the maturity schedule of deposits. If such a maturity schedule can be defined, the fair compensation

5 This is the yield to maturity of a debt.

6 An approximate solution uses a linear approximation of the exact formula: $100 = 40(1 + r_1 - y) + 60(1 + 2r_2 - 2y)$ and the yield is: $y = (40r_1 + 60 \times 2 \times r_2)/(40 + 2 \times 60)$.

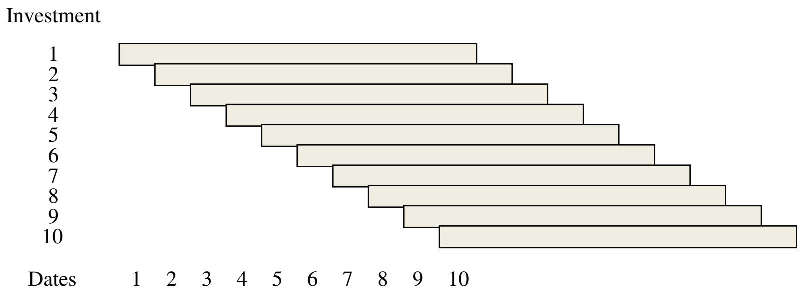


FIGURE 10.4 Notional investment portfolio

of deposits is the return that the bank could make by investing the resources at risk-free rates. This is a fair compensation for collecting resources, which allocates properly the profit created from having such cheap resources. The same investment rates can also be viewed as the return that the bank would have if it invested in risk-free assets instead of lending. The relevant market rates should be those of investments matching the time profile of deposits.

But deposits are non-maturing accounts, with no defined maturity. They are usually broken down into volatile resources and stable resources. The short-term fraction could earn the short-term market rate, which serves as a reference. For the stable fraction, defining a replicating asset requires rules.

Let us assume that the bank is comfortable with a 10-year maturity for stable deposits. It would make sense to use the 10-year rate as the relevant internal rate. But it does not make sense to use the 10-year rate of a single date, as if deposits were entirely invested at this hypothetical date. Rather, the bank should use some average of 10-year rates in order to smooth out the variations of the rate. If all investments have a 10-year maturity, the portfolio return would average the historical 10-year yields.

It is generally considered desirable to smooth the changes of the shape of the yield curve in the long term, and a "ladder" approach is more appropriate. It consists of spreading the investment over several maturities. Each investment is then rolled over when it matures, with the same maturity (Figure 10.4). The resulting yield is a combination of historical market returns for each maturity. With the horizon of 10 years, 10 such layers, invested from one to 10 years, are used. Each year, the maturing layer is reinvested at the prevailing rate. The smoothing effect now extends to the fluctuations of the entire yield curve. If an amortization rule is used for deposits, the approach would replicate the time profile of deposits.

By using several maturities, for example from one to 10 years, the replicating portfolio earns an average of all rates over all 10 years. This policy averages the time series of annual term structures of interest rates up to the management horizon.

10.3.4 Transferring Liquidity and Interest Rate Risks to the ALM Books

The principle of replication removes both interest rate and liquidity risk from loans, by transferring them to ALM. However, for the risk transfer to be effective, transfer prices should not be based on current market rates, but should be historical. For instance, if a fixed-rate asset generates 11% over a transfer price of 9%, the margin is 2%. Recalculating the margin of the same asset with a new rate at a later period would expose the margin to interest rate risk. For example, a reset of the transfer price after origination at 10% would decrease the margin by 1%. For commercial NII to remain interest rate insensitive, the historical transfer price needs to remain attached to the transaction over its life. Using such historical prices for each transaction puts an additional burden on the information system, but this last addition makes the transfer price system comprehensive and consistent.

10.3.5 Economic and Commercial Transfer Prices

Transfer prices are a very effective tool for implementing the commercial policy, as they serve as commercial signals for subsidizing or penalizing the various products offered to clients. Obviously such commercial prices do not match the economic prices, except by chance.

In such a case, using two sets of internal prices makes sense for giving some flexibility to the commercial policy. Any discrepancy between the economic and commercial prices is the cost of the commercial policy. These discrepancies are recorded as mark-ups (penalties) or mark-downs (subsidies) decided by the commercial department. This scheme reconciles the multiple functions of the transfer prices and makes explicit the cost of enforcing commercial policies that are not in line with economic references.

10.4 RISK-BASED PRICING: LENDING

Risk-based pricing consists of charging to the borrower the additional cost of the capital charge for the credit risk. The pricing should be consistent with the target return on capital (ROC) of the bank.

Revenues include interest income and fees. When defining customers' rates, in percentage of loan, an adjustment has to be made for fees. Non-recurring fees create a distortion of revenues across time. Upfront fees increase the return in the first period only. A common practice is to use an annualized all-in-revenue, which averages all revenues, recurring and non-recurring, over the life of the transaction. The all-in-spread is the annualized spread above the transfer price. Since pricing to client is before tax, it is convenient to refer to a target ROC before tax.

The cost of credit risk includes two components: the statistical loss, or average loss due to defaults ("expected loss"), and the cost of losses in excess of average loss. The expected loss is a percentage of the balance of loan, which depends on the borrower's default probability, the recovery rate and the expected loan size. The losses in excess of expected loss are measured by capital, either regulatory (Basel 2) or economic capital from models.[7]

The cost of debt backing the loan, i, is for example 5%, and the cost of equity capital before tax is: $k = 20\%$. The risk premium for credit risk is the incremental cost of substituting capital (K) to debt (D) times the capital charge (Figure 10.5):

$$K(k - i) = K \times (20\% - 5\%)$$

7 The economic capital is a credit value-at-risk, modeled as explained in Chapter 18 on credit portfolio models.

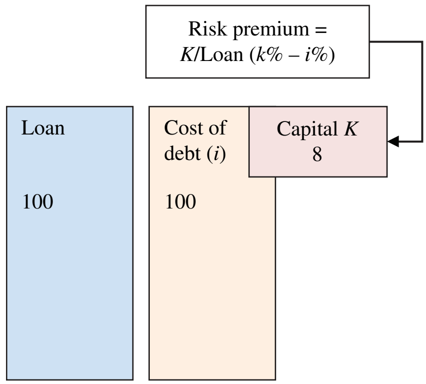


FIGURE 10.5 Risk premium for credit risk

If regulatory capital is 8% of loan 1,000, or 80, the risk premium is:

$$80(20\% - 5\%) = 7.2$$

This is also, in percentage of the loan size: $8\%(20\% - 5\%) = 1.20\%$. The risk premium is the excess of cost of equity over the cost of debt of the bank, weighted by the capital ratio.

The cost of debt, i, could be either the cost of bank's own debt or the transfer price based on the benchmark rates of the bank. Using the bank's benchmark rates means that a single credit spread is charged, that of the borrower. Pricing a single credit spread would be in line with what the market would charge to the borrower. However, this rule does not ensure that the bank meets the target profit, as the earning are calculated over the true cost of debt. Charging this cost of bank's own debt does ensure a mark-up in line with the bank's target profit, but this choice implies that the bank charges two credit spreads to the client: its own spread and the borrower's credit spread.

Under this view, the loan is backed by a mix of equity and debt, which reflects regulatory risk weights. Instead of using this regulatory mix, the loan could be viewed as backed entirely by debt because the capital should not be exposed to risk, hence invested instead risk free. In this alternative view (Figure 10.6), the exposure can be considered as entirely financed with the

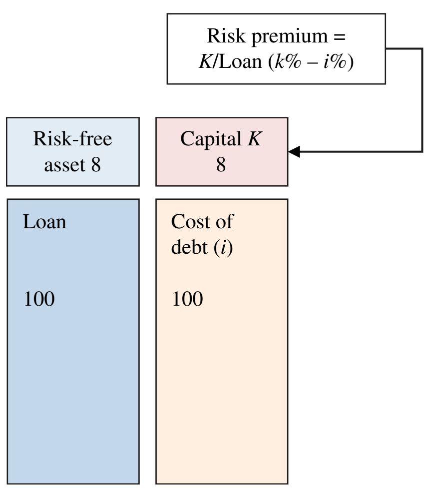


FIGURE 10.6 Financing if capital is invested risk-free

TABLE 10.3 Components of risk-based prices

	%
Cost of funding: transfer price	4.60
Liquidity premium	0.40
Cost of funds	5.00
Operating costs	1.00
Expected credit losses	0.50
= All-in cost in % loan balance	6.50
*Risk premium: 8% × (**20%** – 5%)*	*1.20*
Customer rate, theoretical	7.70

bank's debt and the risk-free revenue from capital allocated is added to earnings. It is easy to show that the risk premium is then calculated over the risk-free rate rather than the cost of the bank's debt.

For deriving the customer's rate, the mark-up over transfer price should add up several items:

- The transfer price, exclusive of a liquidity premium: 4.60%.
- The liquidity premium: 0.40%, as derived from the above measures.
- The expected loss for credit risk: 1% of the size of the loan.
- An option-adjusted spread, when required.
- The operating cost: 0.5% of the loan size.
- The risk premium is the incremental cost of capital over the debt financing, calculated as above: 1.20%.

The target customer rate is 7.70%, and the target spread over the transfer price is, in percentage of the loan, 7.70% – 5% = 2.70% (Table 10.3).

It is easy to check that the return on equity capital is effectively 20% before tax using, for example, a loan size of 1,000 (Table 10.4). The capital base is 80 and the debt is

TABLE 10.4 Top-down calculation of the ROC

Loan	1,000
Capital	80
Debt	920
Cost of funding: 5% × 920	46
Operating costs: 1% × 1000	10
Expected credit losses: 0.50% × 1,000	5
= All-in cost in €	61
Loan interest revenue	77
Earnings before tax EBT	16
ROC: EBT/capital	20%

$1,000 - 80 = 920$, to which the interest rate is applied. The other items follow by applying above percentages to the loan value 1,000. The earnings before tax are 10, a value that matches the 20% return over the capital base 40.

If the loan were risk-free, there would be no capital charge and the statistical loss from default would also be zero. The full cost of the loan would collapse to the sum of financing costs and operating cost.

The calculation of the profit mark-up at the level of individual transaction might be irrelevant, because the bank sells multiple products to the same client. In such a case, the client is a better base for assessing profitability than standalone transactions. This is a strong argument for looking at profitability at the client's level rather than at the transaction level.

11

Returns, Random Shocks and Value-at-Risk

Value-at-risk (VaR) is the magnitude of a potential loss over a given horizon and for a given portfolio of financial assets, with a stated probability. VaR requires modeling the distribution of the random asset values as time passes. These random shocks are measured by variations of the level of prices or by variations relative to initial values, or returns. In the standard market model, a future asset value is derived by compounding a sequence of random returns. These returns are modeled as stochastic processes describing how they behave as time passes.

This chapter explains the value-at-risk concept and how it can be modeled from random market shocks. It covers the standard model of asset returns from which the distribution of market-driven asset values derive.

Contents

11.1 VALUE-AT-RISK

Value-at-risk is a potential loss not exceeded with a given probability, defined as the confidence level, over a given period of time and for a given portfolio of financial assets. For market risk, the VaR measures the magnitude of an adverse market shock on a position, or on a portfolio of positions, over a given horizon. The horizon for measuring gains or losses is the trading horizon required to liquidate the portfolio. The future asset value at the end of the liquidation period is unknown. The algebraic difference between this final value and the today value is the gain or loss over the period. It is equal to the difference, $V_t - V_0$, where V_t is the value at horizon t and V_0 is the known today value. Such algebraic gain/loss is the profit and loss (P&L) of the position.

The VaR is defined as a quantile of the distribution of the variations of value. A quantile is a threshold value of the random variable such that the probability of observing lower values is given. A quantile of a distribution is visualized in Figure 11.1. A quantile, and a VaR, can be defined for any distribution, normal or not normal.

In the figure, X is the random variation of value, $V_t - V_0$, of a portfolio over a given horizon. The α-quantile of the distribution, X_α, is the value of the random variable such that the probability that an occurrence of the variable is lower than this value is: $P(X \leq X_\alpha) = \alpha$. With these notations, the quantile is also the value of the random variable that can be exceeded with probability $1 - \alpha$: $P(X > X_\alpha) = 1 - \alpha$. This probability that X be lower than this lower bound is the area under the curve representing the distribution of the random variable. Typical values of the confidence level α are small, such as 1%. With small probabilities, the cutoff point is on the left-hand side of the distribution.

For market risk, the variable X is the algebraic P&L of an asset or a portfolio. The left-hand tail of the distribution measures losses, L, as negative variations of value. VaR is the expected worst-case loss, L_α. If the loss measures the absolute variation of value, the quantile L_α is the upper bound of loss not exceeded in more than a fraction α of all outcomes. The threshold L_α is such that the negative variations of asset value below $-L_\alpha$ occur with a probability α: $P(V_t - V_0 < -L_\alpha) = \alpha$.

For a normal distribution, the quantiles for various confidence levels are known as multiples of the standard deviation of the distribution. They are represented in Figure 11.2. The normal distribution is often used because it is defined by only two parameters, its mean and its standard

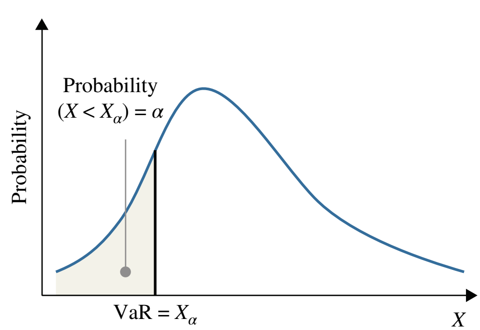


FIGURE 11.1 Probability distribution and quantile

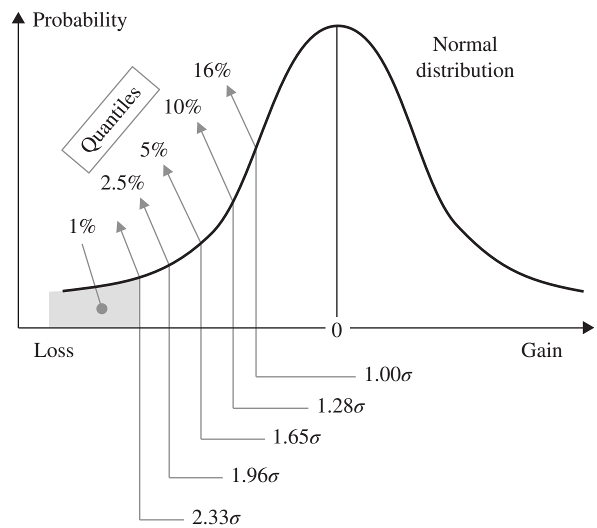


FIGURE 11.2 Normal distribution and one-tailed confidence levels

deviation, or volatility.[1] For each value of α, the corresponding deviation from the mean is expressed as a multiple of the standard deviation. For example, deviations on the left-hand side larger than 2.33 standard deviations do not occur in more than 1% of all possible outcomes. This is no more than two to three days in a year, with around 250 trading days in a year. This 1% confidence level is used in the current regulations for measuring VaR.

The normal distribution is an approximation for market shocks. Actual distributions have distributions that are skewed and with fat tails:[2] they show more frequent losses than gains, unlike the normal curve which is symmetric. Large losses are also more frequent than the normal curve, and actual distributions have fatter tails. Nevertheless, the normal distribution is an acceptable proxy of the distribution of market shocks and is a useful pedagogical tool.

11.2 RANDOM SHOCKS AS ASSET RETURNS

For risk purposes, the random fluctuations of the price of assets from their initial value are relevant. These variations can be measured as variations of the level of prices or as variations relative to initial values. The relative variations are returns. There are two different measures of returns. The discrete returns, or arithmetic returns, $R_{0,t}$ between 0 and t, are percentages of initial value:

$$R_{0,t} = (V_t - V_0)/V_0 \quad \text{and} \quad V_t = V_0(1 + R_{0,t})$$

1 The standard deviation is the square root of variance. Variance is the probability-weighted average of the squared deviations from the mean. The standard deviation measures the dispersion of a random variable around its mean, and is also called volatility.

2 A skewed distribution is not symmetric. Fat tails refer to probabilities of extreme deviations higher than with the normal distribution.

The logarithmic returns, or log-returns $r_{0,t}$, are defined by:

$$r_{0,t} = \ln(V_t/V_0) \quad \text{or} \quad V_t = \exp(r_{0,t})V_0$$

The logarithmic return is a limit of discrete returns when the time interval between compounding dates tends towards zero and the number of compounding periods tends towards infinity (Appendix 1). The two returns have close numerical values for small intervals of time.[3] If the annual returns, discrete and continuous, are, respectively, R and r, for a given discrete annual rate, $R = 10\%$, there is a continuous return r such that the terminal values under discrete and continuous compounding are identical. Over one year, the final value of an asset with initial value 1 is simply: $(1 + R) = 1 + 10\%$. The equivalent log-return is obtained by writing that a continuous compounded value $\exp(r)$ is equal to 1.1. The equivalent rate r is 9.531%, since $\exp(9.531\%) = 1.1$. In general, the continuous return r, equivalent to the discrete return R, is: $r = \ln(1 + R)$.

Compounding varying discrete returns, $R_{i,i+1}$ with i varying from 0 to $t - 1$, over a sequence of time intervals between the dates 0 and t results in a compounded gross return from 0 to t:

$$(1 + R_{0,t}) = (1 + R_{0,1})(1 + R_{1,2}) \dots (1 + R_{t-1,t})$$

With a constant discrete return R over t periods, the compounding formula becomes $(1 + R)^t$. The compounded discrete return is the geometric average of gross intermediate returns.

The compounded log-return from 0 to t is the logarithm of final to initial value, which can also be written as a logarithm of products:

$$r_{0,t} = \ln(V_t/V_0) = \ln[(V_1/V_0)(V_2/V_1)(V_3/V_2) \dots (V_t/V_{t-1})]$$

This compounded return is the sum of intermediate returns because the logarithm of a product is the sum of logarithms:[4]

$$r_{0,t} = \ln\left(\frac{V_t}{V_0}\right) = \sum_{i=0}^{t-1} \ln\left(\frac{V_{i+1}}{V_i}\right)$$

The rules are illustrated with an example of two periods (Table 11.1). The value of an asset is 1.2 times the initial value V_0 after one period and ends up as $V_2 = 1.08V_0$ by the end of the second period. The compounded discrete return is 8% over the two periods. This compounded return differs from the sum of the intermediate discrete returns, 20% and -10%. The compounded logarithmic return is: $\ln(V_2/V_0) = 7.696\%$, and is exactly equal to the sum of the two consecutive logarithmic returns.

When the returns are random, the logarithmic return over the period is the sum of these random returns. If the returns are normally distributed, the compounded return also is. The final value is an exponential function of the compounded return: $V_t = V_0 \exp(r_t)$. The equivalent expression with discrete returns is: $V_t = V_0(1 + R)^t$.

3 The approximate relation between the two returns results from: $\exp(1 + u) \cong u$ for small values of u.

4 The formula is the compact form of the expanded summation: $\ln(V_t/V_0) = \ln(V_1/V_0) + \ln(V_1/V_0) + \ln(V_2/V_1) + \ln(V_3/V_1) + \dots + \ln(V_t/V_{t-1})$.

TABLE 11.1 Compounding discrete and logarithm returns

	V_t	$V_t/V_{t-1} = 1 + R$	$\text{Ln}(V_t/V_{t-1}) = r$
V_0	1.00		
V_1	1.20	20.00%	18.232%
V_2	1.08	−10.00%	−10.536%
		$(V_2/V_0) - 1$	$\text{Ln}(V_2/V_0)$
Cumulative return		8.00%	7.696%

11.3 STOCHASTIC PROCESSES

The distribution of the value after a series of market shocks depends on the time path of the series of intermediate and consecutive random returns. Consider again two time points, t and $t + \Delta t$. The final value is expressed as a function of the continuous return r_t: $V_{t+\Delta t} = V_{t+\Delta t}\exp(r_t \Delta t)$. The continuous return is the logarithm of the ratio of two values distant in time by a very small interval: $r_t = \ln(V_{t+\Delta t}/V_t)$. The distribution of the value at the end of the period derives from the distribution of the return, r_t, which models an instant shock. Over a discrete period, the final value is distributed like the sum of random successive intermediate returns. Random shocks are modeled as log-returns because they best mimic the real world behavior of returns.

In the standard finance model, the instantaneous return and its time path are modeled as stochastic processes. Any variable, the value of which changes in a random way when time passes, follows a stochastic process. When changes can occur at any time, the process is continuous. When the changes occur only at certain times, the process is discrete. The variables can be stock prices, interest rates, returns, or default events. The random variable can have continuous values, which are real numbers, or only discrete values, and the range of discrete values can be bounded or unbounded. For example, a stock price cannot take negative values but can theoretically grow without any upper bound.

The equation characterizing a stochastic process, S_t, makes the variation over a small time interval of the random variable S_t a linear function of the time interval plus a random term, called z_t. The coefficients are generally time dependent. One of the variables is the time t and the second variable is an unpredictable random component.

The discrete form of a stochastic process uses small discrete intervals of time. The horizon T is divided into n equal intervals $\Delta t = T/n$ of time. As of time t, the values of all variables, and the values of the coefficients of the linear equation, are known. Because all is known at t, the process is said to be I_t-adapted. At $t + \Delta t$, the variable $S_{t+\Delta t}$ is unknown. Over the small interval Δt, the variation of the process depends on: $\Delta S_t = S_{t+\Delta t} - S_t$ and $\Delta z_t = z_{t+\Delta t} - z_t$. The equation of a stochastic process is:

$$(\text{S}_{t+\Delta t}\text{-S}_t)/\text{S}_t = a(t)\Delta t + b(t)\Delta z_t$$

When the time interval Δt tends towards zero, the form of the equation uses an alternate notation for continuous variables:

$$dS_t/S_t = a(t)\,dt + b(t)\,dz_t$$

An equivalent form of this equation is:

$$d\ln(S_t) = a(t)\,dt + b(t)\,dz_t$$

This second form makes it explicit that the process applies to log-returns. Depending on the random variable, a stock price, the value of a firm, or an interest rate, the coefficients $a(t)$ and $b(t)$ take various forms. For example, in the case of a stock price, the equation is:

$$d\ln(S_t) = \mu\,dt + \sigma\,dz_t$$

An equivalent form is:

$$dS_t = S_t\mu\,dt + \sigma S_t\,dz_t$$

These equations mean that the log-return has a deterministic component and a random component.

The coefficient of time, $a(t)$, is the drift. The drift measures the deterministic change of the process per unit of time. The drift coefficient can be constant or be a function of time and of the random variable itself. In the case of a stock price, the drift is μS_t. This drift implies that the instantaneous return dS_t/S_t is a constant μ.

Shocks are modeled by the unpredictable component, the innovation, or diffusion term. "Unpredictable" means that, as of time t, the value of the term cannot be inferred from past information. For defining a process, a distribution has to be defined for this random term. The variations of z_t have a form:

$$\Delta z_t = \varepsilon\sigma_t\sqrt{\Delta t}$$

The variable ε is standard normal: it has a mean of zero and a variance 1. The factor σ_t is a constant known at the beginning of the small interval Δt that might, or not, depend on this date. The random term Δz_t has some special properties. Its mean is zero and its variance is $E(\Delta z_t)^2$, where E is the expectation operator. Variance is the size of a typical Δz_t^2 expressed by unit of time. If this squared difference were negligible, the shock would have a zero variance, and the process would be deterministic. According to the form of Δz_t, the random term z_t has a variance proportional to the time interval Δt. The standard deviation, $\sigma_t\sqrt{\Delta t}$, varies like the square root of time.

In the case of a stock price, the instantaneous return combines a constant drift for dS/S plus a random component dz_t normally distributed with variance $\sigma^2\Delta t$. In other words, the log-return is normally distributed around the deterministic drift with a known variance per unit of time.

Other common processes are shown in Appendix 2.

11.4 MODELING RANDOM SHOCKS

A random shock is defined as a particular value of the unpredictable component: $z\sigma\sqrt{t}$. It is an occurrence of a normal variable with a given standard deviation per a unit of time and the shock z is standard normal, with mean 0 and standard deviation 1. After a shock, the new value of the asset becomes:

$$V_t = V_0\exp\left(z\sigma\sqrt{t}\right)$$

Its variation is: $V_t - V_0 = V_0\left[\exp\left(z\sigma\sqrt{t}\right) - 1\right]$.

With the same shock, the equivalent expressions in terms of discrete returns are:

$$V_t = V_0\left(1 + z\sigma\sqrt{t}\right) \quad \text{and} \quad V_t - V_0 = V_0 z\sigma\sqrt{t}$$

The deviation of value can be calculated for various confidence intervals, using the multiples of volatility for the standard normal distribution. For example, the 1%-quantile of the variation of value is obtained with the 1%-quantile of z, equal to −2.3263, and so on for other confidence levels.

Both expressions provide close values for small deviations, less so for large shocks. For example, a stock price has a value of 100, and the annual standard deviation of its return is 20%. The negative shock matching the 1%-quantile is: $-2.3263 \times 20\% = -46.53\%$. With the discrete formula, the value after this shock is 53.47. With the continuous formula, the final value is: $100\exp(-2.3263 \times 20\%) = 62.80$. The negative variation is smaller than in the discrete case. For large deviations, the discrete formulas might generate negative final values, while the continuous formula always results in positive final values.

In common processes the drift is zero and the variance per unit of time is constant. The process is a Wiener process, also called Brownian motion. Because all successive random changes are independent and follow the same distribution, the process is said to be "identically independently distributed", or "i.i.d.". The distribution is stationary when time passes, meaning that it is identical when starting from any time point.

The horizon can be divided into T equal subperiods. The cumulative of the random increments $\Delta z(t)$, from 0 to T, is a sum of T random normal variables with zero expectation. The expectation of the cumulative increment is equal zero. The Brownian motion follows an erratic path around the initial value. At each time point, the process moves from the current value to the next "abruptly", always starting from the initial point. The path is continuous but it is not smooth.

When returns are stationary and independently distributed ("i.i.d."), the variances over successive periods are additive. The variance of the cumulative return, $\ln(V_T/V_0)$, is the sum of the variances of intermediate logarithmic returns $\ln(V_{t-1}/V_t)$, with t varying from 0 to T. Its variance is equal to T times the variance of the instantaneous return, or $\sigma^2 T$. Its volatility is: $\sigma(z_T) = \sigma_1\sqrt{T}$. For example, the volatility over one year, σ_{250} with 250 trading days in a year, is obtained by scaling up the daily volatility with the square root of 250, 15.811. This is the so-called "square root of time rule for uncertainty", which says, loosely speaking, that the uncertainty, measured by the standard deviation of returns, increases like the square root of time. The unit for the standard deviation should be consistent with the unit of time: if T is one year, then σ is an annual standard deviation, and so on.

11.5 VaR CALCULATION FOR A SINGLE ASSET

The calculation of VaR derives from its compounded random return. A random shock is modeled by $z\sigma\sqrt{t}$. The derivation of VaR is relatively simple for a single instrument. Suppose that the asset is the equity index,[5] with annual volatility 30%.

5 It is possible to buy the index by investing though an exchange-traded fund (ETF), which mimics the behavior of the equity index: there is no need to acquire all the shares included in the index in the proportion of the index.

If the confidence interval is 1%, the VaR is a loss not exceeded in more than one day out of 100 trading days, or in more than two or three days in a year. A confidence level of 2.5% represents 2.5 times in 100 days, or once in 40 days, and six to seven times in a year. The VaR depends on horizon and is obtained by using the relevant volatility calculated with the square root of time rule. A daily VaR refers to a daily loss and is derived from the daily volatility of the index. A daily volatility is derived from the annual volatility using the square root of time rule. With an annual volatility of 30%, the daily volatility is:

$$\sigma_1 = \sigma_{250}/\sqrt{250} = 30\%/15.8114 = 1.897\%$$

Under the normal assumption for the equity returns, the daily VaR is equal to the product of 2.33 times the daily volatility, or:

$$\text{Daily } 1\% - \text{VaR} = 2.33 \times 1.897\% = 4.42\%$$

For a 10-day horizon, the relevant volatility becomes: $\sigma_{10\text{ days}} = \sigma_{1\text{ day}}\sqrt{10} = \sigma_{1\text{ day}}$ $3.162\sigma_{1\text{ day}}$. The 10-day VaR is 3.162 times the daily VaR:

$$\text{10-day VaR} = 3.162 \times 4.42\% = 13.98\%$$

11.6 DISTRIBUTION OF VALUE UNDER NORMAL RETURNS

A process specifies how the value at a given horizon is distributed because the terminal value of the asset depends on the compounded return over a period.

With a single period, the distribution of the value at date 1 derives directly from that of the return because $V_1 = V_0(1 + R)$. If the return is normally distributed, so is the final value. The obvious limitation of using this distribution for discrete returns is that it allows the random return to hit values lower than −100%, resulting in non-acceptable negative values of the asset.

With log-returns, the relation between the value at horizon and the return becomes:

$$r_{0,t} = \ln(V_t/V_0)$$

It follows that:

$$V_t = V_0 \exp\left(r_{0,t}\right)$$

The return over the period is the summation of normal algebraic returns over subperiods. The sum is normally distributed as its terms are, and the logarithm of the terminal value is normally distributed. A log-normal distribution is the distribution of a random variable, the logarithm of which is normally distributed: if X is log-normally distributed, then $Y = \ln(X)$ has a normal distribution. Likewise, if Y has a normal distribution, then $X = \exp(Y)$ has a log-normal distribution.

The random asset value at horizon is V_t, such that $V_t = V_0 \exp(r_t)$, where r_t is the random continuous return, t is the horizon and V_0 is the asset value today. The ratio of final to initial value is: $\ln(V_t/V_0) = r_t$. If the return r_t is normal, the distribution of the final value is log-normal. It can be shown that the logarithm of the ratio has expectation $(r - \sigma^2/2)$ and its volatility is $\sigma\sqrt{T}$. The ratio V_t/V_0 is:

$$V_t/V_0 = \exp\left[(r - \tfrac{1}{2}\sigma^2)T + z\sigma\sqrt{T}\right]$$

11.7 FROM SHOCKS ON RISK FACTORS TO SHOCKS ON ASSET VALUE

The above calculations of VaR are relevant for the equity index. The equity index is a risk factor, a market parameter that influences the prices of individual stocks.

If the asset is a stock instead of the index, a shock on the index return triggers a shock on the stock return that depends on how the particular stock responds to a given market shock. Some stocks are highly sensitive to index movements and others much less so. Suppose that a stock amplifies the market movements by a factor of 2.[6] Then a shock on the index of 1% translates into a shock on the stock return of 2%. If the stock return is twice the market shock, its volatility will be twice as large. Its VaR is also twice as large, since the VaR is a multiple of volatility. With the above volatility the stock index, the daily VaR at the same confidence level of 1% is twice as large or: $2 \times 4.42\% = 8.94\%$. This shows that the sensitivities of assets to market shocks are inputs to VaR calculations.

The example shows some properties of VaR. VaR is a "comprehensive" measure of risk because it integrates various components of market risk: the sensitivity of the asset; the volatility of the underlying risk factor; and the distribution of the stock value. Accordingly, there are several steps for deriving the VaR of a single asset:

1 Map an asset to market risk factors.
2 Measure the sensitivity of the asset with respect to each risk factor that has an impact on its value.
3 Input the volatilities of the risk factors for a given horizon. In general, with multiple risk factors, all volatilities of risk factors and how they co-vary together need to be included.
4 Determine a loss percentile at a given confidence level, α, which requires to specify the shape of the distribution of the random values of the asset.

The α-quantile of loss is the VaR at the selected confidence level.

11.8 APPENDIX 1: CONTINUOUS RETURNS AS LIMIT OF DISCRETE RETURNS

Any time interval, h, can be divided into m equal shorter subperiods having an equal length $\Delta t = h/m$. By definition of a discrete return, the final value at t compounds arithmetic return m times: $V_h = V_0(1 + R/m)^m$. In this formula, R/m is the proportional discrete return matching

6 The stock has a β of 2, the statistical average of the ratio of the return of the stock to the return of the equity index.

the constant period $\Delta t = h/m$. The final value is obtained by compounding m times the discrete return R.

As the compounding frequency increases, the compounded value also increases, given a constant R. When compounding once for one year with $R = 10\%$ per year, the terminal value of 1 invested originally is: $1(1 + 10\%) = 1.10$. If the compounding is done twice, over two equal periods of six months, the proportional return becomes $R/2 = 5\%$ and the terminal value of 1 becomes: $1(1 + 5\%)(1 + 5\%) = 1.1025$. Compounding four times a year at a rate $R/4 = 2.5\%$ results in a terminal value: $1(1 + 2.5\%)^4 = 1.1038$.

When the frequency tends towards infinity, compounding becomes continuous over an infinite number of infinitely small subperiods. The formal exact relation between discrete returns and continuous returns is obtained by taking the limit of the compounding term when m grows to infinity and the time interval $\Delta t = h/m$ becomes infinitesimally small. In the formula, the proportional discrete return tends towards zero but it is compounded a number of times as it grows to infinity. When m increases to infinity, the expression within brackets tends towards a limit, according to the mathematical formula:[7] $\text{Lim}(1 + x/m)^m = \exp(x)$. Replacing x by $R, (1 + R/m)^m = \exp(r)$ when m grows to infinity.

11.9 APPENDIX 2: COMMON PROCESSES

A process describes the behavior of a random variable through time. It is used to model market shocks, or the behavior of stock prices, of interest rates, or the random occurrences of default events.

11.9.1 General Form of Processes

In what follows, the process is called S_t and the innovation term is $z = \varepsilon\sigma_t\sqrt{\Delta t}$. The differences between the various processes result from the definitions of the coefficients applied, respectively, to the drift and the diffusion terms.

The simplest process is the Wiener process. The generalized Wiener process includes a drift and a constant standard deviation per unit of time. This simple process considers that the drift, a, and the coefficient of the innovation term, b, are constant. The process is:

$$dS = a\,dt + b\,dz$$

The expectation, $E(dx) = a\,dt$, is the constant drift. The incremental change of value is constant per unit of time. The constant drift represents a trend. The standard deviation is $b\sigma\sqrt{t}$. The time path is a steady trend with random variations at each time point around this trend. The discrete equivalent of this equation serves for simulating the time path of the process. The horizon T is divided into n small intervals of same length $\Delta t = T/n$. The continuous equation

7 A mathematical limit $f(x)$ of an expression $f(x, n)$ exists when there is a value of n big enough so that: $|f(x, n) - f(x)| < \varepsilon$, no matter how close to zero ε is. The mathematical notation is: $\text{limit}[f(x, n)]n \to \propto = f(x)$. For finding the limit of $(1 + x/m)^m$ when m grows to infinity, the expression is written as: $(1 + x/m)^m = \exp[m \ln(1 + x/m)]]$. Using: $[\ln(1 + u)]u \to 0 = u, \ln(1 + x/m)m \to \propto = x/m$. Hence, $\exp[m(x/m)]$, with m large, tends towards $\exp(x)$.

takes a discretized form such that all small variations are related by the same relation:

$$\Delta S = a\,\Delta t + b\,\Delta z$$

A time path for S can be simulated by incrementing the variable by $a\,\Delta t$ at each step and adding a random number, normally distributed with mean zero and standard deviation $\sigma\sqrt{\Delta t}$.

The Ito process is a generalized Wiener process where the coefficients, a and b, are functions of the underlying variable S and of the time t.

$$\Delta S = a(S_t, t)\Delta t + b(S_t, t)\Delta z$$

Drift and innovation are, respectively, $a(S_t, t)\,\Delta t$ and $b(S_t, t)\,\Delta z$, with the variance of the random term being $b^2(S_t, t)\,\Delta t$. The drift and the variance rate are now dependent on S_t and time. The stock price process is a special case of an Ito process.

11.9.2 The Stock Price Process

The stock price process is a special case of the Ito process. For stock prices, the return is randomly distributed normally and i.i.d. because stock price returns are supposed to have a positive trend plus random deviations independent from past deviations. Under the efficiency hypothesis, the return reacts to innovations, not to the past information, which is already embedded in the stock price.

The instantaneous return is dS_t/S_t, where S_t represents a stock price.[8] The return has a constant drift μ and a constant variance σ with drift and diffusion coefficients $a(S_t, t) = \mu S_t$ and $b(S_t, t) = \sigma S_t$. The process for the stock return is:

$$dS_t/S_t = \mu\,dt + \sigma\,dz$$

The return follows a generalized Wiener process, with constants $a = \mu$ and $b = \sigma\sqrt{\Delta t}$. The trend is exponential because it is such that $dS/S = \mu$.[9] The coefficients μ and σ are constant, but the drift and the diffusion terms depend on S_t. They are proportional to the stock price:

$$dS_t = \mu S_t\,dt + \sigma S_t\,dz$$

The variance of S_t increases as the square of S_t.

11.9.3 The Mean-reverting Process and Interest Rates

Mean-reverting processes are used for interest rate models. Unlike stocks, interest rates do not drift away from common ranges of values. The general form of a mean reverting process is:

$$dS_t = \lambda(\mu - S_t)\,dt + \sigma S_t\,dz$$

S_t designates here a process followed by market parameters, typically interest rates.

8 The return, dS_t/S_t, is the derivative of $\ln(S_t)$.

9 If $S = S_0\exp(\mu t)$, $\ln(S) - \ln(S_0) = \mu t$ and the derivative of $\ln(S)$ is $dS/S = \mu$.

When S_t is above the constant μ, the drift term is negative, making negative variations of S_t more likely. The random variable might decline and revert to μ. If S_t is below μ, the drift term is positive and the random variable might increase and revert to μ. The parameter μ represents a long-term trend. When moving away from this long-term trend the asset price tends to revert to it. A mean-reverting process models variables that cannot drift away from some long-term value by a large magnitude and for long periods.

The deviations from the trend last for a period that depends on the value of λ. The higher the λ, the stronger the force making the process revert to the long-term parameter. Hence, deviations last longer when λ is low and are shorter when λ is high. This process applies to interest rates whose deviations cannot be extremely high, as they can be for stock prices.

11.9.4 The "Rare Event" Process and Defaults

Rare events occur unexpectedly and their probability of occurrence increases with the time interval elapsed. Rare events are abnormal jumps of stock prices. For credit risk, rare event processes serve for modeling defaults.

The size of the event is fixed, hence not small as it is with the usual normal distribution, even when the time interval is very small. But its probability becomes zero when the time interval gets near to zero. The occurrence of an event does not depend on past information.

For a process counting rare events, only discrete values can be used. The random term is n_t, instead of the random innovation z_t because n_t is a number. The variation between n_t and $n_{t+\Delta t}$ is either 1, if a rare event occurs, or 0, if not. The probability that the rare event occurs in a small time interval is $\lambda\,\Delta t$. The parameter λ is a positive real number, equal to the expected number of occurrences that occur during a given interval, called intensity or hazard rate. The intensity is the number of events occurring within a unit of time. The probabilities of observing, or not, rare events are:

$$P(n_{t+\Delta t} - n_t = 1) = \lambda\,\Delta t$$

$$P(n_{t+\Delta t} - n_t = 0) = 1 - \lambda\,\Delta t$$

The count of rare events, over a time horizon, is a random variable taking integer values k. The probability of observing k rare events is the Poisson distribution: $f(k, \lambda) = \lambda^k\, e^{-\lambda}/k!$. Rare event processes are the basis of the default-intensity models of default.[10]

10 See Chapter 21.

12

Portfolio Risk and Factor Models

The volatility of the return of a portfolio and its value-at-risk (VaR) are measures of the risk of portfolios of traded assets. The higher the volatility of the return, the higher the chances of observing wider and negative earnings. Both measures depend on how diversified the portfolio is. The variance–covariance approach is the classical method for measuring the dependencies between the returns of individual assets and the extent of risk diversification.

With large portfolios, dealing with all asset covariances would be complex. Factor models allow modeling the diversification from the variance–covariance matrix of factors from which asset returns depend. Because there are fewer factors than individual assets, it is easier to work with factors than with individual assets. When the sensitivities of asset returns to factors are constant and when returns are normally distributed, the VaR is expressed by simple-closed-form formulas. These are the assumptions of the delta-normal VaR.

This chapter explains how the volatility of the portfolio return derives from the variance–covariance of the returns of assets within the portfolio, or from the variance–covariance of risk factors returns. It introduces factor models and shows how they simplify the determination of the volatility of the portfolio returns. Finally, the standard formulas of the delta-normal VaR are derived.

Contents

12.1 PORTFOLIO RETURN VOLATILITY, CORRELATIONS AND COVARIANCES

The basic inputs for measuring portfolio return and its risk are the random returns of individual positions. Each individual return X has an expectation, $E(X)$, and a volatility. The volatility, $\sigma(X)$, is the square root of the variance, $V(X)$, of the return. The variance, $V(X)$, is the sum of the squared deviations from the mean weighted by probabilities of occurrence. The portfolio return is the weighted average of individual returns, with weights calculated from the values of each position at the beginning of the period.

One measure of portfolio risk is the volatility of its return, which depends on how dependent individual returns are from each other. For a pair of positions, the random returns of each are noted X and Y. The correlation is a coefficient in the range -1 to $+1$, which measures the strength of the association, and its direction, positive or inverse. The value $+1$ means that the two variables always vary in the same direction; a correlation of -1 means that they always vary in opposite ways, and a zero correlation means that they are independent. The covariance is an alternate measure of correlation: it is the sum of the products of the deviations from the mean of two returns, weighted by the probabilities of occurrence of each pair of values. The correlation is the ratio of covariance by the product of the variances of X and Y.

The covariance between two random individual returns, X and Y, is $\text{Cov}(X, Y)$, or σ_{xy} in short form. The relation with the correlation coefficient for the two variables is $\sigma_{xy} = \rho_{xy}\sigma_x\sigma_y$, where ρ_{xy} is the correlation between variables X and Y, and σ_x and σ_y are the standard deviations, or volatilities, of variables X and Y.

With a portfolio of assets, the portfolio return, Y_P, over a discrete period, is the weighted average of arithmetic asset returns Y_i, between the same dates, using the initial weights of assets w_i.[1] In general for a portfolio of n assets, the portfolio return is: $Y_P = \sum_{i=1,n} w_i Y_i$. The variance of the return Y_i is the volatility σ_i of the return squared and the covariances are the products of the correlation coefficients by the volatilities of the two returns:

$$\sigma_i^2 = \sigma^2(Y_i) \quad \text{and} \quad \sigma_{ij} = \text{Cov}(Y_i, Y_j) = \rho_{ij}\sigma_i\sigma_i$$

The following calculations apply to a portfolio of two assets. The weights of the assets are 30% and 70%. The portfolio return is: $Y_P = w_1 Y_1 + w_2 Y_2$. The variances and the covariance of the returns of the two assets form a square matrix (2, 2), with variances along the diagonal and covariance off-diagonal. The matrix is symmetric because the covariance does not depend on the order of the two variables. The numerical values of the example are:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} 0.0677\% & 0.0180\% \\ 0.0180\% & 0.0333\% \end{bmatrix}$$

1 The log-returns do not have this property.

The same dependencies can be shown in a correlation matrix:

$$\Sigma = \begin{array}{|cc|} 1 & 37.901\% \\ 37.901\% & 1 \end{array}$$

The variance–covariance matrix, the correlation and the variances are related:

$$\sigma_1^2 = 2.601\%^2 = 0.0677\%$$
$$\sigma_2^2 = 1.826\%^2 = 0.0333\%$$
$$\sigma_{12} = \rho\sigma_1\sigma_2 = 37.901\% \times 2.601\% \times 1.826\% = 0.0180\%$$

The variance of the portfolio return is the variance of the sum of weighted random returns[2] since: $\sigma_p^2 = \sigma^2\left(Y_p = \sum_{i=1,2} w_i Y_i\right)$. For two assets, the formula becomes:

$$\sigma_p^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2\rho_{ij}w_1w_2\sigma_1\sigma_2$$

Using the numerical values of weights and of the variance–covariance matrix, the variance of the portfolio return is:

$$\sigma_p^2 = 30\%^2 \times 0.0677\%^2 + 70\% \times 2 \times 0.0333\%^2 + 2 \times 30\% \times 70\% \times 2.601\% \times 1.826\%$$
$$\sigma_p^2 = 0.0300\%$$

The volatility of the portfolio return is the square root of variance:

$$\sigma_p = \sqrt{0.0300\%} = 1.7317\%$$

These formulas extend easily to a portfolio of n assets. The variance of the portfolio return is lower than the sum of the variances of the weighted portfolio returns because the correlation coefficient is lower than 1. The difference measures the diversification effect. It is equal to the sum of variances only when the correlation coefficient is equal to 1, the case of a perfect correlation between the two returns. The risks of individual returns, measured by variance or volatility of returns, add to less than their arithmetic sum, except in the extreme case where correlation is perfect.

12.2 FACTOR MODELS

The calculation of the portfolio return volatility is difficult to implement when there are thousands of positions. Factor models reduce the dimensionality of the problem. A factor model is a linear relation between the return of each single asset and the returns of a smaller number of factors on which individual asset returns depend. The coefficients of the factors measure the sensitivities of the asset return to the factors.

Relations between asset returns and market parameters can be exact, such as in pricing formulas. For example, the value of a foreign exchange contract, an example used in Chapter 13

2 The general formulas are in Appendix 1.

for illustrating the calculation of VaR, is an exact function of several risk factors. Such pricing formulas are not linear, but they can be approximated by linear functions for small deviations of markets. When there is no exact relation with factors, a linear relation is obtained with statistical techniques, using traditional regression analysis. Such statistical fits explain only a fraction of the variance of asset returns, the rest being the variance of the random error term independent of factors.

Factor models serve for modeling dependencies between individual asset returns because they link individual returns to a set of common factors. When factors vary, they generate co-movements between assets returns. The general risk is the risk explained by factors. The fraction of the variations of asset returns unexplained by factors is the specific risk of the asset. When a factor model is based on regression analysis, the covariances between each factor and the residuals are zero and the expectation of the residual term is zero. Specific risk can be diversified away by combining assets in portfolios, while the general risk cannot because it depends on the same common factors. Factor models are used in risk models because they model the diversification effects.

12.2.1 A Single-factor Model

A traditional example of a single-factor model is the Sharpe's Capital Asset Pricing Model (CAPM), where the return of a single stock is explained by the return on the stock index. Under the CAPM, stock returns depend on the equity index return (r_m). The coefficient of the equity index is the β, which measures the sensitivity of the stock return to the index return. The expected return of a single stock, with index i, is related to the expected equity index return through the equation of the CAPM:

$$E(r_i) = r_f + \beta_i[E(r_m) - r_f]$$

The equation implies that the expected stock return, r_i, should be equal to the risk-free rate r_f plus a risk premium, which is the difference between the expected index return and the risk-free rate, times β_i, the beta of the stock. The β_i is the sensitivity of the stock return to the stock index return. Some stocks have high β and others low β. The response to a standard shock on the equity index return, for example +1%, is the β. If β is higher than 1, for example 2, the stock return should, on average, respond with an increase of +2%. If the β is 0.5, the response to the same shock on equity index would be, on average, +0.5%. The beta equals one by definition for the equity index. The statistical version of the theoretical equation is a one-factor model:

$$r_i = \beta_0 + \beta_1 r_m + \varepsilon$$

The general risk of the stock return depends on the movements of the general market index. The error term is the fraction of the variations of the return independent of the index. The variance of the stock return is the sum of the variance due to general risk and of the variance due to the error term, or specific risk, because they are independent: $\sigma^2(r_i) = \beta_i\sigma^2(r_m) + \sigma^2(\varepsilon_i)$.

The model is simple, with only one factor explaining the variations of stock return. But, in general, asset returns depend on multiple factors. For example, bond values depend on many interest rates of different maturities; option values depend on the underlying, on interest rates and on market volatility; foreign contract values depend on interest rates in different currencies plus the spot exchange rate; etc.

12.2.2 Multi-factor Models

Multi-factor models express the return of each asset as a linear function of multiple factors X_k, with $k = 1, K$, and, with coefficients β_k, plus a residual, ε_i, specific to each asset. With N assets and K factors, the general equation is:

$$Y_i = \beta_{0i} + \beta_{1i}X_1 + \beta_{2i}X_2 + \ldots + \beta_{ki}X_k + \varepsilon_i$$

The N assets have the subscript i, with $i = 1, N$, and the K different factors have the returns X_k. The coefficients relating each asset return to the common factors are the factor loadings. They use double subscripts, the first subscript for the factor and the second for the asset. The factor loadings are the sensitivities of asset returns with respect to each factor. For example, if the first factor changes by 1%, the asset return will change by β_{1i} times 1%. The β_{ik} coefficients measure how the return of asset i responds to a given change of the factor k.

In practice, when combining instruments, there are potentially hundreds of risk factors, such as all interest rates in all currencies, all foreign exchange rates, all equity indices, etc. However, the number of factors is much lower than the number of instruments.

12.3 SENSITIVITIES OF COMMON INSTRUMENTS

Sensitivities of common instruments are well known, and reviewed hereafter for bonds, options and forward contracts.

12.3.1 Bonds and Loans

The values of fixed-rate bonds vary inversely with interest rates since they are the present values of all contractual cash flows using interest rates as discount rates. The sensitivity of a bond to a parallel shift of all rates is the duration. Since bonds depend on interest rates of various maturities, it is convenient to assume a parallel shift of all rates. The sensitivity is a negative number because of the inverse relation between bond values and rates. A common measure of sensitivity for a bond is the basis point "DV01". A basis point ("bp") is 1% of 1%, or 0.0001, and sensitivities are measured by a number or in basis points. The duration formula relates the change of the bond value to the shock on interest rates:

$$\Delta B = -DB\,\Delta i$$

In this formula, B is the bond value, D is the modified duration and Δi is the common shock to all rates. The sensitivity to a shock on any particular interest rate is the variation of value of this discounted cash flow. Using the proxy formula $1/(1 + r) \sim 1 - r$, the variation of value of the bond to a change, Δi_t, of the interest rate of maturity t is approximately: $F_t(1 - t\,\Delta i_t)$. The exact formula is the derivative of this single discounted cash flow at date t with respect to i_t.

12.3.2 Options

The value of options depends upon a number of parameters, as shown originally in the Black and Scholes model:[3] the value of the underlying, the horizon to maturity, the volatility of the underlying, the risk-free interest rate. The option value increases with the underlying asset value, with its volatility, with the maturity, and decreases with the interest rate. Option sensitivities are known as the "Greek letters".

The sensitivity with respect to the underlying asset is the "delta" (δ). The formula for the delta of stock options is: ΔStock option $= \delta\, \Delta S$, where S is the value of the underlying asset. The delta is low if the option is out-of-the-money (underlying asset price below strike) because the holder does not get any payoff under immediate exercise. When exercise provides a positive payoff, δ gets closer to 1 because the payoff of the option increases as the underlying asset value increases.[4]

The variation of δ is the convexity of the option, measured by gamma (γ). Gamma is the change of the slope of the curve representing the option value as a function of the underlying. The delta is always between 0 and 1 (in absolute value); the gamma, which is the change of the delta, is also always between 0 and 1. The option is sensitive to the time to maturity because a longer horizon increases the chances that the stock moves above the strike price. The sensitivity with respect to residual maturity is theta (θ). The Greek letter θ measures the "time decay" of the option value as time passes. The higher the volatility of the underlying asset, the higher the chance that the value moves above the strike during a given period. Hence, the option has also a positive sensitivity to the underlying asset volatility, which is the "vega" (ν). Finally, payoffs can occur only in the future, and the today value requires discounting, which implies a negative sensitivity to the level of the risk-free interest rate. "Rho" (ρ) is the sensitivity to a variation of the risk-free rate.

12.3.3 Forward Contracts and Interest Rate Swaps

For all instruments that can be replicated by a portfolio of other instruments, the sensitivities are derived from those of the components of the portfolio replicating the instrument. Forward contracts can be replicated with static portfolios and can be valued using closed-form formulae. For example, a forward exchange contract depends on three risk factors, the spot rate and the two interest rates in the two currencies. The sensitivities are derived either analytically or numerically using such formulae. Similarly, a forward interest rate contract is replicated from lending and borrowing the same amount in the same currency for differing maturities. It has a long and a short leg, each with a single final cash flow. The sensitivities of each leg are the durations of those contracts with the appropriate signs.[5] Interest rate swaps can be seen as borrowing and lending the same amount in a single currency for the same maturity, with different interest rates. The sensitivity with respect to the fixed rate is that of the fixed-rate leg since the variable-rate leg has zero duration.

3 Call options are introduced in Chapter 8 on the convexity risk of the balance sheet. The Black-Scholes formula is used in Chapter 21, where default models are detailed.

4 See Figure 8.3 in Chapter 8.

5 The example of a forward exchange contract is used to illustrate how the VaR can be calculated for such instruments in Chapter 13.

12.4 NON-LINEAR INSTRUMENTS

Many instruments are related to risk factors by non-linear relations, as the above examples illustrate. They include bonds, of which value discounts cash flows with market rates, or options, of which price is given by the option pricing formula. But any non-linear relation can be approximated, for small variations, by a linear approximation.

The basic formula for understanding sensitivities and their limitations is the Taylor expansion formula. The formula expresses the variation of value of any function $f(x)$ from an initial value, when the argument has initial value x_0, by a series of terms that depend on the difference, Δx, between x and x_0. The formula is primarily used for small variations around x_0; otherwise, it becomes a proxy. The function is the value of an asset, such as an option. The Taylor expansion formula for the function $f(x)$ when x changes from x to $x + \Delta x$ is:

$$f(x+\Delta x) - f(x) == \frac{1}{1!}\frac{\partial f(\Delta x)}{\partial x}\Delta x + \frac{1}{2!}\frac{\partial^2 f(\Delta x)}{\partial x}\Delta x^2 + \cdots + \frac{1}{n!}\frac{\partial^n f(\Delta x)}{\partial x}\Delta x^n$$

For a small change of the variable, the change of the function can be approximated by the first term depending on Δx. The first derivative applied to Δx is the sensitivity, or the first-order term. A better approximation is obtained by using the second-order term. The second-order term is the change of the sensitivity with Δx, or the derivative of the sensitivity. The Taylor series makes explicit the change of value of instruments as a function of the first-, second-, third-order derivatives, and so on.

When the second-order and higher-order terms can be neglected, the instrument is said to be "linear". Linear instruments have variations of values that relate linearly to the risk factors. Options are non-linear instruments because of the "kink" of the payoff when the underlying asset has a value equal to the strike price. Hence, the first-order term is a poor proxy of the variation of the option value. The gamma measures the rate of change of the delta, which is a measure of the convexity.

For non-linear instruments, the first-order sensitivities depend on the prevailing conditions, and are local measures. If the market conditions change, or when time passes, the sensitivities are not constant. When sensitivities cannot be considered as constant, the departure from the linear model can be addressed by using the second-order terms of the Taylor expansion formula. For an option, the adjustment for the convexity results from this equation. For the changes of the underlying asset, S, the Taylor expansion shows that the variation of the option, $f(S)$, is approximated by:

$$f(S) - f(S_0) = \delta\,\Delta S + \frac{1}{2}\gamma \Delta S^2$$

12.5 VOLATILITY OF THE PORTFOLIO RETURN

The volatility of a portfolio return depends on the variance–covariance matrix of asset returns. With portfolios, there is a large number of assets and the dimension of the matrices become unmanageable. However, the general risk, the risk due to factor movements, depends on a much smaller set of common factors. This allows modeling of the dependencies within portfolios by using the variance–covariance matrix of factors rather than the one of assets. Since there are many fewer factors than assets, the matrix has much smaller dimensions and is much easier to manage.

The portfolio return is a weighted average of the asset arithmetic returns. Because the asset returns are linear functions of factor returns, the portfolio return is also a linear function of factor returns. Therefore a portfolio of assets can be treated as a portfolio of factors. The weights of the factors are the weighted averages of factor loadings of the individual assets, using as weights those of assets within the portfolio. Under this approach, the volatility of the portfolio return is a function of the variance–covariance matrix of the factors and of the coefficients of the factor models.

The approach is illustrated with a two-asset portfolio, before the formulas are generalized. The portfolio return, Y_P, is the weighted average of each asset return: $Y_P = w_1 Y_1 + w_2 Y_2$. The portfolio return is a linear function of individual asset returns, asset returns are linear functions of factor returns and the portfolio return is a linear function of the factor returns.

The coefficients of the new relation are the weighted averages of factor loadings, using the weights of assets in the portfolio for weighting the factor coefficients. With a two-asset portfolio and two factors, each asset return is now a linear function of two factors, X_1 and X_2. The factor models for each asset return are, for example:

$$Y_1 = \beta_{01} + 0.5X_1 + 1.0X_2 + \varepsilon_1$$
$$Y_2 = \beta_{02} + 1.2X_1 + 0.8X_2 + \varepsilon_2$$

The constants of each model are: $\beta_{01} = 0.1$ and $\beta_{02} = 0.05$. They can be ignored in what follows because they do not contribute to the random variations of returns. Replacing the asset returns by their expressions as a function of factor returns in the expression of the portfolio return, the portfolio return becomes:

$$Y_P = w_1[\beta_{01} + 0.5X_1 + 1.0X_2 + \varepsilon_1] + w_2[\beta_{02} + 1.2X_1 + 0.8X_2 + \varepsilon_2]$$
$$Y_P = (w_1\beta_{01} + w_2\beta_{02}) + (w_1 0.5 + w_2 1.0)X_1 + (w_1 1.2X_2 + w_2 0.8)X_2 + (w_1\varepsilon_1 + w_2\varepsilon_2)$$

The coefficients are the weighted values of the sensitivities of the two assets. The constant is the weighted value of the constants of the two assets. The residual is the sum of weighted values of the two standalone (single asset) residuals. The factor model for the portfolio return becomes:

$$Y_P = 0.085 + 0.680X_1 + 1.080X_2 + 0.7\varepsilon_1 + 0.3\varepsilon_2$$

The numerical values of the coefficients in this equation are derived as weighted averages of the factor loadings, using as weights the asset weights in the portfolio. The constant is calculated as: $\beta_{0P} = 0.7 \times 0.1 + 0.3 \times 0.05 = 0.085$. The coefficients of factors X_1 and X_2 are:

$$\beta_{1P} = w_1 0.5 + w_2 1.0 = 0.680$$
$$\beta_{2P} = w_1 1.2 + w_2 0.8 = 1.080$$

The portfolio return is a linear function of the factors, using the above weighted coefficients:

$$Y_P = \beta_{0P} + \beta_{1P}X_1 + \beta_{2P}X_2 + \varepsilon P$$

The set of two-factor models, one for each asset, simplifies to a single two-factor model of the portfolio return, with two weighted coefficients, a single weighted constant and a single weighted residual. The expectation of the weighted residual is zero since it combines two residuals with zero expectation.

This shows that the volatility of the portfolio return can be calculated from the variance–covariance matrix of the factors, instead of being derived from the variance–covariance matrix of the assets. It is possible to proceed as if the portfolio of two assets behaved as a portfolio of two factors. From here, the calculation of volatility of the portfolio return due to factors is identical to the calculation of volatility for a two-asset portfolio, except that the weights of the factors are the weighted sensitivities to each factor.

The σ_{ij} are the variances and covariances of the factor returns. The variance–covariance matrix of factors Σ is, for example:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{21} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} 0.25 & -0.01 \\ -0.01 & 0.04 \end{bmatrix}$$

In this matrix, the double index refers only to factors, not to assets. The analytical formula of the variance of the portfolio return, $V(Y_P)$, when the variance of the residual term is ignored, is:

$$V(Y_P) = (\beta_1 \sigma_{11})^2 + (\beta_2 \sigma_{22})^2 + 2\beta_2 \beta_1 \sigma_{12}$$

Replacing β_1 and β_1 by their respective values 0.5 and 1:

$$V(Y_P) = 0.680 \times 0.25 + 1.080 \times 0.04 + 2 \times 0.680 \times 1.080 \times (-0.01) \times 0.04 = 0.148$$

The standard deviation, or volatility, of this portfolio return is:

$$\sigma(Y_P) = \sqrt{V(Y_P)} = 0.384$$

Factor models help in handling the high dimensions of portfolios. For a portfolio, there are as many covariances and correlations as there are pairs of stocks. The direct measures of the variances and covariances of stock returns would imply measuring N^2 terms with N assets.[6] Since the matrix is symmetric, we have $N(N-1)/2$ covariance terms plus N variance terms, a total of: $N + N(N-1)/2 = N(N+1)/2$ different terms. The number of cells is proportional to the square of the number of assets.

A variance–covariance matrix of K factors, with K smaller than N, is easier to handle than the matrix for N assets. With 2 factors, $2(2+1)/2$, or 3, inputs are required. The factor loadings are also required and there are two coefficients for each asset. With N assets and 2 factors, the covariance matrix of factors has 3 terms and there are $2N$ coefficients. The total is $2N + 3$. With only two assets there is not much gain, but the difference is significant when considering a larger number of assets. If $N = 10$, the number of inputs becomes: $2 \times 10 + 3 = 23$. This should be compared with the number of required inputs for the covariance matrix of asset returns, which would be: $10 \times (10 + 1)/2 = 55$.

6 The variance–covariance matrix is squared with dimensions $N \times N$.

12.6 CLOSED-FORM MATRIX FORMULAS FOR THE VOLATILITY OF THE PORTFOLIO RETURN AND FOR VaR

The same result for the volatility of the portfolio return could be obtained from matrix formulas, as expanded below.

Notations for matrix formulas are required. A series of values in the same column forms a column vector and the same series in a single row is the transposed vector. The square table of variances and covariances is a matrix. A transposed vector or matrix uses the subscript "T". The dimensions of vectors or matrices are specified with two numbers, the number of lines and the number of columns, in parentheses. With this convention, a square covariance matrix for two factors is (2, 2); the column vector of weights is (2, 1) and its transpose is the row vector of weights (1, 2).

Instead of using market weights, the weighted sensitivities are used as inputs. The line vector of sensitivities to factors is: (0.680, 1.080), noted S^T. The transpose, S, is the column vector of weights. The covariance matrix is noted Σ (Table 12.1). In the matrix formula, the (1×2) transposed vector of coefficients S^T is multiplied by the square (2×2) variance–covariance matrix of factors Σ. The result is multiplied by the row vector obtained by the (2×1) vector S of coefficients.

With matrix notations, the formula for the variance of the portfolio return collapses to a very simple form:

$$V(Y_P) = S^T \Sigma S$$

The notations can be expanded to make explicit the asset weights, rather than the weighted factor loadings, by isolating the weights of assets within the portfolio from the factor loadings. The asset weights are 0.7 and 0.3, or, in vector notation, $w^T = (0.7, 0.3)^T$. The sensitivities of the portfolio with respect to the factors X_1 and X_2 are obtained above as weighted averages of asset sensitivities, S_1 and S_2, for X_1 and X_2:

$$S_1 = w_1 0.5 + w_2 1.0 = 0.680$$
$$S_2 = w_1 1.2 + w_2 0.8 = 1.080$$

TABLE 12.1 Calculation of the portfolio systematic risk with two-factor models

		X_1	X_2	
		0.25	−0.01	0.680
		−0.01	0.04	1.080
0.680	1.080	0.1592	0.0364	0.148

The equations representing the two-factor models have coefficients that form a matrix:

$$\beta = \begin{bmatrix} \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \end{bmatrix} = \begin{bmatrix} 0.5 & 1.0 \\ 1.2 & 0.8 \end{bmatrix}$$

In this matrix, the coefficients of the two models have double subscripts, the first one for the asset and the second one for the factor.

The portfolio sensitivities to each factor are the function of the factor loadings of each asset and of the asset weights. These sensitivities form a column vector, S, calculated as the product of the above square matrix of the factor loadings, β, by the column vector of asset weights, w:

Weights
0.7
0.3

Matrix of factor coefficients

0.5	1.1	0.680
1.2	0.8	1.080

The column vector of sensitivities is $S = \beta w$. The row vector of sensitivities is the transpose of this column vector, S^T. The rule for transposing a product is that the transpose of a product is the product of the transposed in reverse order: $(\beta w)^T = w^T \beta^T$. With these notations, the compact formula for the calculation of variance is:

$$V(Y_P) = S^T \Sigma S = w^T \beta^T \Sigma \beta w$$

The portfolio volatility is:

$$\sigma(Y_P) = \sqrt{S^T \Sigma S} = \sqrt{w^T \beta^T \Sigma \beta w}$$

These are the standard forms for the calculation of variance and volatility of the portfolio return. They derive from the portfolio sensitivities to each factor and from the variance–covariance matrix of the factors. It is common to designate such sensitivities as market positions, instead of the asset values, within a portfolio because they are the starting points of all calculations.

Finally, the VaR becomes a closed-form formula when returns are normally distributed, because it is calculated as a multiple of the volatility of the portfolio return. For example, the VaR for a confidence level of 1% becomes, with above notations:

$$\text{VaR} = 2.33\sqrt{S^T \Sigma\, S} = 2.33\sqrt{w^T \beta^T \Sigma \beta w}$$

This is the standard formula used in the delta-normal VaR. Under this approach, the sensitivities are considered constant and the distribution of returns is normal. An example of the calculation of the delta-normal VaR is provided in Chapter 13.

12.7 APPENDIX 1: CORRELATION AND VOLATILITY OF A SUM OF RANDOM VARIABLES

A formula extends the calculation of the variance and the volatility to any number N of variables X_i. This is the single most important formula for calculating the portfolio return volatility. The general formula, with m assets, is:

$$\sigma^2\left(\sum_{i=1}^{m} X_i\right) = \sum_{i=j=1}^{m} \sigma_{ii}^2 + \sum_{i=1,m,j=1,m}^{i \neq j} \sigma_{ij}$$

$$\sigma^2\left(\sum_{i=1}^{m} X_i\right) = \sum_{i=j=1}^{m} \sigma_i^2 + \sum_{i=1,m,j=1,m}^{ij} \rho_{ij}\sigma_i\sigma_j$$

In these formulas, σ_i^2 is the variance of variable X_i, equal to the square of the standard deviation σ_i, and σ_{ij} is the covariance between variables X_i and X_j. A similar, and more compact, notation writes that the variance of the random variables summing m random variables is the summation of all covariances and variance terms σ_{ij}:

$$\sigma^2\left(\sum_{i=1}^{m} X_i\right) = \sum_{i=1,j=1}^{m} \sigma_{ij}$$

12.8 APPENDIX 2: MAPPING AN INSTRUMENT TO RISK FACTORS

When mapping a position to risk factors, the exclusion of some factors results in "basis risk", because the retained factors do not track exactly the asset prices any more. Yet, the position has to be mapped to selected factors. For example, if interest rate references do not correspond to exact maturities, it is possible to interpolate a proxy of the rate. Mapping a position to selected risk factors transforms a single position into several positions with respect to each selected factor. The value and volatility of this set of positions should match those of the original instrument. The process requires adjusting the weights allocated to each selected factor to comply with this condition.

For example, an original position is a €1,000,000 zero bond of maturity one year plus three months (1.25 years). Instead of using the 1.25-year interest rate, the selected interest rates are those of the one-year and two-year maturities. Since we have two risk factors distinct from the risk factor matching exactly this position, the original position should be decomposed into two bonds that map to the two selected risk factors. The combined portfolio of the bonds should match the value and the volatility of the original instrument in VaR calculations.

A simple technique consists of interpolating the interest rate for 1.25 years and allocating the position to the one-year time point and the two-year time point. The one-year and two-year interest rates are, respectively, 5.20% and 5.40%. The linear interpolation provides a proxy of the 1.25-year interest rate:

$$i(1.25 \text{ years}) = 5.20\% + (5.40\% - 5.20\%) \times 0.25 = 5.25\%$$

With this rate, the value for the position with this proxy of the 1.25-year interest rate is:

$$1,000,000/(1+5.25)^{1.25} = 938,052$$

The mapping process should preserve volatility. The volatilities for the one-year and two-year bonds are known, and input as percentages of values. Using the available one-year and two-year volatilities, 8% and 10%, the 1.25-year volatility is interpolated:

$$\sigma(1.25\text{ years}) = 8\% + (10\% - 8\%) \times 0.25 = 8.5\%$$

The variance of the 1.25-year rate is $8.5\%^2 = 0.723\%$.

The final step requires the determination of the allocations of the initial position to the two equivalent one-year and two-year positions. The distance to the time points could be used as the allocation rule, allocating 0.25(1.25 − 1) to one year and 0.75(2 − 1.25) to two years. But the variance of such a portfolio will not match the interpolated volatility of 1.25-year position. The correct process consists of finding the allocation percentages that make the variance of the replicating portfolio of two positions identical to the interpolated volatility. The allocations are w to the one-year bond and $1 - w$ to the two-year bond. The resulting variance of a portfolio of these two positions depends on the correlation between these two interest rates, assumed here to be 90%:

$$\text{Portfolio variance} = w^2 \times 8\%^2 + (1-w)^2 \times 10\%^2 + 2 \times 90\% \times w \times (1-w) \times 8\% \times 10\%$$

Matching this variance with the interpolated variance implies a value of w is such that it matches exactly 0.723%. The allocation is $w = 64.1\%$ for the one-year bond and $1 - w = 35.9\%$ for the two-year bond. Such allocation applied to the interpolated value of the 1.25-year position, 938.042, provides the value of each of the two one-year and two-year positions (Table 12.2).

The initial position is equivalent to these two equivalent positions. In the VaR calculation, the initial position is replaced by the two equivalent positions, which depend on retained risk factors only.

TABLE 12.2 Allocating weights and values to each reference position

Allocation to one-year bond w	64.10%
Allocation to two-year bond $1 - w$	35.90%
Total	100.000%
Portfolio variance	0.723%
Value	938,042
Allocation one-year bond	601,285
Allocation two-year bond	336,757

13

Delta-normal VaR and Historical VaR

The delta-VaR methodology applies to portfolios of linear instruments, when sensitivities to risk factors can be considered constant, and asset returns normally distributed. Linear instruments are those of which return is proportional to the returns of risk factors. The principle of the delta-VaR technique is to decompose assets or portfolios as a linear combination of elementary positions mapped, each, to a single factor, of which VaR is derived with the standard formulas.

For non-linear instruments, and notably options, another common method for deriving the VaR is the historical VaR. The historical VaR does not impose the restrictions of delta-normal VaR and is relatively easy to implement. The technique derives the distribution of returns from the historical behavior of risk factors by assessing directly the variations of the values of assets, instead of relying on constant sensitivities. Because of these benefits, it is widely used for non-linear portfolios.

This chapter explains the methodology and assumptions for the determination of the delta-normal VaR and of the historical VaR using the example of a forward exchange contract, which depends on several risk factors, as many assets and portfolios do.

Contents

13.1 DELTA-NORMAL VaR

The simplest technique for modeling VaR for linear portfolios is the delta-normal VaR. The technique relies on sensitivities ("deltas") of positions to risk factor, which are assumed constant for small deviations, and returns are supposedly normally distributed. The sequence of steps of the method is:

- Given sensitivities to risk factors, model the variations of the portfolio value as a linear function of the variations of risk factors.
- Determine the volatility of the portfolio, calculated as the volatility of a linear function of random factor returns.
- For moving from volatility to VaR, the delta-normal VaR relies on the normal distribution and its known quantiles.

The general delta-normal VaR methodology is illustrated with a simple instrument, a forward exchange contract receiving euros and paying dollars in one year. The contract sets the forward exchange rate today. A forward contract has a zero value at inception, but its value fluctuates subsequently as markets deviate. The issue is to determine the VaR of such a contract. The starting point consists of writing down the formula for the value of the contract as a function of risk factors.

13.1.1 A Forward Foreign Exchange Contract

A forward exchange contract can be replicated by borrowing today in dollars, converting the proceeds into euros at the current foreign exchange rate and lending the euros. At maturity, the proceeds from lending euros are used to deliver dollars. The euro amount for one dollar is the forward exchange rate. The forward contract can therefore be replicated by three transactions:

- Borrowing in dollars at the dollar interest rate ($i_\$ = 5\%$) for one year. The borrowed amount is such that, in one year, we have exactly \$1 to pay back: it is equal to $\$1/(1+5\%)$.
- Converting the dollars borrowed into euros at the prevailing exchange rate. The current exchange rate is: = 0.769231 €/\$ or exactly 1/1.30 \$/€
- Lending the proceeds in euros at the euro interest rate for one year: $i_€ = 4\%$. After one year, the euro loan provides: €0.769231$(1 + i_€)$ = €0.76190. This last value is the forward exchange rate in euro per dollar, and the inverse, \$1.3125, is the dollar forward value of one euro.

The forward contract combines a long position in euros and a short position in dollars, the amounts of which are identical at the current spot rate, with same maturity. The implied forward exchange rate is such that: \$1 = €0.76190 or €1 = \$1.3125.

Figure 13.1 shows the cash flows from this combination, fixed since the inception of the contract. If the forward pays \$10,000,000, it receives €7,619,048 one year from now. As time passes, the exchange rate and the interest rates deviate from their initial values. However,

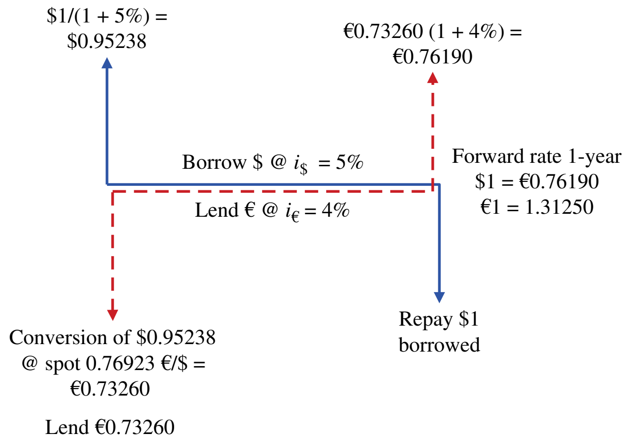


FIGURE 13.1 Borrowing – lending in two currencies

the cash flows remain fixed, which implies that the mark-to-market value of the contract changes. This value depends on the following three market parameters:

Spot rate "S": 1.30 €/$ or 0.76923 €/$.
$i_€ = 4.00\%$ annually.
$i_\$ = 5.00\%$ annually.

These market parameters are the risk factors relevant for repricing the forward contract when time passes.

At any date, t, after initiating the contract and until its maturity, the value of the forward is F_t. Using the euro for valuation, the value is equal to the present value of its long leg (lending in euro) minus the present value of the short leg (borrowing in dollars). The value of the short leg in dollars is converted to euros, using the current foreign exchange rate, S_t €/$, expressed as the euro value of one dollar. The discounting is from current date t up to maturity $T - t$, with T being one year in this example:

$$F_t(€) = -S_t€/\$ \times \$10,000,000/(1 + i_\$)^{T-t} + €7,619,048/(1 + i_€)^{T-t}$$

The value of the contract at initial date is zero. Subsequently, the value fluctuates with the three risk factors according to the above formula, given the fixed cash flows in the two currencies. The foreign exchange contract F_t is an exact non-linear function of these three risk factors. For implementing the delta-normal VaR methodology, it is necessary to move from the non-linear exact relationship between value and factors to a proxy linear relation. The non-linear position is decomposed into elementary positions depending, each, on one single risk factor. The same process applies in general for a portfolio. The difference is that the number of risk factors for a portfolio will be much higher than for a single instrument.

For implementing calculations with the forward, the volatilities and the correlations of risk factors, and their variance–covariance matrix of risk factors are required inputs (Table 13.1).

TABLE 13.1 Correlation and variance–covariance matrix between risk factors

Market parameter	Annual volatilities	Historical correlations: 1-year € interest rate	1-year $ interest rate	Spot €/$ forex
1-year €*i*	0.61%	100%	10%	20%
1-year $*i*	0.58%	10%	100%	10%
S €/$	0.35	20%	10%	100%

Variance–covariance matrix

1-year € interest rate	*1-year $ interest rate*	*Spot €/$ forex*
0.00372%	0.00035%	0.04270%
0.00035%	0.00336%	0.02030%
0.04270%	0.02030%	12.25000%

13.1.2 The Framework of Delta-normal VaR

The conceptual framework for decomposing the variations of value of an instrument into a linear function of elementary positions is the Taylor expansion formula.[1] The formula states that any function of multiple variables can be approximated, for small variations, by a linear function of the changes of value of each argument, the coefficients being the first derivatives of the function with respect to each argument. A small change of value ΔF of the forward contract is approximated by a linear function of the small changes of risk factors, the coefficients being the sensitivities of the value with respect to each risk factor:

$$\Delta F = (\partial F/\partial S)\Delta S + (\partial F/\partial i_\$)\Delta i_\$ + (\partial F/\partial i_€)\Delta i_€$$

This function means that the forward can be replicated by a portfolio of three positions combined in a linear fashion.

In the delta-normal VaR, the sensitivities are considered as constant, and the above linear function is known. The factor returns are normally distributed, so that the linear combination of elementary positions also follows a normal distribution, which is used for determining VaR.

13.1.3 Deriving Sensitivities

The initial value of the forward contract is zero because the present values of the long and short legs of the forward are identical in a common currency: $\$10,000,000/(1 + i_\$) = \$952,380.95$,

1 The Taylor expansion formula is introduced in Chapter 12 on sensitivities.

TABLE 13.2 Initial value of the contract

Short position ($)	10,000,000
Long position (€)	7,619,048
Spot rate S (€/$)	0.7692
Dollar interest rate	5.0%
Euro interest rate	4.0%
Forward value (€)	0.00

with $i_\$ = 5\%$. At the initial spot rate, this is identical to the value of the long leg in euros, or $€7,619,048/(1+4\%)$. These inputs are replicated in Table 13.2.

Sensitivities can be derived analytically, as first derivatives of the function valuing the contract from the risk factors.[2] It is also common to calculate them numerically, by changing the value of the risk factor by some given shock, notably in the general case when there is no closed-form formula. When the residual maturity is set to one year, the analytical formula for the value of the forward contract is:

$$F(€) = -S€/\$ \times \$10,000,000/(1 + i_\$) + €7,619,048/(1 + i_€)$$

The unit changes of risk factors are 0.01 €/$ for the spot rate and 0.01% for each interest rate. The sensitivities are calculated numerically by applying these small variations of the factors starting from their initial values, and changing one parameter at each step. The new value of the forward contract, after each change, is derived with the pricing formula. The difference with the initial value is the sensitivity with respect to the shocked risk factor. Since the initial value is zero, the three sensitivities are equal to each of the three new values obtained (Table 13.3). These revaluations of the contract are done here for the sole purpose of calculating the three sensitivities, later considered constant.

The variation of the forward value can now be written as a linear function of the small changes of each risk factor:

$$\Delta\text{Forward value} = \$-95,238 \times \Delta S\,(€/\$) - €698 \times \Delta i_€ - €704 \times \Delta i_\$$$

TABLE 13.3 Numerical sensitivities

Variation of spot exchange rate: 0.01 €/$	$−95,238
Variation of $i_\$$ by 1 basis point (0.01%)	€+698
Variation of $i_€$ by 1 basis point (0.01%)	€−704

2 The first-order derivatives are: $\partial V/\partial S = -9,523,810/(1+i_€)$; $\partial V/\partial i_€ = S \times 10,000,000/(1+i_€) \times [-1/(1+i_€)^2]$; $\partial V/\partial i_\$ = 10,000,000/(1+i_\$) \times [-1/(1+i_\$)^2]$.

This equation is used to generate variations of values of the contract from variations of the three risk factors. Once the sensitivities are obtained, there is no need to revalue the contract for a change of risk factors from the pricing formula. Instead, the equation making the change of value of the contract a linear function of the changes of risk factors is used. This is a different process from the full revaluation used with historical simulations, as explained in section 13.2. The revaluation of the contract with this method is "partial" since it is inferred from given sensitivities, estimated once, and applied to the variations of risk factors.

13.1.4 Volatility and Delta-normal VaR of the Forward Value

The variance of a linear sum of three random variables weighted by sensitivities results from the usual statistical formulas, using the cross-correlations of the three risk factors:

$$\sigma^2(\Delta V) = \sigma^2(S_s\ \Delta s + S_{i€}\Delta i_€ + S_{i\$}\Delta i_\$)$$

The variance can be calculated by the general algebraic formula of the variance with the correlations of risk factors.[3] But it is easier to derive the volatility of the contract using the matrix formula (Table 13.4). The variance of the portfolio characterized by the vector of

TABLE 13.4 Variance and volatility of the forward contract

			Variance–covariance matrix			
			$i_€$	*$i_\$$*	*Spot €/\$*	
			0.00372%	0.00035%	0.04270%	−7,044,238
			0.00035%	0.00336%	0.02030%	6,977,150
			0.04270%	0.02030%	12.25000%	−9,523,810
S($i_€$)	*S($i_\$$)*	*S(S)*				*Variance*
−7,044,238	6,977,150	−9,523,810	−4,304.1	−1,723.5	−1,168,258.2	11,144,562,180,531
						Volatility
						3,338,347
						Daily volatility
						211,136

3 The formula is: $\sigma^2(\Delta V) = S_s^2\sigma^2(S) + S_{i€}^2\sigma^2(i_€) + S_{i\$}^2\sigma^2(i_\$) + 2\rho_{S,i€}S_s\sigma(S)S_{i€}\sigma(i_€) + 2\rho_{S,i\$}S_s\sigma(S)S_{i\$}\sigma(i_\$) + 2\rho_{\$,i€}S_{i€}\sigma(i_€)S_{i\$}\sigma(i_\$)$, where $\rho_{S,i€}$, $\rho_{S,i\$}$, $\rho_{i\$,i€}$, are the correlations between factors, derived from historical data.

TABLE 13.5 Annual and daily VaR of the forward contract

Confidence level	1%
Multiple	2.330
Annual VaR	7,777,337
Daily VaR	211,110

sensitivities is $S^T \Sigma S$, where S^T is the row vector of sensitivities and Σ is the variance–covariance matrix of the factors.[4]

The VaR is a multiple of volatility matching the selected confidence level. The variance and volatility are annual in the above calculations. For calculating a daily VaR, the volatility, and the VaR, are derived by scaling down the annual volatility with $1/\sqrt{250}$ as scaling factor. The multiple is 2.33 for a 1% one-tailed confidence level under the normal distribution assumption. The annual and the daily VaR of the forward contract are shown in Table 13.5.

13.1.5 Limits of the Delta-normal VaR

The delta-VaR methodology suffers from the constant delta assumption and the normal distribution assumption.

The constant sensitivity assumption does not hold for non-linear, or optional, instruments. In such cases, the instruments cannot be revalued as a linear combination of simple positions, but should instead be revalued from pricing formulas or techniques. This solution is implemented in simulations.

The delta-VaR also uses the property of normal distributions. However, the observed distributions of returns can be skewed or have fat tails. The skewness measures the departure from symmetry, for example when the probabilities or negative returns are higher than the probabilities of high returns. Fat tails exist when the frequencies of large deviations are higher than for a normal curve. Large deviations can arise from "jumps" in the stock prices, or from serial dependence of consecutive returns. Serial dependence means that high returns tend to be followed by large returns (in absolute value, either positive or negative), creating higher frequencies of large variations than if returns were independent. Figure 13.2 compares the shape of the normal curve with a left-skewed curve and a curve exhibiting fat tails (a leptokurtic curve). In both cases, the delta-normal VaR is underestimated.

Skewness and kurtosis are measured by the moments of a distribution. The k central moment of a distribution is the expectation of the deviation from the mean, with power k:

$$\mu_k = E\left[(X - \mu)^k\right]$$

4 With a portfolio, the sensitivities would be calculated as weighted averages of the sensitivities of each position, using the ratios of asset value to portfolio value as weights.

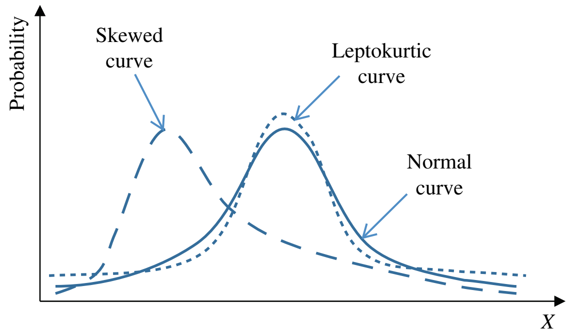


FIGURE 13.2 Skewness and kurtosis of distributions

The expectation is the first moment and measures the central tendency. The variance σ^2 is the second moment and measures the dispersion around the expectation. The skewness and the kurtosis are, respectively, the third and fourth central moments divided by σ^3 and σ^4:

$$\text{Skewness: } \tau = \frac{E\left[(X-\mu)^3\right]}{\sigma^3} = \frac{\mu_3}{\sigma^3} \qquad \text{Kurtosis: } \kappa = \frac{E\left[(X-\mu)^4\right]}{\sigma^4}$$

The skewness is zero and the kurtosis is 3 for the standard normal distribution. The excess kurtosis is the kurtosis minus 3: It serves for comparing the shape of a distribution to the normal curve.

If the moments of actual distributions can be measured, they can be used for addressing the potential underestimation due to non-normal curves, while preserving the conceptual framework of the normal VaR. The Cornish-Fisher expansion formula provides the quantiles of a non-normal distribution given the moments of the distribution.[5]

The simulation methodology provides more freedom for the determination of VaR. In hypothetical simulations, scenarios for risk factor values are generated, consistent with their dependencies, with normal or non-normal distributions. They are detailed in Chapter 14. The historical simulation approach relies on the empirical distributions of risk factors observed over a lookback period. This last technique is now detailed because it is frequently used.

13.2 HISTORICAL VaR: FORWARD CONTRACT EXAMPLE

For non-linear assets, or when the normal assumptions should be rejected, the historical VaR can be used instead of the delta-VaR. The method is widely used because it does not require any restrictive assumption on distributions and sensitivities.

The historical VaR is based on the distribution of historical time series of factor returns. These time series embed historical correlations and volatilities of risk factors without making any assumption on their distributions. The historical variations of risk factors are applied to revalue the current portfolio. The output is a distribution of the portfolio profit and loss (P&L)

5 The Cornish-Fisher expansion is a method for determining the quantiles of a non-normal distribution. See, for example, Holton, G. A. (2003), Value-at-Risk: Theory and Practice, [77].

that is empirical rather than parametric, as the normal distribution is. The historical VaR relies on the full revaluation of asset values with the variations of the factor returns, which bypasses the linear assumption of the delta-VaR technique. The technique is commonly used for optional instruments.

The starting point of historical VaR is the time series of factors observed over a lookback period, for example one year. Suppose that 250 daily values of the three risk factors are observed. The historical values of the risk factors differ from their current values and cannot be used to replicate the current value of the contract. Instead, these historical observations are converted into daily changes. The 250 daily variations of risk factors provide a series of factor values mimicking their historical behavior, with the changes applied to their today values. The instrument is then revalued with the pricing formula at each daily step with the series of factor values obtained by applying historical changes to their current values. The output is a series of daily variations forming an empirical distribution of its P&L. The VaR is then derived from this distribution as a quantile.

The delta-VaR and the historical VaR differ in the sequence and nature of calculations. Under delta-VaR, sensitivities are calculated once and deviations of value follow by plugging these sensitivities into the linear relation between the deviations of value of the instrument and the shocks of the risk factors. This is a partial revaluation method because the calculation does not revert to the pricing formula. Under the full revaluation method, the simulated values of risk factors serve, first, for revaluing the instrument for all daily scenarios. The deviations of value from the current value are then derived from these revaluations.

The process involves the following steps:

- Determine the series of historical values of risk factors.
- Transform this series into percentage changes.
- Apply the percentage variations to the current risk factor values.
- Derive for each daily change of all risk factors a revaluation of the instrument.
- Calculate the daily P&L (daily change of value) of the instrument by subtracting the today value.
- Construct the distribution of daily P&Ls.
- Find the quantile from this empirical distribution.

Table 13.6 shows the historical values of the risk factors, the series of their percentage changes, and the corresponding series of the new risk factor values after these changes are applied to their current values. The table is limited to the first four days of the historical simulation. Percentage variations start from the second day. For example, the "day 2" percentage variation of the spot rate is: $(0.81132 - 0.81251)/0.81251 = -0.146\%$. The last block of the table applies these percentage variations to the current values of the risk factor. For example, the simulated "value 2" of the spot rate is: $-0.146\% \times 0.76923 = 0.7681$.

Using the set of 250 simulated values of risk factors, the forward values are recalculated using the exact formula for pricing the contract for each set of daily values of the risk factors. The daily variations of value of the contract form the empirical distribution of P&L, with 250 points. The loss quantiles are directly derived from the distribution. The loss value exceeded in two to three times, out of 250, corresponds to the 1%-quantile, etc.

One drawback of historical VaR with full revaluation is that the technique is backward looking. It relies on observations of market parameters independently of how representative they are of current conditions. In some cases, it is conservative if historical observations capture

TABLE 13.6 Deriving historical percentage changes of risk factors

Risk factors	*Spot rate S €/$*	*"i $" annually*	*"i €" annually*
Current values	0.76923	0.05000	0.04000
Historical values			
Day 1	0.8125	0.049800	0.038520
Day 2	0.8113	0.049720	0.039521
Day 3	0.8126	0.049102	0.039871
Day 4	0.7925	0.050100	0.040150
...	...	...	...
Daily variations of historical values			
Day 1			
Day 2	−0.146%	−0.161%	2.599%
Day 3	0.153%	−1.243%	0.886%
Day 4	−2.472%	2.033%	0.700%
...	...	...	...
Daily simulations of values			
Value 1	0.7692	0.05000	0.04000
Value 2	0.7681	0.04992	0.04104
Value 3	0.7704	0.04938	0.04035
Value 4	0.7502	0.05102	0.04028
...	...	...	...

market shocks of large magnitude. In others, it might miss the current market conditions if these are outliers compared to historical scenarios.

Historical VaR serves for back-testing alternate VaR methodologies, by making sure that the portfolio values include those that would have appeared if the past were replicated on the current portfolio structure. But the historical VaR cannot serve for stress-testing the VaR since the nature of scenarios is predetermined by past data, which, in general, differ from the desirable stress scenarios.

14

Extensions of Traditional VaR

The conventional value-at-risk (VaR) methodology has several drawbacks. It ignores the losses beyond the VaR. It also does not have the desirable subadditive property of risk measures, which imply that the risk of a portfolio be lower than the summation of the risks of its components. The E-VaR measures the expected loss beyond the VaR threshold and is subaddictive. This metric is proposed by regulators as a substitute for conventional VaR.

The actual distributions of returns deviate from normal distributions, with losses more likely than gains and extreme deviations more frequent than with the normal distribution. The simulation methodology provides additional flexibility in the choice of appropriate distributions of returns. Simulations address hypothetical simulations rather than historical simulations. The technique is relatively simple with normal variables. With non-normal distributions, dependencies are modeled with the alternate copula methodology.

This chapter introduces the E-VaR metric and details the simulation methodology. The simulation process exposed hereafter is general, and also applies for modeling the credit risk of portfolio.

Contents

14.1 E-VaR OR EXPECTED SHORTFALL

The E-VaR, also named "Expected Shortfall" or "Expected Tail Loss", is the expected loss conditional on loss being higher or equal to VaR at a given confidence level, measured as the probability-weighted average of losses exceeding the VaR.

14.1.1 Calculation of E-VaR

The E-VaR is calculated from the distribution of losses exceeding the VaR. By definition, the probability $P(L \leq L_\alpha)$ that the loss L exceeds, in absolute value, the α-quantile L_α is equal to α. According to Bayes' rule,[1] the probability of a loss conditional on being higher or equal to the VaR is the unconditional probability divided by the probability α. For a particular value, l, of the random portfolio loss L, this conditional probability is:

$$P\left(L < l \middle| L \leq L_\alpha\right) = \frac{P(L < l)}{P(L \leq L_\alpha)} = \frac{P(L < l)}{\alpha}$$

The expected shortfall is the probability-weighted average loss conditional on loss being higher than VaR. This definition requires the calculation of the expectation of the distribution of losses truncated at the α-quantile L_α. The E-VaR can be calculated analytically for a normal distribution.[2] In the general case, with an empirical distribution for example, the E-VaR is calculated numerically by taking the average value of all adverse deviations of the daily loss beyond VaR. With n values of losses strictly higher in absolute value than the VaR, each with frequency p_i, the E-VaR is:

$$\text{E-VaR} = \sum_{i=1,n} p_i l_i \bigg/ \sum_{i=1,n} p_i$$

The summation of all p_i, in the denominator, is equal to the confidence level, α, of the VaR.

For example, the tail losses and their frequencies for the forward contract example of the previous chapter, from the historical simulation, are shown in Table 14.1. The VaR at 1% is 11,000,000. There are 10 occurrences greater than or equal to this VaR, or 1% of the number of simulations (1,000). In some instances, there are several occurrences of the same numerical value of the loss. The counts of each numerical value are in the first column of the table, and they sum up to 10. The corresponding values of the loss are in the second column. The percentage frequencies are calculated over the total number of scenarios, and they sum up to 1%.

The E-VaR is the expectation of losses greater than or equal to 11,000,000. It is calculated by assigning a conditional probability to each of these values. The conditional probabilities are

---

1 The Bayes' rule provides the conditional probability $P(A|B)$ of the event A conditional on occurrence of event B as: $P(A|B) = P(A)/P(B)$.

2 It is equal to the density of the standardized normal distribution of the loss l, with density $\varphi(l)$, for a value of loss equal to L_α divided by α: $\text{E-VaR} = \frac{1}{\alpha}\int_{-\infty}^{L_\alpha} l\varphi(l)\text{d}l = \frac{\varphi(L_\alpha)}{\alpha}$. The derivation requires a change of variable.

TABLE 14.1 Calculation of E-VaR (1% VaR = 11,000,000)

Frequency count	*Loss value*	*Frequency %*	*Conditional frequency %*	*Weighted loss value*
3	−11,000,000	0.30%	30%	−3,300,000
2	−12,000,000	0.20%	20%	−2,400,000
1	−13,000,000	0.10%	10%	−1,300,000
1	−14,000,000	0.10%	10%	−1,400,000
2	−15,000,000	0.20%	20%	−3,000,000
1	−18,000,000	0.10%	10%	−1,800,000
10		1%	100%	−13,200,000

obtained by dividing the observed frequencies in percentage by 1%. The sum of loss values weighted by these conditional probabilities is the E-VaR, 13,200,000.

14.1.2 Property of E-VaR

A risk measure should be subadditive, which means, loosely speaking, that the sum of two risks be lower or equal to the risk of the sum.[3] This is the case for the E-VaR, but not for the traditional VaR. If the measure of risk is not subadditive, it becomes possible that breaking down business units would result in a lower VaR than operating them together. For example, with two trading desks, it could be that the sum of the standalone VaRs of each desk be lower than the VaR of the two desks operating together.

For example, consider a portfolio combining two assets with discrete negative payoffs of 0 and −10, respectively, with probabilities 97% and 3%. The 5% VaR of each asset is zero because the loss of 10 occurs only with a probability of 3%, less than the confidence interval. Assuming independence of payoffs, the loss distribution of the portfolio is shown in Table 14.2.

TABLE 14.2 Example of a portfolio with two assets

Event	*Probability*	*Loss L*
Neither *A* nor *B* lose	94.0900%	0
A or *B* lose	5.8200%	10
Both *A* and *B* lose	0.0900%	20
Total	100.0000%	

3 In general, risk measures should be coherent. Coherent measures of risk are defined in: Artzner, Delbaen, Eber, Heath (1997), Thinking coherently, [15]. A discussion of the relative merits of E-VaR compared to traditional VaR is in: Acerbi, C., Tacshe, D. (2002), On the coherence of expected shortfall, [1].

The 95% VaR of the portfolio is such that the probability of occurrence of losses is higher than 5%: it is 10. Under this risk metric, the portfolio is riskier than each of the standalone assets since their corresponding, standalone, VaRs are zero, which shows that the VaR is not subadditive.

The expected shortfall is the expected loss conditional on loss exceeding the VaR, equal to the loss quantile at 5%. The expected shortfall is subadditive, unlike VaR. This can be seen by calculating the expected shortfall of each asset and checking that its sum is higher than the expected shortfall of the portfolio. For calculating the expected shortfall of each asset, the probabilities of losses beyond the VaR at 95% are considered. Each asset has a loss of 10 with an unconditional probability of 3%. The conditional probability of this loss is:

$$P(\text{Loss} = 10)/P(\text{Loss} \geq \text{VaR}) = 3\%/5\% = 60\%$$

The expected shortfall of each asset is: $60\% \times 10 = 6$. The arithmetic sum of expected shortfalls for the two assets is 12.

The expected shortfall for the portfolio is calculated using the probabilities conditional on loss exceeding 10, the 95% VaR of the portfolio. There are two instances where the loss exceeds the VaR of 10. The probability that the portfolio loss is 10 is 5.82% and the probability that the portfolio loss is 20 is 0.09%. The conditional probabilities are obtained by dividing by 5%, and are 98.20% and 1.80%, respectively. The expected shortfall of the portfolio is: 10.18. This shows that the portfolio expected shortfall is lower than the sum of expected shortfalls of the two assets.

14.2 MONTE CARLO SIMULATIONS

Simulations serve to address a comprehensive set of hypothetical scenarios, whether historical simulations are based on observed scenarios only. Simulations consist of generating a large number of portfolio values for obtaining a proxy of the distribution of the portfolio values.

The methodology is more involved than historical VaR. It implies simulating future risk factor values complying with their dependency structure and conducting a revaluation of the portfolio for each simulation. Full-blown simulations are calculation intensive since the full revaluation implies recalculating values of instruments for a large number of scenarios. When no closed-form formulas are available, numerical techniques are used for pricing instruments for each scenario.[4] For each draw of risk factors, the portfolio is revalued and the distribution of its P&L obtained.

When distributions are normal, it is easy to simulate correlated distributions of risk factor returns using their variance–covariance matrix. The technique relies on the Cholesky decomposition. When non-normal distributions of returns are considered, such as the student distribution, which has fatter tails than the normal distribution, the copula technique should be used instead.

4 This leads to "simulations within simulations", which are calculation intensive, such as for look-back options valued with simulations.

14.2.1 Normal Distributions

With normal distributions, the simulation of dependent variables requires simulating independent variables in a first step and constructing the dependent variables as linear functions of these independent normal variables in a second step. The technique relies on the Cholesky decomposition, which derives the coefficients of linear functions from the desired variances and covariances.

This result is well known for two variables. Two normal variables Y_1 and Y_2 follow a standard bivariate distribution with correlation ρ if, and only, Y_2 is a function of two independent normal standard variables Y_1 and X. The linear decomposition follows immediately in this case:

$$Y_2 = \rho Y_1 + X\sqrt{1-\rho^2}$$

In the general case, each one of N dependent variables, Y_j, with $j = 1, N$, should be decomposed into a linear combination of the same number N of independent normal variables X_i. The linear function relating each dependent variable to the independent variables is derived from the variance–covariance matrix of the dependent variables. Their coefficients are found from the decomposition of the variance–covariance matrix Σ into a product of two triangular matrices, L and U, such that: $\Sigma = LU$. The coefficients of the linear functions linking the X_i to the Y_j are the elements of the lower triangular matrix.

The expanded form of the decomposition lower triangular matrix, in Figure 14.1, shows that each dependent variable depends in a linear way on the independent variables: the coefficients of the linear decomposition are the α_{ij}.

The Appendix in section 14.3 details the technique and provides general formulas. The process uses standardized variables, following normal variables with mean zero and standard deviation 1.[5] For simulating K values of N dependent variables Y_j, the first step consists of generating K random draws of N independent normal variables X_i. The K values of the

$$\begin{bmatrix} Y_j \end{bmatrix} = \begin{bmatrix} L \end{bmatrix} \times \begin{bmatrix} X_i \end{bmatrix}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \dots \end{bmatrix} = \begin{bmatrix} \alpha_{11} & 0 & \dots \\ \alpha_{12} & \alpha_{22} & \dots \\ \dots & \dots & \dots \end{bmatrix} \times \begin{bmatrix} X_1 \\ X_2 \\ \dots \end{bmatrix}$$

FIGURE 14.1 Expanded form of the Cholesky decomposition

5 If all variables are normal standardized, the variance–covariance matrix collapses to the correlation matrix, with correlation coefficients varying for each pair of variables and all diagonal terms are equal to 1.

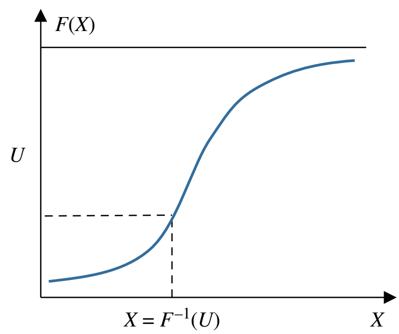


FIGURE 14.2 Relation between a random variable *X* and its quantile $u = P(X \leq x)$

dependent normal variables Y_j are obtained from the linear functions of these independent variables, using the Cholesky coefficients. The simulation of normal variables follows the general process for generating random values of variables with analytical distributions.

14.2.2 Simulations of Random Variables

In many instances, it is necessary to simulate the values of random variables following distributions that can be normal or non-normal. In market risk, distributions with fatter tails than the normal distribution better capture the actual behavior of observed returns. For modeling the credit risk of portfolios of loans, it is also necessary to simulate the dependent times to default of individual loans for reconstructing the loss distribution of the portfolio. The probability distribution of such random times is an exponential function (Chapter 21). Generating random values of variables following given distributions is a common building block for such simulation processes.

The process starts with the simulation of random uniform numbers. The values of uniform standard variables are real numbers within the [0,1] interval. Such numerical values can always be seen as quantiles of any distribution. The simulation of a variable X following a given distribution function, $F(X)$, starts with the simulation of uniform random numbers. Their values are converted into values of X by inverting the cumulative distribution function $F(X)$ of X. The cumulative distribution function has an S shape and is monotonously increasing with values being between the upper and lower bounds 0 and 1. For each value of the cumulative function, U within [0,1], there is a single matching value of the random variable, such that $X = F^{-1}(U)$ and $U = F(X)$ (Figure 14.2). The variable X follows the distribution with cumulative function F. By repeating the process, as many values of X as needed can be simulated.

14.2.3 Market Simulations of Non-normal Variables

For market risk, the simulation method allows using other distributions than the normal distribution, such as the Student distribution. The variables are dependent factor returns. With non-normal variables, the simple Cholesky method for generating dependent variables does not apply and an alternate technique should be used instead. The current practice relies on copula functions for non-normal distributions. For market risk, the technique could be applied for simulating dependent Student variables.

The principle of the copula technique is to make uniform standard variables dependent. A copula function describes how these uniform variables are dependent on each other. For two variables, a copula density function is the joint probability that variables are below given quantiles, u and v, of their marginal distributions. Since the copula dependence applies to uniform variables, it can be used with any distribution function, normal or not. The copula transformation provides the quantile v (of Y) conditional on the quantile u (of X). For example, the Gaussian copula transformation is written as:

$$v = \Phi\left[\rho\Phi^{-1}(u) + \sqrt{1-\rho^2}\Phi^{-1}(w)\right]$$

In this equation, Φ^{-1} is the inverse of the standard normal distribution Φ. Therefore, $\Phi^{-1}(u)$ is a value of the variable X. The variable w is a uniform standard variable independent of u, the inverse of which is seen as the value of an intermediate standard normal variable, independent of X. The combination in brackets is a linear combination of two independent standard normal variables. It is also normal and is standardized since it has a zero expectation and a variance of one. Applying the standard normal function, Φ, to this linear combination results in a quantile, v, of another standard normal distribution. The value of v is seen as a quantile of Y, which is conditional on the quantile u of X. This dependence between u and v makes the variables X and Y dependent.

Using the normal standard function in the copula transformation does not imply that the underlying variables X and Y are normal. The function only says that the dependence between X and Y is such that their quantiles are related in the same way as the quantiles of two normal variables. In other words, dependence is modeled by making two uniform variables dependent, with the dependence being the same as normal variables.

The simulation process starts with the two independent uniform variables. A third variable, w, uniform and independent of u and v, is used to make these two variables dependent using the copula transformation. A numerical application is shown in Chapter 25 on credit loss distributions, where random dependent times to default are simulated for constructing the loss distribution of a credit portfolio. The same technique can be implemented for addressing non-normal market returns.

When the purpose is to model the dependence of extreme returns, different copula functions can be used. With the Gaussian copula, large deviations are not frequent, and the joint frequency of simulated large deviations is also low relative to the number of simulations. A Student copula function can be implemented for increasing the joint frequencies of tail deviations.[6]

6 Extreme VaR techniques are also used to model tail deviations. They differ from simulations in that they attempt to model extreme situations, using a distribution for the fat tail. The distribution is statistically fitted with the highest loss values. A specific distribution is used for fitting the tail, separately from the rest of the distribution. The technique involves "smoothing the tail" to obtain better estimates of the quantiles.

14.3 APPENDIX: CHOLESKY DECOMPOSITION

According to the Cholesky decomposition, N dependent normal variables can be decomposed into linear functions of the N independent normal variables, and the coefficients of the functions are obtained from the variance–covariance matrix of the dependent variables. The decomposition makes it easy to simulate dependent variables by generating K draws of N independent normal variables and constructing the values of the dependent variables as linear functions of these independent variables.

The principle of the Cholesky decomposition is easy to illustrate and detail in the case of two variables. The extension to several variables follows the same process for determining the coefficients. When two dependent variables Y_1 and Y_2 are linear functions of two independent normal variables, X_1 and X_2, the derivation of the coefficients follows a sequential process. The two functions to be specified are:

$$Y_1 = \alpha_{11} X_1$$
$$Y_2 = \alpha_{12} X_1 + \alpha_{22} X_2$$

The issue is to determine the coefficients α_{ij} such that the two new variables, Y_1 and Y_2, are dependent with correlation ρ. The two relations can be written in matrix format. Let us define the matrix L with the coefficients of the linear combinations in rows:

$$L = \begin{array}{|cc|} \hline \alpha_{11} & 0 \\ \alpha_{12} & \alpha_{22} \\ \hline \end{array}$$

The system of two equations is: $Y = LX$. The matrix L is a lower triangular matrix because all items above the first diagonal are zeros. Since the variables Y_i are standardized, they have unit variance, which implies: $V(Y_1) = \alpha_{11}^2 = 1$. This requires $\alpha_{11} = 1$ and $Y_1 = X_1$. The variance of the second variable Y_2 should also be one. The covariance of the variables Y_1 and Y_2 is given. These conditions result in two other equations:

$$\text{Cov}(Y_1, Y_2) = \text{Cov}(\alpha_{11} X_1, \alpha_{12} X_1 + \alpha_{22} X_2) = \alpha_{11}\alpha_{12} = \rho_{12}$$
$$V(Y_2) = \alpha_{12}^2 + \alpha_{22}^2 = 1$$

Since $\alpha_{11}\alpha_{12} = \rho$, α_{12} should be equal to the correlation, ρ, with $\alpha_{11} = 1$. The condition on the variance serves for finding the last coefficient: $\alpha_{22} = \sqrt{1-\rho^2}$. This result shows that the two variables Y_1 and Y_2 are linear functions of standardized independent normal variables X_1, X_2:

$$Y_1 = X_1$$
$$Y_2 = \rho Y_1 + \sqrt{1-\rho^2}$$

In the general case, the Cholesky technique decomposes the square variance–covariance matrix as a product of two triangular matrices. A triangular matrix is such that the off-diagonal terms, on one side of the diagonal, are zeros. The triangular matrix is called "lower triangular", or L, when the zero terms are above the diagonal. It is "upper triangular", or U, when the zero terms are below the diagonal. The elements of the lower triangular matrix are the coefficients of the linear functions linking the dependent variables to the independent variables.

The decomposition process is easily illustrated with only two variables. Consider two normal standardized variables X_1 and X_2, which are independent. The variance–covariance matrix of the variables, Y_1 and Y_2, to be simulated, with correlation ρ, is given:

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

The matrix L is:

$$L = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}$$

It can be checked that the variance–covariance matrix is the product LU where U is the transpose of L:

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \times \begin{bmatrix} 1 & \rho \\ 0 & \sqrt{1-\rho^2} \end{bmatrix}$$

The functions linking the dependent variables to the independent variables can be written in matrix format using L:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \times \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

If X represents the column vector (2, 1) of the two independent X_i variables, and Y the column vector (2, 1) of the dependent Y_j variables, the relation between the two sets of variables is: $Y = LX$.

Generalizing to N variables requires a squared variance–covariance matrix. L and U stand, respectively, for lower and upper triangular matrices. In the triangular decomposition, the U matrix is the transpose of the L matrix: $U = L^T$. By construction, $LU = LL^T$ is the original variance–covariance matrix Σ. The column $(N \times 1)$ vector Y of the Y_i variables is derived from the vector column $(N \times 1)$ vector X of X_i independent variables by the product in that order: $Y = LX$. L is the lower triangular matrix of the α_{ij} coefficients.

The functional matrix relation between the Y_j variables and the X_i variables follows:

$$\begin{bmatrix} Y_j \end{bmatrix} = \begin{bmatrix} L \end{bmatrix} \times \begin{bmatrix} X_i \end{bmatrix}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \end{bmatrix} = \begin{bmatrix} \alpha_{11} & 0 & \cdots \\ \alpha_{12} & \alpha_{22} & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix} \times \begin{bmatrix} X_1 \\ X_2 \\ \cdots \end{bmatrix}$$

The coefficients of the set of linear equations are the unknown. The decomposition of the original variance–covariance matrix follows from matrix algebra rules. The algebraic determination of the unknown coefficients follows the above sequential process. A first equation imposes $\alpha_{11} = 1$. Correlations are imposed sequentially between pairs of variables: for the first pair of variables, for the next pair of variables, and so on. A unit variance constraint also applies for each Y_i.

The linear system of equations is:

$$Y_j = \sum_{m=1,j} \alpha_{im} X_i$$

A first set of equations imposes a unit variance of Y_j, equal to the sum of the unit variances of the X_i weighted by the squares of α_{ij}, for i varying from 1 to j. There are j equations, one for each Y_j, with j varying from 1 to N:

$$V(Y_j) = V\left(\sum_{m=1,j} \alpha_{im} X_i\right) = \sum_{m=1,j} \alpha_{im}^2 = 1$$

The index m varies from 1 to j, the number of independent variables of which Y_j is a linear combination. The index j of Y_j varies from 1 to N, the number of dependent variables equal to the number of independent variables X_i.

A second set of equations imposes the correlation structure ρ_{ij} for all i different from j. The covariance of Y_j and Y_k is:

$$\text{Cov}(Y_j, Y_k) = \text{Cov}\left(\sum_{m=1,j} \alpha_{mj} X, \sum_{j=1,k} \alpha_{jm} X\right) = \rho_{ij}$$

This imposes a constraint on the coefficients:

$$\sum_{j \neq m} \alpha_{jm} \alpha_{jk} = \rho_{ij}$$

The problem is to find the set of α_{jm} for each pair of different values of j and m. The two sets of equations provide the coefficients of L.

15

Volatility

Volatility is a critical measure of uncertainty in risk models and using different estimates of volatilities would result in very different VaRs. Historical volatilities are derived from observed values as standard deviations of returns. This metric of volatility depends on the lookback observation window. Techniques have been implemented to adjust the measure to the current context and use a forecasted measure of volatility. The simplest technique for addressing time-varying volatilities is to use a moving time window of observations. The exponentially weighted average scheme assigns more weight to the more recent observations. A more complex technique consists of modeling the stochastic process of the volatility, as the Garch family of models does.

This chapter details the measure of historical volatility and how alternate techniques, the moving average estimate and the simple version of the Garch model, allow estimating time-varying volatilities.

Contents

15.1 VOLATILITY

The volatility is the standard deviation of a market variable and can be measured by using historical data. The implied volatility is the volatility embedded in the value of options. It is obtained by inverting the Black-Scholes formula, and finding the value of volatility such that the formula matches observed prices. Implicit volatilities look forward by definition, as market prices do, but can be very unstable and dependent on the specific supply and demand of particular options. Risk models rely instead on realized, or observed, volatilities.

Measuring historical volatility over long periods would provide the long-term estimate of volatility, or the "unconditional" volatility. Because volatilities are time varying, such long-term estimate of volatility might not be representative of current conditions and shorter time windows are used. For other periods, the time series of the values of a market return depend on the time window over which observations are collected historically and the frequency of observations. If the distribution of returns were stable over time, volatility would depend only on frequency but not on the look-back period.

The standard deviation calculated over a sample of observations is only an estimate of volatility. The true value of volatility is not observable. The best estimate over a sample is obtained by dividing the sum of squared deviations over n observations by $n - 1$ instead of n, as the theory of sampling implies. The number of observations cannot be too small; otherwise the sampling error would be too large. A common practice is to use daily observations over a year, or around 250 trading days. The number of observations is large enough to reduce the sampling error and the period is short enough to eliminate outdated information.

Measuring daily volatilities is convenient, and volatilities over different periods than a day can be obtained from the square root of time rule. This is in line with intuition, which suggests that the instability of a random variable should increase when the frequency of observations gets lower because the chance of observing large deviations within a longer interval gets higher. The standard market model assumes that the logarithmic returns are identically and independently distributed (i.i.d.). The assumption allows scaling daily volatility over any time horizon, at least as long as the i.i.d. assumption holds.

The rule has limitations. For a long horizon, the volatility derived from the square root of time rule would become infinite, which is unacceptable. Many phenomena are "mean-reverting", such as interest rates. Mean reversion means that, when the deviation from a long-term value grows larger, another random term tends to offset partially such deviations and makes them more likely to revert to zero. Successive returns might also be serially correlated instead of independent. Serial correlation means that high returns tend to be followed by high or low returns. Such correlation would dismiss the square root of time rule. Finally, the underlying volatility can be stochastic.

Market variables have time-varying volatilities, low in some periods and spiking in other times. If the volatility per unit of time cannot be considered constant, the constant variance and the stationary assumptions do not hold. This implies that realized volatilities are conditional on the current conditions.

The simplest way to address time-varying volatilities is to use moving averages. For example, 250 daily observations are used but the window of observations moves as time passes. The technique is simple and allows updating the volatility. The formula of the equally weighted volatility over an observation window uses the squared returns. If σ_t as the volatility of a market variable on day t as estimated from day $t - 1$, the square is the variance σ_t^2. Let r_i be the continuous return between day $i-1$ and day i: $r_i = \ln(S_i/S_{i-1})$. Using a sequence of observations, the equally weighted estimate of the variance is:

$$\sigma_t^2 = \frac{1}{k-1} \sum_{i=1,\,k} \left[r_{t-i} - E(r) \right]^2$$

When considering shocks over short horizons, the distribution of asset value has a drift, which is usually ignored because it is much smaller than the standard deviation. The reason is that the drift is inversely proportional to time while the standard deviation is inversely proportional to the square root of time, hence much larger. For example, consider a process using the following inputs:

Initial price: $S_0 = 100$.
Time interval: $\Delta t = 1$ day.
Expected return: 10% annualized.
Volatility of the random component: 30% annualized.

The expected return over one day is: $10\%/250$ per day $= 0.040\%$. The volatility is: $30\%\sqrt{250} = 1.897\%$ daily, and is much larger. With this assumption, the equally weighted estimate of volatility is:

$$\sigma_t^2 = \frac{1}{k-1} \sum_{i=1,\,k} r_{t-i}^2$$

The square of the return becomes the last estimate of the variance.

Historical volatilities have drawbacks. Assume that there is a spike at some date. As long as the spike is within the observation window, it will affect the historical volatility. As soon as the spike moves out from the windows, the volatility estimate decreases. Such effects might be desirable to the extent that the spike should increase volatility. But the drawback is that the volatility can decrease abruptly when the spike moves out of the window. This limitation results from the equal-weighting scheme of all observations. Other techniques have been designed to better capture the stochastic behavior of volatility.

15.2 EXPONENTIALLY WEIGHTED MOVING AVERAGE MODEL (EWMA)

The exponentially weighted moving average, or EWMA, assigns higher weights to most recent observations. Old spikes in volatility count less than with equally weighted averages since they have a lower weight and the technique allows smoothing out the abrupt variations. The EWMA methodology was introduced with RiskMetrics from JP Morgan.[1] The formula relies on a

1 The technical document was updated in 2006: Zumbach, G. (2007), The RiskMetrics 2006 Methodology, [139].

parameter, λ, within 0 and 1. The formula for estimating variance over the horizon t and moving back from t to date 0 is:

$$\sigma_t^2 = (1r_{t-1}^2 + \lambda r_{t-2}^2 + \lambda^2 r_{t-3}^2 + \ldots + \lambda^{t-1} r_0^2)/(1 + \lambda + \lambda^2 + \ldots + \lambda^{t-2})$$

In the formula, there are t terms, since the value at t depends on all previous squared returns from $t-1$ to 0. The coefficient applied to past observations is lower than for recent observations and the weights decline when moving backward because $\lambda < 1$. When extending the horizon to infinity, the formula simplifies because the sum of a geometric progression with λ is equal to $1/(1-\lambda)$:

$$(1 + \lambda + \lambda^2 + \ldots + \lambda^{t-2} + \lambda^{t-1} + \lambda^t + \ldots) = 1/(1-\lambda)$$

The expressions of σ_t^2 and σ_{t-1}^2 sum up terms from $t-1$ to 0 and from $t-2$ until 0, respectively:

$$\sigma_t^2 = (1r_{t-1}^2 + \lambda r_{t-2}^2 + \lambda^2 r_{t-3}^2 + \ldots + \lambda^{t-2} r_1^2 + \lambda^{t-1} r_0^2)(1-\lambda)$$
$$\sigma_{t-1}^2 = (1r_{t-2}^2 + \lambda r_{t-3}^2 + \ldots + \lambda^{t-3} r_1^2 + \lambda^{t-2} r_0^2)(1-\lambda)$$

Taking the difference between these two series, $\sigma_t^2 - \lambda\sigma_{t-1}^2$, the intermediate terms, except the first term of σ_t^2, r_{t-1}^2, cancel out: $\sigma_t^2 - \lambda\sigma_{t-1}^2 = (1-\lambda) r_{t-1}^2$. The formula of the estimate of the variance of returns as of t becomes:

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda) r_{t-1}^2$$

According to this formula, updating the variance estimate requires only the previous estimate of variance plus one additional observation.

This leads to a simple interpretation of λ. When λ is high, the estimate reacts less to the last recent observation of return r_{t-1}^2, while the previous estimate of variance σ_{t-1}^2 counts more. Conversely, a low λ puts more weight on the most recent observations of return and less weight on the last estimate of variance. Since the weight of the variance term is lower than one, high variances are followed by decreasing variances until a next spike occurs, a scheme that is commonly observed. Finally, any old spike of return will have declining weights as time passes.

EWMA gets closer to the equally weighted estimate in some conditions. At the limit, $\lambda = 1$ makes the EWMA exactly equal to the equally weighted estimate. It is also possible to consider that EWMA is similar to an equally weighted average when the decay parameter is lower than one. Distant lagging terms will quickly fall down to zero making the calculation closer to a historical volatility estimate over a time window, moving forward when time passes. This shows that the method still relies on sampling to estimate the variance and on the stationary assumption. The RiskMetrics document suggests relatively high values of λ, of 0.94 and 0.97. Such values are sufficient to provide estimates significantly different from the equally weighted scheme with daily data.

15.3 GARCH MODELS

The Garch family of models[2] aims at modeling the time behavior of volatility. Volatilities seem to follow a mean-reverting process. High values can be persistent but they tend to fade out after a while. The model attempts to capture these patterns assuming that the variance calculated at t depends on the variance of $t-1$, over the same horizon, plus the latest available observations.

The Garch model differs from EWMA in that it combines a long-term variance with the last estimate of the variance and the last return. Garch (1, 1) indicates that σ_t^2 is based on the most recent observation of the squared return and the most recent estimate of the variance rate, without adding more lagged terms. In other words, (1, 1) refers to the same lag of one period for the squared return and for the variance estimate. The formula for the common Garch (1, 1) model is:

$$\sigma_t^2 = \gamma\sigma_L^2 + \beta\sigma_{t-1}^2 + \alpha r_{t-1}^2$$

The distribution of the return is not identically distributed anymore since its variance now depends on the date t. The variance σ_t^2 is a "conditional variance", dependent upon past observations. The long-term variance is unconditional.

Since the variance changes, sampling returns, weighted or not, cannot be used for estimating the variance. The appropriate methodology for estimating the variance is the maximum likelihood function expanded in the next section.

Comparing the EWMA model with the Garch model, we see that, for EWMA, $\gamma = 0, \beta = \lambda$ and $\alpha = 1 - \lambda$. There is a single parameter, λ, to estimate. Estimating the EWMA or the Garch (1, 1) model relies on the same technique but estimating the EWMA is simpler to implement.

The Garch (1, 1) model is subject to a number of constraints. In the special case where σ is constant and equal to the long run variance $\sigma_L^2 = \sigma$, all terms can be replaced by this long-term variance. This variance is unconditional since it does not depend on any factor that influences the variance at date t. The long-term variance is identical in all terms of the above equation, which results in a first constraint:

$$\sigma_L^2 = \gamma\sigma_L^2 + \beta\sigma_L^2 + \alpha\sigma_L^2$$

Therefore, $\gamma + \beta + \alpha = 1$, and $\gamma = 1 - (\alpha + \beta)$. The constraint $(\alpha + \beta) \leq 1$ is required for the weight of the long-term variance to be positive. It has to be factored in when estimating the coefficients.

The formula can be used iteratively to show that the Garch model is consistent with declining weights for lagged terms as we move backward in time. For example, if we substitute the lag estimate of the variance at $t-1$ as a function of lagged estimates of the variance and the squared return as of $t-2$, the estimated variance at t becomes:

$$\sigma_t^2 = \gamma\sigma_L^2 + \alpha r_{t-1}^2 + \beta\alpha_{t-1}^2 = \gamma\sigma_L^2 + \alpha r_{t-1}^2 + \beta(\gamma\sigma_L^2 + \alpha r_{t-2}^2 + \beta\sigma_{t-2}^2)$$

2 A review of issues in forecasting volatility can be found in Poon, S-H., Granger, C. (2003), Practical issues in forecasting volatility, [115]. A review of the literature by the same authors is in [114, 2003]. The current presentation deals with the simplest form.

The coefficient of the r_{t-2}^2 is $\beta\alpha$ and that of lagged σ_{t-2}^2 is β^2. The weights applied to the squared returns and variance lagging by two periods decline exponentially. The recursive operation shows that the current variance, σ_t^2, becomes a function of lagged squared returns and of variances with various lags higher than one.

15.4 MAXIMUM LIKELIHOOD METHODOLOGY

A fitting method, other than the usual statistical methods, is required for calibrating the Garch family of models. Since EWMA is a special case of Garch (1, 1), the methodology can be used for both models, but is simpler when only the single decay parameter, λ, should be estimated. The maximum likelihood methodology is expanded here and an example follows.

The maximum likelihood methodology involves choosing parameters that maximize the chances of observing the data. Assume that we can assign probabilities to each observation and derive the probability of observing the entire set of data. For example, when observations are independent, the probability of observing k values is the product of the probabilities of observing each one. The maximum likelihood methodology finds the parameters of the probability distribution that best match the data set. The methodology can be illustrated for finding the maximum likelihood estimate of a constant variance, as an intermediary step. This would apply to EWMA.

Assume that the returns r_t have a constant variance σ. For assigning a probability, it is assumed that returns follow a normal distribution with mean zero and variance σ. The density of the distribution is:

$$-\frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-r_t^2}{\sigma^2}\right)$$

The probability of observing the k observations of returns is the product of the probabilities of observing each one, since all r_t are independent:

$$\prod_{t=1,k}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-r_t^2}{\sigma^2}\right)$$

The best estimate of σ^2 is the value that maximizes this expression. Maximizing the product is equivalent to maximizing the logarithm of the product:

$$\ln\left[\prod_{t=1,\,k}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-r_t^2}{\sigma^2}\right)\right]$$

The logarithm is a simple summation, leaving aside the constant $1/\sqrt{2\pi}$:

$$\sum_{t=1,\,k}\ln\left(\frac{1}{\sigma^2}\right)-\sum_{t=1,\,k}\ln\left(\frac{r_t^2}{\sigma^2}\right)=-k\ln(\sigma^2)-\sum_{t=1,\,k}\ln\left(\frac{r_t^2}{\sigma^2}\right)$$

The maximum likelihood problem is:

$$\text{Max}\left(-k\ln(\sigma^2) - \sum_{t=1,\,k}\frac{r_t^2}{\sigma^2}\right)$$

For maximizing the function, the derivative with respect to σ^2 is set to zero:

$$-k/\sigma^2 = -\sum_{t=1,\,k}(r_t^2/\sigma^4)$$

The variance estimate is:

$$\sigma^2 = \frac{\sum_{t=1,\,k} r_t^2}{k}$$

The maximum likelihood estimate is the average of squared returns, which matches the statistical calculation of variance over the entire set of return data.[3]

15.5 ESTIMATING EWMA VOLATILITY

Under the Garch model, the return at date t is normally distributed and the variance σ_t^2 is conditional on time. Since returns are normally distributed and independent, the same equation as above applies, except that the conditional variance changes with time. Observations are indexed with subscript t. The probability of observing k data is:

$$\prod_{t=1,\,k}\frac{1}{\sqrt{2\pi\sigma_t^2}}\exp\left(\frac{-r_t^2}{\sigma_t^2}\right)$$

For estimating the conditional variance, we maximize the likelihood as above:

$$\text{Max}\sum_{t=1,\,k} 1/\sqrt{2\pi\sigma_t^2}\exp\left(\frac{-r_t^2}{\sigma_t^2}\right)$$

The process requires calculating the time series of returns, deriving the squared returns, then the variance estimate. The sequence of calculations is expanded after the empirical findings are graphed in Figure 15.1. The fit of the EWMA model over the French index CAC 40 provides the EWMA variance for the CAC 40 from March 1, 1990 and up to February 25, 2009. Spikes are observed, and the spike for the recent period is much higher. Squared returns measure the updating term in the Garch equation. The graph shows the typical time pattern of variance, with variance declining slowly after each spike. The biggest spike is in the recent crisis period, as expected.

3 Note that the best statistical estimate would be to divide by $k - 1$ instead of k.

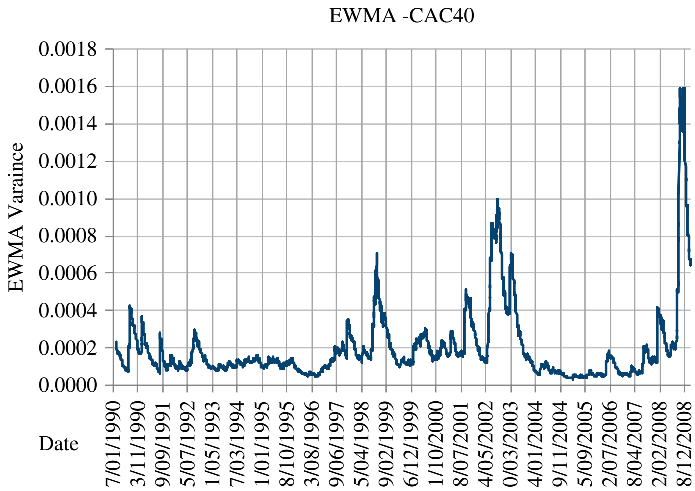


FIGURE 15.1 EWMA estimates of the equity index (CAC 40)

Estimating the model for the recent period might not be very meaningful because of the turmoil of the market created very large deviations. The EWMA fits smoothly to time-varying volatilities because the simple weighting scheme depends on a single parameter.

The process for estimating EWMA is identical to the process for estimating the Garch (1, 1) model. The difference is in the parameters estimated. They are, in the Garch (1, 1) model, the weighted long-term variance $\omega = \gamma\sigma_L^2$ and α and β. The constraints are: $\alpha > 0$, $\beta > 0$ and $(\alpha + \beta) \leq 1$. For comparison, the EWMA model requires estimating the parameters: $\gamma = 0$, $\beta = \gamma$ and $\alpha = 1 - \lambda$. The steps are similar, except for the coefficients. The calculations below are conducted with the EWMA model.

The sequence of calculations is expanded in Table 15.1. The full table looks like the one below, but has longer time series of observations: only the first lines are shown, where all calculations start. Dates are moving forward, starting from the oldest date, March 1, 1990. The final date is February 25, 2009.

First, the daily returns are calculated as logarithmic returns $\ln(I_t/I_{t-1})$, where t is the time and I_t is the equity index value at date t.[4] The first return, as of date 2, $r_2 = (I_2 - I_1)/I_1$, uses the two first observations, I_1 and I_2. This first logarithmic return is calculated in the second line (of column "r_t") as: $r_2 = \ln(1,860/1,832) = 0.015168$. The first estimate of variance is the square r_2^2 of this return. It is the first value in the σ_t^2 column, 0.0001255.

The next lines of this third column use the formulas of the EWMA model. The calculation requires plugging in numerical values of the coefficients. The calculations shown in the table use these best estimates of the parameters, obtained from the maximization of the likelihood function. Under Garch, the formula would be:

$$\sigma_3^2 = \omega + \beta\,\sigma_2^2 + \alpha\, r_2^2$$

4 The logarithmic returns are very close to the proportional discrete returns $(I_t - I_{t-1})/I_t$, since both are very close for small variations.

TABLE 15.1 Calculation of log likelihood function of the EWMA or Garch models

Date	*CAC 40*	r_t	σ_t^2	$-ln(\sigma_t^2) - r_t^2/\sigma_t^2$
		0.000591	0.0001255	13,138.6
3/1/90	1,832	$\ln(r_t/r_{t-1})$	$\sigma^2 = r_1^2$	
3/2/90	1,860	0.015168	$\omega + \alpha r_{t-1}^2 + \beta\sigma_{t-1}^2$	
3/5/90	1,874	0.007499	0.0002301	8.133
3/6/90	1,872	−0.001068	0.0002250	8.394
3/7/90	1,880	0.004264	0.0002185	8.346
3/8/90	1,917	0.019490	0.0002126	6.669
3/9/90	1,921	0.002084	0.0002175	8.413
3/12/90	1,912	−0.004696	0.0002113	8.358

TABLE 15.2 Parameters of the EWMA model

ω	0.00000000
α	0.02918817
β	0.97081183

Under EWMA estimates, the Garch parameters collapse to: $\omega = 0, \beta = \lambda, \alpha = 1 - \lambda$ and the formula is:

$$\sigma_3^2 = \lambda\,\sigma_2^2 + (1 - \lambda)r_2^2$$

The table uses the general Garch notations for the parameters. In this example, α and β have been determined with the log-likelihood function maximized. They are shown in Table 15.2.

These values are used for moving down in Table 15.1 with the estimated values of the volatility. Starting at the third observation, the new estimate of the instantaneous variance, σ_3^2, is calculated. The second value of the estimate of instantaneous variance is: 0.0002301. For all subsequent dates, the same formula is used step by step until we reach the end of the series of observations.

The log-likelihood function sums k times, $\sigma_t^2 - \sum_{t=1,\,k}(u_t^2\;/\sigma_t^2)$, where k is the total number of observations of the instantaneous variance: $-k\ln\left(\sigma_t^2\right) - \sum_{t=1,\,k}(u_t^2/\sigma_t^2)$.

With a total of $t + 2$ observations of returns, the sum adds up t times this log-likelihood function. The value of log-likelihood function appears on top of the last column, as 13,138.6. The corresponding value of the parameter is $\alpha = 0.9708$.[5] This is in line with usual estimates.

5 This value, in Table 15.1, is the best estimate, used for obtaining the numerical values in the table.

Since the function may have several local maxima, the issue is to find the maximum of these maxima. In the example, the solver function of Excel was used, which does not guarantee finding the single maximum.

The same procedure can be implemented for estimating the Garch (1, 1) model, by imposing the following constraints on the maximum algorithm:

$$\alpha \geq 0; \beta \geq 0; \alpha + \beta \geq 0; \omega \geq 0$$

The long-term variance should be positive or zero. It is possible to reduce the number of parameters to be estimated by setting a value for the long-term variance α_L^2.[6]

6 If so, we have: $\gamma\sigma_L^2 = \omega$ and $\gamma = 1 - (\beta + \alpha)\cdot$ Using $\sigma_L^2 = \omega/\gamma$, two parameters remain, with $(\beta + \alpha) < 1$ since $(\beta + \alpha) = 1 - \gamma\cdot$

16

Simulation of Interest Rates

Modeling the term structure of interest rates is a critical input for the value-at-risk (VaR) of fixed-income portfolios. It is also a sensitive issue for managing the interest rate risk of the banking book. Interest rate models used for pricing purposes are not so helpful for such simulations.[1] The statistical methodology based on principal component analysis (PCA) is convenient because of its simplicity and because it is capable of modeling the entire term structure of rates with a very small number of factors.[2]

Under a market VaR perspective, using standard factor models requires the simulation of term structures of interest rates, which embed the consistency constraints across interest rates. Under an asset-liability management (ALM) perspective, the issue is to explore what could happen with simple scenarios other than standard shifts of all interest rates. The simulations of the term structure interest rates provide the faculty of investigating the outcomes of a comprehensive set of scenarios, which can include back-tests or stress-tests.

This chapter explains how simulations of interest rates can be conducted using the PCA, and how the technique applies to market VaR and ALM policies. The simulation of the principal factors is used to model the distribution of the profit and loss (P&L) of a fixed income portfolio and derive its VaR. The same model of interest rates is also used for modeling the distribution of the net interest income of the banking book.

1 There are many models of interest rates, which need to be calibrated on market data if used for pricing purposes. The purpose of this chapter is not to provide an overview of all models. For interest rate models, see Hull, J. (2014), Options, Futures and Other Derivatives, [78], or Brigo, D., Mercurio, F. (2007), Interest Rate Models – Theory and Practice: With Smile, Inflation and Credit, [36].

2 The method is presented in: Frye, J. (1997), Principals of risk: Finding VaR through factor-based interest rate scenarios, in VaR: Understanding and Applying Value at Risk, [65].

Contents

16.1 INTEREST RATES AND FACTOR MODELS

The PCA serves to explain a set of highly correlated variables by a small number of factors, hence is well adapted to interest rates. The factors are called principal components. The number of factors that contribute significantly to the explained variance decreases with the correlations between variables. Unlike the usual factor models, PCA relies on independent factors, which are much easier to handle. Orthogonal factor models have the same linear form as the general factor model. Any standard factor model can be transformed into an orthogonal factor model with the same number of factors because the principal components are linear combinations of standard factors.

The generic form of the PCA models is:

$$Y_i = \eta_{0i} + \eta_{0i}P_1 + \eta_{0i}P_2 + \ldots + \varepsilon_i$$

The P_i are the principal components and the η_{0i} are the factor loadings.

A factor model allows the sensitivities of a portfolio with respect to the factors, instead of interest rates, to be determined. The volatility of the return of a fixed income portfolio, or the volatility of interest earnings of a banking book, is a function of the variance–covariance matrix of the factors. With the independent factors of the PCA technique, the variance–covariance matrix of factors collapses to a diagonal matrix with all off-diagonal terms equal to zero, which simplifies greatly the derivation of volatilities.

A major application of PCA is the modeling of the term structure, because it is observed that very few factors, usually no more than three, are sufficient for explaining most of the variance of interest rates. It is not obvious, in general, to provide an interpretation of what these factors represent. The interpretation depends on how the factors relate to explained variables. For example, with interest rates, a factor that tends to increase all interest rates would represent a parallel shift of the term structure.

16.2 MODELING THE TERM STRUCTURE OF INTEREST RATES WITH PRINCIPAL COMPONENT ANALYSIS

For presenting the technique, an example is fully detailed. As usual for factor models, the historical data are the starting point of the analysis. The observation window should be selected according to the goal. For market VaR, recent periods are most relevant. For ALM applications, it might make sense to uses various periods with different behaviors of the term structure of rates, such as for stress-testing purposes.

The data set used in the example includes 911 daily observations of euro interest rates for one month, two months, six months, nine months, and 1, 3, 5, 7, 10, 15, 20 and 30 years. These are zero rates. The different rates are designated as: E_m1, E_m2, E_m3, E_m6, E_m12, E_y1, E_y3, E_y5, etc., where "m" stands for month, "y" for year, and "E" represents the euro currency. Observations are from January 2, 2004 until July 12, 2007, a period of roughly 3.5 years. Over the period, the yield curve was upward sloping at the beginning of the period and progressively flattened out in the last years. The period is characterized by a high volatility of short-term rates and the flattening of the yield curve. Therefore we do not expect to have any "bump" in the curve, but we should capture its variations of level and steepness over the observation period.

The correlations between the interest rates and the short-term rate E_m1 are higher than 0.9 up to the three-year maturity, with long-term rates having a declining correlation with E_m1, from 0.847 for E_y5 down to 0.064 for E_30y. Long-term rates vary almost independently of the short-term rates, while intermediate interest rates co-vary strongly with it.

For implementing PCA analysis, standard statistical techniques provide the coefficients of the model, or factor loadings, for normal standard factors. The explained variables are the 11 interest rates selected. The PCA analysis provides a total of 11 factors, but, in this example, only the first two factors contribute significantly to explaining the variance. The factors are numbered 1 to 11 and ranked according to the percentage of variance explained. Table 16.1 shows the first five factors, with the two first factors only explaining most of the variance of interest rates. The two main principal components explain 99.31% of the variance of the interest rates: P_1 explains 81.39% and P_2 17.91% of the total variance.

The interpretation of principal components is based on the coefficients of the factors for the various interest rates, provided in Table 16.2. For each factor, there are as many coefficients as there are interest rates. The factor loadings for each interest rate are those of the two main principal components only, as shown in the columns of the table.

The first factor is positively correlated with all interest rates. A shift of this factor would increase all interest rates, hence represents a parallel shift of all rates. The second factor has negative correlations with the short-term rates and positive correlations with long-term rates. A positive shift of that factor would increase the slope of the curve, making short-term rates lower and long-term rates higher. The factor represents the slope of the term structure.

In general, when using long periods, three factors account for the changes of the term structure. The first two represent a shift of the curve and a change of its slope and the third factor represents the "bump" of the term structure of the curve. The "bump" factor correlates positively with intermediate maturity interest rates and negatively with the short end and the long end of the curve. In this example, the third "bump" factor does not appear.

TABLE 16.1 Applying PCA to the term structure of interest rates

Factor	*% of variance*	*Cumulative %*
1	81.392	81.392
2	17.914	99.306
3	0.632	99.938
4	0.044	99.983
5	0.012	99.994

TABLE 16.2 Factor loadings of the principal components

Factor	*1*	*2*
E_m1	0.1840	−0.0872
E_m3	0.2079	−0.1065
E_m6	0.2343	−0.1220
E_m12	0.2492	−0.1102
E_y3	0.1186	0.0709
E_y5	0.0431	0.1449
E_y7	0.0060	0.1721
E_y10	−0.0195	0.1866
E_y15	−0.0377	0.2016
E_y20	−0.0489	0.2155
E_y30	−0.0590	0.2272

The mean and standard deviations of the two factors, P_1 and P_2, are obtained by fitting the model to historical data. All variables of the model are standardized, both the factors and the interest rates. For example, the three-month standardized interest rate value is 0.2079 times the standardized P_1 and −0.1065 times the standardized P_2 (second row of the table). These factor loadings serve for determining the variations of each standardized rate as a function of the factor variations. The same process applies with all rates, making them a function of only two principal components.

16.3 INTEREST RATE SIMULATIONS

The PCA allows simulating the values of interest rates. The variations of all interest rates are linear functions of the variations of the two first factors, ΔP_1 and ΔP_2, ignoring residuals since the two factors explain most of the variance of interest rates. For example, considering the 3-month Euribor, E_m3, its standardized value is a linear function of the factors:

$$\text{E\_m3(standardized)} = 0.2079P_1 - 0.1065P_2$$

Applying similar relationships between each interest rate and the same common two factors, all standardized interest rates are linear functions of P_1 and P_2, with coefficients given in Table 16.2. Generating random variations of P_1 and P_2 generates random variations of all 11 standardized interests. The technique allows generating scenarios for the entire term structure of the 11 interest rates by varying only two factors randomly.

Generating random values of P_1 and P_2 is simple since those two factors are independent and standardized principal components. The usual process requires the generation of two series of random values of a uniform standard function $U(0, 1)$ and taking the normal standard inverse function of these random values. For moving back from factor values to interest rates, it is necessary to transform the standardized values of interest rates into

TABLE 16.3 Mean and standard deviation of interest rates

	μ	σ
E_m1	2.612	0.678
E_m3	2.687	0.714
E_m6	2.776	0.756
E_m12	2.922	0.790
E_y3	3.306	0.639
E_y5	3.554	0.521
E_y7	3.751	0.454
E_y10	3.974	0.410
E_y15	4.200	0.395
E_y20	4.320	0.397
E_y30	4.384	0.401

unstandardized values,[3] using the historical estimates of mean and standard deviation of each rate (Table 16.3).

Table 16.4 reproduces the first five simulations.[4] For example, in the first simulation run the random values of the independent P_1 and P_2 are 0.340 and −1.313. The standardized value of 3-month Euribor results from the linear function with principal components:

$$\text{E\_m3} = 0.2079P_1 - 0.1065P_2$$

The standardized value of the 3-month Euribor is obtained by replacing the factors by the first couple of simulated values of P_1 and P_2. The unstandardized value of the 3-month Euribor is:

$$\text{E\_m3} = (0.2079P_1 - 0.1065P_2)(0.714) + 2.687$$

TABLE 16.4 Simulations of the principal components

Simulations	P_1	P_2
1	0.340	−1.313
2	−0.464	−0.335
3	1.533	−0.898
4	−1.471	2.871
5	0.837	0.399

3 The standardized variable derived from the non-standard variable X is $X_S = (X - \mu)/\sigma$ since X_S has mean zero and σ equal to 1. Reversing the equation, we obtain the non-standard variable $X = X_S\sigma + \mu$.

4 As usual, two independent uniform standard variables are generated and converted into standard normal variables using the normal standard inverse function.

TABLE 16.5 Simulations of interest rates from PCA

P_1	P_2	E_m1	E_m3	E_m6	E_m12	E_y3	E_y5	E_y7	E_y10
0.340	−1.313	2.732	2.838	2.957	3.103	3.272	3.462	3.649	3.871
1.085	−0.325	2.766	2.873	2.998	3.164	3.374	3.554	3.728	3.940
1.252	2.745	2.606	2.663	2.744	2.930	3.526	3.789	3.969	4.174
0.380	1.244	2.586	2.648	2.728	2.889	3.391	3.656	3.849	4.066
−1.226	−2.128	2.585	2.667	2.755	2.866	3.116	3.366	3.581	3.821

The first term within parentheses is the standardized value of the first component. The unstandardized value is obtained by multiplying this value by the volatility of the 3-month Euribor and adding its mean. The simulated value of the 3-month Euribor, in percentage, is calculated by replacing P_1 and P_2 by their values, respectively, 0.340 and −1.313:

$$\text{E\_m3} = (0.2079 \times 0.340 - 0.1065 \times -1.313) \times 0.714 + 2.687 = 2.838$$

Repeating the process for all simulated values of the factors and for all interest rates generates a series of term structures of these rates (Table 16.5). The five first simulations of the rates, E_m1 to E_y10, are in rows in the table.[5] The simulated factor values are in the first two columns and the three-month interest rate is the column labeled "E_m3". Figure 16.1 shows the first 20 random term structures of interest rates.

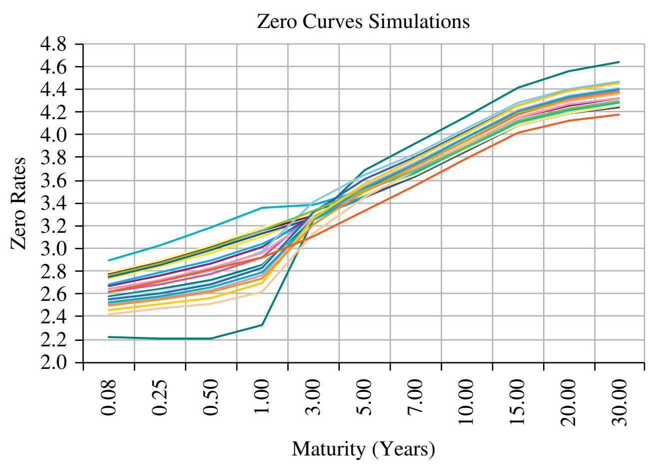


FIGURE 16.1 Simulation of zero interest rates

5 Since the PCA analysis relies on normal distributions, the calculation could result in negative interest rates. Such values should be eliminated by setting a floor to zero, although this is an ad hoc method.

16.4 APPLICATION TO MARKET VaR

For calculating a VaR of a fixed income portfolio, the assets are mapped to the relevant factors, the sensitivities to these factors are determined and the variations of value of the portfolio are simulated by generating a series of term structures, as above. It is assumed that the non-zero sensitivities are those relative to the three-month rate, the five-year rate and the 10-year rate. The sensitivities to interest rates represent the change of the position value due to a unit shock of the interest rate. The sensitivities to factors are the sensitivities to interest rates weighted by the products of factor loadings and interest rate volatility.[6] The calculation of the sensitivities to the factors is decomposed in Table 16.6. The table shows the factor loadings of the two principal components for each interest rate, the sensitivity of the portfolio to each interest rate and the weighted sensitivities of these positions to the two factors in the last two columns.

The sensitivities of the portfolio to interest rates are replaced by the sensitivities to the principal components. The sensitivities to the factors of the entire portfolio are S_{P1} and S_{P2}, respectively equal to -977 and $1,312$. They are obtained by summing up the sensitivities to factors over the three positions of the portfolio. These two weighted sensitivities are sufficient for calculating the variations of the P&L according to the linear equation:

$$\Delta(\text{P\&L}) = S_{P1}\,\Delta P_1 + S_{P2}\,\Delta P_2 = -977\,\Delta P_1 + 1,312\,\Delta P_2$$

A distribution of the P&L of the portfolio can be obtained by simulating a series of random shocks on the factors. For generating standard normal factors, two independent uniform standard variables are simulated. The normal inverse of the values of uniform variables are the shocks on factor values. The random variations on the P&L of the portfolio derive from the above equation for each pair of random values of ΔP_1 and ΔP_2. The first five simulations, out of 100, are shown in Table 16.7.

The distribution of the daily variations of the P&L can be generated from 100 simulations (Figure 16.2). The distribution would get closer to the normal distribution with a larger number of simulations. The mean, the standard deviation and the loss quantiles derive from the distribution of the P&L. The variance of the P&L is the sum of the squared variances of the

TABLE 16.6 Portfolio sensitivities to interest rates

	Factor loading for P_1	*Factor loading for P_2*	*Sensitivity to interest rate*	*Volatility of interest rate*	*Sensitivity to factor P_1*	*Sensitivity to factor P_2*
E_3m	0.208	−0.107	−10,000	0.7140	−1,485	764
E_y3	0.119	0.071	7,000	0.6389	532	318
E_y10	−0.020	0.187	3,000	0.4102	−25	230
		Weighted portfolio sensitivities			−977	1,312

6 The multiplication by the volatility of the interest rate is required to obtain a non-standardized variation of the position since the factors apply to standardized interest rates.

TABLE 16.7 Simulations of the portfolio P&L

Simulations	P_1	P_2	*P&L*
1	2.073	0.219	−1,738.2
2	1.212	−1.089	−2,612.1
3	−1.568	0.167	1,751.7
4	0.639	0.971	649.2
5	−1.239	−0.065	1,126.9

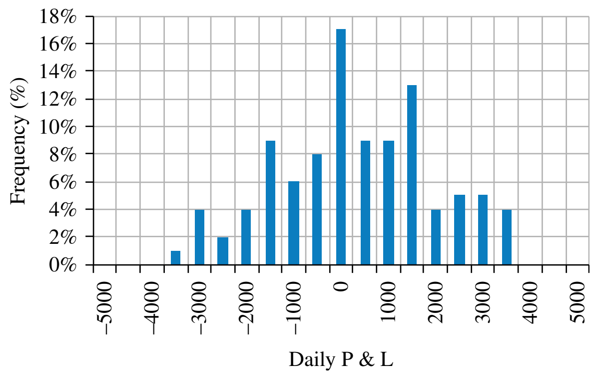


FIGURE 16.2 Distribution of the portfolio daily P&L

factors weighted by the squared sensitivities since the factors are independent:

$$\sigma^2(\text{P\&L}) = S_{P1}^2\sigma^2(\Delta P_1) + S_{P2}^2\sigma^2(\Delta P_2)$$

It simplifies into the sum of the squared sensitivities since the principal components have a unit variance. The daily volatility of the P&L is the square root:

$$\sigma(\text{P\&L}) = \sqrt{S_{P1}^2 + S_{P2}^2} = 1,636$$

The portfolio delta-normal VaR at confidence level 1% would be $-2.33 \times 1,636 = -3,812$, ignoring the mean of the daily P&L.

16.5 SIMULATIONS OF INTEREST RATES FOR THE BANKING BOOK

For ALM purposes, the simulations of interest rates allow capturing long-term movements of the interest rates and simulating the net interest income or the economic value. Such simulations allow to model the effect of historical scenarios on the net interest income or the sensitivity of the value of the balance sheet. The same process can be used for back-testing against historical movements, or for stress-testing, by plugging in scenarios matching crisis periods. The financial managers can select the observation period for modeling interest rates with the PCA analysis. A short observation period ending today would match the current conditions. Another period of unstable interest rates would better fit stress-testing needs.

For modeling the effect of shocks of interest rates on the net interest income, the balance sheet exposures are mapped to interest rates. The scenarios for the curve of interest rates are simulated using the factor model. Finally, the distribution of the net interest income is calculated from the sensitivity of the banking book to the interest rates. The "earnings-at-risk" metric is analogous to VaR for earnings. It is a quantile of the distribution of earnings at a given confidence level. The quantiles are derived from the distribution and represent the worst-case adverse deviations of earnings at selected confidence levels.

In the following example, the balance sheet has exposures to the short-term 3-month Euribor, for short-term deposits, and the three-year interest rate for loans. The example is representative of a bank having a structural position, lending long with short-term resources, and having mismatch risk. The sensitivities of the net interest income to these interest rates are measured by the standard interest rate gaps, supposed to be $-10,000$ and $+7,000$, for the short and the three-year interest rates, respectively.

The scenarios are generated by simulating values of the independent principal factors, P_1 and P_2. Since interest rates are a linear function of the factors, it is easy to move from the sensitivities of earnings to interest rates to the sensitivities of earnings to factors. For net interest income, daily volatilities are not relevant, and annual volatilities of interest rates are used instead. The scaling factor required for moving from the initial daily volatilities to annual volatilities is: $\sqrt{250} = 15.811$. For example, the annual volatility of the interest rate E_m3 is: $15.811 \times 0.714 = 11.29$. Table 16.8 shows the factor loadings of the two principal components for each interest rate, the sensitivity of the portfolio to each interest rate and the weighted sensitivities of these positions to the two factors in the last two columns.

The sensitivities to the factors of the entire portfolio are S_{P1} and S_{P2}, respectively equal to $-15,066$ and $+17,100$, obtained as the algebraic summation of the sensitivities to the factors over the portfolio. The deviations of the net interest income are expressed as a linear function of the variations of the factors, using the numerical values of the sensitivities to the factors:

$$\Delta(\text{NII}) = S_{P1}\,\Delta P_1 + S_{P2}\,\Delta P_2 = -15,066\,\Delta P_1 + 17,100\,\Delta P_2$$

TABLE 16.8 Sensitivities of the ALM portfolio to the factors

	Factor loading for P_1	*Factor loading for P_2*	*Sensitivity to interest rate*	*Volatility of interest rate*	*Sensitivity to factor P_1*	*Sensitivity to factor P_2*
E_m3	0.208	−0.107	−10,000	11.29	−2,3481	12,079
E_y3	0.119	0.071	7,000	10.10	8,415	5,021
		Weighted portfolio sensitivities			−15,066	17,100

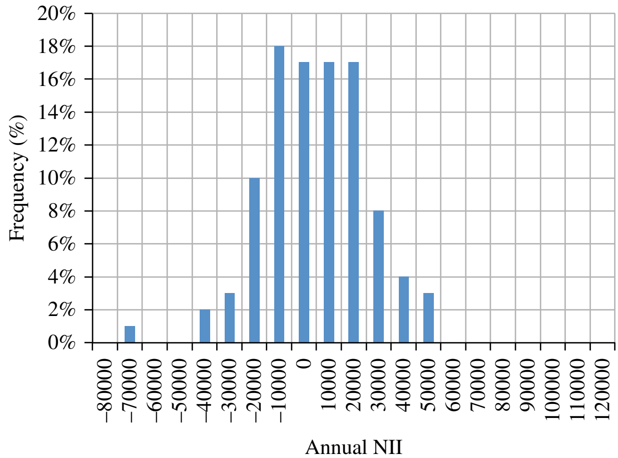


FIGURE 16.3 Distribution of ALM NII

The variance of the net interest income (NII) follows as the sum of squared sensitivities:

$$\sigma(\text{NII}) = \sqrt{S_{P1}^2 + S_{P2}^2} = 22,790$$

The simulation of the distribution of annual NII is shown in Figure 16.3.

The VaR of the NII, or earnings-at-risk, at the 1% confidence level, would be: $2.33 \times 22,790 = 53,101$. The cumulative probability that the NII is negative is around 31%.

17

Market Risk Regulations

The market risk regulations were introduced in 1996–1997. They are being revised from successive regulations. A "Revision to the Basel 2 Market Risk Framework" was published in February 2011, as a number of measures to make banks more resilient to shocks in the banking system and financial markets were introduced. In 2013, the Basel Committee published a consultative paper for an overhaul of the market risk framework. The rules are not yet enforced, but the proposal outlines the directions that the Basel Committee is aiming at.

This chapter reviews the market risk regulations, from the Basel 2 comprehensive accord of 2006 up to the proposed update, which introduces radical changes to the framework.

Contents

17.1 THE MARKET RISK REGULATIONS AS OF JUNE 2006

The initial market risk accord of 1996–1997 was revised in the Basel 2 document: "International convergence of capital measurement and capital standards: A revised framework – Comprehensive version", published in June 2006.[1]

Market risk is defined as the risk of losses from balance sheet and off-balance-sheet positions arising from the movements of market price. The document refers to the positions in

1 The comprehensive 2006 Basel 2 Accord is in [21].

fixed income instruments and equities of the trading book and to the foreign exchange and commodities risk. The trading book boundary is based on the management intention. Two approaches are allowed: the standardized approach and the internal model approach.

17.1.1 The Standardized Approach

The standardized methodology is differentiated across instrument classes, interest rates, equity, foreign exchange and commodities.

In the interest rate and equity classes, the specific risk and the general market risk are calculated separately. The specific risk is defined as supervisory percentages of exposures by rating class or for unrated positions. For example, unrated positions would receive an 8% capital charge for specific market risk.

For interest rate positions, the methodology for general market risk uses a ladder approach in which positions are mapped to bands of maturity or duration. Positions within a slot are weighted, then weighted long and short positions within a band can be offset. The regulatory weights account for the differing sensitivities of positions to interest rates. Some offsetting is allowed across bands. Derivatives are replicated by a combination of long and short positions, such as an interest rate swap that has receiving and paying legs, and replaced by these positions.

Equity positions are subject to general and specific risk capital charges. The capital charge for general risk allows offsets of long and short positions. The specific risk is based on the total equity exposure without netting long and short positions (gross exposure), unless the positions are exactly matched. The capital charge for specific risk is 8% of the exposure, or 4% for a liquid and diversified portfolio. There is no charge for specific risk for derivatives.

For foreign exchange positions, the exposure is broken down by currency. The capital charge is 8% of the overall net open position by currency. For commodities, the risk combines the spot price risk, the basis risk and the interest rate risk. Long and short positions are slotted to time bands, and regulatory capital charges apply to net positions of each time band.

17.1.2 The Internal Model Approach

The internal model approach is allowed for market risk, based on value-at-risk (VaR) measures of risk. The methodology was refined in the comprehensive document of 2006.

The focus of most internal models is a bank's general market risk exposure. Specific risk, the market risk of exposures to specific issuers of debt securities or equities, is measured largely through separate risk measurement systems. A separate capital charge for specific risk applies to banks using a model to the extent that the model does not capture specific risk.

The principles for the usage of internal proprietary market risk models are as follows. Market risk is the risk of loss during the minimum period required to liquidate transactions in the market. The regulatory liquidation period is 10 days. The loss is the 99th loss percentile (one-tailed) for market risk VaR models. The 99% refers to the probability that losses are lower than VaR in absolute value, or, equivalently, that the probability that the loss exceeds the VaR is 1%. A multiplication factor, between 3 and 4, applies to this modeled VaR. It accounts for potential weaknesses in the modeling process or exceptional circumstances.

Banks calculate a daily VaR and extend the horizon to 10 days by using the square root of time rule. The 10-day VaR is derived from the daily VaR by applying a constant scaling factor, the square root of 10, to the daily variations of values of instruments.[2]

Models should incorporate historical observations over at least one year. The VaR is calculated over the most recent data. The daily VaR is VaR_t and its average over the last 60 days is VaR_{avg}. The VaR retained for capital charge is the maximum of these two VaRs:

$$\text{Capital} = m \times \max(\text{VaR}_t, \text{VaR}_{\text{avg}})$$

The formula prevents a sudden decline of VaR to serve as a base for the calculation of capital. The coefficient m is a multiplier between 3 and 4, which is based on a back-test of the model. Reliable models get a "premium" in capital with a lower multiplier of VaR. The multiplier is imposed based on the number of P&L observations that exceed the VaR over the period of observation. The number of outliers, the daily losses exceeding the VaR, is obtained from a back-test of the VaR.

17.1.3 Back-testing Value-at-Risk

Back-testing is the process used for checking that the number of daily losses in excess of VaR is, or is not, in line with the quantile of the VaR model. With a quantile of 1%, daily excess losses observed over the 250 trading days within a year should be around two to three.

For implementing a back-test, the portfolio composition is fixed and daily shocks of risk factors are imposed, resulting in daily P&Ls, measured as variations of values from one day to the next. These daily variations are compared to the VaR, with the confidence level 1%. If the model is accurate, two or three losses should be observed in excess of the modeled VaR. In practice, the number of excess losses can vary around these expected numbers. If there are many more exceptions than two or three, chances are that the model is not accurate.

Basel defines threshold values of the number of exceptions for qualifying the model as more or less accurate. The values of the multipliers applied to the internal VaR calculations are mapped to these threshold values, and increasing in steps with the number of exceptions observed in the back-test.

A series of N daily observations provides the total number of occurrences, X, such that VaR is exceeded by daily (negative) P&Ls. The number is a realization of a binomial variable depending on q, the quantile used for VaR. The null hypothesis is that the model has the desired coverage. For testing that the model is accurate, upper and lower bounds of values of X should be defined, such that the probability of falling outside the range is lower than some low probability.

A type 1 error occurs if an accurate model is rejected, and a type 2 error occurs if accepting an inaccurate model as a correct model. Type 1 and type 2 errors depend on the number, X, of outliers. The Basel reference document provides the chances of observing various particular numbers of outliers with an exact model for 250 observations, and a 1%-quantile for VaR. The probability of observing five or fewer outliers with an accurate model is 89.2%. The probability

2 The rule is acceptable for linear instruments. For non-linear instruments, an adjustment should theoretically apply because the 10-day variations of values are not a linear function of the one-day variations of value.

of falsely rejecting an accurate model with 250 independent observations with five outliers is the complement to 1, or 10.8%. The number of exceptions serves to classify the bank's models into regulatory-defined zones, called green, yellow and red zones, which are mapped to the value of the multiplier applied to VaR for determining the capital charge.

17.2 REVISION TO THE MARKET RISK FRAMEWORK (2011)

The revision to the market risk framework of February 2011[3] introduced the stressed VaR (S-VaR) and the incremental risk capital charge. Banks need to add to a standard VaR another S-VaR, calculated over a one-year observation period relating to significant losses. The incremental risk charge is an additional capital required for losses due to migration and default risk of credit products subject to the market risk framework. It is introduced because the VaR modeled from the movements of market risk factors does not include default and migration losses.

17.2.1 Internal Model Approach

A stressed VaR component (S-VaR) is added to the standard VaR calculation, for defining the capital charge under the internal model approach. The additional S-VaR component should allow the cyclicality of market risk capital charges to be reduced and help to ensure that regulatory capital is sufficient in periods of significant market stress. The standard VaR calculations based on recent observations are procyclical because they are based on the price volatility of underlying assets using historical data. As the volatility expands, the VaR and the capital charge also do, making capital charges cyclical.

The S-VaR is computed on a 10-day 99% confidence basis, but with inputs taken from times of significant financial stress relevant to the firm's portfolio. The capital requirement for general market risk is based on the maximum VaR calculation between the last date calculation (VaR_t) and the average VaR over 60 days preceding the current date (VaR_{avg}), according to the following formula:

$$\text{Capital} = \max(\text{VaR}_{t-1}, m_c\ \text{VaR}_{\text{avg}}) + \max(\text{S-VaR}_{t-1}, m_s\ \text{S-VaR}_{\text{avg}})$$

The multiplication factors applicable to the VaR and to the S-VaR should be at least 3, and higher, depending on how the VaR model performs, a performance assessed by back-testing the unstressed VaR.

17.2.2 Securitizations and Credit Products

The securitizations positions have a specific risk charge, which is calculated as if they were in the banking book, using the rating-based approach or the supervisory formula from the Basel 2 credit framework (Chapter 26). A new correlation trading portfolio is introduced for credit default swaps and securitizations other than retail portfolios. The minimum capital requirement

3 Basel Committee on Banking Supervision (2011), Revisions to the Basel II market risk framework, [25].

is expressed in terms of two separately calculated charges, one applying to the specific risk of each security and the other to the interest rate risk in the portfolio (general market risk) where long and short positions in different securities or instruments can be offset.

17.2.3 Specific and Incremental Risk Charge

In 2009,[4] the VaR for the trading book was supplemented with an incremental risk capital charge, which includes default risk and migration risk for credit products unrelated to securitizations.

The incremental risk charge (IRC) is intended to complement the VaR modeling framework. The original VaR framework ignores differences in the underlying liquidity of trading book positions and does not fully reflect large daily losses that occur less frequently than two to three times per year, or the potential for large cumulative price movements over periods of several weeks or months.

The IRC represents an estimate of the default and migration risks of credit products, other than securitizations, over a one-year horizon and depending on the liquidity horizons of individual positions. The IRC captures default risk, losses from credit migrations, widening of credit spreads and the loss of liquidity. The regulators did not specify how the IRC should be modeled. Banks are expected to develop their own models for the calculations. The methodology of credit VaR, along the CreditMetrics™ model (Chapter 18), can be used for deriving the IRC.

The IRC model is based on an assumption of a constant level of risk over the one-year capital horizon. This constant level of risk assumption implies that a bank rebalances, or rolls over, the trading positions over the one-year horizon in a manner that maintains their initial risk level. Positions whose credit standing migrates over the liquidity horizon are replaced with positions with the same risk as the original positions. The frequency of the rebalancing is based on the liquidity horizon for a given position.

17.2.4 Overall Capital Charge Revisions

When all additions to the framework for market risk are considered, the capital charge should include the VAR plus a measure of specific risk plus the incremental capital charge:

$$\text{Capital} = m \times \text{VaR} + m \times \text{Stressed VaR} + \text{Specific risk} + \text{IRC}$$

17.2.5 Counterparty Credit Risk and Credit Value Adjustment (CVA)

Under the Basel 3 global regulatory framework of 2010 [24], banks are also subject to a capital charge for potential mark-to-market losses associated with the deterioration in the credit-worthiness of the counterparties of the bank. The addition refers to the counterparty credit risk of derivatives (Chapter 22).

4 Basel Committee on Banking Supervision (2009), Guidelines for computing capital for incremental risk in the trading book, [23].

The CVA is the difference between the risk-free value and risky value of a derivative trade. The Basel 2 standard covers the risk of a counterparty default; but it does not address the risk of variations of the CVA. The Basel Committee observed that two-thirds of the losses on over-the-counter instruments, during the 2008 crisis, appeared to be due to adverse migration risk, in the absence of default.[5] The losses of value were observed when the creditworthiness of players was significantly impaired.

The new framework imposes a capital buffer against adverse deviations of the credit adjustment, to be modeled as a CVA VaR, which is derived from the variation of credit spreads. The CVA VaR is added to the market VaR. As the CVA definition relates to the counterparty credit risk, the definition of CVA and the regulations for the CVA VaR is deferred to Chapter 22 on counterparty credit risk.

17.3 REVISIONS TO THE BASEL 2 MARKET RISK FRAMEWORK

In 2013, the Basel Committee published a consultative paper for a radical revision of the market risk framework: "Consultative document: Fundamental review of the trading book: A revised market risk framework", October 2013 [27].

The document is consultative, but the proposal outlines the orientations of the Basel Committee. The proposal addresses a number of major issues: the definition of the trading book, the calibration of both standardized and internal model approaches to stressed periods; the incorporation of market risk illiquidity; a new standardized approach, with a better convergence with the internal model approaches; and a shift from the traditional VaR measure to a new unique stressed expected shortfall measure.

17.3.1 The Trading Book/Banking Book Boundary

The boundary between the trading book and the banking book is being revised. Traditional boundaries are defined in terms of management intent or in terms of valuation principle. The new regulation considers an instrument-based approach and would provide explicit guidelines on which instruments should be included in the trading book.

17.3.2 Treatment of Credit-related Products

The calibration of capital charges against default risk in the trading book will be more closely aligned to the banking book treatment. The standardized method remains the only method used to capture the risk of securitizations. Internal models will continue to be allowed for other exposures but will be subject to a separate incremental default risk (IDR) charge. The CVA risk capital charge will continue to be calculated as a standalone capital charge.

17.3.3 Approach to Risk Measurement

The regulation will move to the calibration of the internal model approach and the standardized approach on periods of market stress. It also proposes to move from the VaR metric to the

5 Basel Committee on Banking Supervision (BCBS) (2011), "Capital treatment for bilateral counterparty credit risk finalized by the Basel Committee".

expected shortfall metric. The expected shortfall measures risk by considering the size and the likelihood of losses above a certain confidence level. The confidence level will be 97.5% for the expected shortfall under the internal model-based approach. Instead of using two metrics, VaR and S-VaR, as in Basel 2.5, the revised framework would rely on a single stressed metric.

17.3.4 Incorporation of the Risk of Market Illiquidity

The liquidity horizon would be integrated in the market risk metrics. A liquidity horizon is defined as "the time required to execute transactions that extinguish and exposure to a risk factor, without moving prices in stressed market conditions". The liquidity horizon should be related at the level of risk factors, with five liquidity horizon categories. Capital add-ons against the risk of jumps in liquidity premium would be introduced for desks trading illiquid complex products.

17.3.5 Treatment of Hedging and Diversification

The current regulatory capital framework is viewed as too reliant on banks' internal models. The current discrepancies between the capital charges under the internal model-based approach and the standardized approach make it difficult to fall back on the standardized approach when internal models have a poor performance. The Basel Committee proposes to narrow down the differences between the capital charges resulting from the two approaches. A mandatory disclosure of standardized capital charges would be imposed, on a desk-by-desk basis. The possibility of imposing a floor equal to the capital of the standardized approach is considered.

17.3.6 Revised Standardized Approach

A revised standardized approach should achieve three objectives: provide a method to calculate capital requirements; provide a credible fall-back option in the event that a bank's capital is deemed inadequate; facilitate transparent and consistent reporting of market risk across banks.

Under the revised standardized approach, market positions are allocated to asset classes, which include: general interest rate risk, foreign exchange risk, equity, commodities, credit spreads and default, with distinct treatments for securitizations and non-securitizations. Within each asset class, there are buckets grouping instruments with close risk characteristics. Any instrument can be decomposed in individual exposures by asset class and by bucket. A single risk weight will apply to all positions assigned to a risk bucket. Hedging and diversification benefits are captured through regulatorily-determined correlation parameters defined for each bucket. The correlations are lower for positions with opposite signs than for positions with the same sign. This approach prudently captures the risk to perceived hedging and diversification benefits that arises due to the unstable and time-varying nature of correlation parameters, particularly in times of stress.

For example, a bond would be decomposed in several positions referring to different risk factors. The value of a bond is exposed to default risk. The bond cash flows are also mapped to the various interest rates and to the credit spreads. The present values of the bond cash flows are assigned to vertex points (0.25yr, 0.5yr, 1yr, 2yr, 3yr, 5yr, 10yr, 15yr, 20yr and 30yr), separately for each currency. Where a cash flow falls between two vertices it is assigned on a proportional basis.[6] Each position by vertex is assigned a capital charge. The capital charge is relabeled as "risk weight" for market risk.[7] These positions are aggregated using a formula that recognizes offsetting between cash flows at different vertices in the same currency:

$$K = \sqrt{\sum_i RW_i^2 MV_i^2 + \sum_i \sum_{j \neq i} \rho_{ij} RW_i MV_i RW_j MV_j}$$

where K is the capital charge; MV_i is the present value of the net cash flow at vertex i; RW_i is the risk weight assigned to vertex i; and ρ_{ij} is the correlation parameter between vertices i and j, which is regulatorily defined. The formula is analogous to the calculation of volatility of a portfolio with sensitivities being replaced by risk weights. As the capital for a portfolio of positions is lower than the sum of the capital charges of individual positions if the correlation is not perfect, the formula recognizes diversification effects.

17.3.7 Internal Model Approach

The model approval process would be broken down into smaller, more discrete steps, including trading at desk level.

The expected shortfall replaces the traditional VaR calculations for the determination of capital. It is calculated for a 97.5th percentile, one-tailed confidence interval. It must be computed on a daily basis, bank-wide and for each trading desk that a bank wishes to include within the scope of the internal model for regulatory capital purposes.

In calculating the expected shortfall, instantaneous shocks equivalent to an n-business day movement in risk factors are to be used. The number n is defined based on the liquidity characteristics of the risk factor. These shocks must be calculated based on a sample of n-business day horizons with overlapping observations[8] over the relevant sample period.

The expected shortfall measure must be calibrated to a period of stress. The measure should replicate an expected shortfall charge that would be generated on the bank's current portfolio if the relevant risk factors were experiencing a period of stress. This calibration would rely on a reduced set of risk factors. Banks should specify which reduced set of risk factors is relevant for their portfolio, with a sufficiently long history of observations.

The approach is applied to eligible desks. All desks must also provide a standardized measure of their risk, as if standalone.

The ICR takes care of defaults that are not modeled from the movements of risk factors. It can be measured as a credit VaR, with correlations based on listed equity prices. All positions subject to the market risk framework are subject to the default risk model, except for commodities and foreign exchange. Default risk is measured over a one-year horizon using

6 For example, if it has a tenor of 2.25yr, 75% of its value is assigned to the two-year vertex and 25% of its value is assigned to the three-year vertex.

7 Risk weights for credit risk apply to risk-weighted assets (RWA) and capital is derived from RWA with the capital adequacy ratio. Hence, RWA = 12.5 capital. In the proposed terminology for market risk, the product of risk weights and a position is the capital charge.

8 Say that n is one month. Twelve non-overlapping observations would extend over one year. Overlapping shocks can be observed from each trading day and up to one month, which allows for a larger number of observations.

real-world default probabilities. The bank assumes constant positions over this one-year horizon. To avoid double counting of the risk from mark-to-market loss and the risk of loss from default, the model may assess default risk from the perspective of the incremental loss from default in excess of the mark-to-market losses already taken at the time of default.

The "fundamental review" also imposes qualitative documentation on the trading strategies of desks. The documentation should describe, for example, the risk of the desk, the hedges and the hedging strategies, specifies whether the hedges are fully effective or not. Although such documentation should exist internally, the requirement is not yet mandatory.

The aggregate capital charge for market risk under the internal models approach is equal to the aggregate capital requirement for eligible trading desks plus the standardized capital charge for ineligible trading desks:

$$\begin{aligned}\text{Aggregated capital} &= \text{Aggregated capital for eligible desks} + \text{IDR} \\ &\quad + \text{Standardized capital of ineligible desks}\end{aligned}$$

17.3.8 Back-testing the VaR

Back-testing the firm-wide risk model will be based on a VaR measure calibrated at a 99th percentile confidence level. The back-test remains based on a VaR model because it is not possible to back-test an expected shortfall. Back-testing the expected shortfall is not feasible since there are too few statistics for losses beyond the VaR, the losses of which the expected shortfall is a probability-weighted average.

18

Credit Risk

Credit risk is the risk of loss resulting from an obligor's inability to meet its obligations. It refers to the risk that a borrower defaults on any type of debt by failing to make required payments. The potential loss from default includes lost principal and interest and increased collection costs. The loss may be partial when the lender is able to recover a fraction of the amount due. Or, in a less extreme situation, the credit quality of a counterpart may deteriorate so that a loan becomes more and more risky.

Traditional credit analysis is the method by which one evaluates the ability of a company to honor its financial obligations, examining the ability of a potential borrower to repay the loan as well as non-financial considerations such as the management's track record, the purpose of a loan, the guarantees and environmental conditions. As regulators imposed standards in credit risk management and measurement, and the development of credit derivatives expanded, credit risk entered an era of quantification and modeling. Sound practices now involve measuring the credit risk at facility level and portfolio level to determine the amount of capital that banks need to hold as a cushion against potentially extreme losses. Credit risk modeling describes the analytical due diligence a bank performs to assess the risk of borrowers.

This chapter introduces the breakdown of credit risk into its basic credit risk components, required for quantification purpose. The metrics implemented to model credit risk including credit value-at-risk (VaR), are also discussed.

Contents

18.1 CREDIT RISK COMPONENTS

The factors that have an incidence on the potential loss from credit risk are called "credit risk components".[1]

Default is the situation when an obligor is unable to make the required payments on its debt obligations. Default risk refers to the likelihood of such an event. The default likelihoods at the level of a borrower or at the level of a facility differ depending on how a facility is protected from borrower's default. Exposure is the amount potentially lost at the time of default. Exposure risk refers to the uncertainty with respect to the future amount that can be lost at the unknown time of default. The loss under default is the loss incurred after collection efforts. It is lower than the amount due because of recoveries arising from the work-out process initiated after default, including recoveries from third party guarantees or liquidation of assets pledged to the lender. Such recoveries are usually specific to each facility, not to the borrower. The measures of these factors are called credit risk components, and became mandatory with the Basel regulations. They are the default probability (DP), the exposure at default (EAD) and the loss given default (LGD).

The credit risk components characterize the current credit state of a borrower or a facility. Migration risk refers to the potential deterioration of the credit standing of borrowers as time passes. A DP characterizes a credit state and changes as the borrower migrates to new credit states. The risk of adverse migrations towards lower credit quality is the migration risk. The migrations are documented by changes of ratings, which are credit assessments assigned by rating agencies or by banks, of a firm or facility.

The risk related to migrations is the spread risk of credit products traded in markets. Credit spreads are the spreads between the return required from investors for risky assets and the risk-free rate. Spread depends on the credit class of the instrument. The variations of spreads are market driven and are correlated with migrations across credit states. Spread risk refers to the potential changes of these spreads from migrations. For interest-bearing traded assets, such as bonds, these variations of spreads trigger gains or losses of prices.

18.1.1 Default Event

Default risk is the risk that borrowers fail to comply with their contractual payment obligations. Various events potentially qualify as default:

- Delaying payments temporarily or indefinitely;
- Restructuring of debt obligations due to a deterioration of the credit standing of the borrower;
- Bankruptcies.

Restructuring of debt is very close to default if it results from the inability of the borrower to face payment obligations unless its debt structure changes. Defaults are not permanent if they are corrected within a short period of time as the borrower resolves cash deficiencies.

1 The terminology is used in the Basel 2 Accord of 2006, [21].

The definition of default relies on rules. Rating agencies consider that default occurs from the very first day of defaulting by a single dollar on a payment obligation. For regulators, a default is the absence of a payment due extending over at least 90 days. If payment occurs before the end of the period, the default is a transitory delinquency.

18.1.2 Default Probability and Default Event

A probability of default measures the likelihood of a borrower's default. With a stable credit state, the chances of defaulting increase with the horizon as the credit state and the DP changes as the credit standing migrates. Basel requires the usage of annual DPs.

DPs, given credit state, depend on prevailing economic conditions. The chances of default increase when the current conditions deteriorate. Point-in-time DPs are probabilities conditional on the economic conditions and vary as time passes. "Through-the-cycle" DPs are probabilities averaged through the ups and downs of the economic cycle and are measures of the long-term unconditional DPs.

Estimating DPs is a challenging task. With large volumes of clients, statistics allows counting the default events and deriving their frequencies over various periods. This is the method used in retail banking by banks and by rating agencies for all rated firms of various rating classes. In other cases, statistics are not significant. In general, there is a variety of methodologies and data sources, which banks may use for mapping DPs to internal grades. The three broad approaches include the usage of data based on a bank's own default experience, the mapping of internal defaults to external data (Chapter 19) and the usage of default models (Chapter 21).

18.1.3 Exposure and Exposure Risk

Exposure is the size of the amount at risk with an obligor. Exposure risk refers to the randomness of the size. The EAD is an estimate of the potential size under default, which is generally unknown at the current date.

The contractual exposure for loans derives from the repayment schedule. For a term loan with a contractual amortization schedule, this schedule is known. However, the effective amortization schedule differs from the contractual schedule, for example in the case of prepayment of mortgages. In other instances, the related cash flows are stochastic. Interest payments are driven by market indexes for floating-rate loans. Or the cash flows are driven by the clients' behavior for lines without maturities, such as credit card loans. For corporations, banking facilities include commitments of the bank to provide funds at the initiative of a borrower. With committed lines of credit, the lender has the obligation to lend up to a maximum amount up to a contractual maturity. The drawn amount is the cash effectively borrowed. The undrawn amount is the remaining fraction of the committed line of credit, which is an off-balance sheet commitment. Both future amounts are unknown and draws on the credit line are contingent upon the borrower's willingness to borrow.

For derivatives traded over the counter, the amount at risk is market driven, because the size of the cash flows is generally indexed on market parameters. The consequence is that the EAD is defined by conservative rules or models (Chapter 22).

18.1.4 Recovery Risk and Loss Given Default

The LGD is the fraction of the exposure at risk that is effectively lost under default, after work-out efforts and recoveries from guarantees. The recovery rate is the percentage of exposure recovered after a default, and is the complement to one of the LGDs expressed in percentage of exposure. The LGD is a major driver of credit losses and the capital charge for credit risk is proportional to the final losses post default.

Because of the uncertainty attached to LGD, Basel imposed percentages under certain approaches. Senior claims on corporates, sovereigns and banks not secured by recognized collateral are assigned a 45% LGD and all subordinated claims on corporates, sovereigns and banks are assigned a 75% LGD. Own estimates of LGD by banks are allowed only in the advanced approach. Otherwise, supervisory rules have to be applied.

18.1.5 Credit Risk Mitigation: Collateral

Common practices for credit risk mitigation include posting collateral and guarantees of third parties. Pledging assets as collateral transforms the credit risk of exposure into an asset risk. Third party guarantees change the likelihood of default attached to the credit exposure.

The assets pledged are sold in the event of default and the proceeds used as partial or full repayment. Good quality collateral is made of assets that are easy to liquidate without incurring losses from reduced liquidity or from movements of markets. When financial assets are pledged, the lender can liquidate the assets if the obligor defaults on its payment obligation. Collateral includes cash and securities, often bonds of good credit quality.

When a transaction is collateralized, the liquidation value at the time of default is unknown. The value of securities is driven by market factors, such as interest rate movements for bonds. The proceeds from the sale of assets also depend on how liquid the assets are at the time of sale. The consequence is that the credit risk of the collateralized fraction of a facility is transformed into an asset price risk.

Because of the uncertainty with respect to the liquidation value of the assets posted as collateral, the recognized value of the collateral is lower than its current value. The difference is the haircut (Figure 18.1). The haircut serves as a buffer against the fluctuations of values of securities posted as collateral and against potential adverse price variations arising from the liquidation.

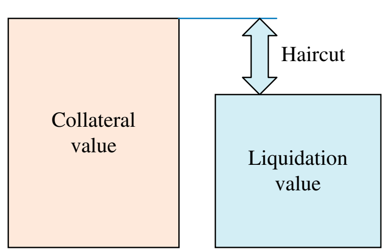


FIGURE 18.1 Haircut

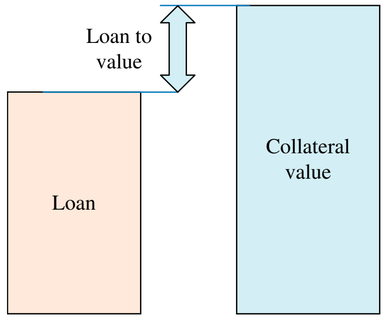


FIGURE 18.2 Loan-to-value ratio

Haircuts apply notably in repurchase transactions (or "repos"), when securities are sold with a commitment to repurchase them later at terms fixed at the inception of the transaction. The seller of the security is a borrower of cash. These transactions are similar to lending the securities in exchange for cash and the securities serve as collateral for the cash lent. The benefit of the transaction for the borrower of cash is that the cost of borrowing is lower than the cost of an unsecured debt. In securities lending and borrowing, the process is similar but securities are exchanged instead of cash. In such instances, the lender of cash faces the risk that the values of securities posted as collateral declines and requires a haircut.

Another common practice is to impose a cap on the amount of loan backed by collateralized assets, or, equivalently, a floor of the value of collateral function of the loan size. The loan-to-value ratio is frequently used in mortgages, as the home is pledged to the lender. When the collateral is made of financial securities, and a loan-to-value ratio is imposed, the debt might become a function of the collateral value (Figure 18.2). For example, funds borrow to buy securities and the amount of borrowing are capped to a fraction of the value of collateral. If the market prices move down, the funds have to increase the collateral by pledging new assets or to reduce the debt.

A haircut measures the deviation of the collateral value over the period elapsing before new collateral can be posted by the borrower. In many instances, margin calls are required by the lender whenever the collateral value falls below a minimum level. When the lender is required to post additional securities, a minimum period is required before additional collateral can be posted, or for repaying a fraction of the debt, otherwise the collateral is liquidated.

The period includes the notification period plus a grace period for posting additional collateral. If no collateral is posted, there is a liquidation period elapsing after the end of the grace period. Suppose that the total period between the margin call and effective liquidation of existing collateral is T, in days. The worst-case situation is that the collateral value drifts down over this period.

The potential variations of value of the collateral depend on the market volatilities, σ_m, and the sensitivity, S, of the securities pledged. The daily volatility of the collateral value is: $S\sigma_m\sqrt{T}$. The drift from an adverse market shock over a period T is derived from the distribution of the collateral value for given confidence levels (Figure 18.3). For example, if the collateral value follows a normal distribution, the potential downward deviation of the collateral value at the 1% confidence level is: $-2.33S\sigma_m\sqrt{T}$. This is the value of a haircut for having no collateral deficiency over the period required for a margin call, with a given confidence level.

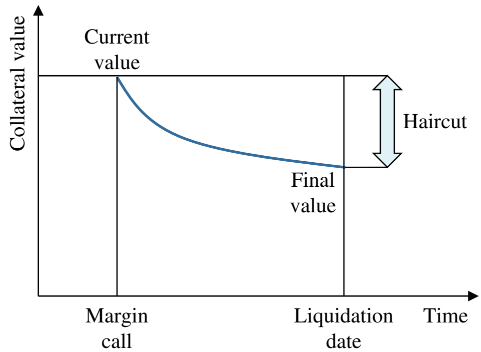


FIGURE 18.3 Collateral value and haircut

18.1.6 Credit Risk Mitigation: Third Party Guarantee

When a third party provides a guarantee, the lender has a claim to the guarantor for defaulted payments. Guarantees are binding commitments to honor the debt payment obligation in the event of a default of the direct lender, similar to an insurance given by the guarantor to the lender. Credit derivatives are instruments that pay to the lender the loss under default if the debt issuer defaults on its obligation. The insured debt is the underlying of the instrument. The difference with standard insurance contracts is that these derivatives are traded: their prices are market driven and a function of the credit risk of the underlying debt.

Third party guarantees mitigate the likelihood of default. As the guarantor acts as a substitute to the lender in the event of default, the risk can be seen as transferred to the guarantor as if it were the direct borrower. Under this view, a guarantee has a value if and only if the credit standing of the guarantor is better than that of the direct borrower. The actual effect of a guarantee on the likelihood of default is different. The DP is neither that of the direct borrower nor that of the guarantor. A default occurs only when both the borrower and the guarantor default and the likelihood of default is the joint probability of default of both entities. A joint DP is generally different, and lower than the DPs of the borrower or the guarantor. The default likelihood in the presence of a third party guarantee depends on how dependent the guarantor and the borrower are. The issue is addressed in the Chapter 23 on dependencies between firms.

18.2 CREDIT RISK MODELING

When a facility remains in the same credit state, a credit loss occurs only when there is a default, and is the LGD. Under this view, the loss of a single loan has only two values: the zero loss under no default and the LGD. When credit risk is seen as limited to default only, it is analyzed under "default mode".

This view is restrictive because it assumes that the borrower stays in the same credit state over the life of the facility. In reality, the credit standing of a borrower changes as time passes, and a facility can become riskier or safer. A credit state is characterized by the likelihood or default, and this probability changes as the borrower migrates towards other credit states.

When the risk metrics capture such potential migrations of the credit standing, credit risk is analyzed under "full valuation mode".

Common metrics of credit quality are credit ratings or DPs. A credit rating is an ordinal measure of the credit quality of a debt or a borrower, relative to scale ranking credit standings. DPs are quantified measures of the likelihood of default. The migrations across credit states are widely documented by the observed transitions of credit ratings, of a given facility, or borrower, as time passes. The public statistics on migrations are detailed in Chapter 19.

Migrations of credit quality raise a valuation issue. If a facility migrates towards a lesser credit quality, the DP increases. The book value of a loan is unchanged as long as there is no default but the loan has less economic value since its likelihood of default increased. For the value to reflect the change of risk, it has to be mark-to-market. In listed markets, bond values are calculated by discounting contractual payments at the market yield. Risky yields embed a credit spread above the risk-free rate, which increases if the credit state deteriorates and vice versa. The bond gains value when it migrates towards a better credit quality and loses value when its credit state deteriorates.

There are as many potential deviations from the current value as there are migrations, including to the default state. Under this view, the gains and losses due to credit risk result from the revaluation of facilities according to their future credit states, and the modeling of credit risk is under full valuation mode.

Credit risk losses (or gains) depend on credit events, default and migrations, between the current dates and horizon, that affect the value of facilities. A valuation at horizon is required to assess credit loss statistics, the expected loss, and the loss distribution for credit risk.

The valuation methodology, introduced by CreditMetrics™,[2] considers a discrete number of credit states at horizon matching the possible rating classes, for determining the potential values of a facility. The probabilities of reaching each of the final states are migration probabilities, which can be inferred from historical data. When the facility defaults, the facility is valued as its recovery value, its face value minus the loss under default. When the facility migrates to another credit state, it is revalued using the credit spreads mapped to each final state. Because credit spreads vary in general with rating and residual maturity, they form in a matrix. This revaluation technique is sometimes referred as "matrix valuation" and is a common building block for several credit portfolio models.

The value of the facility is the present value of the contractual cash flows from horizon to maturity. With migrations, there is a distribution of values at a future horizon. The distribution of values allows various credit risk metrics to be derived, including a credit risk VaR as a quantile of the distribution.

If the migration matrix technique can be applied to a single facility, it provides only a small number of final values matching all credit states from possible migrations (Figure 18.4). When applied to a portfolio, each facility potentially migrates to the same states, but as they are many facilities, the distribution becomes closer to a continuous distribution. An example of a full valuation model of a single facility follows.

In the example, the horizon is one year (date 1). There is no need to consider changes of interest rates between now and end of year if the purpose is to assess the effect of credit risk only. The discounting applies to cash flows posterior to the future date 1 up to maturity T.

2 Gupton, G. M., Finger C. C., Bhatia, M. (1997), CreditMetrics™: Technical Document, [74].

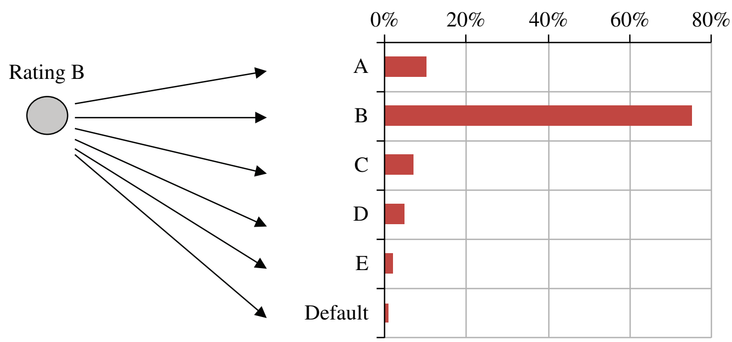


FIGURE 18.4 Migration risk and distribution of future values

The future risky rates as of date 1 for a cash flow of date t for the credit state k are $y_{1,t,k}$. The value V_1 of the bond at date 1 in the credit state k is the present value of all cash flows F_t, posterior to the horizon, at the rate $y_{1,t,k}$:

$$V_1 = \frac{\sum_{t>1,T} F_t}{(1 + y_{1,t,k})^t}.$$

The return $y_{1,t,k}$ represents the market yield applicable at the future date 1, for a maturity t and a credit state k.

The valuation is applied to a facility that has a two-year maturity, with face value 1,000 and a coupon of 6.5%. Recoveries under default are 50% of the contractual amount due, inclusive of interest earned of 65 (at end of year), or 532.5. The facility initial rating is "B", within a scale from A to E, and the corresponding credit spread is 1%. The risk-free rate is 5% and the current market rate applicable for valuation of the facility is 6%. Revaluation is conducted at date 1, when the residual maturity is one year, and after the payment of the cash flows at date 1. The transition matrix, starting from rating B, is given in Table 18.1 with the corresponding credit spreads at horizon.

TABLE 18.1 Migrations, final credit states and credit spreads

	Transition probabilities	*Spread*	*Risky rate at date 1*
Rating A	10.0%	0.1%	5.1%
Rating B	75.0%	1.0%	6.0%
Rating C	7.0%	2.0%	7.0%
Rating D	5.0%	4.0%	9.0%
Rating E	2.0%	8.0%	13.0%
Default	1.0%		
	100.0%		

TABLE 18.2 Distribution of value at horizon

	Transition probabilities	*Rate*	*Value at horizon*
Rating A	10.0%	5.1%	1,013.3
Rating B	75.0%	6.0%	1,004.7
Rating C	7.0%	7.0%	995.3
Rating D	5.0%	9.0%	977.1
Rating E	2.0%	13.0%	942.5
Default	1.0%	Recovery	532.5

The current value of the facility is the present value of the next two cash flows, 65 and 1,065, at the credit risk-adjusted market rate 6%:

$$65/(1+6\%) + 1065/(1+6\%)^2 = 1{,}009.2$$

The current value is above par because the bond generates a spread over risk-free rate in excess of the market credit spread.

For each final credit state, the remaining flows are discounted at the risky rate matching the credit state. It is assumed in the calculations of Table 18.2 that the intermediate payoff of 65 is already paid at date 1. If the facility is defaulted at date 1, the payoff is 532.5.[3] Otherwise, the value at horizon, conditional on survival at date 1, discounts the final flow at end of year 2, 1,065, at the risky yield matching the final rating: $V_1 = 1{,}065/(1 + y_{1,2,k})$, where k is the final rating. The distribution of the values at the horizon is obtained from the revaluation over all final credit states. Using the current market rates for revaluation at horizon is feasible and is a simplifying assumption, under which the market rates at 1, seen from now, are identical to current rates.[4]

The final value when the rating B does not change is: 1,004.7. This value is conditional on no migration from initial state. It is lower than the current value 1,009.2. The variation of value as time passes, conditional on no migration, depends on the remaining cash flows and on the excess spread of the facility over the market yield used for discounting.[5]

The calculation of gains and losses due to credit risk depends on how the zero-loss point is determined. The initial value is not appropriate since value drifts with time even in the absence of any migration. One possible reference point is the value at horizon under no migration (1,004.7). A more appropriate reference is the expected value under no default. The approach is more involved. This expected value, $E(V|\text{No default})$, is the average of values for migrations other than default weighted by the transition probabilities conditional on no default.

---

3 This value is calculated assuming that the default occurs only at end of the first year, when the full annual interest is accrued.

4 The rates at date 1 are forward rates, which are equal to current rates if the yield curve is flat.

5 The value of a coupon bond, V, discounts at the market risky yield y all future interest payments, with coupon rate r, plus the final repayment of face value, N, over the residual maturity T: $T : \frac{V}{N} = \left(1 - \frac{r}{y}\right)\frac{1}{(1+y)^T} + \frac{r}{y}$. A "rich" facility, the coupon of which is above the market yield, loses value at constant risk because it generates less high interest flows. The reverse applies to a relatively "poor" facility.

TABLE 18.3 Distribution of values at horizon

Rating	*Value at horizon*	*Gain (+) & loss (−)*	*Transition probability*	*Cumulated probability*
A	1,013.3	11.1	10.0%	10.0%
B	1,004.7	2.4	75.0%	85.0%
C	995.3	−6.9	7.0%	92.0%
D	977.1	−25.2	5.0%	97.0%
E	942.5	−59.8	2.0%	99.0%
Default	532.5	−469.8	1.0%	100.0%

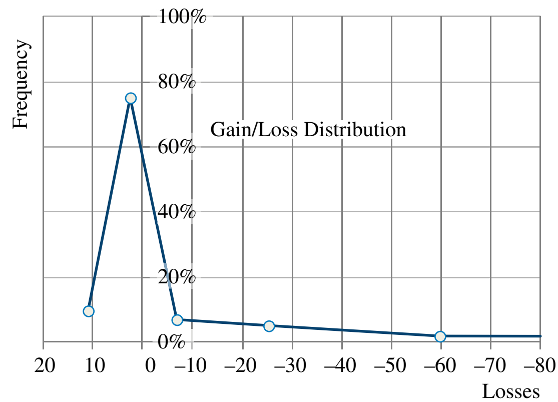


FIGURE 18.5 Distribution of variations of value

Such expected value is not identical to the unconditional expected value of the facility because the probabilities attached to each state are conditional on no default. The conditional probabilities follow from the Bayes' rule: $P(\text{Migration}|\text{No default}) = P(\text{Migration})/P(\text{No default})$, where the probability of survival in this example is 1%. These conditional probabilities apply to non-default states and sum up to 100%.[6]

The conditional expected value is: 1,002.27. It is higher than the unconditional expected value, 997.57, because it ignores the migration to default, which is when the facility incurs the greatest loss. Table 18.3 shows the variations of value from this zero-loss point, 1,004.7, and the cumulative probabilities for determining the quantiles of the distribution.

The variations of value, in Figure 18.5, are measured relative to the value under no migration at horizon. The horizontal axis is truncated at −80 and does not show the largest decline of value occurring at default.

6 For example, the unconditional migration probability to the credit state A at horizon is 10%. The probability of migrating to A conditional on no default is: 10%/99% = 10.10%. All other conditional migration probabilities derive from the unconditional probabilities by dividing by 99%. They sum up to 100%. The expected value under no default is the probability-weighted average over all values, ignoring the default value.

The credit risk metrics are derived from this distribution. Notably, the example illustrates how a credit VaR can be defined as a quantile of the loss distribution under full valuation mode. For example, the negative variation at the 99% confidence level is −59.8. The expected loss is the probability-weighted average of these deviations, 4.70. The mechanism for determining a credit VaR would extend to the whole portfolio and should include dependencies between defaults. Credit VaR is the economic capital for credit risk.

18.3 LOSS DISTRIBUTIONS FOR CREDIT RISK

For portfolios of facilities, loss distributions can be defined in default mode and in full valuation mode. Under default mode, the distribution for a single facility collapses to the two points, default or survival, but the distribution of the aggregated portfolio loss would appear near continuous. In full valuation mode, each single facility has a distribution of values.

In both cases, the distribution is highly skewed from the left with small portfolio losses being the most frequent (Figure 18.6). The mode of the distribution is the most frequent loss. The expected loss is the mean of the distribution. Because of the skew to the right, the expected loss is higher than the mode. Unexpected losses show on the right-hand side, above the expected loss. The capital for credit risk is derived as a loss quantile capturing the unexpected loss at a given confidence level and for a given horizon.

Under default mode, some credit metrics are derived in a relatively simple way. For a single loan, the value is the book value of the loan or is the LGD if the borrower defaults. The expected loss, EL, of the loan is the product of the default probability, DP, by the loss given default, LGD. When the LGD is measured as a percentage of the exposure, the EL formula is:

$$\text{EL} = \text{LGD} \times \text{EAD} \times \text{DP}$$

For a portfolio, the expected losses are additive and derived by summing up individual expected losses. The regulations impose that such expected losses be provisioned, so that the

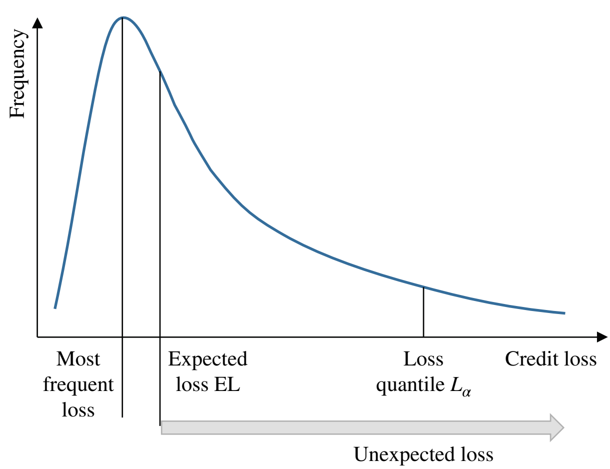


FIGURE 18.6 Credit loss distribution, statistics and capital

regulatory capital measures unexpected losses in excess of expected loss. The concept is relevant at portfolio level. For a single loan, the expected loss never materializes as the loan value is either its book value or its LGD. However, since no one knows which loan will default in the future, it makes sense to set aside a fraction of each loan to absorb the aggregated expected loss of the portfolio.

Under full valuation mode, there are losses other than default when migrations are adverse, without reaching the default state, and the credit risk metrics get more involved. The potential losses of the current portfolio are relative to its future values. A relevant point for determining loss is the expected future value of the portfolio conditional on no prior default (as in the above example). This is the conditional expectation of the value, V, of the portfolio, written as: $E(V|\text{No default})$. The expected loss is identical, by definition, to the difference between the expected value and the expected value under no default: $\text{EL} = E(V) - E(V|\text{No default})$.

Once an expected loss derived, the credit VaR is a quantile of the loss distribution. The economic capital for credit risk serves as a buffer against the unexpected loss in excess of the expected loss. It is the difference between the α-quantile L_α of loss, with confidence level α, and the expected loss EL: $K_\alpha = L_\alpha - \text{EL}$.

The confidence level has an immediate implication for a bank. If the capital is defined as the 1%-quantile, it means that unexpected losses will be larger than capital in 1% of all potential outcomes. The 1% confidence level represents therefore the likelihood that the bank becomes insolvent, and is its DP. The DP of a bank is a key parameter for a bank. It determines its rating, and beyond, with whom it is eligible to do business, and its cost of financing. For good ratings, the DP has to be well below 1%.[7] Hence, an appropriate confidence level should be very low. The implication is that the tail of the distribution should be modeled as accurately as possible.

The assessment of unexpected loss depends on loss distributions, and loss distributions depend on horizon. For market risk, short horizons make sense because a traded portfolio can be liquidated in the market. For credit risk, much longer horizons are relevant because the chances that a credit event materializes are very small over a very short period.

Under a management rationale, the relevant horizon is the minimum time required to take corrective actions. If there is a capital deficiency relative to the credit risk of a bank's portfolio, the bank needs to raise new capital in line with requirements, or, alternatively, to restructure the portfolio, for example by enforcing new limits, securitizations, or usage of credit derivatives. These actions take time. A relevant horizon potentially extends over one, two or three years. From a practical standpoint, a one-year horizon is often used.

7 The default frequencies for firms rated by rating agencies can be found in Chapter 19.

19

Credit Risk Data

Statistics on defaults, transitions across risk classes and recoveries under default based on historical observations are necessary to calibrate models and serve as inputs for complying with Basel credit risk rules.

Banks can collect data on their large retail portfolios, which makes it relatively easy to expand their statistical base. This is not so for large corporations, banks and sovereigns, when default events are much lower. Rating agencies have published historical data on credit events by rating class for a relatively long time. They include counts of defaults, migrations across rating classes and estimates of recoveries. The default statistics are a valuable reference for assessing the default frequencies of firms of different ratings and over varying horizons.

Default frequencies vary with the horizon because of migrations. The default probabilities applying to a period between two future dates, seen from today, are marginal, or forward, probabilities. The cumulative probabilities, of default or of survival, depend on the series of marginal default probabilities over prior periods.

This chapter provides an overview of credit risk data that is available publicly for corporations, addressing default data, recovery data and migration data. It also explains how default frequencies vary as the horizon extends and how migrations alter the default probabilities between futures dates, the forward default probabilities.

Contents

19.1 DEFAULT STATISTICS

When considering public statistics from rating agencies, care should be taken that such statistics apply to rated entities only, generally large and very large corporations, banks and sovereigns. Applying such statistics to the pool of loans of banks is necessary in the absence of other alternatives. Basel provided major incentives for developing internal data pools within banks and stimulating data sharing across banks for those counterparties for which external data are scarce, such as specialized lending and public sector entities.

Agencies report default statistics as counts, or as percentages, of a population of rated entities by rating class over several years. The percentages are either arithmetic or weighted by size of debt issues. The historical default rates are the ratios of the number of defaults in a given period over the total of rated firms at the beginning of the period. These default statistics are annual and cumulated over several years for a cohort of rated firms defined at the initial date. The tables below show the ratings and default frequencies from Moody's Investors Service.

19.1.1 Annual Default Rates

Annual default rates are ratios of the number of defaulted firms to firms within a pool assembled at the beginning of the year. Table 19.1 shows the magnitude of annual default rates for the six-notch simplified rating scale of Moody's. The data are averaged across the years of the observation period. The table shows the first-year default rates, defined as the ratio of the number of defaults to the total number of rated firms at beginning of the year, and over a period of one year after a rating is assigned or confirmed. The annual default rates are close to zero for the best ratings. They increase to around 21% for the lowest rating class. The top three ratings characterize investment-grade borrowers. The three other classes are called speculative grade. For investment-grade borrowers, the annual default rate is below 0.1%. For speculative-grade borrowers, it ranges from 0.2% to 21% a year.

TABLE 19.1 Ratings and default statistics: average default frequencies in percentage for the first year (average 1983–2005)

Rating	*Year 1*
Aaa	0.000
Aa	0.010
A	0.022
Baa	0.213
Ba	1.307
B	5.693
Caa-C	20.982
Investment Grade	0.075
Speculative Grade	5.145
All Corporates	1.737

Source: Moody's Investors Service, 2007.

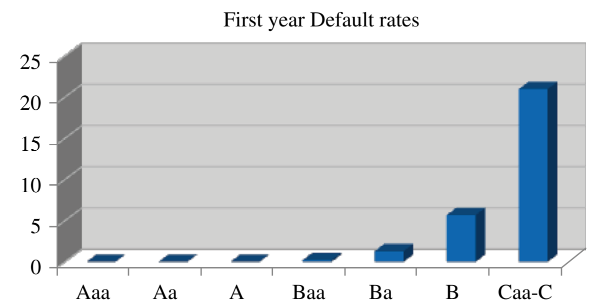


FIGURE 19.1 Annual default rates and ratings (*Source:* Moody's Investors Service, 2007.)

The plot of these numbers shows the typical "exponential-like" shape of the default frequency across credit ratings along the simplified scale (Figure 19.1).

Ratings drive the cost of borrowing for banks and their eligibility for participating to certain activities. Improving the rating of a bank is therefore of major importance. The size of steps of default frequencies for gaining one rating class is much higher in the lower grades of the scale than in the upper grades. Since economic capital calculation is based on a confidence level that represents the default risk of the bank, the table shows that targeting investment grade ratings implies very tight confidence levels, way below 0.1%. The importance of modeling the fat tail of credit loss distributions follows.

19.1.2 Cumulative Default Rates

Cumulative default rates are based on default frequencies over several years. Figures 19.2 and 19.3 show the characteristic shapes of the time profiles of default frequencies, by rating, for the

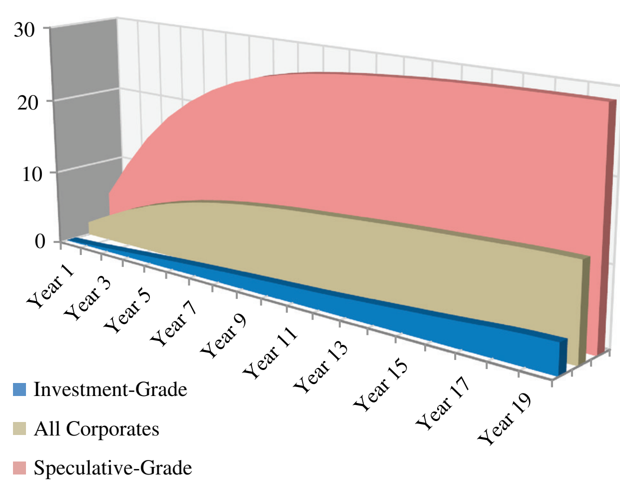


FIGURE 19.2 Average cumulative default rates by letter rating from 1 to 20 years (%) – 1983–2005 (*Source:* Moody's Rating Services, 2007.)

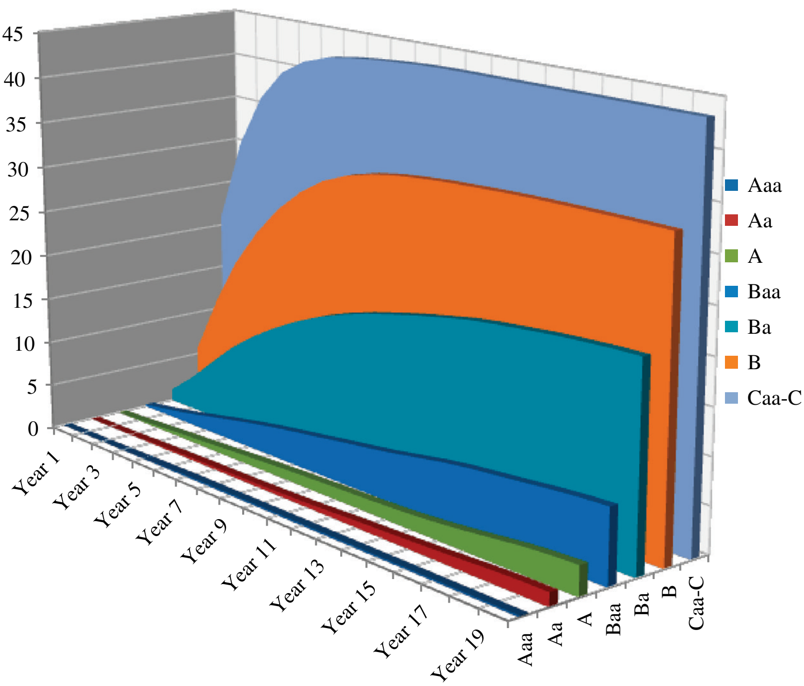


FIGURE 19.3 Average cumulative default rates by letter rating from 1 to 20 years (%) – 1983–2005 (*Source:* Moody's Investors Service, 2007.)

period 1983 to 2005. The first figure shows cumulative default rates for speculative grades, all grades and speculative grades, in that order. The second figure shows similar curves for a simplified rating scale of Moody's.

The longer the period, the higher the chances of observing a default. But the growth of default rates with horizon is not proportional. For high ratings, or low default rates, the increase is more than proportional. For low ratings and high default rates, it is less than proportional. High-risk borrowers upgrade their risk as time passes, when they survive, over long horizons. Safe borrowers tend to have a deteriorated credit standing as time passes.

Annualized rates are cumulative default frequencies divided by the number of years. Due to the shape of the cumulative default rates, annualized rates are higher than the first-year rate when default rates increase less than proportionally to horizon. The opposite holds when default rates increase less than proportionally with time, for speculative-grade borrowers. For regulatory purpose, the first-year default rate only is relevant because ratings are supposed to be reviewed annually.

19.2 RECOVERY STATISTICS

Available recovery statistics are based on the prices of defaulted bonds and loans, and calculated as the ratio of the value of the defaulted issue to its par value. Recovery rates vary by seniority levels, as shown on Table 19.2. There is a wide variation around the mean, so

TABLE 19.2 Recovery rates by seniority level of debt

	1982–2005
Bank loans	
Sr. Secured	70
Sr. Unsecured	57.6
Bonds	
Sr. Secured	51.9
Sr. Unsecured	36
Sr. Subordinated	32.4
Subordinated	31.8
Jr. Subordinated	23.9
All Bonds	35.9
All Debt Instruments	37.7

Source: Moody's Investors Service, 2007.

that the ranges of recovery values across types of debts widely overlap. The last column shows the average 1982–2005. Another empirical fact pointed out by Moody's is that recovery rates seem correlated with economic conditions. In worsening conditions, chances are that recoveries also decline.

Recoveries determine loss under default. Therefore, it is not surprising that the regulators insist on building up recovery data. Recovery data by seniority level, types of products and nature of guarantees would help differentiate more significantly average recoveries for these subclasses.

19.3 TRANSITION MATRICES

When time drifts, the risk either improves or deteriorates, as illustrated by cumulative default rates. These shifts are captured by the transition frequencies between risk classes. Within a given period, transition percentages between classes are transition frequencies divided by the number of original firms in each risk class.

A transition matrix is organized as shown in Table 19.3. Each row and each column is a rating class. In each cell, the percentages are the frequencies of transition between the initial rating (in rows) and the final rating (in columns). Next to the last row provides the frequency of defaulting per risk class. The last row shows the percentage of withdrawn ratings within the period, for which no final state can be assigned. All transition probabilities, plus the percentage of withdrawn ratings, sum up to one across rows because each row lists all possible final states. Transition matrices are subject to constraints:

- Cells across rows sum up to one because they group, together with the default state, all possible migrations;
- All migration probabilities are positive and lower than one.

TABLE 19.3 Average one-year whole letter rating transition rates, 1983–2005 (% of issuers)

Beginning of year rating	*End of year rating*									
	Aaa	*Aa*	*A*	*Baa*	*Ba*	*B*	*Caa-C*	*Default*	*WR**	*Total*
Aaa	94.65	5.21	0.00	0.00	0.00	0.00	0.00	0.00	0.15	100
Aa	4.43	92.28	1.38	0.00	0.00	0.00	0.00	0.00	1.90	100
A	0.00	2.90	91.73	3.19	0.52	0.00	0.00	0.00	1.66	100
Baa	0.00	0.00	8.88	79.02	6.85	0.85	0.00	0.00	4.40	100
Ba	0.00	0.00	0.00	4.28	85.93	8.29	0.73	0.62	0.15	100
B	0.00	0.00	0.00	0.00	3.50	87.18	2.09	3.89	3.34	100
Caa-C	0.00	0.00	0.00	0.00	0.52	28.18	53.12	18.18	0.00	100

* Withdrawn Rating
Source: Moody's, 2007.

The transitions between neighboring rating classes are more frequent than between distant rating classes, and their frequencies have higher values along the first diagonal of the matrix.

Rating classes can be mapped with credit spreads for valuation purposes. However, credit spreads do not depend on rating class only. Statistical analyses on credit spreads show that they depend, at a minimum, on rating class, industry and geographic region, for a given asset class, such as corporates.

19.4 CUMULATIVE AND MARGINAL DEFAULT PROBABILITIES

The cumulative default frequencies and the marginal default frequencies applying to the subperiods are related by consistency constraints.

Historical frequencies are calculated from counts of defaults over an initial population of firms formed at date zero. Defaults are counted from the initial date up to a final date t and between dates t and $t+1$. The ratio of defaults from 0 to t to the initial number of firms is a cumulative default frequency $D_{0,t}$. The ratio of defaults from $t-1$ to t to the population of surviving firms at t is a marginal default frequency, $D_{t-1,t}$.

Similar ratios apply to surviving firms: $S_{0,t}$ is the ratio of firms surviving at date t to the number of firms at the initial date. The marginal surviving frequency is the ratio of firms surviving at date t to the number of firms surviving at the preceding date $t-1$.

Forward defaults and survival probabilities apply to dates posterior to the current date. Marginal default rates, or survival rates, derive from the cumulative rates. The process is illustrated to two periods and then generalized. The cumulative survival probability from 0 to 2 is the probability of not defaulting between dates 0 and 1 and between 1 and 2:

$$S_{0,2} = (1 - D_{0,1})(1 - D_{1,2}) = S_{0,1}(1 - D_{1,2})$$

The forward default probability $D_{1,2}$ is conditional on survival between 0 and 1:[1] $D_{1,2} = 1 - (S_{0,2}/S_{0,1})$. The cumulative default probability is: $1 - S_{0,2}$. The formula can be generalized to any horizon t. The cumulative survival probability is the product of the survival probabilities in all periods, each equal to one minus the marginal default probability of the period:

$$S_{0,t} = (1 - D_{0,1})(1 - D_{1,2}) \ldots (1 - D_{t-1,t}) = S_{0,t-1}(1 - D_{t-1,t})$$

The marginal default probability from $t - 1$ to t can be isolated and expressed as a function of the survival probabilities from the initial date to $t - 1$ and t:

$$D_{t-1,t} = 1 - \frac{S_{0,t}}{S_{0,t-1}}$$

The relation shows how forward default probabilities are derived from cumulated survival probabilities.

Table 19.4 illustrates the calculation when using cumulative default probabilities as inputs. The notations are the same as above, with two consecutive periods and three dates. The marginal default probabilities derived from cumulative probabilities are 6% and 7% in the second and third periods, respectively.

For moving from marginal probabilities to cumulative probabilities the calculations rely on the same basic formulas, as illustrated in the example of Table 19.5. The inputs are now the marginal default probabilities supposed to be constant. The table shows that the cumulative default probabilities, $D_{0,2}$ and $D_{0,3}$, increase less than proportionally with time.

TABLE 19.4 From cumulative to marginal default probabilities

$D_{0,1}$	5.00%	
$D_{0,2}$	10.70%	
$D_{0,3}$	16.95%	
$S_{0,1} = 1 - D_{0,1}$		95.00%
$S_{0,2} = 1 - D_{0,2}$		89.30%
$D_{1,2} = 1 - S_{0,2}/S_{0,1}$		6.00%
$S_{0,2} = 1 - D_{0,2}$		89.30%
$S_{0,3} = 1 - D_{0,3}$		83.05%
$D_{2,3} = 1 - S_{0,3}/S_{0,2}$		7.00%

1 When some random event X is dependent upon some random scenario, the probability of occurrence of X depends on whether S occurred or not. This probability of X is conditional on S. The basic formula for conditional probabilities is: $P(X, S) = P(X)P(S|X) = P(Y)P(X|S)$. In the present case, the probability of defaulting over the interval (1, 2) is the joint probability of surviving until 1 and defaulting between 1 and 2. This is the product of these two probabilities.

TABLE 19.5 From marginal probabilities to cumulative probabilities

$D_{0,1}$	5%	
$D_{1,2}$	5%	
$D_{2,3}$	5%	
$S_{0,2} = S_{0,1}S_{1,2}$		90.25%
$D_{0,2} = 1 - S_{0,2}$		9.75%
$S_{0,3} = S_{0,1}S_{1,2}S_{2,3}$		85.74%
$D_{0,3} = 1 - S_{0,3}$		14.26%

19.5 MIGRATION MATRICES AND CUMULATIVE PROBABILITIES

Cumulative and marginal default probabilities depend on migrations. Migration matrices provide discrete transition probabilities across rating classes. Table 19.6 shows an example with three credit states, A, B and D, inclusive of the default state. The migration matrix is squared with initial credit states, or ratings, in rows and final credit states in columns. The default state is absorbing: once the firm migrates to the default state, it stays there.

The highest transition probabilities are along the diagonal and the matrix is not symmetric. The column for the default state shows the default probabilities assigned to each credit state. The migration probabilities out of the default state are zero. The matrix is invariant with time: a transition probability depends only in the credit states and does not change with time.

The marginal default probabilities of successive periods differ because of migrations across credit states. A cumulative default probability over several periods depends on the default probabilities characterizing each credit state and on the transition probabilities across credit states.

Table 19.7 provides the transitions when the initial state is A. Over the two periods, the firm defaults either in period 1 or in period 2. The cumulative default probability at the end of the second period is the summation of the default probabilities for each time path, weighted by the probabilities of each path starting from A. The default probability in period 1 is weighted 100%, since this is the probability of staying in the default state. The cumulative default probability over the two periods is:

$$0.80 \times 0.05 + 0.15 \times 0.10 + 0.05 \times 1 = 10.5\%$$

TABLE 19.6 Sample transition matrix

	A	*B*	*D*	*Total*
A	0.80	0.15	0.05	100%
B	0.10	0.80	0.10	100%
Default	0	0	1	100%

TABLE 19.7 Cumulative default probability and transitions

		Period 2			
Initial state	*Period 1*	*Transition to:*	*Transition probability*	*Default probability*	*Weighted probabilities*
A	Survival	*A*	0.80	0.05	0.040
		B	0.15	0.10	0.015
	Default	*D*	0.05	1	0.050
		Cumulative default probability			0.105

The process shows how cumulative default probabilities depend on the marginal default probabilities of each credit state and the transitions across credit states. It can be generalized for determining the cumulative transition probabilities of ending in default when the initial state is B instead of A.

A transition probability has two subscripts, the first one for the initial state and the second for the final credit state: the transition probability from initial credit state i to the final credit state j is: T_{ij}. Decomposing the calculation, the default probability over the two periods, when initial state is A, or $D_{02}(\text{A})$, is:

$$D_{02}(\text{A}) = T_{\text{AA}}T_{\text{AD}} + T_{\text{AB}}T_{\text{BD}} + T_{\text{AD}}T_{\text{DD}} = 10.5\%$$

Similar calculations apply for the state B, or even default.[2]

The default probabilities are the elements of the matrix product MD, where M is the square (3×3) migration matrix, and D is the (1×3) column vector of default probabilities of A, B and D, identical to the last column of the matrix. The default probabilities over two periods are the elements of the (1×3) column vector MD.

0.05
0.1
1

0.8	0.15	0.05
0.1	0.8	0.1
0	0	1

0.105
0.185
1

2 When starting in the default state, at end of period 1, all transitions probabilities out of the default state are zero and the probability of staying there is one: $\text{DP}(D) = T_{\text{DA}}T_{\text{AD}} + T_{\text{DB}}T_{\text{BD}} + T_{\text{DD}}T_{\text{DD}} = 0 + 0 + 1$.

TABLE 19.8 Cumulative probabilities from migration matrices

	A	B	D
	0.8	0.15	0.05
	0.1	0.8	0.1
	0	0	1

	A	B	D
A	0.8	0.15	0.05
B	0.1	0.8	0.1
D	0	0	1

0.655	0.24	0.105
0.16	0.655	0.185
0	0	1

The process can be generalized to other final states. Starting in A, the probability of ending in B is: $T_{AA}T_{AB} + T_{AB}T_{BB} + T_{AD}T_{DB}$. A similar calculation provides the probability of ending in A. This shows that the transition matrix over two periods is simply the matrix product $MM = M^2$, as shown in Table 19.8. Generalizing this result, the transitions over N periods are simply given by the matrix M^N.

20

Scoring Models and Credit Ratings

The assessment of the credit quality of borrowers is a critical input for banks that needs to measure the credit risk of their portfolios and comply with the internal rating-based approach of Basel. Banks generally use rating schemes for large entities and rely more on statistics in retail banking.

Credit scoring uses techniques for finding the criteria that best discriminate between defaulters and non-defaulters. Scoring does not rely on conceptual models, but on statistical fits of "scoring functions". These functions provide a score derived from observable attributes, which is a number. Scoring applies best for retail portfolios.

For corporations, credit ratings, based on quantitative and qualitative data, combined with expert judgments are preferred. Credit ratings are assessments of the credit standing of borrowers or of specific debt issues, assigned to large corporations, banks, insurance companies, sovereign or public sector entities. The assessments are materialized by ranks ordering the credit quality, of a particular debt or of a borrower, relative to others.

The chapter explains how the scores used in retail banking and the internal ratings of corporations are assigned. For scoring purposes, it covers the principles of models capable of discriminating between defaulters and non-defaulters, based on observable attributes. For credit ratings, a mix of qualitative and quantitative criteria is used for assessing the credit standing of obligors, along with expert judgments. A last section discusses the issue of mapping scores or ratings to default frequencies, as required by regulations.

Contents

20.1 SCORING

There are several generations of models of credit risk and default probabilities, starting from the early statistical models relating ratings or default frequencies to financial characteristics of firms, up to elaborated econometric techniques and neural network models.

The principle of scoring is to use a metric for dividing "good" and "bad" credits into distinct classes, using observable characteristics, or attributes, of borrowers. For firms, the technique uses current and past values of observable attributes such as financial ratios, profitability, leverage and size. For individuals, income, age and professional activities, for example, are related to their credit standing.

The usage of such models implies that default be properly defined. In retail banking, this is not a trivial exercise. For example, a default of a member of a group can trigger a default for the whole group for individuals belonging to the same family. Moreover, a single individual usually has several products from the bank, such as deposit accounts and loans. It can be that a loan payment is honored from an excess overdraft from the deposit account. The loan apparently does not default, but the account does. In such cases, the default in one product extends to other products of the same borrower. For properly defining a default in retail banking, contagion rules across groups of individuals and across products are a prerequisite. Statistics of default events depend directly on such rules.

20.1.1 Scoring Functions

Scoring functions are statistical models discriminating between defaulters and non-defaulters within a population of individuals. The function is a linear combination of selected discriminating variables that are observable, called attributes. A scoring function serves for ranking scored entities according the level of credit risk. The values of the discriminant function are the scores. The coefficients of the attributes in the scoring function serves for interpreting which attributes best explain the classification. Once defaults are defined, a scoring function is obtained from a statistical fit to statistics of default.

For firms, scoring models are based on observable attributes, such as accounting ratios. There are a number of classical studies explaining scoring systems and providing numerical results. As an example, the Altman Z-score model for corporations is well known.[1] The scoring function is a combination of selected ratios and categorical measures. The Z-score models were applied to a large variety of corporate borrowers, from small to medium size firms, non-manufacturing and manufacturing firms. In the simple Z-score function, a small number of financial variables are used for discriminating between obligors.[2]

For individuals, in retail banking, there is a large volume of default statistics. Scoring functions rely on such observable attributes as: income per dependent in the family, renting or owning a home, marital status, occupation, etc.

1 See Altman, E. I. (1968), Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, [8]. Other references are Altman, E. I., Haldeman, R., Narayanan, P. (1977), ZETA Analysis: A new model to identify bankruptcy risk of corporation, [10] and, for small and medium enterprises, Altman, E. I., Sabato, G. (2007), Modelling credit risk for SMEs: Evidence from the US market, [11]. See also Altman, E. I., Saunders, A. (1997), Credit risk measurement: Developments over the last 20 years, [12].

2 The variables in Altman [8] are: working capital/total assets, retained earnings/total assets, EBIT/total assets, market value of equity/book value of liabilities, sales/total assets.

The primary usage of scoring is to discriminate between defaulters and non-defaulters, in the existing population of clients, or for new clients. A type 1 error occurs when bad credits are accepted; a type 2 error occurs when good credits are rejected. For individual decisions, type 1 errors result in a full loss of principal and interest if the risk materializes. Type 2 errors have an opportunity cost, which is the income lost because a wealthy borrower is rejected. This cost of a type 2 error is lower than that of a type 1 error.

20.1.2 Logit Models

The models dedicated for predicting default or non-default include the linear probability models and the more adequate Logit and Probit models. These models use the multivariate regression technique.[3] The explanatory variables are the observable attributes.

The simple linear probability model illustrates the principle. The probability P of the event, default or no default, is a linear function of several attributes X_i. The explained variable is Y, which takes the values 0 for default and 1 for default. If X is the vector of attributes,[4] the model is:

$$Y = \beta_0 + \sum\nolimits_{i=1,n} \beta_i X_i + \varepsilon$$

The expectation of Y is the default probability, $P(Y = 1)$, since:

$$E(Y) = 1P(Y = 1) + 0[1 - P(Y = 0)]$$

Taking the expectation from the equation of the model, and given the zero expectation for the error term of the regression:

$$P(Y = 1) = \alpha + \beta E(X)$$

The model provides the value of the default probability as a function of the vector of attributes X. However, the linear regression provides coefficients such that Y values, given the X values, do not necessarily fall within the 0 to 1 range. The modeled values of Y should be truncated to avoid such outliers. The Logit model avoids this drawback.

The Logit model resolves the problem of imposing that Y be in the [0,1] range of values. The variable Y is again expressed as a linear function of the attributes X_i: $Y = \alpha + \beta X + \varepsilon$. But, instead of mapping directly Y to a probability, the Logit model uses Y as the argument of a cumulative distribution function $F(Y)$. The cumulative function has values between 0 and 1 and is monotonously increasing. A value of the cumulative function maps to a single value of Y and any value of Y corresponds to a probability using a known distribution function.

The Logit model uses for $F(Y)$ the cumulative logistic probability distribution. The probability $P(Y)$ is then expressed as:

$$P(Y) = \frac{1}{[1 + \exp(-Y)]} = \frac{1}{\{1 + \exp[-(\alpha + \beta X + \varepsilon)]\}}$$

3 See Altman, E. I., Saunders, A. (1997), Credit risk measurement: Developments over the last 20 years, [12].

4 Several explaining variables are used, representing the various attributes used to predict default or no-default states.

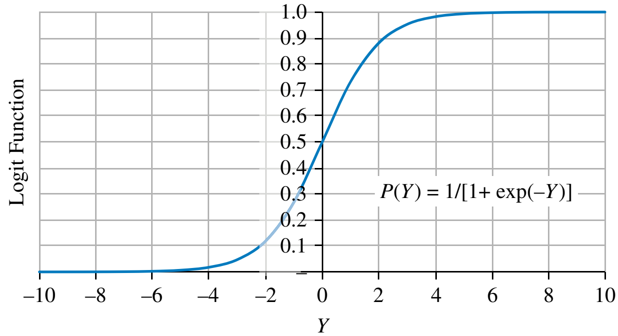


FIGURE 20.1 Function *P(Y)*

When Y gets very high, the exponential is close to zero and $P(Y)$ tends towards 1. When Y gets very negative, the exponential becomes very high and $P(Y)$ tends towards 0. Any value of $P(Y)$ can be seen as a probability, and is a monotonous function of Y. Since $P(Y)$ declines monotonously from 1 to 0 when Y increases from negative values to positive values, $P(Y)$ can be seen as a number characterizing the credit standing of the borrower. For interpreting $P(Y)$ as a probability, it is convenient to transform the relation:

$$\exp(Y) = \frac{P(Y)}{1 - P(Y)}$$

Taking the natural logarithm:

$$\ln\left[\frac{P(Y)}{1 - P(Y)}\right] = Y = a + bX + \varepsilon$$

The argument of the logarithm is the odds ratio, and is equal to $\exp(Y)$. The odds ratio is the ratio of the probability of belonging to a group to the probability of not belonging to that group. The logarithm of the odds ratio is the “Logit”. $P(Y)$ is a monotonously increasing function of Y, and the odds ratio also is. The relationship between Y and $P(Y)$ is shown in Figure 20.1. When $Y = 0$, the logarithm is 1 and the odds ratio is also 1: there are as many chances of belonging to one group as there are belonging to the other, implying a common value $P(Y = 0) = {}^1\!/_2$.

In short, the Logit model provides a value of the logarithm of the odds of a credit event from a linear model. The explained variable is a linear function of observable attributes, its value is the logarithm of the odds ratio and its exponential is the odds ratio itself.

The Probit model is analogous to the Logit model, but uses the normal distribution instead of the logistic distribution. The Logit and Probit models give very close results. The only difference is that the Logit form is easier to handle.

20.1.3 Scoring in Retail Banking: Behavioral versus Origination Models

Statistical scoring models (usually “Logit models”) relate the status “Default” – “No Default” of individuals with observable variables, such as revenues, age of accounts, etc. There are several types of models according to the details of the "credit history" that is available to measure attributes used as predictors of the client's credit standing.

"Behavioral models" attempt to model the behavior of existing clients, when there is no new event that would change the debt level, given historical data of the account of the client. Historical data make it easier to deal with existing clients than for new clients for which there is no credit history. "Origination models" describe the credit standing of new clients, or the credit standing of existing clients when new products are originated. The information set is not the same for new clients and for existing clients with new products.

20.1.4 Implementation of Scoring in Retail Banking

The implementation of a scoring model faces a number of statistical challenges. Some of them are illustrated with an example.

20.1.4.1 Comparing Attributes for Defaulters and Non-defaulters

For behavioral models, a history of six months is a good compromise for measuring attributes. These include, for example:

- Value of cash flows, both negative and positive, averaged over a period.
- Maximum number of debit days over the past six months.
- Number of transactions suspended by credit officers.
- Count of incidents over the past six months.
- Liquid savings, as end of month average of balances.
- Age of account, since origination of the account.

Such variables would the X_i in the Logit model. The outputs are Y and $P(Y)$, the latter being interpreted as a score.

For determining which attributes best discriminate between defaulters and non-defaulters, it is interesting to compare their average values between defaulters and non-defaulters. For example, a large difference of the numbers of debit days for the two populations suggests that the variable should be considered in the Logit model. For comparing means across populations, tests of statistical significance should be used. In retail banking, the volume of observations is usually high, as well as the number of defaults, which facilitates the statistical tests. For example, in a sample population of 1,000,000 individuals, with 1% defaulted, the subpopulation of defaulters is 10,000.

In the example used for illustrating the process, the number of debit days and the age of accounts are identified as relevant attributes. The average number of debit days for non-defaulted individuals is five over six months, and increases to 20 for defaulted individuals, a difference statistically significant. Similarly, the mean of the age of defaulted accounts is 81 months versus 123 months for non-defaulters.

20.1.4.2 Non-linear Relationships between Attributes and Credit Standing

Figure 20.2 shows the frequency distributions of the number of debit days and the age of the account for defaulters and non-defaulters. For low numbers of debit days, there are many more

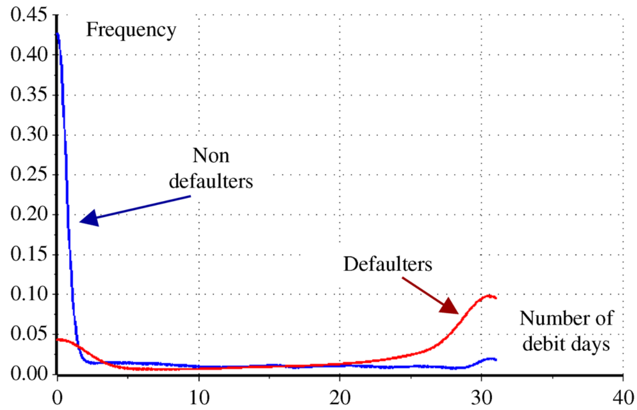


FIGURE 20.2 Distribution of the number of debit days over six months for defaulters and non-defaulters

non-defaulters than defaulters. For large numbers of debit days, there are also many more defaulters than non-defaulters. In the interval, the frequencies of the subpopulations are similar for the two groups. These findings show that the significant differences appear only over some subranges of the values of the number of debit days.

Similar characteristics appear with the frequency distributions of the age of account for defaulted individuals and non-defaulted individuals (Figure 20.3). In general, non-defaulters have older accounts than defaulters.

The relation between these attributes and the binary variable Y, default or no default, are not linear. For example, the number of debit days between 10 and 20 days does not differentiate defaulters and non-defaulters. It becomes a better predictor of default when debit days are 20 or higher. Hence, a linear relation such as $Y = \alpha + \beta X + \varepsilon$, where Y is a binary variable of default and X is the number of debit days would not work too well.

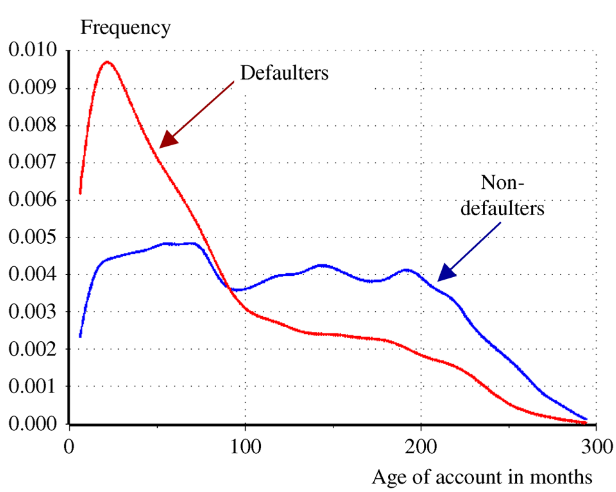


FIGURE 20.3 Frequency distribution of age of account for defaulters and non-defaulters

TABLE 20.1 Discretization of continuous variables

Number of debit days	*% of all accounts*	*% of all defaults*	*Default frequency %*
0	62.83	19.31	0.36
[0; 12]	15.64	8.11	0.60
[12; 22]	10.39	11.27	1.25
>22	11.14	61.31	6.36
Total	100	100	1.16

The standard technique for fitting a non-linear relationship between two variables is to break down the continuous variables, the number of debit days or the age of accounts, into several discretized variables. An example of discretization is shown in Table 20.1. The intervals are selected empirically according to the gap between frequencies of defaulters and non-defaulters.

The table shows that the number of debit days above 22 days over a six-month period corresponds to 11% of all accounts and to 61% of all defaults, with an average default frequency of 6.36%. Conversely, 63% of all accounts have zero debit days over the same period and correspond to 19% of all defaults, with a much lower default frequency of 0.36%. The number of debit days is divided into binary variables taking the values 0 or 1 according to whether the number of debit days is within the interval or not. Each of these binary variables, N_i, is defined as an indicator function of the number of debit days. The Logit model uses the new variables as attributes, instead of the actual number of debit days:

$$Y = \alpha + \sum_i \beta_i N_i + \varepsilon$$

20.1.4.3 Scoring Models and Default Frequencies

The model should be fitted to various samples of the population. A first sample provides a first fit. An "out-of-the-sample" fit, over another sample of the population, serves for checking that the same model provides similar results. "In-" and "out-of-the-sample" fits are conducted over the same reference period.

There should be a good match, in terms of ranks, between the score and historical frequencies of defaults. The process for mapping scores to default probabilities involves intermediate steps. The population is divided into subgroups according to ranges of values of $P(Y)$ ranked in ascending order. For each group, a historical default frequency is calculated. The default frequency D_T over a period, T, is the ratio of the number of defaults to the initial number of individuals as of the starting date 0. T can be set to 1 for having annual default frequencies.

The default frequencies of each subgroup are then mapped to the score values. The process should be repeated for various segments of the retail portfolio, defined by nature of products or type of borrowers: mortgages or consumer loans, and individuals or very small firms, etc. The models differ across segments since the relevant attributes are not the same across segments. With multiple scores, a single "master scale" of default frequencies should be used. The scores of the various segments are then mapped to the single default frequency scale.

There are technical difficulties for assigning default frequencies to ranges of values of scores. For example, the mapping process should ensure that the ranks of default frequencies derived from scores are identical to the ranks of average scores when moving along the scale from low-risk obligors to high-risk obligors. Such consistency of between score and default frequencies is not guaranteed once individuals are grouped according to score values.[5]

20.2 ACCURACY OF SCORING MODELS: THE "CAP"

A standard technique for visualizing and quantifying the accuracy of any default risk measure is the usage of a "power curve", also called cumulative accuracy profile or "CAP".[6] This is a back-testing methodology. It applies to various default risk measures, from scoring to ratings, and to more involved measures such as the Moody's KMV Credit Monitor EDFs.[7]

The CAP (Figure 20.4) plots the fractions of the population of defaulted individuals with the fractions of the entire population ranked by decreasing risk, risk being measured with a model. Obligors are ranked by risk score, from riskiest to safest, along the horizontal axis. For example, 10% on the horizontal axis represents the 10% worst individuals ranked in terms of credit standing of the entire population; 20% represents the 20% most poorly rated, etc. The maximum value is 100% of the population rated by the score.

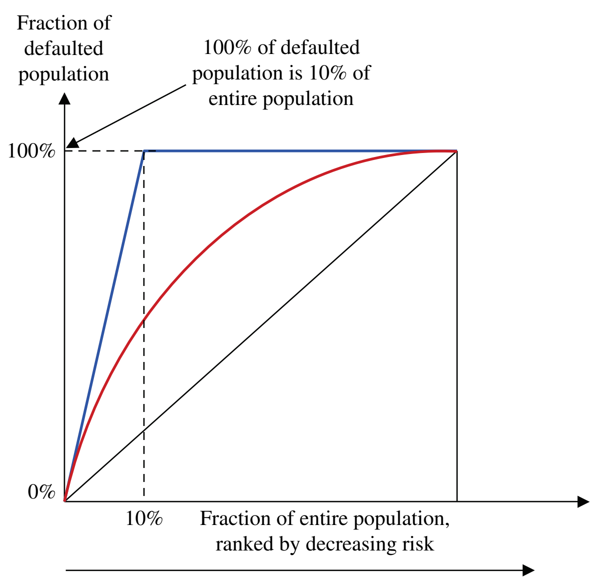


FIGURE 20.4 Cumulative accuracy profile

5 Consider two groups A and B. A is supposed to be less risky than B in term of scores. All individuals of A have a higher score than those of B. Accordingly, the historical default frequency of A should be higher than the historical default frequency of B. But a mismatch of ranks between scores and default frequencies might occur. Some algorithm is needed to segregate groups, by varying the size of each group, until the historical default frequency varies monotonously from one group to the next one.

6 See Engelmann B., Hayden E., Tasche, D. (2003), Testing rating accuracy, [61] and Satchell, S., Xia, W. (2006), Analytic models of the ROC curve, applications to credit rating model validation, [122].

7 The Credit Monitor model is detailed in Chapter 21.

The vertical axis shows percentages of the defaulted population only. Assume, for example, that the default frequency of the population sample is 10% and the size of the sample is 1,000 individuals. The population defaulted over the observation period represents 10% of the entire population. A 50% value on the vertical axis refers to only 5% of the entire population.

When moving from left to right along the horizontal axis, the fraction of defaulted population increases but it cannot exceed 100%. If the scoring model was perfect, the first 10% of individuals of the entire population would be those who actually defaulted. The 10% worst ranked in terms of credit score, on the horizontal axis, would match exactly the 10% of total defaults along the vertical axis. Moving beyond 10% along the horizontal axis, the fraction of defaulted population measured along the vertical axis remains 100%, or, equivalently the fraction of entire population that defaulted remains constant at 10%.

The upper bound of the CAP is reached only with perfect models. It increases from 0% to 100% when moving left to right from 0% to 10% on the horizontal axis. The upward sloping straight line represents the fraction of individuals that defaulted. A second horizontal line, at 100% on the vertical axis, and starting at 10% on the horizontal axis, is reached when the model is perfect. The overall upper bound of the CAP is made of those two lines.

If the score were independent of actual defaults, it would not discriminate between defaulted and non-defaulted individuals. Whatever its value, there would be a 50%/50% chance of default or no default. The fraction of the defaulted population will be the same as the fraction of the entire population ranked according to the model. This is the first diagonal in the graph. It is the lower bound of the CAP.

If the model does provide information on who defaults and who does not, the line showing the fraction of defaulted population as a function of the entire population is below the upper bound and above the first diagonal. This implies that individuals who do not default were ranked as risky – a type 2 error of rejecting "good" individuals. Conversely, some of the individuals, ranked as safer than the fraction of defaulters, actually default – a type 1 error.

A scoring technique is back-tested with the CAP profile, by measuring the actual ratios to defaulted population with the scores. A score is more accurate when the actual CAP is closer to the upper boundary lines. The Gini coefficient is the ratio of the area between the actual CAP line and the first diagonal to the area below the upper bound and the same first diagonal:

$$\text{Accuracy measure} = \text{Area below upper bound of CAP}/\text{Area below actual CAP profile}$$

The area below the actual CAP profile is shown from the actual curved CAP line, which has a continuous shape, and monotonously increasing when moving to the right. This area is purely empirical, and the area under the upper bound made of the two straight lines can be calculated easily. Using as lower bound the first diagonal, the accuracy measure has a minimum value of 0, meaning that the model does not discriminate the risk. The maximum value is 100%. The higher the Gini coefficient, the better the model. Acceptable values start at 60% and can go up to values much closer to 100%.

20.3 CREDIT RATINGS

Internal ratings are a critical input in banking for complying with Basel regulations. They make up an essential piece of the "internal ratings-based" approach, which allows banks to assess the capital charge for credit risk. Internal ratings generally differ from a statistical assessment of the credit quality of borrowers, used in retail banking in that they include judgmental components in addition to other qualitative and quantitative information.

20.3.1 External and Internal Credit Ratings

The philosophy underlying credit ratings differs for external rating agencies and for regulators. External ratings provide the necessary information for investors on the credit quality of a debt issue. A sample of the broad descriptions of letter-grade credit ratings is in the Appendix in section 20.4. Credit ratings from agencies apply to debt issues, rather than to the credit standing of issuers. A same issuer usually has several debt issues, which differ by seniority levels and guarantees. Ratings assigned to senior unsecured debt are close to issuer ratings since these debts default only if the issuer does.

Banks need to rely on their own internal rating schemes for differentiating the risk of their exposures to these counterparties. Internal credit ratings are assigned to borrowers, not debt issues. In credit risk regulations, these are inputs required for assessing the default probability of an issuer. The risk of facilities is differentiated by the severity of loss in the event of default, or loss given default. With some transactions, such as project financing, the credit rating of the entity is irrelevant and covenants govern the behavior of the transaction. When the specific risk of a transaction is the dominant risk, special criteria should be used.[8]

The internal rating process generally uses rating grids, which list the criteria used for obtaining a final assessment of the credit quality of borrowers. Ratings require some judgmental inputs, unlike scoring models. In retail banking, the large volume of statistics allows assessing the risk from statistical models. The process is mechanical and the effect of some attributes can be offset by the positive impact of some others. In the rating process, there is no mechanical effect, as some criteria can dominate, without compensation, other criteria.

Rating schemes use various criteria, from qualitative factors, such as strengths or weaknesses of firms, up to financials of corporate borrowers. Internal rating scales have a varying number of levels, typically from 10 up to 30, like the rating agency scales, simplified or detailed. Moody's simplified rating scale uses three levels for investment-grade ratings and the next three levels are speculative grades. Detailed scales include around 20 levels, excluding the near-default states, for Moody's, S&P and Fitch.

Both banks and agencies are supposed to continuously review their ratings on a periodical basis or under occurrence of contingent events that affect the credit standing of an issuer. It is often mentioned that credit ratings from agencies are "sticky" and do not adjust as frequently as they should. The view of rating agencies is that the long-term ratings should precisely be relatively stable as they assess the long-term drivers of the credit standing of a corporation.

20.3.2 The Bank's Counterparties

Internal credit ratings apply to all counterparties with whom the bank is at risk. The population of such counterparties extends beyond that of direct borrowers. All counterparties with legal links to a direct borrower are involved since they have a potential impact on the borrower's credit standing. For example, firms extending a guarantee to the direct borrower should be

8 Some samples of credit risk criteria, other than issuers' ratings, are provided in the Basel 2 Accord, in appendix, [21], as a guide for banks for assigning ratings to such transactions.

considered as well. Such indirect exposures make it necessary to assess their credit standings and require extending the rating process to all involved entities.

Whenever a direct borrower depends on a group, its credit risk depends on the potential support of the group. The lender has therefore an implicit indirect exposure to the group's risk. Without considering support, a small and medium size enterprise (SME) belonging to a group would be assessed in the same way as a standalone SME that does not belong to any group. An implication is that legal links should be recorded and the credit risks of all related entities should be assessed.

The hierarchy of counterparties, which records how different counterparties are linked, is a critical piece of the rating system. Counterparties are organized in groups of entities, such as a holding company and all its subsidiaries. Monitoring hierarchies of entities is not a trivial task as groups keep merging and making acquisitions. Modern systems have a module dedicated to the management of such hierarchies of entities and help analysts to define with whom the bank is at risk.

The system requires adequate identifiers of corporations. Names are not adequate. For example, Continental could apply to an airline with that name or to a manufacturer of automobile products. Or IBM and International Business Machines are the same entity. Common identifiers include the CUSIP and ISIN codes or tickers, for example. The counterparty management system should include a permanently monitored mapping between internal identifiers of the bank's counterparties and these several external identifiers.

If hierarchies are known, the assignments of exposures to counterparties follow risk transfer rules. Credit exposures are allocated to corporations and, eventually, to entities other than the direct borrower, if the head of group is more relevant. Risk transfer rules define with which firm the lender is at risk and allocate exposures accordingly. The exposure is transferred internally to the entity that the bank considers as most adequate for allocating the amount at risk. Without adequate transfers, exposures on some entities that effectively concentrate the risk, although they are not direct borrowers, can be underestimated. Credit risk transfers also apply when a substitution rule is implemented for credit risk mitigation. A substitution rules applies when the relevant credit risk is that of a guarantor or when it depends on the credit quality of issuers of securities pledged as collateral.

Other rules relate to contagion effects across counterparties. For example, the default of a group might trigger the default of all or some subsidiaries, even though these do not default. Contagions are not only legal but some also follow internal rules. The bank that has a relationship with a group might judge that the default of the head of a group is actually dependent on default events occurring within the group, with contagion to the whole group. Contagion is legal when the obligor's debt is subject to covenants that trigger default if other debts within a group default.

Such interdependencies between counterparties are related to the notion of support. Support refers to financial assistance provided by a head of group to subsidiaries. Support, unlike legal guarantees, is not legally binding. It depends on the willingness of the head of group to assist subsidiaries by providing guarantees or funds to affiliated companies. It is an implicit commitment, effective or not, of the group for supporting a subsidiary. If support is ignored, a small subsidiary is rated as if it is standalone. Since the head of a group usually has a better credit standing than its subsidiaries, ignoring potential support would be misleading. For differentiating these two cases, the standalone credit standing of the firm and its credit standing when considered as part of a group should be assessed.

When a subsidiary benefits from an effective and positive support from a head of group, the subsidiary will not default unless the supporting entity does. The risk is similar to a formal guarantee that allows a risk transfer to the supporting entity. But support is not necessarily positive. Negative support occurs when it increases, rather than decreases, the default probability. For example, the default of the supporting entity might trigger the default of the borrower, perhaps because a subsidiary is highly dependent on the holding company business, or because the holding company has access to the assets of the subsidiary if it faces difficulties. A negative support deteriorates the credit quality of the direct borrower.

Negative support depends on the legal framework. For example, when a holding company is not allowed to access the assets of the subsidiary, negative support might not exist. The relation can be both ways, between the supporting entity and the supported entity. Or the credit standing of the head of group might depend on the subsidiary's default because the subsidiary's survival is critical for the core business of the holding company. The borrower's default would then trigger default of the supporting entity. The reverse is also possible. This ambiguous effect of support has a direct effect on the determination of the default probability of a subsidiary subject to positive or negative support.[9]

There are sources of indirect exposures other than groups' hierarchies. Exposures to securitizations or funds are exposures to the underlying issuers of assets within the portfolio. A "look through" policy identifying the indirect exposures to the underlying obligors raises difficulties, either because the information is not disclosed or because it is overly complex to manage.

20.3.3 Rating Grids

Internal credit ratings grids have several components synthesized in the final rating assigned to a direct counterparty. The intrinsic rating assesses the borrower's risk as a standalone entity. The overall rating depends on a group's support within the hierarchy of the firm, if any. For assessing support, it is necessary to assess the credit standing of the borrower, that of the supporting entity and the "strength" of the support. Combining support assessment with intrinsic rating provides the borrower overall rating.

Issuers' rating criteria include qualitative assessment of the counterparty's credit standing plus quantitative indicators. Criteria and their weights vary across industries and across counterparty types, such as corporations versus financial institutions. Moreover, the rating "model" is not compensating. Some criteria might be critical and drive the rating independently of other criteria. For example, the legal guarantee of a sovereign entity might be enough to assign a sovereign rating rather than the standalone rating of the direct borrower. In other cases, criteria might compensate within the rating grid, meaning that favorable and unfavorable items offset to a certain extent. In some instances, statistical models are used for small firms, but larger corporations require some judgmental component through rating grids.[10]

The intrinsic analysis is followed by an assessment of support. The final rating is the synthesis of the intrinsic rating, plus an assessment of support and a rating of the supporting entity, if any. In the example of corporations, the three building blocks that make up the final rating accumulate certain information, as illustrated below with simplified grids (Tables 20.2, 20.3 and 20.4).

9 The calculation of default probabilities of two dependent entities is addressed in Chapter 23.

10 From a statistical standpoint, "rating models" might fit to the data when trying to explain ratings by various observable variables. Empirical findings show that external ratings are sensitive to operating profitability, measured as return on assets, size and financial leverage (debt to equity ratio), and market to book value, when available, which presumably capture the efficiency of capital markets in assessing the strengths and weaknesses of firms.

TABLE 20.2 Intrinsic rating criteria

1. Intrinsic rating	*Corporations*
Fundamentals	
Industry	Growth potential, cyclicality, technology, capital, competition, regulations, regional dynamics, barriers to entry
Position and market share, business model	Size and market share, diversity and stability of revenues (flow business), business model, reputation, alliances
Management and governance	Experience and track record, achievement of goals, efficiency
Financials	
Financials	Key ratios: debt:equity; management ratios; profitability (return on equity, return on assets)
Financing	Access to various sources of financing, existence of other committed lines of credit
Repayment ability	Cash flow and revenue generation, flexibility of costs, debt cover ratios
Intrinsic rating	C

TABLE 20.3 Support assessment criteria

2. Support *Existence of a supporting entity (holding company): "support" does not imply any legal guarantee*	
Core business of holding company	Yes
Track record of support	Good
Economic and technology links	Some
Support is financial	No
2. Assessment of support	Strong
3. Supporting entity	
Rating of supporting entity	B

TABLE 20.4 Final rating assessment

4. Synthesis	
Intrinsic rating	C
Rating of supporting entity	B
Quality of support	Strong
Existence of a country cap rating	No
Final rating	**B**

20.3.4 Mapping Ratings to Default Probabilities

In the internal ratings-based approaches of the regulations, internal ratings should be mapped to default probabilities. In many instances, notably for large corporations, internal default statistics are limited. With such "low-default portfolios", external statistics can be used to assess default frequencies.

A common technique to derive default frequencies is to map internal ratings with external ratings and use the correspondence between external ratings and default probabilities to obtain the missing link from internal ratings to default probabilities (Figure 20.5). Because external ratings apply to debt issues, not to issuers, internal ratings are comparable only to external senior unsecured ratings (of issues). The principle is simple, but its implementation might raise practical difficulties or rely on unrealistic assumptions.

The usage of external default frequencies has biases. Relatively few corporations are rated and the bank's portfolio does not have the same composition as the portfolio of rated companies. Using external ratings implies that the portfolio of corporations rated externally is considered as representative of the bank's portfolio, which is generally not true. For example, firms of relatively small sizes are often considered as of lower credit standing than large firms. If so, firms of lower size should map to the lower ratings of the scale. This depends on how internal rating grids are designed.

The correspondence between internal ratings and external ratings for the same firms is generally not exact. In some cases, the internal rating grid will result in the same rating as that assigned by agencies, and in others it will not.

Comparing internal and external ratings is feasible only if the internal and external rating scales are similar. A cross-tabulation of the internal ratings and external ratings for the same firms might reveal inconsistencies. Some discrepancies might arise from the differences between the dates of ratings. These would be corrected by using ratings that are not too distant in time. In addition, there might some systematic bias between internal and external ratings, depending on how conservative, or not, the credit analysts of the bank are relative to rating agencies. Such biases should be corrected. For example, if the bank ratings are consistently below the external ratings for the same firms, the internal ratings should map to higher external ratings for assigning the default frequencies of agencies. In general, the cross-tabulation of internal and external ratings should show a concentration of matches along a diagonal of the matrix, and the diagonal could be used for mapping purposes, for lack of a better proxy.

The basic assumption of external ratings is that a given rating matches a certain default probability, across all rated entities, banks and other financial entities. The standard argument for considering that a same rating measures the same risk across counterparty types is that expert judgments correctly assess credit standing in the same way. In that case, ratings and default probabilities map one to one. This is an acceptable assumption for large corporations. However, considering that bank, municipalities and corporations have the same default frequency if they have the same rating is an assumption given the low number of defaults in some categories of counterparties.

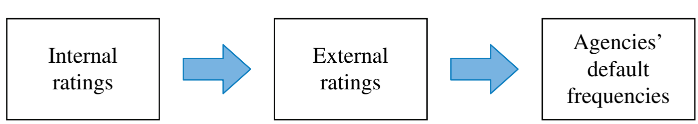


FIGURE 20.5 From internal ratings to external ratings and default statistics

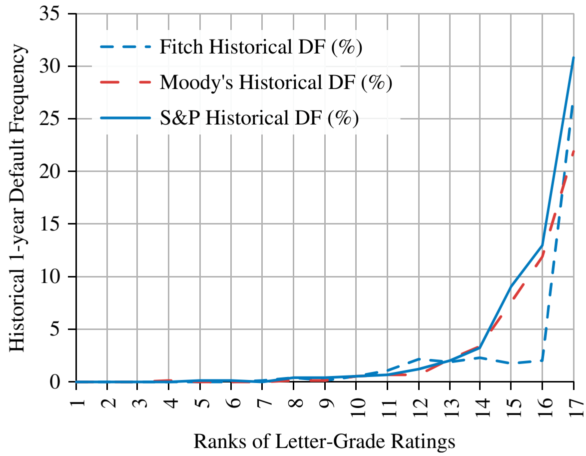


FIGURE 20.6 Letter-grade ratings and historical default frequencies

There are other statistical difficulties when linking ratings to default frequencies. For external ratings, the curve showing annual default frequencies by rating class is not necessarily monotonously increasing, at least when using a detailed rating scale. There are mismatches between the ranks of ratings and the rank of default frequencies. Smoothing the curve is an ad hoc methodology.

Figure 20.6 shows such mismatches. Each rating class is assigned a rank: "1" is the highest rating, or AAA for Standard & Poor's and Fitch ratings and Aaa for Moody's, and "17" is the lowest grade considered, CCC/C for the first two rating agencies and Caa/C for Moody's. The figure plots the ranks of ratings with the historical averaged one-year default frequencies, averaged from 1983 until 2004. At first sight, the curves look pretty smooth. But zooming in, the highest grades, in Figure 20.7, show that default frequencies are not a monotonous increasing function of the ranks.

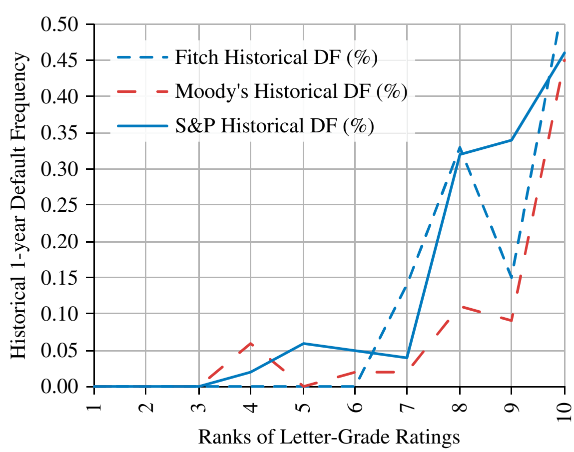


FIGURE 20.7 Letter-grade ratings and historical default frequencies: zooming in on investment grade ratings

The mapping process is very common in spite of such limitations, with the two usual steps for low-default portfolios, from internal to external ratings and from external ratings to historical default frequencies.

There are not many alternatives for low-default portfolios. Default probability models provide point-in-time default probabilities derived from equity prices (Chapter 21). These market-implied default probabilities are significantly higher than historical default statistics.[11] Moreover, the model allows comparing the distributions of default probabilities by rating class. It can be that the distributions of default probabilities of firms by rating class significantly overlap between adjacent rating classes.

20.4 APPENDIX: RATING SCALES OF RATING AGENCIES

S&P		
AAA	Highest	A1
AA	Strong	A2
A	Still strong	A3
BBB	Adequate but sensitive adverse conditions	B
BB	Major uncertainties	C
B	Strong uncertainty	D
CCC	Vulnerable	
CC	Highly vulnerable	
C	Still facing obligation despite bankruptcy or like action	
D	D is non-performing. Default materialized	

Moody's		
Aaa	Highest	P-1
Aa	High grade	P-2
A	Upper grade with some risks	P-3
Baa	Medium grade. Not outstanding; speculative	
Ba	Speculative elements. Risk exists	
B	Small insurance of making payments	
Caa	Poor standing. Possibly, some issues defaulted	
Ca	Speculative to a high degree	
C	Extremely poor prospects	

11 See, for example, Moody's Analytics (2011), Moody's market implied ratings description, methodology, and analytical applications, [104].

21

Default Models

Two families of models of default probabilities have been developed. In the reduced-form models, default events are seen as a random process, where defaults occur as unexpected surprises. In the structural model, a causal framework explains default probabilities as an outcome of the value of the firm, which can be estimated from market prices.

The default intensity models see default events as a stochastic process driven by the intensity of default, a parameter measuring the marginal default probability per unit of time. This default intensity is calibrated from historical default frequencies.

Under the structural model, a default event occurs when the value of the assets of a firm is lower than the contractual obligations of its debt. Such interpretation of default requires assumptions on the instability of the firm's return on assets and its financial structure. The structural model of default can be implemented to assess the market-implied default probabilities. This approach is intuitive because it refers to economic fundamentals, such as the structure of the balance sheet of the firm.

This chapter covers both families of models. The default intensity model is used for deriving the theoretical relation between market spreads and default probabilities and the distribution of times to default. The conceptual framework of the structural default model is detailed. Its application for deriving market estimates of the expected default probability of firms follows. The technical calculation of the default probability from the structural model is provided in the Appendix in section 21.3.

Contents

21.1 DEFAULT INTENSITY MODELS

The "reduced-form model" of default sees default as a stochastic process. The intensity models serve for modeling the time elapsed between rare events. They have been implemented for capturing "jumps" in market prices, and they apply as well to defaults in the credit risk universe. The default probability is supposedly known and measured by the intensity of default, which represents the number of defaults per unit of time. The intensity is also called "hazard rate".

21.1.1 The Intensity Model

Default intensity, or hazard, models are continuous-time models. Over the small interval Δt, the intensity is the marginal probability of default:

$$D_{t,t+\Delta t} = \lambda_t \, \Delta t$$

The default intensity is λ_t, which is in general time dependent. It is measured per unit of time: if the unit is one year, the intensity should also be the probability of default over one year. The default intensity can be calibrated from historical default statistics. The survival probability over the same time interval is: $1 - \lambda_t \, \Delta t$. In continuous time, the small interval Δt tends towards zero and is replaced by the infinitesimal dt. The model becomes identical to the so-called rare-event process, or jump process. In a jump process, the size of an event is independent of time but the probability of occurrence increases with the length of the time interval. The increments in the number of events are either zero or one.

The time to default follows an exponential distribution. The horizon, from 0 to t, can be decomposed into n equal intervals of length $\Delta t = t/n$. If λ is a constant default rate,[1] the probability of survival from 0 to t is the product of survival probabilities in each time interval: $(1 - \lambda \, \Delta t)^n$. The survival from 0 to t has a limit[2] when n increases:

$$S_{0,t} = \exp(-\lambda t)$$

Because of the minus sign in the exponent, the survival probability is a decreasing function of time.

The forward default probability at t, as seen from an initial date, is the probability of default during Δt conditional on survival until t, equal to the product of the probability of defaulting within Δt and the survival probability until t:

$$P[\text{Default over}(t, t + \Delta t) | S_{0,t}] = \lambda \, \Delta t \exp(-\lambda t)$$

This is the equivalent, over small intervals, of the discrete forward default probability.

1 An intensity varying with time can be approximated intensities that are constant over discrete time periods but variable across time periods.

2 The survival probability is $[1 - \lambda \times (t/n)]^n$. When n increases, the expression $(1 + x/n)^n$ tends towards $\exp(x)$.

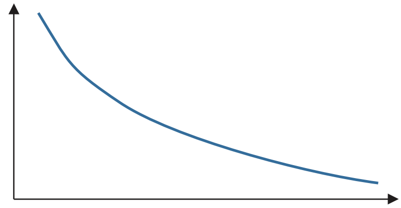


FIGURE 21.1 The exponential distribution of time to default

The probability of default between two distant dates, a and b, is the summation of the default probabilities over all intervals between a and b, which is an integral of the probability density:[3]

$$D_{a,b} = \exp(-\lambda a) - \exp(-\lambda_b)$$

The probability of a default occurring between 0 and t is obtained by replacing a by 0 and b by t:

$$D_{0,t} = 1 - \exp(-\lambda t)$$

The probability of the time to default, t, is the probability of surviving until t:

$$S_{0,t} = \exp(-\lambda t)$$

The time to default follows an exponential distribution (Figure 21.1). The survival probability tends towards zero as the time tends to infinity. The expectation of the time to default is simply $1/\lambda$ and is equal to its standard deviation.[4]

Default intensities are calibrated to observed discrete default frequencies. For example, if the one-year default probability is 4%, the survival probability over one year is: $S_{01} = \exp(-\lambda 1) = \exp(-0.04) = 96.1\%$. The default probability in continuous time is: $1 - 96.1\% = 3.9\%$. Inversely, the default intensity matching the discrete default probability of 10% is: $1 - S_{01} = 90\% = \exp(-\lambda 1)$, which implies $\lambda = 10.54\%$.

The default intensity matching a discrete default probability, D_{0t}, is such that: $1 - D_{0t} = \exp(-\lambda t)$. It is:

$$-\lambda t = \ln(1 - D_{0t}) \quad \text{and} \quad \lambda = -\ln(1 - D_{0t})/t$$

With the last example, where $D_{01} = 10\%, \lambda = -\ln(1 - D_{0t})/t = -\ln(0.9)/1 = 0.1054$. The expectation of the time to default is: $1/\lambda = 1/0.1054 = 9.488$ years. It is also the standard deviation of the time to default.

3 The probability is: $\int_a^b \lambda \exp(-\lambda t)dt$. The integration between two dates of the instantaneous forward probability uses the variable change $u = -\lambda t$: $du/dt = -\lambda, du = d(-\lambda t) = -\lambda\, dt$. The integration can be done with respect to u instead of t: $\int_a^b \lambda \exp(-\lambda t)\, d(-\lambda t)$. The bounds of the integral change when we replace u by $-\lambda t$ and become: $-\lambda a - \lambda b$. Substituting: $\int_{-\lambda a}^{-\lambda b} \exp(u)du = \exp(-\lambda a) - \exp(\lambda b)$.

4 This can be seen by using the expression of expectation for a continuous variable: $E[t \exp(-\lambda t)] = \int_0^\infty t \exp(-\lambda t)dt = \frac{1}{\lambda} \int_0^\infty \lambda t \exp(-\lambda t)d(\lambda t)$.

The constant intensity is a restrictive assumption. Piece-wise intensity models use a constant intensity over a period but allow the intensity to vary across time periods. The periodical intensities are additive over the horizon. With two periods, from 0 to t_1 and t_1 to t_2, with marginal default intensities over each period being λ_1 and λ_2, respectively, the cumulative default probability over the two periods is the sum of the marginal default probabilities applicable to each period:

$$P(0 < T \leq t_2) = \exp(-\lambda_1 t_1) + \exp(-\lambda_2 t_2) = \exp(-\lambda_1 t_1 - \lambda_2 t_2)$$

Piece-wise intensity models can be calibrated over the corresponding marginal default probabilities from historical term structures of default frequencies.

21.1.2 Credit Spread, Implied Default Intensity and Recovery Rate

Credit spreads are related to default probabilities and loss given default. The relationship results from the two alternate methods for valuing a risky bond: discounting expected payoff at the risk-free rate, or discounting the contractual payoffs at the risky rate. There is a set of probabilities such that the two valuations provide the same value. These probabilities embed the risk aversion of market investors and are the risk-neutral probabilities. The default intensity model can be used to make explicit the relation between market spreads and the risk-neutral probabilities of default and survival.

The discount factor, in continuous time, applying to a future date t is: $D(t) = \exp(-yt)$, where y is the discount rate. A bond value discounts the promised payoffs at the risky yield to maturity y. Under risk-neutral valuation, the value also discounts the expected cash flows at the risk-free rate. Expectations are calculated under risk-neutral probabilities. The risk-neutral intensity of default is $\lambda(t)$ and the recovery rate is R, in percentage of amount due.

Consider valuation over a small time interval in a continuous time setting.[5] The small period of time Δt is defined by two time points, t and $t + \Delta t$, such that Δt tends towards zero. The default probability from t to $t + \Delta t$ is: $\lambda\,\Delta t$. The probability of not defaulting in the same interval is: $1 - \lambda\,\Delta t$. The payoff at any intermediate date t is F_t under no default and the payoff under default is R. Over Δt, the value of the bond at the beginning of the interval $[t, t + \Delta t]$ discounts the payoffs under survival and the recoveries under default at the risk-free rate y_f:

$$V(\text{bond}) = F_{t+\Delta t}\exp(-y_f \Delta t)[(1 - \lambda\Delta t) + \lambda R \Delta t]$$

Since the argument of exponential is small, its first order approximation is: $\exp(-y_f \Delta t) \cong (1 - y_f \Delta t)$. When expanding the bracket terms, the second order terms, Δt^2, are ignored when the small interval Δt tends towards zero, and the expression simplifies to: $1 - \Delta t[y_f + \lambda(1 - R)]$.

The relation with the credit spread is obtained from the alternate valuation of the bond with the risky rate. The value of the bond at t is the discounted value of the payoff at the end of

5 This is inspired by the simple presentation of Duffie, D., Singleton, K. J. (2003), Credit Risk: Pricing, Measurement and Management, [58]. The default intensity is constant over a small time internal, and its subscript is dropped.

interval $F_{t+\Delta t}$. The discount rate is the risky yield, which is equal to the risk-free rate plus the credit spread s. The bond value, $V(\text{bond})$, is:

$$V(\text{bond}) = F_{t+\Delta t}\exp\left[-(y_f + s)\right]\Delta t = F_{t+\Delta t}\left[1 - \Delta t\left(y_f + s\right)\right]$$

The two alternate valuations are:

$$\text{Risk neutral valuation: } V(\text{bond}) = F_{t+\Delta t}\left\{1 - \Delta t\left[y_f + \lambda(1 - R)\right]\right\}$$

$$\text{Risky yield valuation: } V(\text{bond}) \cong F_{t+\Delta t}\left\{1 - \Delta t\left(y_f + s\right)\right\}$$

Eliminating the contractual payoff and equating those two values, the credit spread s is:

$$s = \lambda(1 - R) = \lambda\text{LGD}$$

The credit spread is the product of the risk-neutral default intensity and loss given default (LGD). It is identical to the expected loss under risk-neutral probabilities. The relation is used to derive the risk-neutral default intensity implied in credit spreads.

21.2 THE OPTION THEORETIC APPROACH TO DEFAULT

A common view of a firm's default is that the firm does not generate enough free cash flow to face debt obligations. Such view of default is not sufficient for modeling default. If the cash flow is not high enough to repay debt over a certain period, or is negative, new debt or equity can be raised. Raising new funds should be feasible as long as future cash flows are high enough to repay both current and additional debts. The cash flow model of default is flawed because it has a limited horizon: persistent cash flow deficiencies make default likely, but temporary deficiencies do not.

The economic value of the firm, also called the firm's asset value, or simply the asset value, is the present value of future free cash flows at the cost of capital. Under corporate finance theory, the enterprise value is the discounted values of all future cash flows, the relevant discount rate being the weighted average cost of capital.[6] The firm's value discounts all future free cash flows. The free cash flow is the cash flow generated by operations, net of mandatory capital expenditures, hence available for shareholders and lenders. This implies that the value of the firm is identical to the sum of the values of debt and of equity:

$$\text{Firm value} = \text{Equity value} + \text{Debt value}$$

The option theoretic approach views default as an economic event triggered by a value of assets insufficient compared to the debt payment obligation over a given horizon. The approach to default modeling follows the principles set up by Merton in his seminal paper

6 The weighted average cost of capital is the weighted average of the required return on equity and of the cost of debt, using the market values of equity and debt as weights.

of 1973.[7] Default is seen as an option, held by equity holders, to sell assets to lenders with a strike price equal to the debt face value. The face value of debt is the amount to be repaid, which is distinct from the value of debt. The default option results from the limited liability feature of equity. The underlying of the put option is the random firm's value.

21.2.1 The Option Theoretic Framework

Both equity value and debt value vary with risk but they sum up to the firm's value. However, the debt contractual repayment at some horizon does not change and is the strike of the default option. If the asset value is well above the debt payment due, the put option is far out-of-the-money and has no value. If the asset value gets closer to the debt obligation, the put gains value. A credit-risky loan is seen as a risk-free loan minus the value of the put option to default given by lenders to equity holders. The credit risky component of debt is the put option to default. For the firm, the value of the liability is lower than a risk-free debt by the value of the option to default.

Credit-risky debt = Risk-free debt – Put option to default

The value of the risk-free debt is constant if there are no interest rate movements and when asset value changes. Figure 21.2 shows how the value of the risky debt changes when asset value varies. When the asset value is well above the debt payment, the value of debt tends towards the value of a risk-free debt. When asset value declines, and gets closer to debt obligation, the put option gains value. The value of the risky debt declines, down to zero as asset value tends towards zero.

From the shareholder's standpoint, the put option to default is the equivalent of a call option on asset value. Shareholders hold a positive value as long as the asset value is above the face value of debt. Their value is the asset value minus the debt payment. Under this view, shareholders have a call option on asset value with strike equal to the debt face value (Figure 21.3). When asset value declines below the face value of debt, the value of equity gradually declines to zero. When asset value is well above the debt contractual obligation, the value of the call option is the asset value plus the time value of the option. The time value is the

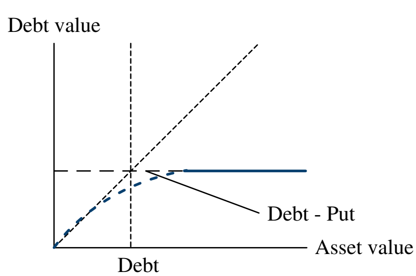


FIGURE 21.2 Debt value equals risk-free debt minus the put option to default

7 The original reference is Merton, R. C. (1974), On the pricing of corporate debt: The risk structure of interest rates, [101].

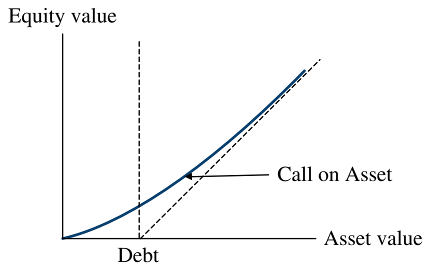


FIGURE 21.3 Equity as a call on asset value

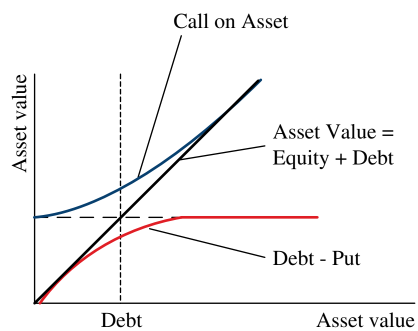


FIGURE 21.4 Asset value: sum of a call on asset value plus risk-free debt and minus the option to default

additional value of higher gains due to the chances that the asset value gets even higher than today.

Accordingly, there are two views on equity. Shareholders can be seen as holding the assets plus the put option to sell them with strike equal to debt payment. Alternatively, shareholders can be seen as holding a call option on asset value with the same strike. The two views are consistent because the call option is the mirror image of the put option netted from the value held by lenders. Asset value always sums up equity value and debt value. In Figure 21.4, the asset value is obtained by adding up vertically the values of the risky debt and the equity.[8] The first diagonal represents the asset value in this figure.

21.2.2 Implementing the Structural Model of Default

Moody's KMV Corporation implemented the structural model of default to estimate the default probability.[9] The process requires modeling the distribution of asset value at horizon. A default occurs when asset value is below debt obligation at horizon. The probability of asset value

8 This is the put–call identity as: Call + Put = underlying, with underlying being the asset value of the firm.

9 The principles of the Moody's KMV model are in Kealhofer, S. (2003), Quantifying credit risk I: Default prediction, [87] and in Kealhofer, S. (2003), Quantifying credit risk II: Debt valuation, [88].

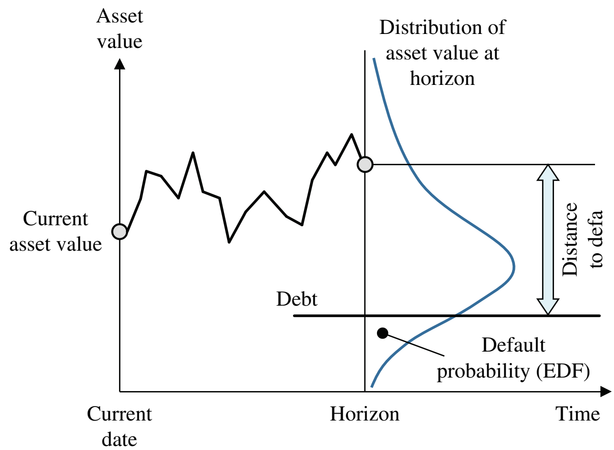


FIGURE 21.5 Moody's KMV modeling of EDF

being lower than debt payment is represented by the area under the curve and the horizontal line representing the debt face value (Figure 21.5).

The distribution of asset value is generated by random time paths of asset returns starting from the current asset value. The probability that the asset value falls under the debt contractual obligation at horizon is the theoretical EDF. The time path of the asset value is stochastic. There are multiple time paths, each one reaching a specific value along the distribution at horizon. Since the asset return over a short period is normally distributed, the asset values at horizon follow a log-normal distribution. If default occurs only at horizon, the distribution defines the probability of default.

The distance to default is the gap between the asset value at horizon and the horizontal line representing the debt face value. Higher distances to default at horizon mean that the probability of staying above the debt obligation is higher in subsequent periods. Hence, the distance to default is a measure of the credit state of the firm at horizon.

In Moody's KMV Credit Monitor, the horizon is fixed and the debt obligation combines short-term debt plus a fraction of long-term debt. The default probability is called "EDF", for "Expected Default Frequency". The EDF is forward looking, unlike historical data, because it is implied from equity prices that are forward looking.

The asset value is not observable, only equity value is. The distribution of the asset value at horizon has to be modeled from the distribution of equity prices for listed companies. The derivation of the dynamics of asset values from equity prices and the volatilities of equity returns is feasible because the equity is seen as a call option on asset value. Under this view, the value of equity and the volatility of equity returns are both functions of the underlying asset value and of the asset volatility. These two equations allow deriving the unobservable asset value and the asset volatility. In compact notations:

$$E = f[D_T, A_0, \sigma(A), T]$$

$$\sigma(E) = g[D_T, A_0, \sigma(A), T]$$

In these equations:

T is the horizon.
D_T is the fixed book value of debt at horizon.
E is the market value of equity.
A_0 is the unobservable market value of assets at date 0.
$\sigma(A)$ is the volatility of the asset return.
$\sigma(E)$ is the volatility of the equity return.

The Black-Scholes formula applies to equity, as a call on asset value, with exercise price equal to book value of debt and maturity T:

$$E_0 = A_0 N(d_1) - D_T\, e^{-rT}\, N(d_2)$$

D_T is the debt value at horizon.
r is the risk-free rate.
$N(d_1)$ and $N(d_2)$ are the Black-Scholes formula coefficients, which depend on asset volatility.

The first equation is the formula for a call on asset value. A second equation derives the volatility of asset value $\sigma(A)$ from the one of the equity value $\sigma(E)$.[10]

21.2.3 The EDF Model

The commercial version of the structural model is "Moody's KMV Credit Monitor". In Credit Monitor, default occurs at maturity of the facility if it is either shorter than horizon or at horizon. The calibration process relies on proprietary default databases. It is necessary to have a sufficient number of defaulted companies for mapping distances to default to actual default frequencies. Periodical fits are necessary to ensure that the model actually fits the data.

Once could expect that the Credit Monitor EDFs approximately match the default probabilities mapped to the rating of a company. In fact, the market-implied EDFs tend to be higher than the average default frequencies by rating class, but the model does discriminate between firms that default and firms that do not. The EDFs are point-in-time measures where default frequencies are averaged over long horizons. In theory, the market-based EDFs should anticipate rating changes. Ratings tend to lag somewhat the changes in credit standing of issues, simply because they are long-term oriented. Unlike ratings, which are adjusted at discrete points in time, the EDF and its drivers are continuously monitored.

Moody's KMV Credit Monitor provides all information to compare EDFs and historical default frequencies. An EDF can be explained by the variations across time of the asset value, of asset volatility and of the default point. Providing the time series of these parameters allows interpretation of why a firm defaults or not at a given date. Figure 21.6 shows some typical time profiles of asset values and debt face values. The continuous line represents the asset value and

10 In order to derive this relation, the Ito lemma is used, which provides the drift and the volatility of the stochastic process of a function $A = f(E)$ when both the asset value A and the equity value E follow stochastic processes. The relation between the volatilities of equity and asset value is: $\sigma(E) = A_0\sigma(A_0)N(d_1)/E_0$. The asset volatility is: $\sigma(A) = E_0\sigma(E)/A_0 N(d_1)$.

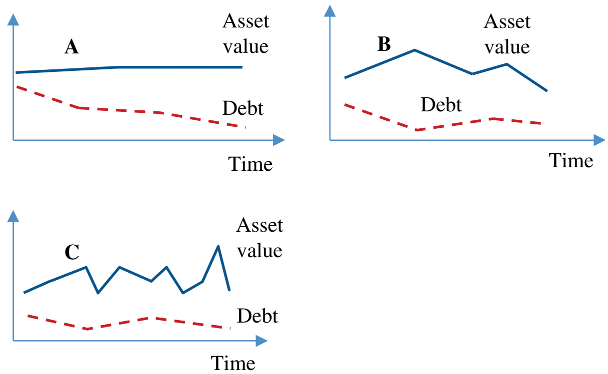


FIGURE 21.6 Time profiles of asset values and debt

the dashed line the debt payment, as a function of time. In the A case, the EDF declines because leverage decreases. In the B case, the asset value declines while leverage is constant, and the default probability increases. In the C case, the asset volatility increases as time passes, resulting in a higher default probability. All such variables can also be used to anticipate default. The EDF should increase when asset value declines, or when asset volatility increases, or when the default point or leverage increases. How those drivers interact when time passes can be used as an early-warning system of adverse rating migrations and defaults.

Several studies compared the Moody's KMV EDFs with similar results obtained with other techniques, such as scoring and using credit ratings. They rely on the cumulative accuracy profile (or CAP[11]) to show how well the model discriminates between defaulters and non-defaulters. Published results suggest that the accuracy of Credit Monitor is relatively high.

A limitation of Credit Monitor is that it applies to public companies since its uses equity prices. Moody's KMV Corporation provides another model, the "Private Firm Model", which necessitates minimal information on a private company to model default risk. The basic information is in financial statements. The usage of the model requires documenting the industry and region of the firm. The principle is to proxy, through this information, what the EDF should be.

21.2.4 Theoretical Values of the Option to Default and the EDF

The calculation of the theoretical EDF is relatively easy. The process requires the formulas of the put option to default plus the usual assumptions on the asset returns leading to the log-normal distribution of the asset value at horizon. The option formula is used to make a sample calculation in the Appendix in section 21.3.

The today asset value is A_0. The random asset value at horizon is A, such that $A = A_0 \exp(rT)$, where r is the random continuous return and T is the horizon. It follows that: $\ln(A/A_0) = rT$. If the return r is normal, the ratio of final to initial asset value is log-normal. The logarithm of the ratio has an expectation $(r - \sigma^2/2)$ and its volatility is $\sigma\sqrt{T}$.

11 See Chapter 20.

The ratio A/A_0 is:

$$\frac{A}{A_0} = \exp\left[(r - 1/2\sigma^2)T + Z\,\sigma\sqrt{T}\right]$$

In this expression, the variable Z is standard normal, and the corresponding distribution of the ratio A/A_0 is such that the logarithm has the desired mean and standard deviation.

The asset value A at horizon is obtained by assigning a value to the normal standard variable Z. If the default probability is 1%, the asset value hits the default point when the value of Z is −2.33. The number −2.33 measures the downside variation of Z required for hitting the default point. The absolute value of Z, 2.33, can be interpreted as the distance to default of the standard normal variable Z. It is the number of standard deviations of the downside deviation of the asset value required for a default to occur.

The minimum value of the ratio for hitting a default state with a given probability, d, is the d-quantile of the distribution. This d-quantile is such that $Z < Z_d$, where Z_d is the d-quantile of the standard normal distribution, noted: $d = \Phi(Z_d)$. If the default probability is 1%, the value of Z_d is: $Z_d = \Phi^{-1}(d) = -2.3263$.

With a risk-free continuous return of 10%, an annual volatility of the return of 30% and an horizon of one year, the 1%-quantile of the ratio A_d/A_0 is derived from the above equation by setting Z to −2.3263:

$$\frac{A_d}{A_0} = \exp\left[(10\% - 30\%^2/2) \times 1 - 2.32635 \times 30\% \times 1\right] = 40.7\%$$

The firm defaults with a probability of 1% when the ratio is 40.7%: $P(A_d/A_0 \leq 40.7\%) = 1\%$. The value of the threshold A_d is also the value of the debt that would trigger default with a probability of 1%. The probabilities of higher values of the ratio are migration probabilities to non-default states.

Mathematically, the relation between the ratio and the threshold value of Z_d of Z is obtained by replacing Z_d by $\Phi^{-1}(d)$:

$$\frac{A_d}{A_0} = \exp\left[(r - 1/2\sigma^2)T + \Phi^{-1}(d)\sigma\sqrt{T}\right]$$

Figure 21.7 shows the ratio of final to initial asset value, using the same data as the above example, as a function of the quantiles of asset value. The difference between the ratio and 40.7% is a measure of the distance to default. The figure provides the default probabilities of final credit states for values of the ratio higher than 40.7%. Inversely, the figure shows the default probabilities matching the various debt thresholds expressed in percentage of initial asset value. For example, if the default point is such that the ratio A_d/A_0 is 100%, the default probability is close to 70%.

The default probability d from a given threshold point measured by A_d/A_0, in percentage of initial asset value, is given by:

$$\Phi^{-1}(d) = \frac{\ln(A_d/A_0) - (r - \sigma^2/2)T}{\sigma\sqrt{T}}$$

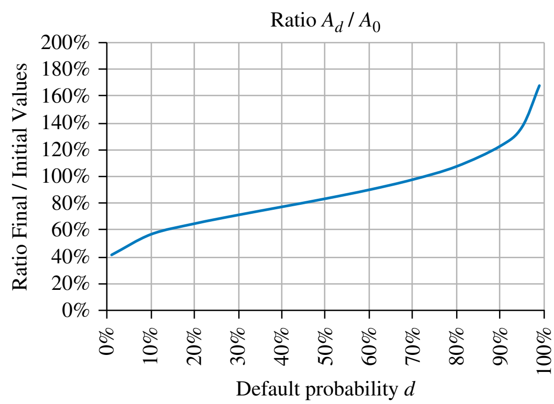


FIGURE 21.7 Quantiles of the ratio of final asset value to initial asset value (expected return 10%, volatility 30%, horizon 1 year)

The default probability, given A_d, is:

$$d = \Phi\left[\frac{\ln(A_d/A_0) - (r - \sigma^2/2)T}{\sigma\sqrt{T}}\right]$$

21.2.5 Variations of Merton's Model

Some variations of Merton's model implying default when asset value goes under a preset value of debt exist, and others are extensions extending the scope of the model to other variables.

One benefit of the Merton model is that it looks forward. Another is that asset value captures the present value of the entire stream of future free cash flows. Hence, a single negative cash flow cannot fail the firm, and cumulated cash drains cannot either, as long as the firm has sufficient upside potential.

Moody's KMV implementation has limitations. Default occurs only at a given horizon, where asset value is below the debt value. In reality, default can occur at any time when the time path of asset value crosses the default point. Real situations are also more complex than considering debt as a bullet bond. In fact, both the asset value and the debt value, linked to interest rates and to the default risk, follow a stochastic process. In addition, the value of debt is a combination of short-term debt and a fraction of long-term debt in the commercial implementation of the model. This is only a proxy of the actual time structure of debt repayments.

21.3 APPENDIX: VALUATION OF THE PUT OPTION TO DEFAULT

The standard Black-Scholes formula applies for valuing the call on asset or the put of equity holders. The Black-Scholes formulas for the call and the put with the same underlying, strike D and horizon is T are:

$$C_0 = A_0 N(d_1) - D\exp(-rT)N(d_2) = E_0$$

$$P_0 = D\exp(-rT)N(-d_2) - A_0 N(-d_1)$$

In these equations, A_0 is the initial asset value, D is the debt payment, E_0 is the equity value, r is the risk-free rate and d_1 and d_2 are:

$$d_1 = [\ln(A_0/D) + (r + \sigma^2/2)T]/\sigma\sqrt{T}$$

$$d_2 = [\ln(A_0/D) + (r - \sigma^2/2)T]/\sigma\sqrt{T} = d_1 - \sigma\sqrt{T}$$

In these formulas, $N(x)$ stands for the standard normal distribution, ln is the Neperian logarithm, and the volatility σ is the volatility of the random asset return. It is useful to note that $N(d_1) = 1 - N(1 - d_1)$ and that $N(d_2) = 1 - N(1 - d_2)$.

The example uses an asset value of 100 and a debt of 50, resulting in a leverage ratio (debt/equity) of 1. The risk-free rate is 10% with discrete compounding. The continuous rate equivalent to this discrete rate is 9.531%. The asset value at horizon is:

$$A = A_0 \exp\left[\left(r - \tfrac{1}{2}\sigma^2\right)T + \sigma\sqrt{T}Z\right]$$

The default condition implies that Z is lower than a value Z_d of Z, such that:

$$Z_d = \frac{\ln(D/A_0) - (r - \tfrac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$$

The variable Z_d is identical to $-d_2$.[12] The distance to default is equal to d_2, the argument of the Black-Scholes formula. The probability of default is that Z is lower or equal to the distance to default. Since Z is standard normal, it is also:

$$(Z \leq Z_d) = N(-d_2)$$

12 This can be seen by changing the sign in the equation:

$$-Z_d = \frac{-\ln(D/A_0) - (r - \sigma^2/2)T}{\sigma\sqrt{T}} = \frac{\ln(A_0/D) - (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_2$$

TABLE 21.1 Equity value as a call option on assets value and default put

Elements of the EDF formula	*Numerical values**
$\ln(A_0/D)$	0.59784
$(r + \sigma^2/2)T$	0.13479
$(r - \sigma^2/2)T$	0.05583
$d_1 = [\ln(A_0/D) + (r + \sigma^2/2)T]/\sigma T^{0.5}$	2.60733
$d_2 = [\ln(A_0/D) + (r - \sigma^2/2)T]/\sigma T^{0.5}$	2.32635
$N(d_1)$	0.99544
$N(d_2)$	0.99000
Call on asset $= A_0N(d_1) - D\exp(-rT)$	50.04374
Put $= D\exp(-rT)N(-d_2) - A_0N(-d_1)$	0.04374

* Values rounded to the fifth decimal digit.

Because $N(-d_2)$ depends only on leverage, it is insensitive to the absolute sizes of asset and debt. The value of the put, P_0, can be transformed into:

$$P_0 = [D - A_0N(-d_1)N(-d_2)]N(-d_2)$$

The first term in brackets can be interpreted as the present value of the cost of default for lenders. It is the product of the default probability $N(-d_2)$ times a value of debt lower than the book value D. The difference is the expected loss under default, and is equal to the second term of the equation above. This expected loss is the expected value of asset, conditional on default.[13] Since the asset value triggering default is equal to the debt at date 1, it is still positive, so that lenders recover the expected value of assets conditional on default.

Using these equations, a sample calculation of the EDF and of the value of the put option to default can be conducted (Table 21.1). The book value of debt is 50, the value of assets is 100 at date 0 and the strike price is 50 at $T = 1$. The volatility of equity is 28.099%. This value results in an EDF of 1%, once calculations are done (Table 21.2). This is sufficient to derive all the intermediate values, shown in Table 21.1.

The theoretical value of equity E_0, valued as a call option with the above inputs, is 50.04374. It is higher than 50, the equity value without the put, because of the potential upside on asset value. The gain for stockholders equals the expected loss from debt holders, which is identical to the value of the put, 0.04374.

Various credit statistics are of interest. The distance to default is $d_2 = 2.32635$. The EDF is $N(-d_2)$, or 1%. There is an expected recovery under default because the asset value is still positive when default occurs. The expected recovery is the second term of the put value,

13 It is similar to the E-VaR or the expected value of loss conditional on asset value lower than debt.

Table 21.2 Edf and loss on debt

Distance to default = d_2	2.32635
Probability of default EDF = $N(-d_2)$	1.000005%
Expected recovery on debt	45.62565
Loss given default (LGD = 50 – recovery)	4.37435
Expected loss: $N(-d_2) \times 45.625651$	0.04374

or $A_0 N(-d_1)/N(-d_2)$. The loss given default is the difference between recoveries and debt, 50, or 4.3744. The expected loss is the loss given default times the default probability 1%.

The debt value is the book value minus the put value at date 0: $50 - 0.04374 = 49.95625$. The implied equity value has a volatility that results from the relationship:

$$\sigma(E) = A_0 \sigma(A) N(d_1)/E_0 = (100 \times 28.099\% \times 0.99544)/50 = 55.9408\%$$

The equity volatility is much higher than the asset volatility. This is due to the leverage effect. This calculation shows how to work backward from equity value and volatility to asset value and asset volatility. There is only a couple of values of these parameters that comply with the structure of all equations.

The model provides the value of the risky debt, lower than the risk-free value by the amount of the put. This allows the theoretical credit spread that applies to the risky debt to be determined.

With the risk-free rate at 10% (discrete compounding), a debt with face value 50 would require a payback of 55 at 1. There is a discount rate that makes the present value of 55 at date equal to the value today, adjusted for default risk, 49.95626. The gap between the higher discount rate and the risk-free rate is the credit spread. The risky yield making the 55 equal to the today debt value, using continuous discounting, is such that:

$$49.95626 = 55 \exp(-yT)$$

The risky yield is $y = 9.6185\%$. This rate is higher than the risk-free continuous rate, equivalent to the 10% discrete rate, 9.5310%. The difference is the credit spread, or 0.0875% in continuous compounding, or 8.7 basis points.

22

Counterparty Credit Risk

Counterparty risk is a type of credit risk that exists when both parties of a contract are exposed to a loss if any one party defaults on its obligation. Counterparty risk arises notably with over-the-counter derivatives contracts, but also exists for financing transactions backed by securities. The credit risk from loans is one-way and the amount at risk is not market driven. For derivatives, the credit risk is two-way and the amount at risk is market driven. A swap contract is the typical example. The party who receives more than it pays is at risk with the other party. But who holds the positive value, and how large is this positive value, are driven by market factors.

The determination of exposure for derivatives requires that the potential exposure of derivatives be modeled. This chapter discusses exposure uncertainty and the measurement of potential exposure for derivative contracts. A simplified example shows how the potential exposure of an interest rate swap can be estimated with a given probability at a given horizon. The rules implemented under Basel, for exposure and for the credit value adjustment, are summarized at the end.

Contents

22.1 CREDIT EXPOSURES

Counterparty risk exists,[1] in general, with exposures that are market driven, for example when securities are exchanged between two parties. Repurchase transactions involve the exchange of securities in order to borrow cash, with the agreement of repurchasing the securities at a future date. The borrower of cash is essentially pledging securities to back up its loan, which the lender of cash should return on completion of the transaction. The lender of cash is the borrower of securities and the borrower of cash is the lender of securities. A haircut is applied to the value of securities to mitigate the chances that the value of the securities falls below the amount of cash exchanged. Although the securities are liquid and have a relatively stable value, there is counterparty risk since the borrower of cash is exposed to the risk that the lender of cash fails to return the securities. In securities lending and borrowing, securities are exchanged for other securities and a similar situation arises.

But the bulk of counterparty risk results from the volume of derivatives. Derivative contracts are often initiated over the counter (OTC) on a bilateral basis. Derivatives include currency and interest swaps, options and any combination of these. Swaps exchange interest flows based on different rates, or flows in different currencies. Options allow buying or selling an asset at a stated price. In the case of swaps, the exposure shifts from one party to the other depending on market movements. In all cases, the size and the sign of the exposure is driven by the market movements.

An interest rate swap can have a positive or a negative value depending on whether the party holding the swap receives more or less than it pays to its counterparty. Suppose that a bank holds a swap receiving a fixed interest rate and paying a variable rate. If the floating rate is low enough, the bank is a net receiver of interest flows. The swap has a positive value and is similar to an asset. Conversely, the swap has a negative value, and is similar to liability, for the other party. In the event of default, the party with the positive exposure faces a loss equal to the cost of replacing the contract. Depending on market movements, the value of the contract for one party can shift in sign and in size.

Options have a potential positive value for the buyer only. The buyer faces the risk that the seller does not comply with its commitment and has credit risk to the extent that the option is in-the-money. For the seller of an option, the contract is a contingent liability and the exposure is always zero.

Counterparty risk can be reduced by various means. Netting and collateral agreements have been common tools to achieve this. The ISDA (International Swap Dealer Association) master agreements allow netting the individual exposures and exchanging only net flows due to or owed by each party. For instance, if A has two transactions with B of which liquidation values are 50 and −20, the risk exposure is the netted value, or +30. Without netting agreement, only positive liquidation values are at risk and the risk would be +50 and zero for the second transaction.

Under collateral agreements, cash or securities are pledged to the party having a positive exposure. Collateral reduces, in theory, the exposure to the period required to adjust the collateral level. Trading in organized markets such as central clearing houses eliminates most of the counterparty risk. Daily margin calls limit the risk to daily variations, since any gain or loss of value is posted with the clearinghouse. In OTC markets, the risk exists up to maturity.

1 There are various references on counterparty credit risk, for example, the book of Gregory, J. (2010), Counterparty Credit Risk, [72].

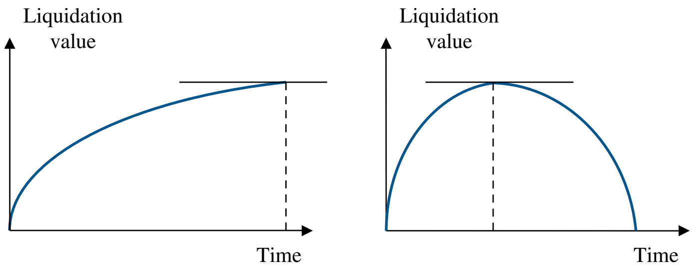


FIGURE 22.1 Time profiles of PFE

22.2 POTENTIAL FUTURE EXPOSURE

Whether from options or swaps, the future exposures have to be modeled. The current risk is the current mark-to-market value of the instrument if it is positive but the future exposure is unknown. The issue is to estimate the potential positive values of the derivative, conditional on scenarios of market parameters.[2]

The potential future exposure, or PFE, is the time profile of the upper bounds of positive exposures, at a given confidence level, at all future time points until maturity. From a risk standpoint, potential exposures are estimated by assuming that the drift of risk factors will be adverse and maximize the derivative potential value. The counterparty should do the same.

The shape of the time profile of a potential exposure depends on the nature of the contract. When there is a unique exchange of flows at maturity, the drift of risk factors follows a diffusion process and has a maximum impact at maturity. This is the case for a currency swap, as the drift of exchange rates potentially increases with time. For other contracts, such as interest rate swaps, there is an amortizing effect, as the number of remaining flows decreases as time passes. The diffusion and the amortizing effects tend to offset each other. The potential exposure increases with time when amortization is at an early stage and the value starts declining when the amortization effect dominates. At maturity, the liquidation value becomes zero. The maximum of deviations of the value of an interest rate swap occurs somewhere between now and maturity. The peak is between one-third and one-half of the residual life. Figure 22.1 shows the time profiles of potential future exposures without and with amortization effect.

22.3 CREDIT RISK FOR DERIVATIVES: METHODOLOGY

The G30 group, in 1993, defined the framework for monitoring OTC derivative credit risk. The G30 report recommended techniques for better assessing uncertain exposures, and proposed a framework for modeling the potential deviations of value for portfolios of derivatives over long periods. The report recommended that the potential and positive exposures be modeled as upper bounds of the value of the portfolio at selected time points until maturity of the

2 See Cooper, I., Mello, A. (1991), The default risk of swaps, [47]. Pykhtin, M., Zhu, S. (2007), A guide to modelling counterparty credit risk, [117], provides a useful review of the method.

instruments. The potential drift of the value was defined as a value not exceeded in more than a small fraction of outcomes, the confidence level.

The approach provides a time profile of potential positive values of the derivatives, which is updated continuously. Such time-varying exposures can be treated as "credit equivalents", or as if they were loans. The approach is the current standard practice for banks who can model the drift of derivative values as a function of the underlying risk factors.

According to standard terminology, the current value is the current risk, and the potential risk measures the potential upward deviations of positive exposures. The overall risk at any point in time is the sum of the current and the potential risks.

A potential exposure is defined as the upper bound of liquidation values, at each time point until maturity, and given a confidence level.[3] The construction of potential exposure requires modeling the distribution, at various time points, of the future value of the instrument, from which such worst-case exposures are derived.

The valuation applies to netting sets, which are the portfolios of derivatives within the scope of the netting contract. Within a netting agreement, the negative and positive values of the instruments are netted. Portfolios depend on many risk factors and the deviations of instruments are interdependent. The variations of the risk factors should comply with their dependency structure and the future value of the portfolio should be modeled, as for market value-at-risk (VaR).

The adequate methodology requires simulations, including full revaluations of non-linear instruments. The technique generates a distribution of the values of the portfolio for each future date. The upper bounds at a given confidence level result from this distribution. The composition of the portfolio changes over time. Some derivatives reach maturity and their value vanishes at that time. Discrepancies of maturities results in non-monotonous time profiles with ups and downs.

For an individual transaction, the issue is simpler to address, although some instruments might depend on several factors. When there is a single risk factor that has the most significant influence on the value of an instrument, such as the floating rate of an interest rate swap, the exercise is simple. The upper and lower bounds of the future values of the derivative, at the given confidence level, are obtained by revaluation of the instrument using the upper or lower bound values of the risk factor. The bounds of risk factors are derived from the diffusion process that is relevant for the factor. For standard processes, without mean reversion, the volatility increases like the square root of time and the distribution of its value at future dates is log-normal (Figure 22.2). Once the distribution is known, it is a simple matter to revalue the instrument and find the upper bound of its value given a confidence level.

22.4 CALCULATING THE POTENTIAL FUTURE EXPOSURE FOR AN INTEREST RATE SWAP

The determination of PFE is conducted in the case of an interest rate swap under simplifying assumptions. The example is a four-year interest rate swap, notional 1,000, receiving the fixed rate and paying the floating. The yield curve is flat. The fixed rate is 10% and equal to the floating rate at the initial date. The value of the interest rate swap at inception is zero. Dates refer

3 The guidelines are published in the Basel regulations, for example in the Basel 2 document, [21].

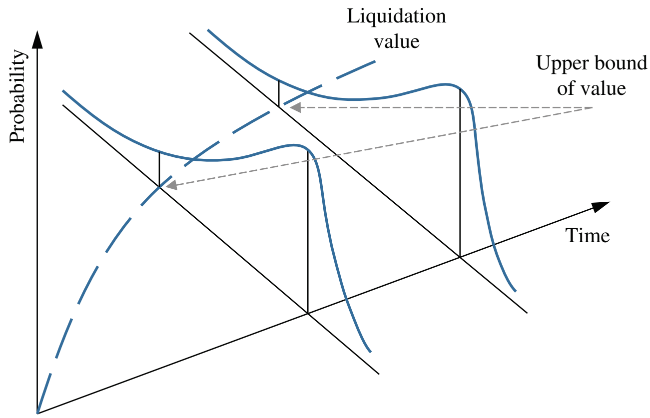


FIGURE 22.2 Distributions of future values and its upper bounds

to end of year. After one period, the floating rate declines to 8% and the swap has a positive current value.

The swap is valued as a portfolio of two bonds, long in the fixed-rate bond and short of a floating-rate bond with same maturity. The principal payments of the two bonds offset at inception and at maturity, and only the interest flows are exchanged. Interests are paid at end of years. The primary source of uncertainty is the floating rate.

If the interest rate changes, only the value of the fixed-rate leg changes and the floating leg remains at face value. For revaluing the long leg, the future cash flows, posterior to the current date, are discounted at the prevailing interest rate. The swap is revalued by deducting the constant value of the floating-rate leg.[4] Its value is zero at origination. The swap gains value if the floating rate moves down and the upper bounds of its positive value correspond to the lower bounds for the floating rate.

In the example, it is assumed that the interest rate is 8% when reaching date 1. The values of the swap are calculated at all dates with this new interest rate (Table 22.1). In the calculation,

TABLE 22.1 Time profile of the swap value at current rate*

Dates	*1*†	*2*	*3*	*4*
Fixed leg cash flows	100	100	100	1100
Current interest rate	8%	8%	8%	8%
Fixed leg value	1,051.54	1,035.67	1,018.52	0
Swap value (fixed leg value – 1,000)	51.54	35.67	18.52	0

* The values obtained are those of the net value of a fixed-rate bond minus the par value, 1,000, of the floating-rate bond. The valuation is identical to discounting only interest flows at the current rate.
† The valuation at date 1 discounts the flows posterior to date 1, of dates 2, 3 and 4. The same rule applies for other dates.

4 The swap can also be revalued by discounting at the new rate the interest flows from the fixed leg minus the interest flow of the floating leg calculated with the new rate. Both calculations are equivalent.

TABLE 22.2 Time profile of lower bound of interest rates

Dates	*1*	*2*	*3*	*4*
Square root of time (as of 1)		1.00	1.41	1.73
Volatility $\sigma\sqrt{t}$		20.00%	28.28%	34.64%
Shock ($z\sigma\sqrt{t}$, with $z = -1.96$)		−39.20%	−55.44%	−67.90%
Lower bound of interest: 8% $\exp(z\sigma\sqrt{t})$	8.0%	5.41%	4.60%	4.06%

the cash flows of the current period are assumed already exchanged and only the subsequent flows should be discounted.[5] The swap has a positive value of 51.54 at date 1. Later, the value, seen from date 1, changes because the swap amortizes and fewer flows are discounted. The time profile of these values is the current exposure of the swap.

The swap positive values increase when the floating rate declines. For obtaining the potential exposure, a downward drift of interest rates should be applied, based on volatilities applicable to dates 2, 3 and 4. Starting at date 1, when the floating rate is 8%, future values are derived by applying market shocks to the rate. If the shocks are normally distributed, the distribution of the future rates is log-normal.[6]

For deriving the lower bound of the interest rate, it is assumed that the volatility of the random shocks is 20%, annually, in percentage of the current rate. The confidence level for determining a lower bound of interest rates is 2.5%. The volatility of the random shock follows the square root of time rule if we ignore the mean-reversion of rates for simplicity. A random shock is: $z\sigma_1\sqrt{t}$. The value of z is the multiplier of volatility at date t matching the confidence level of 2.5%, or −1.96. The lower bound of the interest rate at date t is: $i_t = i_0\ \exp(z\sigma\sqrt{t})$.

Table 22.2 shows the time profile of the lower bounds of interest rates. The annual volatility is applied from date 2 since the current date is 1. For date 2, the shock on the rate is: $1.96 \times 20\% = -39.20\%$. The corresponding lower bound of the interest rate is: 8% $\exp(-39.20\%) = 5.41\%$.[7]

The variations of the value of the swap are those of its fixed-rate leg. Table 22.3 shows the values of the fixed legs using the lower bound values of the market rate. The upper bound of the swap values are the netted values of the fixed leg minus the floating-rate leg, valued 1,000. The calculations of the table show an unchanged value of 51.54 at the current date, but the values at subsequent dates are above the profile of the current risk. This new set of values forms the time profile of potential exposure. The potential risk has the typical shape of an IRS. It starts increasing, and then declines to zero (Figure 22.3).

Actual simulations would require determination of the entire yield curve at different time points, relying on interest rate models and would use a mean-reverting process of short-term rates.

5 The exposure increases before the interest payment and jumps down by the amount of interest exchanged immediately after.

6 This is assuming no mean reversion. The log-normal distribution keeps the interest rate positive whatever the magnitude of the shock applied.

7 See Chapter 11 for the expression of random shocks when the value is log-normally distributed.

TABLE 22.3 PFE of the swap

Dates	*1*	*2*	*3*	*4*
Fixed leg cash flows	100.00	100.00	100.00	1,100.00
Current interest rate	8.0%	5.41%	4.60%	4.06%
Fixed leg value	1,051.54	1,084.94	1,051.67	0
Swap value (fixed leg value – 1,000)	51.54	84.94	51.67	0

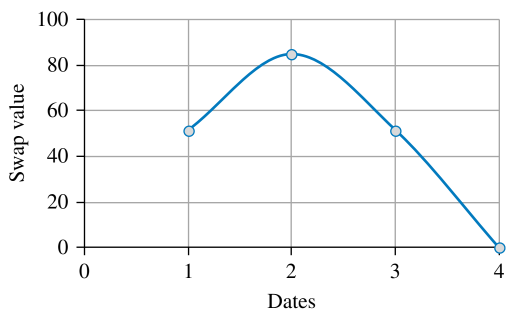


FIGURE 22.3 Time profile of the swap potential exposure

22.5 REGULATORY RULES FOR COUNTERPARTY CREDIT RISK

The regulations for counterparty credit risk define the measure of potential exposure used as credit equivalent in the Basel 2 credit risk framework and to the credit value adjustment (CVA) implemented in market risk regulations in "A global regulatory framework for more resilient banks and banking systems", December 2010, revised in June 2011 [24].

22.5.1 Exposure

Under the current exposure method, the exposure is defined as the current value of the derivative plus an add-on, which measures the potential positive deviations from the current value. The add-on is calculated as a percentage of the notional value. The regulatory add-ons depend on the nature of the underlying and on maturity and are defined in percentages of notional. Offsetting effects can be recognized within a netting agreement. Summing up the add-ons within a netting agreement and across netting agreements does not capture the dependencies between risk factors.

Under the Internal Model Method, banks are allowed to model the potential exposure using simulations and accounting for the dependencies of risk factors. The potential future exposures are the upper quantiles of the distribution of modeled exposures, at each time point, for a confidence level, typically 97.5%. When the exposure is modeled, various risk metrics are used for regulatory purposes.

The expected exposure at a time point is the expected value of the distribution truncated to positive values only, at any date. The expected positive exposure (EPE) used for regulatory purposes is a different statistic. It is the time-weighted average of such expected exposures. For example, if the exposure is E_1 over a time interval t_1 and E_2 for another time interval t_2, the EPE would be: $\text{EPE} = (E_1 t_1 + E_2 t_2)/(t_1 + t_2)$. The EPE should be calculated over one year. The effective expected positive exposure (EEPE) is the maximum of the exposure over two consecutive time points: $\text{EEPE}_t = \text{Max}(\text{EPE}_t, \text{EPE}_{t-1})$. The rationale of this definition is that the exposure should not decline over the calculation period due to an amortization effect. This rule is equivalent to assuming a rollover of short-term exposures over the one-year calculation period.

Finally, the EEPE is increased by a multiplier of 1.4 or a modeled multiplier with a floor of 1.2. The EEPE is a number measuring exposure as if it were deterministic, and replacing random exposures. The multiplier measures the excess exposure over EEPE due to the random movements of underlying exposures.

The expected exposure definition of the Basel 2 Accord is retained under Basel 3,[8] but potential exposures should now be modeled with parameters, such as volatility and correlation, using the data from a one-year stressed period, selected within a three-year horizon. This requirement addresses concerns about capital charges becoming too low during periods of low market volatility, and will raise the capital backing these exposures. It should reduce procyclicality by referring to stressed periods instead of current conditions. It should also provide incentives to move OTC derivative contracts to central counterparties and exchanges.

22.5.2 Credit Value Adjustment

The standard practice to value derivatives has been to discount all future cash flows at the Libor curve without taking the counterparty risk into account. But a true value must incorporate the possibility of losses due to the counterparty's default. An entity holding an instrument with a positive value would only get the recovery fraction if the other party defaults. The value of the instrument is the value without counterparty risk minus an adjustment, which is the CVA.[9]

A CVA charge is an expected loss, which should be accounted for as an upfront reserve for compliance with accounting standards. The expected loss is the present value of all expected losses over all periods of the remaining life of the contract. Over a future period, it is the product of default probability, exposure and loss given default (LGD).

$$\text{CVA charge} = \text{Discounted expected exposure} \times \text{DP} \times \text{LGD}$$

Derivatives have a modeled exposure, which varies across time points. The expected loss is the present value of expected losses of each period between two consecutive time points. The default probabilities can be derived from credit spreads: $\text{Spread} = \text{DP} \times \text{LGD}$. When using

8 Basel III: A global regulatory framework for more resilient banks and banking systems, 2010, [24].

9 For a detailed analysis of CVA, see Gregory, J. (2012), Counterparty Credit Risk and Credit Value Adjustment: A Continuing Challenge for Global Financial Markets, [73], or Brigo, D. (2012), Counterparty risk FAQ, credit VaR, PFE, CVA, DVA, closeout, netting, collateral, re-hypothecation, WWR, Basel, funding, CDS and margin lending, [35]. An extensive treatment of counterparty credit risk and pricing is found in Brigo, D., Morini, M., Pallavicini, A. (2013), Counterparty Credit Risk, Collateral and Funding: With Pricing Cases for All Asset Classes, [38].

spreads, the corresponding probabilities are those inferred from market prices. These market-implied probabilities are risk-neutral probabilities.

22.5.3 Regulations for Credit Value Adjustment

The credit adjustment varies with the market price of credit risk, measured by credit spreads. A CVA VaR accounts for the potential adverse deviations from the expected value. The potential variations of value of exposures are modeled by applying the volatility of credit spreads. The CVA VaR is the potential loss of value due to spread widening, when only the spreads are considered random, keeping the exposure and LGD fixed.[10]

The new CVA capital add-on is based on a one-year horizon and a 99% confidence level. The exposure level in the CVA calculation is the total exposure at default (EAD) of the counterparty, and the maturity is the longest dated instrument in the netting set with the counterparty. The counterparty exposure is represented as a zero coupon bond. This VaR model is restricted to changes in the counterparties' credit spreads and does not model the sensitivity of CVA to changes in other market factors, such as changes in the value of the underlying asset, commodity, currency or interest rate of a derivative.

When the bank is not allowed to use an internal model, a supervisory formula for the CVA add-on should be used. The new regulations also include a capital charge for wrong-way risk. Wrong-way risk occurs when the exposure increases if the credit standing of the counterpart deteriorates.[11] Wrong-way risk cannot be ignored for derivatives, as it was the source of losses during the financial crisis. It materializes when the counterparty default risk and the exposure to the counterparty can move together. A notable example was the exposure of the buyer of credit risk insurance, in the form of credit default swaps, from AIG Financial Products before it collapsed. As AIG built up its portfolio of sold credit insurances, the likelihood of not being able to face the corresponding huge commitments towards its buyers increased. The buyers had an increasing exposure to AIG, as the credit risk of the industry inflated, while, simultaneously, AIG's own default probability increased. When AIG defaulted, the wrong-way risk of AIG materialized.

For including wrong-way risk in the CVA calculation, the regulators imposed an adjustment to the capital charge. The adjustment obtains by adjusting the EAD with multiplier called alpha, set at 1.4, and no less than 1.2 if modeled by banks.

10 The CVA regulations were first introduced in the Basel 3 document (2010), [24].

11 For wrong-way risk, see Hull, J., White, A. (2012), CVA and wrong-way risk, [80], or Rosen, D., Saunders, D. (2012), CVA the wrong way, [121].

23

Credit Event Dependencies

Credit events are dependent. Modeling their dependencies is a requirement for credit portfolio models and deriving loss distributions for credit risk. At the level of individual facilities, or borrowers, a similar approach is necessary to assess the value of third party guarantees.

Dependencies between credit events can be measured by their correlation or by the probability that two obligors default jointly. Under the structural approach, the asset value of the firm is the factor that drives dependencies between the credit standing of firms. The approach provides a way to infer dependencies between credit events.

This chapter explains how the probabilities of joint defaults can be derived when the correlation between credit events is known, and how the structural model provides an elegant solution to the measure of dependencies.

Contents

23.1 JOINT DEFAULT PROBABILITY USING DISCRETE VARIABLES

The issue of dependence between the credit risks of two firms arises for third party guarantees, between the guarantor and the direct borrower. In the substitution rule, the risk is transferred to the guarantor if the credit standing of the guarantor is better than the one of the borrower.

This rule implies that the relevant default probability is that of the guarantor. The economic reality is different. Default occurs only when both entities default and the relevant default probability is the joint default probability of the guarantor and the borrower. If the credit risks of both entities are independent, the joint default probability would simply be the product of the two default probabilities. In general, credit events are positively correlated because they depend on the economic or industry conditions or because there are economic links between the two entities. Their joint default probability is higher than under independence and depends on the correlation between defaults.

The risk mitigation effect of third party guarantees results from the "transformation" of the default probability assigned to a direct borrower into a joint default probability of borrower plus guarantor, or "double default". The mitigation effect results from a lower probability of the double default than the standalone default probability of the borrower. However, the probability of a double default increases when the positive dependence is higher. For example, a guarantee from a firm in the same industry as the borrower might have less value than a guarantee from a firm in a different industry.

In general, a joint probability P_{XY} is the probability that two random events X and Y occur simultaneously. A conditional probability, noted $P_{Y|X}$, is a probability of observing Y given that an event X occurred. The relationship between the joint probability that both events X and Y occur and the conditional probability of Y given occurrence of X is given by Bayes' rule: $P_{XY} = P_X P_{Y|X}$. When X and Y are independent, the joint probability collapses to the product of their standard probabilities, and the conditional probabilities are equal to the standard unconditional probabilities.

The probability of double defaults can be quantified from conditional default probabilities or from the correlation between credit events. There is a positive correlation between the default events of the firms, A and B, when the probability that A's default occurs is higher if B's default occurs.

Consider two firms A and B, with default probabilities P_{A} and P_{B}. The joint default probability is P_{AB} and the default probability of B conditional on A's default is $P_{\mathrm{B|A}}$. Assume that the unconditional probabilities of default of A and B are 1% and 2%, respectively, but that the probability that B defaults when A defaults increases to 10%. This conditional probability is much higher than the unconditional probability of B's default, because the situation of B depends on the situation of A. In the notation of conditional probability: $P_{\mathrm{B|A}} = 10\%$. The joint default probability becomes:

$$P_{\mathrm{AB}} = P_{\mathrm{B|A}} P_{\mathrm{A}} = 10\% \times 1\% = 0.1\%$$

This is much higher than 0.02%, the joint default probability if the defaults were independent.

This situation implies a strong correlation between A and B default events. Correlation implies that conditional probabilities differ from unconditional probabilities, and vice versa. Correlation and conditional probabilities relate to each other, but they are not identical. Conditional probabilities are within the (0, 1) range, while correlations are within the $(-1, +1)$ range.

For a pair of obligors, A and B, their joint default probability depends on the correlation between variables representing their default events. The default variable for A is defined as $1_{\mathrm{A}} = 1$ if A defaults and 0 otherwise with probabilities P_{A} and $(1 - P_{\mathrm{A}})$, respectively. A similar default variable is 1_B for B, with values 1 and 0 with probabilities P_B and $(1 - P_B)$. It is shown below that the joint default probability of A and B is:

$$P_{AB} = P_A P_B + \rho\sqrt{P_A(1 - P_A)p_B(1 - P_B)}$$

The formula shows that the joint default probability is larger than the product of the default probabilities whenever the correlation is positive.

The derivation of this expression relies on standard statistical formulas. The joint default probability P_{AB} is the expectation of the product of the two default variables, the values of which are 0 or 1 because a double default occurs only when both have a value of 1. A general formula for the expectation of the product of two variables X and Y is: $E(XY) = \text{Cov}(X, Y) + E(X)E(Y)$. Using 1_A for X and 1_B for Y, the following holds: $E(1_A) = P_A$, $E(1_B) = P_B$, $\text{Cov}(A, B) = \rho\sigma(1_A)\sigma(1_B)$ where $\sigma(1_A)$ and $\sigma(1_B)$ are the standard deviations of 1_A and 1_B and ρ is their linear correlation. These standard deviations are $\sqrt{P_A(1 - P_A)}$ and $\sqrt{P_B(1 - P_B)}$. The expression of the joint default probability follows.

The joint default probability increases with the positive correlation between default events but there are boundary values for correlations and for unconditional default probabilities. Intuitively, a perfect correlation of +1 would imply that A and B default together, and this is possible only when the default probabilities are identical. If a firm never defaults, there is no correlation at all with any other risky counterparty. This implies that both the correlation and the joint default probability collapse to zero. The consequence is that the positive correlation cannot take all values in the range 0 to +1.

The dependence between default events can be modeled either as a joint default probability or as a default probability conditional on default of the other firm. If the correlation is known, the conditional probabilities follow. For example, two firms A and B have a default correlation of 10% and their default probabilities are: $P_A = 7\%$ and $P_B = 5\%$. Their joint default probability is calculated from the above formula: 0.9061%.

If the default events were independent, the joint default probability would be: $7\% \times 5\% = 0.35\%$. The default probability of B conditional on A's default now differs from the unconditional default probability of B according to:

$$P_{B|A} = P_{AB}/P_A = 0.9061\%/7\% = 12.944\%$$

It is significantly higher than the unconditional default probability of B, which is 5%.

23.2 SUPPORT

Support refers to financial assistance provided by a head of group to subsidiaries but, unlike legal guarantees, is not legally binding. It depends on the willingness of the head of group to assist subsidiaries by providing guarantees or funds to affiliated companies. But support is a form of credit dependency between firms.

A positive support that is always effective is similar to a guarantee. The default probability becomes, as for formal guarantees, the joint default probability. However, support can also be negative when a subsidiary's survival is dependent on the survival of the holding company.[1]

1 See also Chapter 20.

In this case, a lender is exposed to credit risks of both the direct borrower and the supporting entity because a default occurs when the borrower defaults or when the supporting entity's default triggers the default of the borrower. The two events are not exclusive as both firms can also default jointly. The probability of default is the sum of their default probabilities, minus the probability that both occur.[2] If B is the direct borrower and S is the supporting entity, the default probability is:

$$P(B \text{ or } S \text{ default}) = P(B \text{ defaults}) + P(S \text{ defaults}) - P(\text{Both } B \text{ and } S \text{ default})$$

Suppose that the default probabilities, P_S and P_B, are 2% and 3%, respectively, and that the joint default probability, P_{BS}, is 1%. This joint default probability implies a positive dependence between defaults since it is higher than the product of the default probabilities (0.06%). With negative support, the default probability becomes:

$$P(B \text{ or } S \text{ default}) = 2\% + 3\% - 1\% = 4\%$$

The default probability under negative support is much higher than any one of the standalone default probabilities. The usage of the above relationship is subtle. If a single default, either that of B or S, triggers the default of the other, the default correlation is necessarily one. In that special case, both default probabilities should be equal, and equal to the joint default probability. In practice, the relationship is neither deterministic nor necessarily symmetrical. It is possible that the supporting entity's default triggers that of the borrower while the reverse is not true. A correlation coefficient does not reflect this asymmetry.

23.3 MODELING JOINT DEFAULTS WITH THE STRUCTURAL MODEL

Deriving joint probabilities when dependencies are known is a relatively simple matter. The issue is to find a way to measure the default correlations. Statistics do not help much as joint defaults are rarely observed, and even more so for large corporations. With "low default portfolios", portfolios of firms that have very low default probabilities, statistical methods become ineffective and default events, and their dependencies, need to be modeled rather than observed.

The structural model of default provides an instrumental framework for deriving default correlations. The principle is to model the joint probability of the assets of two obligors of falling under the thresholds triggering default. Default events occur when asset value becomes lower than debt. Two obligors default jointly when both assets' values fall below their respective debt levels. This probability is a joint default probability, which embeds the correlation of asset values.

The situation of double default is illustrated in Figure 23.1. The figure shows the distributions of the asset values of two firms, where the asset value of firm X is on the horizontal axis and the asset value of firm Y is on the vertical axis. For firm X, the distribution

2 This is the probability rule for two non-exclusive events.

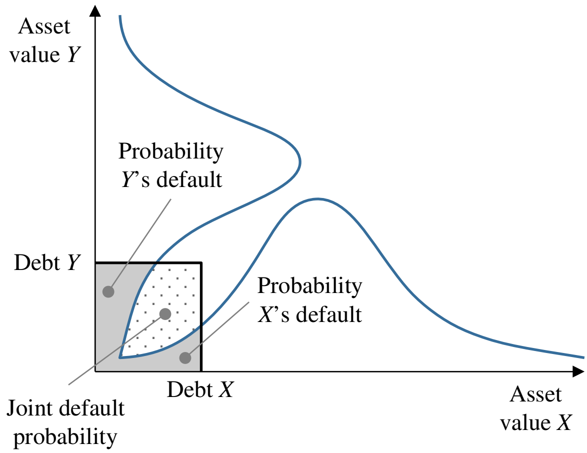


FIGURE 23.1 Joint default probabilities of two obligors and assets correlation

shows the probability vertically, and for the firm Y, the probability is measured horizontally.[3] Two lines represent the debt levels of each firm at some defined horizon: the vertical line refers to the debt of X and the horizontal line to the debt of Y.

The joint probability of default is within the area of the rectangle between the origin of the axes and the two debts of two obligors X and Y. In this area, the asset value of X falls below its debt and the same happens to Y. The areas under the distribution curves represent the default probabilities for X and Y. If the asset values are correlated, there will be more observations in the rectangle than if they are independent.

The correlation effect on the joint default probability is visualized in Figure 23.2, which plots the asset values of two obligors on two axes. The cloud of dots is more or less round depending on correlation. With a zero correlation, it is a circle. When the positive correlation increases, it extends along the first diagonal. The density of dots in the rectangle is the count of joint default events, and this number represents the frequency of joint defaults.

The framework of the structural model allows modeling of the dependencies between defaults, whether they occur or not, because the asset values and their correlations can be

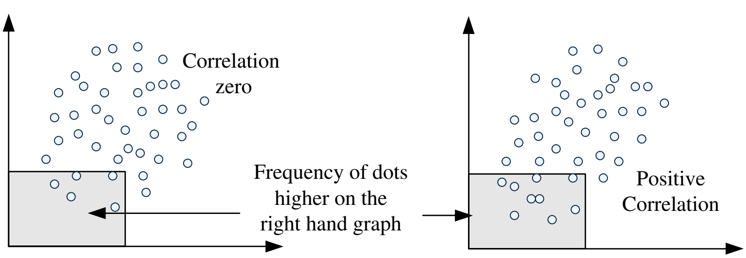


FIGURE 23.2 Effect of assets' correlation on the joint default probability

3 The graph is simplified. There should be two horizontal axes and two vertical axes, one for the asset value and the other for the probability. For this intuitive visual representation, we omit the two probability axes.

inferred from equity prices. For estimating the default probability, the observation of equity returns, of their volatilities and of their correlations is sufficient. The asset values derived from equities are also correlated, which allows modeling joint defaults.

The correlation between asset returns is not identical to the correlation between discrete default events, but both are directly related. Since asset returns are normally distributed, the joint default probability derived from asset values is obtained from their bivariate distribution, given the default thresholds matching the standalone default probabilities of each firm. For example, if asset values were normally distributed, the probability could be obtained analytically from the bivariate normal distribution. Once the joint default probability is obtained, it is possible to trace back the correlation between the discrete default events matching this joint probability from the formula for joint defaults.

Correlations across pairs of firms are available in Moody's KMV Portfolio Manager™ and in CreditMetrics™. CreditMetrics derives the correlations from those of equity returns, using a multifactor model of equity returns. By contrast, Moody's KMV Portfolio Manager extracts the correlations from the modeled asset values under the option theoretic approach and uses a multifactor model with orthogonal factors for making modeled asset values dependent. The difference is that asset returns are not affected by the leverage of firms, while equity returns are.

23.4 JOINT MIGRATION PROBABILITIES

The structural model applies to migrations derived from asset values. Migrations to states other than defaults are modeled as positive distance to default in the structural model. Higher distances to default result in lower default probabilities. For each firm, migration events are defined as asset values falling below a threshold matching the default probability assigned to a credit state other than default.

For two obligors, the thresholds can be defined such that they result in distances to default (noted DD) corresponding to their final credit states. The count of firms that both have asset values below these thresholds is the frequency of joint migrations to their final credit states. Migration thresholds are above the default thresholds. Figure 23.3 uses a similar presentation to the above. Each positive distance to default matches a credit state other than default.

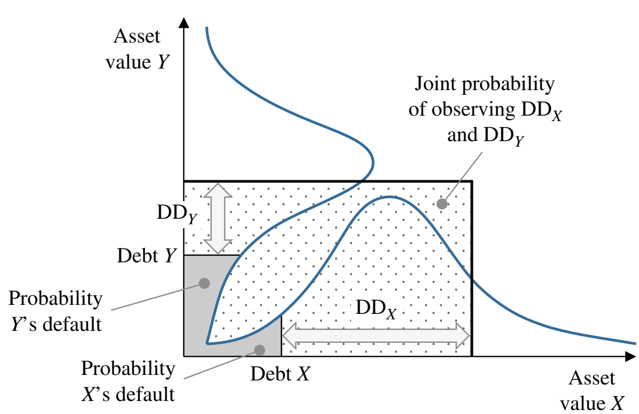


FIGURE 23.3 Joint migration probabilities of two obligors and asset correlation

The rectangle between straight lines represents the joint probability that the X and Y distances to default are lower or equal to those threshold values.

The transition matrices of rating agencies are often determined over long periods; hence unconditional on the state of the economy and reflect historical data. They provide the probabilities than an obligor in risk class i moves to risk class j during the period. Joint transition matrices provide the probability that a pair of obligors moves from one pair of credit states to another pair of credit states by tabulating the final states of two obligors and assigning probabilities to them.

However, dealing with several credit states implies consistency constraints. When adjusting a transition probability within a transition matrix, all transitions still have to sum up to one along rows. If we shift upward a transition probability within a row due to correlation, some other transition probabilities will decline to comply with the same 100% total of transition probabilities along the same row. This implies that the correlation between migration probabilities varies along a single row.

This makes the usage of joint transition matrices more difficult to handle. One credit portfolio model, Credit Portfolio View,[4] resolves the problem by shifting the conditional migration probabilities proportionally to the shift of the default probability from its standalone unconditional value. The model refines the adjustment by making speculative-grade firms more sensitive to the general factors that affect the default probabilities and investment-grade firms less sensitive.

4 The model "Credit Portfolio View" can be found in Wilson, T. (1997), Portfolio credit risk I and Wilson, T. (1997), Portfolio credit risk II, [135 and 136].

24

Credit Portfolio Risk: Analytics

Loss distributions for portfolios of loans are a key building block for addressing the credit risk capital. The default probability of a bank is identical to the confidence level assigned to capital. Banks target a credit rating in the investment-grade range for which the default probability is extremely low. For such tight confidence levels, the tail of the distribution, showing the largest aggregated losses, is mostly important. Loss distributions can be modeled, in a simplified framework, or obtained from simulations. An important finding of such distributions, common to all approaches, is that the tail of the loss distribution is highly sensitive to the correlation between defaults.

This chapter reviews the analytical techniques for modeling loss distributions with dependent defaults across obligors. The main approach considers that defaults are dependent on a single factor, the state of the economy, and derives loss distributions from a simplified version of the structural model. Applications include stressing default probabilities and modeling the default loss distribution of granular and homogeneous portfolios.

Contents

24.1 INDEPENDENT DEFAULT EVENTS: THE BINOMIAL DISTRIBUTION

When default events are independent, and have the same probability, the distribution of the number of defaults in a portfolio is the binomial distribution. Consider a uniform portfolio, with loans having the same exposure, same loss under default, same default probability and with the same correlation between all asset values of firms. Defaults can occur between today and a finite horizon at least equal to the maturity of loans. The loss under default for a single loan is equal to exposure X when the loss under default is 100%.

Each random default is a Bernoulli variable with value 1 in the event of default, with probability d and value 0 in the event of no default, with a probability $1 - d$. The variable representing the credit state of each firm is the indicator function: 1 (default) equals 1 under default and 0 if no default. If all defaults are independent, the distribution of the number of defaults is the binomial distribution. The number of defaults has an expected value Nd.

If all loans have the same loss given default X, the expected loss for a single obligor "i" is: $E(L_i) = Xd + 0(1 - d) = Xd$. For the portfolio, the count of defaults is the sum of individual defaults and the losses equal the number of defaults times the uniform exposure X.

The individual borrower's loss variance is $X^2d(1 - d)$. The portfolio loss has a variance equal to the sum of all individual default variances, or $NX^2d(1 - d)$. The volatility of the portfolio loss is the square root, or: $X\sqrt{Nd(1 - d)}$.

With N exposures, which unit size is $X = 2$, and $d = 10\%$, the unit variance in value is 10% $(1 - 10\%)2^2 = 0.36$ and the variance of the number of defaults is: $0.36N$. The volatility is the square root, or $\sqrt{0.36N}$. With a large number of obligors, the binomial distribution tends to the normal distribution.

Unfortunately, the binomial distribution cannot capture the effect of size discrepancies and correlations. The independence assumption also leads to unacceptable errors, as it can be seen when introducing dependencies based on a simplified version of the structural model.

24.2 THE STRUCTURAL MODEL

A simplified default model for a single firm assumes that the default probability embeds all relevant information about the underlying asset value and uses a normal standardized asset value. This is a standard approach commonly used for modeling dependent defaults because it is extremely simple to correlate standardized normal asset values.

If the asset value follows a standardized normal distribution, the default points corresponding to given default probabilities are in the tables of the standardized normal distribution. If the default probability is 2.5%, the corresponding standardized asset value is -1.96; if the probability of default is 1%, it becomes -2.33. There is a one-to-one correspondence between the asset value A_d triggering the default and a default probability d (Figure 24.1). The financial interpretation is that the debt value is lower than the asset value by 2.33 standard deviation units. The standardized model requires a single input, which is the default probability d.

The probability that asset values fall below the default point A_d is $\Phi(A_d) = d$ and implies that $A_d = \Phi^{-1}(d)$. The absolute values of A_d and the corresponding default probabilities are shown in Figure 24.2.

Table 24.1 shows the standardized distances to default matching the historical frequencies of defaults from Moody's simplified scale. Negative numbers correspond to downside variations of asset value.

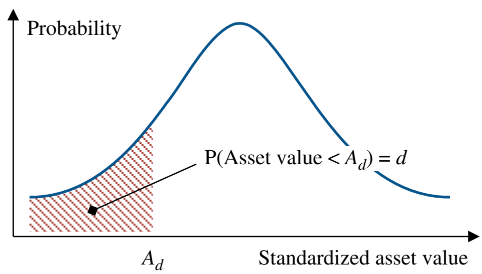


FIGURE 24.1 Asset value and default probability

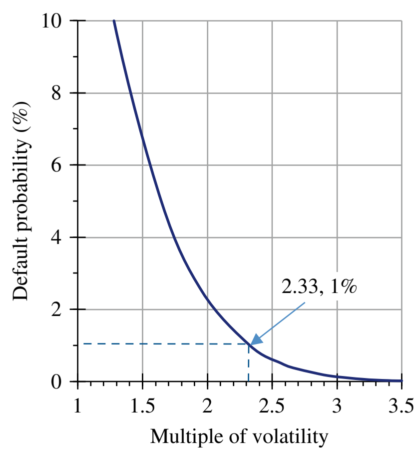


FIGURE 24.2 Multiples of volatility and default frequency (standard normal distribution)

TABLE 24.1 Standardized shocks matching historical default frequencies

Rating class	d	Φ^{-1}(d)
Aaa, Basel 2[(1)]	0.03%	−3.432
Aa	0.05%	−3.291
A	0.08%	−3.156
Baa	0.20%	−2.878
Ba	1.80%	−2.097
B	8.30%	−1.385
Default	20.00%	−0.842

[(1)] For Moody's, the annualized default probability is 0%, but Basel 2 imposes a floor of 0.03%.

24.2.1 A Single Obligor Dependent on the State of the Economy

The credit standing depends on general economic conditions, which can be measured by a single factor. The default probability of an obligor becomes conditional on a given state of the economy.

For a single obligor, the asset value A depends on two independent variables, normal standard, the common factor, Z, with correlation ρ, and an independent normal standard variable X representing the specific risk independent of the economic conditions:

$$A = \rho Z + \sqrt{1 - \rho^2} X$$

The unconditional default probability, d, is given. It represents the long-term default probability averaged across all states of the economy. The default point is the threshold of asset value, depending on the default probability d, and is called A_d. This default point is such that the probability of asset values being below this threshold equals the default probability d: $P(A \leq A_d) = d$. The value of the threshold is: $A_d = \Phi^{-1}(d)$, because the unconditional distribution of A is normal standard.

On the other hand, the conditional default probability varies as the state of the economy changes. If the state of the economy improves (Z increases), the distribution of the asset values shifts to the right and the default probability, conditional on Z, declines.

The distribution of A and Z is a bivariate normal distribution, which assigns a joint probability of observing a pair of values for each of the variables. When one of these two variables, the state of the economy Z, is fixed, the other variable follows a normal distribution, which depends on Z, entirely defined by its expectation and variance. This conditional distribution of asset value can be viewed as a vertical tranche of the bell-shaped bivariate normal distribution, defined by the value of Z. When the state of the economy varies, there are a series of such distributions, shown in Figure 24.3, as vertical tranches for different values of Z.

Each of these conditional curves is a normal distribution, but it is not standard. When the conditioning factor Z has a particular value, the mean of the distribution of the asset value becomes ρZ. The volatility of asset value does not depend on Z and is equal to $\sqrt{1 - \rho^2}$. The distributions have different means, increasing with Z, and a constant variance depending on the correlation.

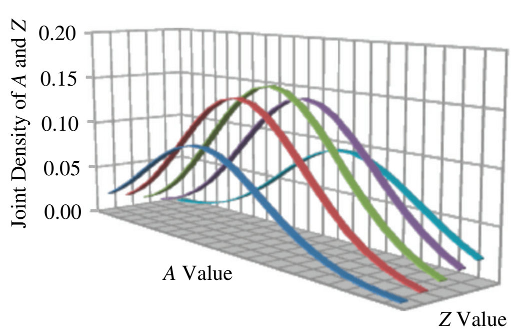


FIGURE 24.3 Conditional normal distributions

The standardized normal asset value, conditional on Z, is identical to X:

$$X = \frac{A - \rho Z}{\sqrt{1 - \rho^2}}$$

X follows a standard normal distribution, $\Phi(X)$, with mean 0 and standard deviation 1. The asset value A, conditional on Z, follows a non-standard normal distribution $N(\rho Z, \sqrt{1 - \rho^2})$.

The neutral state is $Z = 0$. When Z becomes positive, the entire distribution of A shifts upward while keeping a constant volatility. The default point A_d does not change because it is defined as a given threshold that is independent of the state of the economy. But the chances of hitting the default point, from higher asset values, are lower. Therefore, the conditional default probability moves down if the state of the economy is up.

24.2.2 The Default Probability Conditional on the State of the Economy

The default probability conditional on Z is d_Z and the asset value conditional on the state of the economy is A_Z. The default condition is:

$$d_Z = P(A_Z \leq A_d)$$

This condition applies to the standardized asset value, given the mean ρz and the volatility $\sqrt{1 - \rho^2}$ of asset value:

$$d_Z = P\left[\frac{A - \rho Z}{\sqrt{1 - \rho^2}} \leq \frac{A_d - \rho Z}{\sqrt{1 - \rho^2}}\right]$$

Since the condition applies to a standard normal variable, the default probability is the quantile of the standard normal distribution:

$$d_Z = \Phi\left[\frac{A_d - \rho Z}{\sqrt{1 - \rho^2}}\right]$$

The default point A_d is given and is the inverse of the standard normal function: $A_d = \Phi^{-1}(d)$. Replacing A_d in the above expression provides the conditional default probability as a function of Z and ρ:

$$d_Z = \Phi\left[\frac{\Phi^{-1}(d) - \rho Z}{\sqrt{1 - \rho^2}}\right]$$

As an example, the conditional default probabilities are calculated below for a range of values for the state of the economy. The default point, for a given unconditional default probability of 1%, is: $A_d = -2.3263$. When the state of the economy changes, the conditional default probability varies inversely with Z. Using $Z = 2$, the conditional default probability is

TABLE 24.2 Default probability conditional on the state of the economy

Correlation, ρ	30%	30%	30%	30%	30%
State of the economy, Z	−2.0	−1.0	0.0	1.0	2.0
Expected asset value, ρZ	−0.6	−0.3	0	0.3	0.6
Volatility asset value	0.954	0.954	0.954	0.954	0.954
Z	−2.0	−1.0	0.0	1.0	2.0
Default point	−2.326	−2.326	−2.326	−2.326	−2.326
Conditional d_Z	3.52%	1.68%	0.74%	0.30%	0.11%

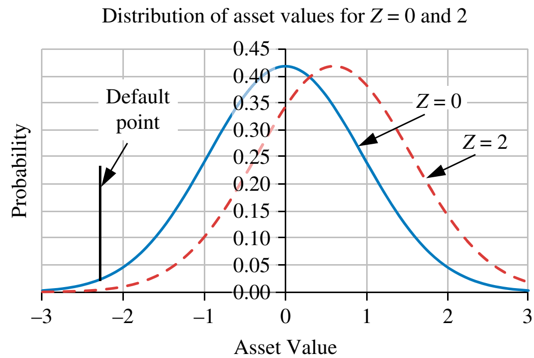


FIGURE 24.4 Asset value distributions: quantile 1%, Z equals 0 and 2

0.108%, much lower than 1%. Table 24.2 shows what happens when Z varies between −2 and +2 by steps of one, keeping the correlation constant at 30%. When Z increases from −2 to 2, the mean of the distribution of the asset value shifts to the right, increasing from −0.6 to 0.6. The volatility is constant because it depends only on the correlation at 0.954.

The shift of the conditional distribution of the asset value, when Z increases from zero to 2, is shown in Figure 24.4. The mean of the distribution increases and its volatility is unchanged. The same default point −2.326 applies for both curves. Because the curve shifts to the right, the default probability declines when Z increases.

Finally, Figure 24.5 plots the inverse relation between conditional default probability and the state of the economy.

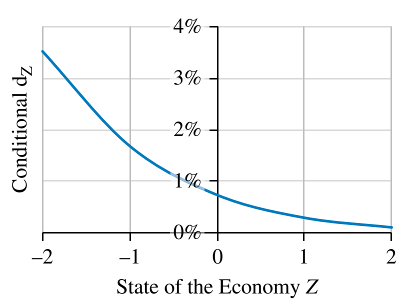


FIGURE 24.5 Conditional default probability on the state of the economy

24.3 APPLICATION: THE STRESSED DEFAULT PROBABILITY UNDER BASEL 2

The risk weights of the Basel 2 regulations for credit risk apply to the exposure less recoveries, or loss given default, for determining the capital charge for credit risk. The risk weights use as input, among others, the "through-the-cycle", long-term default probabilities of obligors. The term, within the general formula, which includes the default probability as input has the form:

$$\Phi\left[\frac{\Phi^{-1}(\mathrm{DP}) + \rho\Phi^{-1}(0.999)}{\sqrt{1-\rho^2}}\right]$$

This term can be interpreted as a stressed default probability.

The formula is similar to the expression of a default probability conditional on a single factor, noted here as d_{ZB}. Some transformations are required in order to show that the Basel formula matches the general form of a conditional default probability. The Basel formula uses the term $\rho\Phi^{-1}(0.999)$ instead of $-\rho\Phi^{-1}(0.001)$. But, with a normal distribution $\Phi^{-1}(1-\beta) = -\Phi^{-1}(\beta)$, where β is a confidence level, because the distribution is symmetric. Hence, the formula is the same as the expression of a conditional default probability, using a value of Z equal to $-\Phi^{-1}(0.001)$.

$$d_{ZB} = \Phi\left[\frac{\Phi^{-1}(\mathrm{DP}) - \rho\Phi^{-1}(0.001)}{\sqrt{1-\rho^2}}\right]$$

Under this form, the equation is that of a default probability stressed by the factor Z representing the state of the economy. With the 0.001 confidence level, the selected value of Z is the 0.1% quantile. The negative sign implies that economic conditions are stressed, as the conditional default probability. The inverse function used for Z is that of a standard normal distribution.

The stressed value, Z_B, of Z is such that the probability of lower values is 0.1% when Z follows a standard normal distribution:

$$Z_B = \Phi^{-1}(0.1\%) = -3.0902$$

The stressed default probability follows from the stressed Z_B value. For example, with an unconditional default probability of 1%, the stressed default probability, d_{ZB}, matching this confidence level for Z is:

$$d_{ZB} = \Phi\left[\frac{\Phi^{-1}(d) - \rho Z_B}{\sqrt{1-\rho^2}}\right]$$

The correlation is an input in the calculation. Using the same correlation of 30% as in the previous calculations, the stressed default probability is 7.12%.

The formula from risk weights in Basel 2 is discussed in Chapter 26. It is shown above that the complex term is a conditional default probability, derived from the long-term default probability of an obligor.

24.4 MODELING DEFAULTS IN A UNIFORM PORTFOLIO: THE LIMIT DISTRIBUTION

The limit distribution refers to the loss distribution of a granular and homogeneous portfolio. In this portfolio, all firms have the same exposures and all exposures are small compared to the portfolio aggregated exposure. The correlation between defaults is also supposedly uniform across obligors, and it results from the systematic risk due to a single common factor.

Each obligor is characterized by a standard normal asset value and a default threshold. The weights of each loan are identical. The default probability is identical for all obligors and all pairs of asset values have the same uniform correlation. This default probability is equal to the expected loss, if loss given defaults are also identical. The approach was originally presented by Vasicek.[1]

The basic idea of the limit distribution is that the expected value of the percentage of the portfolio loss converges to the conditional default probability when the number of obligors is large. Therefore, finding the percentage portfolio loss sums up to modeling the expected portfolio loss, in percentage of total loss, or, equivalently, modeling the distribution of the conditional default probability.

If all individual loans have the same loss under default and the same default probability, and if all losses were independent, the distribution of the portfolio loss would be the binomial distribution, with parameter equal to the default probability. It would converge, following the central limit theorem, to a normal distribution as the portfolio size increase. Because defaults are not independent, the loss distribution is not normal. The issue is to find out the portfolio loss, in percentage of portfolio exposure, as a function of the unconditional default probability d and the uniform correlation.

The limit distribution expresses the quantiles of the distribution of the portfolio loss in percentage of total exposure as a function of the uniform parameters of the portfolio. The key point is that the percentage portfolio loss, in percentage of portfolio exposure, given the state of the economy Z, converges by the law of large numbers to its expectation, d_Z, when N tends towards infinity. Assuming that the default probability is d_Z, and is determined once Z is fixed, is equivalent to assuming that the diversification of the portfolio is large enough for the default probability to become almost certain. This implies that the portfolio loss volatility is entirely due to its sensitivity to the factor and that the specific risk tends towards zero. In that case, there is a one-to-one correspondence between the portfolio loss and the Z value.

The default probability conditional on Z is given from the above formulas. The value of Z is obtained by inverting the expression of d_Z. The quantile, x, of the portfolio loss in percentage of the total portfolio exposure can be found by replacing Z by its expression and stating that Z is now the threshold of the x-quantile:

$$P(l_P \leq x) = \Phi^{-1}\left[\frac{\Phi^{-1}(x) - \Phi^{-1}(d)\sqrt{1-\rho^2}}{\rho}\right]$$

The equation is the x-quantile of the distribution of the portfolio loss in percentage, l_P, given the uniform correlation, ρ, and the uniform default probability, d. For generating the full distribution, x varies between 0% and 100%. The derivation of the quantile is detailed in the Appendix in section 24.5.

1 See Vasicek, O. (1991), Limiting loan loss probability distribution, [133] and Vasicek, O. (2002), Loan portfolio value, [134].

A numerical application uses as inputs the default probability, 5%, and the uniform correlation, 30%. The formula for the distribution of the portfolio loss serves for finding the x-quantile of the portfolio loss, for example the 1%-quantile. Replacing the above inputs in the general expression of the quantile:

$$P(l_P \leq 1\%) = \Phi^{-1}\left[\frac{\Phi^{-1}(1\%) - \Phi^{-1}(5\%)\sqrt{1 - 30\%^2}}{30\%}\right]$$

The probability that the percentage portfolio loss does not exceed $x = 1\%$ is 2.78%.

The value of the factor Z such that the loss is 1% is:

$$Z = \frac{\Phi^{-1}(d)\sqrt{1 - \rho^2} - \Phi^{-1}(x)}{\rho}$$

The value of Z is found by replacing the inputs by their values:

$$Z = \frac{\Phi^{-1}(5\%)\sqrt{1 - 30\%^2} - \Phi^{-1}(1\%)}{30\%}$$

using $\Phi^{-1}(5\%) = -1.6449$ and $\Phi^{-1}(1\%) = -2.3263, Z = 6.6983$.

Accordingly, the default probability conditional on this value of the common factor should be 1%, equal to the portfolio loss in percentage. The general formula for the conditional default probability is:

$$d_Z = \Phi\left\{\frac{\Phi^{-1}(d) - \rho Z}{\sqrt{1 - \rho^2}}\right\}$$

Replacing the default probability by 5%, the correlation by 30% and using the above value of the factor Z, the formula provides 1%. This shows that the portfolio loss is effectively equal to the conditional default probability, as it should be.

Figure 24.6 shows the distribution of the portfolio loss in percentage of total exposure as a function of various values of the quantile using the same data as in the example.

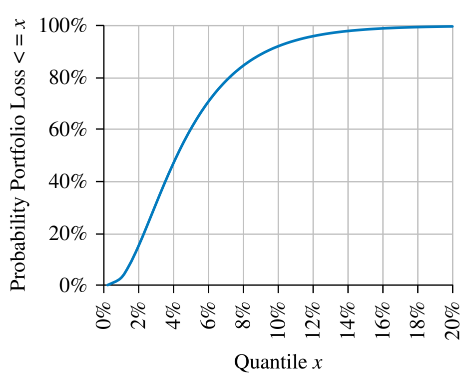


FIGURE 24.6 Portfolio loss distributions in percentage of exposure

24.5 APPENDIX: THE LIMIT DISTRIBUTION

The uniform portfolio is a limit case, but it is a proxy for a widely diversified granular portfolio of which exposures and correlations are similar. Loans are assumed to have the same maturity, which allows modeling losses under default mode, ignoring migrations.[2]

The portfolio is made of N loans, each with subscript i. All asset values, A_i, follow normal standard distributions correlated with a single factor Z, with a uniform correlation ρ:

$$A_i = \rho Z + \sqrt{1-\rho^2} X_i$$

X_i is the specific risk of the obligor "i" independent of Z.

The individual facilities have the same loss given default,[3] L. The default event is a Bernoulli variable 1_i, with value 1 when an obligor defaults and 0 under no default. The random loss of firm i is $1_i L$. The portfolio loss, L_P, is the sum of losses of individual obligors, or $L_P = \sum_{i=1,N} 1_i\, L$. The common unconditional probability of default is d. The expected value of each individual loss, $1_i L$, is equal to default probability times the loss:

$$E(1_i L) = dL + (1-d)0 = dL$$

The portfolio loss L_P has an expectation equal to the sum of the expectations of individual loans: $E(L_P) = \sum_{i=1,N} E(1_i L) = N\,dL$.

The loss, in percentage of total portfolio exposure, is l_P, equal to the total portfolio loss divided by the total of individual exposures, $\sum_{i=1,N} L_i = NL$. It follows that the expectation of this percentage portfolio loss is equal to the default probability:

$$E(l_P) = \sum_{i=1,N} E(1_i L)/NL = N\,dL/NL = d$$

If the default probability, d_Z, is conditional on the state of the economy, Z, the expectation of the portfolio loss, in percentage of total exposure, becomes equal to this conditional default probability:

$$E(l_P|Z) = d_Z$$

The conditional default probability, d_Z, varies inversely with Z:

$$d_Z = \Phi\left[\frac{A_d - \rho Z}{\sqrt{1-\rho^2}}\right] = \Phi\left[\frac{\Phi^{-1}(d) - \rho Z}{\sqrt{1-\rho^2}}\right]$$

2 Such restrictive assumptions can be relaxed by using Monte Carlo simulation. For example, single factor models allow generating correlated asset values without making any restrictive assumptions on the default point, or default probability, and size of exposure. The Cholesky decomposition can impose a variance–covariance structure on an obligor's asset values. In Chapter 25 on modeling loss distribution for credit portfolios, the equal exposure and equal default probability assumptions are relaxed.

3 The distinction between exposure and loss given default is skipped in this example.

A quantile of the distribution of the portfolio loss, in percentage, is a value x such that the probability that the loss is lower or equal to x is given. The percentage portfolio loss is identical to the conditional probability, d_Z, which is an inverse and monotonous function of Z. For a given value x of this conditional probability, there is a value $z(x)$ of Z such that:

$$P(d_Z \leq x) = P[Z \geq z(x)] = P[Z \leq -z(x)]$$

The last term uses the symmetry property of the standard normal distribution: $P(Z \geq z) = P(Z \leq -z)$. Therefore, the probability that the portfolio loss l_P in percentage remains below or equal to a particular value x is:

$$P(l_P \leq x) = P[Z \leq -z(x)]$$

This equation means that the conditional default probability, equal to the percentage loss of the portfolio, is lower than a threshold, x, when the factor Z is higher than a value of Z depending on x.

The corresponding value of Z is given by the general expression of the default probability conditional on Z:

$$l_P = \Phi\left\{\frac{\Phi^{-1}(d) - \rho z(x)}{\sqrt{1-\rho^2}}\right\}$$

The value $z(x)$ of the factor Z matching the quantile x for the portfolio loss in percentage is:

$$z(x) = \frac{\Phi^{-1}(d)\sqrt{1-\rho^2} - \Phi^{-1}(x)}{\rho}$$

This relation also shows that the quantile x of the loss distribution matches the quantile $z(x)$ of the factor distribution:

$$P(l_P \leq x) = P[Z \leq -z(x)]$$

These two probabilities are equal to $\Phi[-z(x)]$. The quantile of the loss is found by replacing $z(x)$ by its expression:

$$P(l_P \leq x) = \Phi^{-1}\left[\frac{\Phi^{-1}(x) - \Phi^{-1}(d)\sqrt{1-\rho^2}}{\rho}\right]$$

The equation is the x-quantile of the distribution of the portfolio loss in percentage l_P, given the correlation and the unconditional default probability d. For generating the full distribution, x varies between 0% and 100%.

25

Credit Portfolio Risk: Simulations

For modeling the distribution of credit portfolio losses, the standard technique is the simulation. There are two main types of simulations for assessing portfolio risk:

- Simulations of asset values of firms, defaults being triggered when asset values are equal to or lower than a threshold that matches the default probability assigned to each individual facility. The rationale follows the principles of the "structural default model". The technique is used by Moody's KMV Portfolio Manager™.
- Simulation of times to default, along the lines of the "reduced form" model of default. The technique was originally detailed by Li.[1]

This chapter explains the simulation methodologies with the two standard techniques from detailed examples where the simulation process is broken down into sequential steps. The simulated loss distribution show the effect of dependencies on portfolio risk. The dependency between the firms of the portfolio used as an example is modeled with a single factor model.

Contents

1 Li, D. (2000), On default correlation, a copula approach, [90].

25.1 SIMULATIONS OF DEPENDENT DEFAULT EVENTS WITH THE ASSET MODEL

The simulations generate default events using the standardized model of asset value. In this approach, the default probability embeds all information relevant to default and on unobservable asset values. The model operates in default mode only. Asset values are standard normal and the default point is a function of the default probability of each firm. The simulation generates dependent asset values. Default occurs when the asset value is equal to or lower than the threshold level mapped to the default probability. The count of defaults within the portfolio, for a simulation, is the sum of the dummy variables representing default.

In the general case, firms might differ by exposure size, default probability and correlation with the rest of the portfolio. In the simulations of this chapter, correlation is uniform, there is a unique default probability and the exposure is the same across firms. These assumptions are not restrictive and the same technique applies when default probability and exposure vary across firms.

For a single firm with the default probability d, the default point A_d of the random asset value A is such that: $\Phi(A_d) = d$, or $A_d = \Phi^{-1}(d)$, where Φ is the cumulative standardized normal distribution. For instance, with $d = 1\%$, the default point is: $A_d = \Phi^{-1}(1\%) = -2.33$, and so on. Generating a default event requires that the asset value be lower than A_d. A dummy variable, 1, represents the default event. It has a value of 1 if the asset value is below the threshold (default), and zero otherwise (non-default):

In this model, all dependent asset values depend on a common factor Z. The asset value of the firm i, Y, has the form:

$$Y_i = \rho Z + \sqrt{1-\rho^2}X_i$$

In this case, the constant correlation with the common factor is also the correlation between all dependent variables.[2] All variables Y_i are dependent and the factor is the first variable. The variables have unit variance and the correlation for any pair Y_i and any Y_j is the uniform correlation.

For conducting each simulation, the factor Z is first generated. Then, N independent variables X are generated and combined with Z using the above coefficients. The combination provides a set of dependent asset values Y_i, for the obligors in the portfolio. Defaults are triggered by comparing asset values with the default point. For each value of the factor Z, we obtain a count of portfolio defaults. The process is repeated with another value of Z for generating a distribution of portfolio defaults.

There are N obligors in the portfolio. Each simulation generates as many asset values as there are obligors, or N asset values. For each obligor, the discrete default variable 1_i takes the values 0 or 1. The sum of all 1_i is the number of defaults of the portfolio for the simulation. The simulations are repeated K times. For each simulation, N asset values, N values for the dummy variables are generated, and the count of portfolio defaults follows. With multiple simulations, there are as many counts of defaults within the portfolio as there are simulations. The number of

2 $\text{Cov}(Y_i, Y_j) = \text{Cov}[\rho Z + \sqrt{(1-\rho_i^2)}X_i,\ \rho Z + \sqrt{(1-\rho_i^2)}X_j] = \rho^2\sigma^2 + \text{Cov}(X_i, X_j)$. If we ignore the correlation between residuals, the correlation for any pair of variables equals the correlation with the common factor.

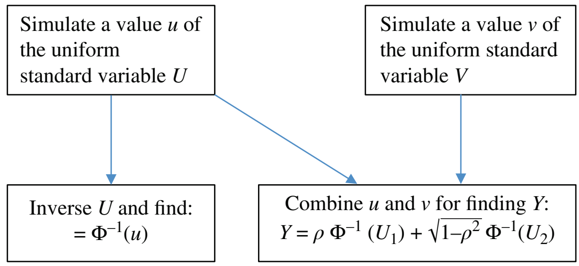


FIGURE 25.1 The simulation algorithm for two normal dependent variables

defaults is proportional to the portfolio loss if all exposures are identical. From the distribution, all statistics of defaults are derived.

The required inputs are the correlation between asset values, ρ, and the default probabilities of each obligor. The common default probability is $d = 5\%$, and the corresponding threshold point of asset value is: $\Phi^{-1}(d) = -1.6449$.

The sequence of simulations of two dependent normal variables is represented in Figure 25.1.

This sequence of the simulation also appears in the rows of Table 25.1, which shows a single simulation.[3] The first line corresponds to the common factor. The subsequent lines refer to each of the obligors in the portfolio and there are as many lines as there are obligors. The full simulation uses a portfolio of 100 firms. The calculations are illustrated for the first three firms only. A first uniform standard variable, U_1, is equal to $\Phi(Z)$, the common factor. All firms are dependent on this common factor. This means that the value of the first uniform variable, U_1, in the first column is the same for all firms. In this single simulation, $U_1 = 0.5$ and the common factor is $Z = \Phi^{-1}(U_1) = 0$.[4] Those two values appear in all lines of the portfolio. The second uniform variable, U_2, is independent of the first one. It is an intermediate variable used to generate the variable Y, a normal variable, with correlation ρ with Z, calculated with the formula:

$$Y = \rho\Phi^{-1}(U_1) + \sqrt{1-\rho^2}\,\Phi^{-1}(U_2)$$

TABLE 25.1 Simulation of the asset values for the first firms of the portfolio

	$U = U_1$	$Z = \Phi^{-1}(U_1)$	U_2	$X = \Phi^{-1}(U_2)$	Y
Factor	0.500	0.000	0.567	0.170	0.162
Firm					
1	0.500	0.000	0.039	−1.765	−1.684
2	0.500	0.000	0.471	−0.072	−0.068
3	0.500	0.000	0.511	0.027	0.026

3 The method is equivalent to the Cholesky decomposition for normal variables.

4 Since U_1 represents a cumulative probability of a normal standard variable, 0.5 matches the mean 0 of the standard normal variable.

TABLE 25.2 From asset value to default event

Firm	Φ^{-1}(d)	*Default*
1	−1.645	1
2	−1.645	0
3	−1.645	0

The variable Y represents the asset values of firms conditional on Z in the last column.

For each simulation, the sequence of steps is: deriving the asset values, Y_i; comparing this value with the threshold point; assigning a default whenever the asset value hits the default point.

For modeling defaults, the simulated asset values have to be compared to the default point. Table 25.2 calculates whether there is a default or not. Each line of this table corresponds to a firm, hence to the same line of the previous table. The default column has only 0 or 1 values. The dummy variable of default is derived from the asset value, in the last column. It equals one when this asset value is equal to or lower than the default point. This default point is −1.645 for a default probability of 5%. In the example, a default occurs for the first firm since the asset value is $-1.684 > -1.645$. For the entire portfolio, the number of defaults sums up the values of the default variables across firms. For running K simulations, the same table is recalculated K times.

The simulations can be used to model distributions with an increasing correlation. The asset correlation varies from 0% to 50% in steps of 10%, generating five loss distributions. A simulation software was used for running 1,000 simulations with a portfolio of 100 lines.

Figure 25.3 shows distributions of portfolio defaults for the different correlation values. The highest mode corresponds to the curve simulated with the highest correlation. The lowest mode corresponds to a zero correlation, when the number of defaults follows a binomial distribution. For highlighting the differences between the distributions of the number of defaults, the X-axis shows only the most frequent losses, between 0 and 10 defaults, although the fat tail extends further to the right.

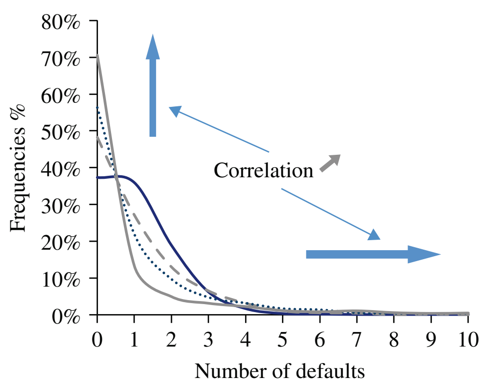


FIGURE 25.2 Loss distributions and loss correlation

The modes of the distributions shift to the left when the uniform correlation increases. Simultaneously, the fat tail extends to the right. The curves are asymmetric and highly skewed. The standard deviation, the skewness and the kurtosis of the distribution increase with the correlation. In the extreme case of a correlation of one, the distribution would have only two points: zero default and 100 defaults. The mean of all distributions is common to all curves since the expectation does not depend on correlation and should always be around five defaults (5% of 100 lines).

Getting to the far end of the right-hand tail implies a much greater number of simulations to stabilize the tail for reducing simulation noise. The simulation noise is the variation of the count of portfolio defaults across different simulation runs. It decreases proportionally to the square root of the number of the simulations. The loss quantiles, measured by the count of defaults in this example, are derived from these distributions. They serve for measuring the credit value-at-risk (VaR) and deriving capital as a loss quantile minus expected loss.

None of the simplifying assumptions are restrictive. Differentiating exposures is not an issue as the losses are obtained by multiplying each default variable by the loss given default assigned to each line. The portfolio loss sums up individual losses of each firm. For differentiating the default probability across firms, it is sufficient to differentiate the default point: a default point of -2.3263 matches a default probability of 1%, and another, -1.96, would match 2.5% default probability, for example. For using different correlations across firms in the portfolio, the factor models of each line of the portfolio could use different correlations with the common factor across lines.

25.2 SIMULATIONS OF TIMES TO DEFAULT

Under the reduced model of default, the distributions of defaults are exponential. Times to default are exponentially distributed for all firms. The dependency between times to default is modeled using the copula technique. A default is triggered when the simulated time to default is shorter than or equal to the horizon of the simulation, for example one year. The simulation generates correlated times to default of all firms. For those firms defaulting within the horizon, a Bernoulli variable takes the value 1 for default and 0 if no default. The count of defaults within the portfolio is the sum of the dummy variables representing default. For the portfolio, the times to default are dependent. Since the distribution of the random variable is exponential and not normal, the methodology relies on the copula function.

25.2.1 Simulation of Time to Default for a Single Obligor

The distribution of times to default is an exponential distribution. The intensity of default λ measures the probability that a discrete event, a default, occurs during a small time interval dt, or:

$$P(t \leq \text{default time} < t + dt) = \lambda\, dt$$

The probability of survival until t is $\lambda \exp(-\lambda t)$ and the probability that time to default is inferior or equal to t is, by integration from 0 to t, the exponential function:

$$F(t) = 1 - \lambda \exp(-\lambda t)$$

TABLE 25.3 Simulating first time to default for a single obligor

U(0, 1)	Time to default
0.39460	10.037
0.19907	4.440
0.36490	9.079
0.72696	25.963
0.62970	19.869
0.17176	3.769
0.05818	1.199
0.71579	25.161
0.45865	12.274
...	...
...	...

The first random time to default is T_1. The simulation of a single series of time to default values uses a first $U_1(0, 1)$ uniform standard variable and derives the values obtained for $F(T_1)$, the survival exponential distribution, with $F(t_1) = U_1$. Each value t_1 of T_1 is the inverse function of F, the exponential distribution:

$$t = F^{-1}(U) = -\ln \frac{1 - U}{\lambda}$$

This procedure allows constructing the frequency distribution of times to default for a single obligor. The example uses $\lambda = 5\%$ for one year, from the default frequency attached to the obligor (Table 25.3).

For example, the first time to default is:

$$t = -\ln \frac{1 - U}{\lambda} t = -\ln\left[(1 - 0.3946)/5\%\right] = 10.037 \text{ years}$$

The curve of Figure 25.3 is the cumulative density function of the time to default generated with 100 simulations. The cumulative probability, $F(T_1)$, between 0 and 1, is along the horizontal axis.

The distribution of Figure 25.4 shows the frequency of times to default. The frequencies are the counts of default events over consecutive constant intervals of times to default, each with duration five years. The time to default in years is on the horizontal axis and the counts of simulated default events in each interval are on the vertical axis.

25.2.2 The Simulation Algorithm for Times to Default

The simulation of dependent times to default was introduced first by Li [90], and has been widely implemented in the industry for credit portfolio models.

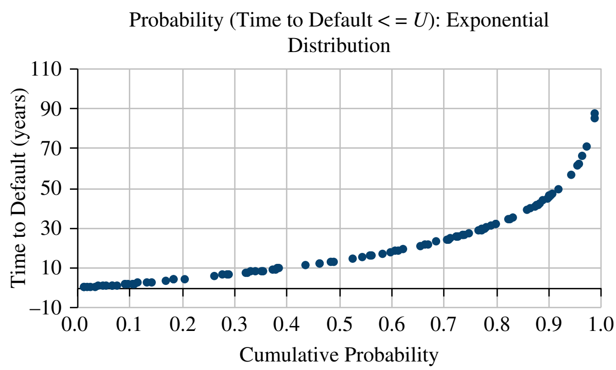


FIGURE 25.3 Cumulative distribution of the first time to default of a single obligor

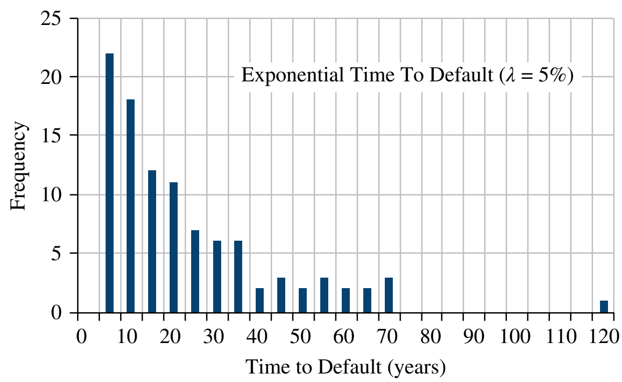


FIGURE 25.4 Frequency distribution of simulated first time to default for a single obligor

For simulating times to default, which are exponentially distributed, the decomposition technique used for normal variables is not applicable and the copula technique is used instead.[5] The two dependent variables are X and Y, with cumulative distribution functions, non-normal, $F_x(X)$ and $F_y(Y)$. The problem is to find two dependent quantiles, respectively, u and v for X and Y. The first quantile is an independent variable, uniform standard. By definition, $u = F_X^{-1}(X)$. The second quantile, v, is conditional on U. It is obtained from the Gaussian copula transformation, written as:

$$v = \Phi\left[\rho\Phi^{-1}(u) + \sqrt{1-\rho^2}\Phi^{-1}(w)\right]$$

In this equation, w represents the quantile of a uniform standard variable independent of u. The correlation is an input. Once v is obtained, the variable Y is simply the inverse of v, using the cumulative function F_y.

5 The principles of copula functions are explained in Chapter 14, section 2.3.

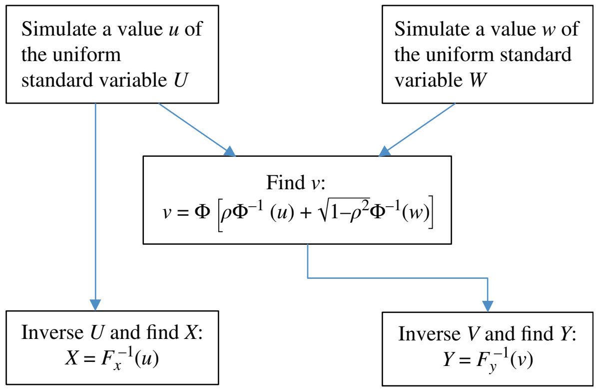


FIGURE 25.5 The simulation algorithm for two dependent variables

The simulation process is detailed in Figure 25.5. For simulating the values of two dependent variables, X and Y, the initial phase of the sequence consists in generating two independent uniform variables representing the quantiles u and w. The Gaussian copula provides v, the quantile of Y. The value, u, of the first uniform variable is the cumulative function of X. The simulated value of X is the inverse of u, or $F_x^{-1}(u)$. The value of Y is derived by taking the inverse of v: $Y = F_y^{-1}(v)$.

For illustrating this simulation process the intensity of default is set to 5%, and the uniform correlation is set to 30%. The purpose of the simulation is to generate times to default of all firms dependent on a common factor.

The first simulations are in Table 25.4, which has a similar structure as the table used for the simulations of asset values. The difference is that the dependent times to default are derived as the inverse function of the exponential distribution. For preserving the notations of the simulations of asset values, U_1 represents the first uniform variable U and the variable U_2 represents W. The sequence of calculations for all firms within the portfolio, for each simulation, is in rows. The first row corresponds to the common factor. The subsequent rows correspond to each firm. The table shows a single simulation, out of 100, with the first three firms.

The first uniform standard variable U_1 represents the common factor, which has a single value for all firms in a simulation run. The second uniform variable, U_2, is the independent

TABLE 25.4 Simulation of times to default for the first firms of the portfolio

	$U = U_1$	U_2	$\Phi^{-1}(U_2)$	$\Phi_2(U, U_2, \rho)$	$V = \Phi(\Phi_2)$	$T_1 = F^{-1}(U_1)$	$T_2 = F^{-1}(V)$
Factor	0.500	0.092	−1.326	−1.265	0.103	13.863	2.171
Firm							
1	0.500	0.382	−0.300	−0.286	0.388	13.863	9.805
2	0.500	0.011	−2.299	−2.194	0.014	13.863	0.285
3	0.500	0.997	2.718	2.593	0.995	13.863	106.951

TABLE 25.5 From time to default to default event

Firm	*Default*
1	0
2	1
3	0

intermediate variable. The bivariate variable $\Phi_2(U_1, U_2, \rho)$ is the copula transformation embedding the dependency. The uniform standard variable V is dependent on $U = U_1$.

The times to default are obtained from U and V values by inverting the exponential distribution. The cumulative distribution of the time to default, T, of each firm is: $F(T) = 1 - \exp(\lambda T)$. The inverse function of the exponential distribution is: $T = -\ln[1 - F(T)]/\lambda$. The values of T_1 and T_2 are derived from U and V from this relation, with U and V replacing $F(T)$. The times to default are in the last two columns. The first time to default matches the factor and is the same across rows. The second time to default is that of the firm, in rows.

Table 25.5 shows the values of the dummy variable representing the default of each firm. Each row of this second table corresponds to the same firm, or row, of the previous table. The values of dummy variables are derived from the firm's time to default, in the last column. Default occurs when the time to default is lower than or equal to 1, which happens for the second firm in this sample simulation.

Each simulation run provides the number of defaults for the portfolio by summing defaults across firms. The frequency distribution of the portfolio defaults is obtained by running a large number of simulations. For K simulations, the same table is calculated K times. For each simulation, U_1 is common to all firms of the portfolio.

The simulations are conducted by generating as many tables as above, with 100 lines. For reducing the simulation sample error in the fat tail of the distribution of portfolio defaults, the number of simulations is increased until the portion of the tail matching some given default quantile is stable enough. Figure 25.6 shows the frequency distribution of defaults of

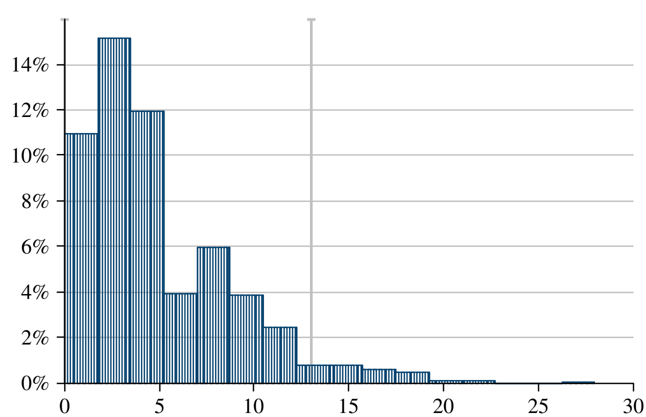


FIGURE 25.6 Distribution of portfolio defaults

the 100-line portfolio over a horizon of one year. The mean of the distribution is 4.84. The 5%-quantile is two defaults and the 95%-quantile is 13 defaults.

None of the simplifying assumptions is restrictive. The default intensity and the exposure could vary across firms. The loss under default is the product of the dummy variable and the exposure attached to each firm. The sum across lines of portfolio would provide the portfolio loss instead of the count of defaults. The correlations with the factor can also be differentiated.

25.3 CREDIT PORTFOLIO MODELS

A number of credit portfolio models has been developed, which rely on the simulation technique for generating credit loss distributions. Well-known models are:

Moody's KMV Portfolio Manager™
CreditMetrics™
CreditPortfolioView™ (CPV)
CreditRisk+

The two first models, Moody's KMV Portfolio Manager and CreditMetrics are vendor models that can be readily implemented by end-users. They are full valuation models that rely on Monte Carlo simulations for generating a distribution of portfolio values at a final horizon, from which loss statistics are derived. The other models provide instead open frameworks, where inputs need to be provided by end-users. CreditPortfolioView is an econometric model, where default rates are modeled from observable economic and country-industry factors. CreditRisk+ relies on actuarial techniques and provides an analytical distribution of default, within an open framework.

A comparative view of models is in Gordy, M. B. (2000), A comparative anatomy of credit risk models, [69]. CreditMetrics is detailed in a technical document from the RiskMetrics Group: Gupton, G. M., Finger C. C., Bhatia, M. (1997), Creditmetrics: Technical Document, [74]. This technical document provides a number of details on credit risk modeling. CreditPortfolioView is presented in Wilson, T. (1996), Portfolio credit risk, I and Wilson, T. (1996), Portfolio credit risk, II, [135, 136]. CreditRisk+ is detailed in Credit Suisse (1997), CreditRisk+: A credit risk management framework, [48].

26

Credit Risk Regulations

The Basel 2 Accord, enforced in 2007, made the capital charges for credit risk more risk sensitive, recognized various forms of credit risk mitigation, provided various enhancements to the former Basel 1 Accord and added capital requirements for operational risk. The Accord combines several approaches and subapproaches differentiated by asset class. The treatment of credit risk for Basel 2 was not amended in subsequent regulations, except for marginal adjustments, and is the current regulatory treatment of credit risk.

This presentation covers the main building blocks for credit risk, as in the final version of the Basel 2 Accord (2006) "International convergence of capital measurement and capital standards: A revised framework – Comprehensive version" [21]. The subsequent documents addressed primarily other issues, such as sound balance sheet management and a revised approach to market risk. The Basel 2 Accord also included a capital charge for operational risk, which is briefly referred to in this chapter for comprehensiveness (see Appendix in section 26.10).

Contents

26.1 THE BASEL 2 ACCORD

The Basel 2 Accord covers three pillars: Pillar 1: Minimum Capital Requirements; Pillar 2: Supervisory Review Process; Pillar 3: Market Discipline.

For pillar 1, the capital requirement for credit risk, the Accord provides three main approaches: the "Standardized" approach and the "Foundation" and "Advanced" approaches, which are internal ratings-based ("IRB") approaches. The foundation and the advanced approaches use internal credit ratings assigned by the bank to all counterparties. Basel allows the recognition of guarantees for reducing the capital charge, under the credit risk mitigation ("CRM") approaches. The standardized approach is used only for banks that have no eligible credit rating system. The capital charge is calculated as the capital adequacy ratio times the risk-weighted exposure.

The entry point for the determination of capital treatment is the asset class. The five main Basel 2 asset classes are:

- Corporate
- Banks
- Sovereign
- Retail
- Equity

Within the corporate asset class and the retail asset class, there are subclasses. Insurance companies are treated as a "corporate" asset class. Counterparties treated as sovereigns include governments (and their central banks), certain public sector entities (PSEs) and development banks. Equity exposures are direct and indirect ownership interests in other corporations, detained by the bank.

The retail portfolio is characterized by a large number of small exposures to individuals. Retail facilities include revolving credits and lines of credit (e.g. credit cards, overdrafts and retail facilities secured by financial instruments) as well as personal term loans and leases or consumer lending. Mortgages secured by residential properties are included in the retail portfolio. Retail facilities are "granular", without size concentrations and a wide diversification across a large number of clients. Small and medium-size enterprises (SME) can be considered as "retail" provided that they are not dealt with on an individual basis when granting credit and are subject to a cap on size. When lending is based on a specific credit analysis, the loan falls in the corporate asset class.

In the standardized approach, the risk weights are defined by the regulators. In the IRB approaches, the regulator provides risk-weight functions, differentiated according to asset classes and subclasses. For specialized lending activities and securitization exposures, risk weights are defined by specific rules.

26.2 THE STANDARDIZED APPROACH

In the standard approach, risk weights are supervisory, and depend on external ratings, when applicable, from credit agencies. The regulatory risk weights are defined by asset class, corporate, sovereigns and banks. They range from 20% for AAA-AA rating up to 150% below the BB rating class for the corporate asset class.

Retail exposures are eligible for a top-down approach, which allows them to be treated as a pool rather than individually as for corporate exposures. The risk weight of general retail exposures is 75%. Lending fully secured by mortgages on residential property is risk weighted at 35%.

26.3 INTERNAL RATINGS-BASED FRAMEWORK

The IRB approaches rely on internal ratings assigned by banks to all counterparties. The capital charge is determined from the risk-weight functions applying to each supervisory portfolio.

Under the foundation approach, banks provide their own estimates of probability of default (DP) and rely on supervisory estimates for other risk components. Regulatory recovery rates are 55% for senior debt and 25% for subordinated debt. Under the advanced approach, banks provide their own estimates of the credit risk components: DP, loss given default (LGD) and exposure at default (EAD). For the retail portfolio, there is no foundation approach and the treatment is either standard or advanced.

26.3.1 Credit Risk Components

The risk weights, under the IRB approaches, are functions using as inputs the credit risk components: DP, EAD, LGD and CCF (credit conversion factors). The "risk-weight functions" define the applicable risk weights.

The regulatory definition of a default event is a non-payment of debt obligation for at least 90 days. Default probabilities are based on internal credit ratings in the IRB approaches. A bank can use the approaches as long as it maps, in a sound manner, its own assessment of ratings with default probabilities. The estimate of the DP should represent a conservative view of a long-run average and annualized default probability, or "through the cycle", rather than a short-term assessment.

The EAD measures the potential amount that can be lost under default. Such amount is often unknown as of current date. Sources of uncertainty with respect to EAD are numerous. For example, for lending products, the fraction drawn on a committed line of credit depends on the borrower's willingness to use the line. For derivatives traded over the counter, the exposure is market driven, and can determined either using regulatory rules or using models of potential future exposures.

The LGD is the fraction of the exposure at risk that is effectively lost under default, after work-out efforts and recoveries from guarantees. Under the foundation approach, senior claims on corporates, sovereigns and banks not secured by recognized collateral are assigned the 45% and 75% LGD for senior and subordinated claims, respectively. Own estimates of LGD by banks are allowed only in the advanced approach.

CCF are used to weight the size of some items that are not cash commitments, such as off-balance sheet commitments. Direct credit substitutes and general guarantees, including standby letters of credit and acceptances, receive a credit conversion factor of 100%. Other short-term commitments have a lower CCF (20%), such as short-term self-liquidating trade letters of credit arising from the movement of goods (e.g. documentary credits collateralized by the underlying shipment).

26.3.2 Capital Calculation and Risk-weight Functions for Corporates

In the IRB approaches, the capital is calculated as the product of the capital ratio times the risk-weighted assets: Capital = 8% × Risk-weighted asset (RWA). The risk-weight functions convert the "risk components" into risk weights for calculating capital:

$$\text{RW} = f(\text{DP}, \text{EAD}, \text{CCF}) \times \text{LGD}$$

The core term of the risk-weight function for corporates is:

$$\text{RW} = \text{LGD}\left\{ N\left[\frac{N^{-1}(\text{DP})}{\sqrt{1-R}} + \sqrt{\frac{R}{1-R}}N^{-1}(0.999)\right] - (\text{DP LGD})\right\}$$

The first term of this complex formula can be summarized as the LGD times a "stressed DP" and minus the expected loss. The expected loss is calculated as DP × LGD. The stressed default probability is derived from the default probability assigned to the exposure when stressed for a factor representing the general economic conditions and under a 0.1% confidence level.[1]

The expanded formula uses other adjustments. The correlation R is a complex function of DP and size of obligor. A firm size adjustment applies for exposures to SME borrowers. The formula is multiplied by a factor with a maturity adjustment. DP and LGD are measured as decimals, and EAD is measured in currency (e.g. euros), except where noted otherwise. $N(x)$ denotes the cumulative distribution function for a standard normal random variable.

For the retail portfolio, the assessment of risk components is made at the segment level rather than at the individual exposure level, unlike other asset classes.

The Basel 2 document lists some examples of risk weights referring to the case of a three-year loan (Table 26.1), with different values of default probabilities, an LGD of 50%. For a DP of 0.7%, the RW is 100%, and the maximum risk weight, for a DP of 20%, reaches 625%. This value is a cap for all maturities and all default probabilities.

A comparison between the former Basel 1 capital charges and the Basel 2 capital charges under the foundation approach is shown in Figure 26.1. The bars are the capital charges for the corporate asset class. The horizontal axis shows the ratings corresponding to the Basel 1 and Basel 2 capital charges. The chart was calculated using a mapping between ratings and default probabilities from historical data.

TABLE 26.1 Sensitivity of risk weights with maturity: benchmark case (three-year asset, 50% LGD)

DP (%)	0.03%	0.7%	20%
BRW (%)	14%	100%	625%

1 As explained in Chapter 24, section 3.

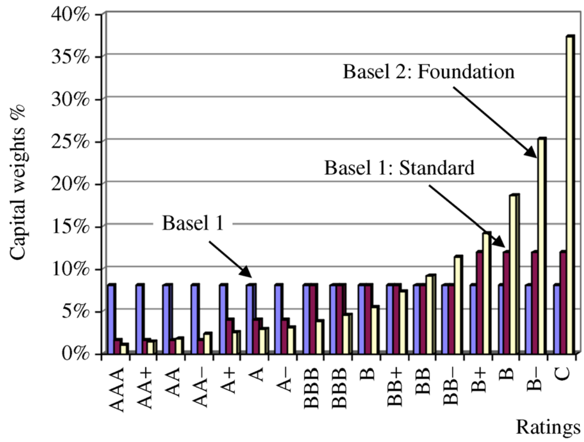


FIGURE 26.1 Capital charges by rating class: Basel 1, Basel 2 standard, Basel 2 foundation

26.4 CREDIT RISK MITIGATION

The Basel 2 Accord recognized the CRM effect of eligible collateral and credit guarantees in the CRM approaches. The CRM rules are common to all approaches, and include several options. When guarantees are recognized, the risk weight and the LGD differ for the collaterized portion of an exposure and the unsecured portion.

For third party guarantees, the credit protection is recognized for sovereign entities, PSEs, banks with a lower risk weight than the counterparty and for other entities rated A- or better. The substitution approach simply substitutes the risk of the guarantor, or of the issuer of securities pledged to that of the borrower, if they have a better credit standing than the borrower.

For collateral-based transactions, it is possible to offset a fraction of the exposure by a value assigned by the collateral. For any collateral-based exposure, a portion is exposed to the collateral and the remaining portion of exposure is exposed to the counterparty. The secured fraction receives the risk weight of the collateral and the remaining fraction receives that of the borrower. The remainder of the claim is unsecured and is assigned the LGD and risk weight of the counterparty. In the presence of multiple collaterals, the exposure should be divided into fractions, each being assigned a single CRM treatment.

For securities pledged as collateral, their value is adjusted by haircuts. Depending on the approach, haircuts are defined with regulatory rules or based on banks' own estimates of collateral volatility. The rules maximize the gap between exposure value and collateral value in order to take into consideration adverse market movements, using a combination of haircuts.

The exposure subject to capital charge is the difference between exposure value and the collateral value to which an add-on is applied. The add-on sums up the various haircuts, assuming that the exposure value can grow by the amount of the haircut and that the collateral value can decrease by the amount of the haircut. The combined effect is to maximize the gap between exposure and collateral values. The general formula for calculating netted exposure to which the risk weight of the counterparty is applied is as follows:

$$E^* = \max\{0, [E(1 + H_e) - C(1 - H_c - H_{fx})]\}$$

where:

E^*: The exposure value after risk mitigation.
E: Current value of the exposure.
H_e: Haircut appropriate to the exposure.
C: The current value of the collateral received.
H_c: Haircut appropriate to the collateral.
H_{fx}: Haircut appropriate for currency mismatch between the collateral and exposure.

26.5 SPECIALIZED LENDING

Under Basel 2, the specialized lending (SL) activities have dedicated capital treatments due to the specifics of these transactions. The five subclasses of SL are: project finance, object finance, commodities finance, income-producing real estate and high-volatility commercial real estate. For SL transactions, the primary source of repayment is the income generated by the asset(s), rather than the independent capacity of a commercial enterprise.

Project finance is a method of funding in which the lender looks primarily to the revenues generated by a single project, both as the source of repayment and as security for the exposure. This type of financing is usually for large projects that might include, for example, power plants, mines, transportation infrastructure and telecommunications infrastructure. In such transactions, the lender is paid from the money generated by the facility's output. The borrower is usually an SPE that is not permitted to perform any other function than developing, owning and operating the installation.

Object finance refers to the financing of the acquisition of physical assets, such as ships, aircraft, satellites, railcars and fleets. The repayment of the exposure depends on the cash flows generated by the specific assets that have been financed and pledged to the lender.

Commodities finance refers to short-term lending to finance reserves, inventories or receivables of exchange-traded commodities, where the exposure will be repaid from the proceeds of the sale of the commodity. The exposure's rating reflects its self-liquidating nature and the structuring of the transaction rather than the credit quality of the borrower. In such cases, the value of the commodity serves as risk mitigant rather than as the primary source of repayment.

Income-producing real estate refers to a method of providing funding to real estate (such as office buildings or industrial space). The repayment and recovery on the exposure depend primarily on the cash flows generated by the properties. The source of these cash flows is generally lease or rental payments, or the sale of the asset.

High-volatility commercial real estate lending refers to the financing of commercial real estate that exhibits a higher loss rate volatility compared to other types of SL. High-volatility commercial real estate includes: commercial real estate exposures; loans financing land acquisition, development and construction phases. The source of repayment is either the future sale of the property or cash flows from the properties.

For SL categories, banks may use the "supervisory slotting criteria approach", mapping their internal risk grades to five supervisory categories. Each supervisory category is associated with a specific risk weight.

26.6 SECURITIZATIONS

Securitizations refer to the sale of the credit risk of portfolio of loans or bonds. The assets are then transferred to a vehicle that finances the acquisition and issues bonds, backed by the assets, in the capital markets.[2] Banks can hold such asset-backed bonds in their banking portfolio or in their trading portfolios.

The treatment of securitization exposures was amended in Basel 3. Notably, under the 2010 revision to the Basel 2 market risk framework,[3] the standardized approach imposes new risk weights for securitization exposures. The exposures in the market portfolio have a specific risk charge, which is calculated as if they were in the banking book.

Under the rating-based approach, supervisory risk weights depend on external ratings, plus risk weights for unrated positions. The risk weights also apply under the IRB approach when external ratings are used. When no ratings exist, a supervisory formula is applied. Under the supervisory formula, the capital charge depends on whether the capital charge for the underlying assets is greater or lower than the level of credit enhancement provided by the securitization vehicle.

26.7 COUNTERPARTY CREDIT RISK

The definitions of exposures under Basel 2 for the measurement of counterparty credit risk have been retained in the new framework of December 2010. But the new framework requires that exposures be modeled using a one-year period of stress. The definitions are found in the related Chapter 22. Under the new framework, a credit value adjustment (CVA) is introduced as an additional charge, addressing the potential variations of value arising from adverse credit risk migrations. A CVA VaR is also imposed as a capital charge.

26.8 INTEREST RATE RISK

The Basel 2 Accord treats interest rate risk in the banking book under pillar 2, rather than defining capital requirements. This implies no capital load, but an enhanced supervisory process. The guidance on interest rate risk considers banks' internal systems as the main tool for the measurement of interest rate risk in the banking book. To facilitate supervisors' monitoring, banks should provide the results of their internal measurement systems using standardized interest rate shocks. If supervisors determine that a bank is not holding capital commensurate with the level of interest rate risk, they can require that the bank reduces its risk, or holds an additional amount of capital, or combine the two.

2 Chapter 30 details securitizations.

3 Basel Committee on Banking Supervision (2010), Basel III: A global regulatory framework for more resilient banks and banking systems, December 2010, revised in June 2011, [24].

26.9 PILLAR 2: SUPERVISORY REVIEW PROCESS AND MARKET DISCIPLINE

The second pillar of the Basel 2 framework aims at ensuring that each bank has sound internal processes to assess the adequacy of its capital based on a thorough evaluation of its risks. Supervisors are responsible for evaluating how well banks are assessing their capital needs relative to their risks. The risk-sensitive approaches developed by the Accord rely extensively on banks' internal methodologies, giving banks more discretion in calculating their capital requirements.

The regulators regard the market discipline through enhanced disclosure as a fundamental part. Separate disclosure requirements are prerequisites for the supervisory recognition of internal methodologies. Given the influence of internal methodologies on the capital requirements established, regulators consider that a comprehensive disclosure is important for market participants to understand the risk profile of an institution. Accordingly, the usage of internal approaches is contingent upon a number of criteria, including appropriate disclosure.

26.10 APPENDIX: OPERATIONAL RISK IN BASEL 2

The Basel 2 Accord introduced a capital charge for operational risk. The Basel Committee adopted a standard industry definition of operational risk: "The risk of direct or indirect loss resulting from inadequate or failed internal processes, people and systems or from external events."

Operational risks relating to internal processes include events such as:

- Inadequate procedures and controls for reporting, monitoring and decision-making.
- Inadequate procedures on processing information, such as errors in booking transactions and failure to scrutinize legal documentation.
- Organizational deficiencies.
- Risk surveillance and excess limits: management deficiencies in risk monitoring, such as not providing the right incentives to report risks, or not abiding by the procedures and policies in force.
- Errors in the recording process of transactions.
- The technical deficiencies of the information system or the risk measures.

Operational risk raises essentially data and methodology challenges. Modeling operational risk requires setting up a classification of risk events, plus the assessment of their frequencies and monetary impacts. There are various sources of historical data on incidents and their costs, which serve for measuring the number of incidents and the direct losses attached to such incidents. Beyond statistics, other sources on operational events are expert judgments, questioning local managers on possible events and what would be their implications, pooling data from similar institutions and insurance costs that should relate to event frequencies and costs.

The Basel Committee estimated operational risk at 20% of minimum regulatory capital as measured under the Basel 1 Accord. Basel proposes a range of three approaches to capital requirements for operational risk of increasing sophistication:

- Basic indicator;
- Standardized;
- Internal measurement.

The "basic indicator approach" links the capital charge for operational risk to a single indicator that serves as a proxy for the bank's overall risk exposure. For example, if gross income is the indicator, each bank should hold capital for operational risk equal to a fixed percentage ("alpha factor") of its gross income.

The "standardized approach" builds on the basic indicator approach by dividing a bank's activities into a number of standardized business lines (e.g. corporate finance and retail banking). Within each business line, the capital charge is a selected indicator of operational risk times a fixed percentage ("beta factor"). Both the indicator and the beta factor may differ across business lines.

The "internal measurement approach" allows individual banks to rely on internal data for regulatory capital purposes. The technique requires three inputs for a specified set of business lines and risk types:

- An indicator of exposure to operational risk;
- The probability that a loss event occurs;
- The losses given such events.

Together, these components allow construction of a loss distribution for operational risks.

27

Capital Allocation and Risk Contributions

The risk allocation issue stems from the diversification effect of portfolios. The risk of the sum of the portfolios is lower than the sum of the risks of the subportfolios. The allocation problem is consists of distributing this diversification benefit of the portfolios to its facilities, or subportfolios, in a fair manner, for assessing their risk within the global portfolio.

The risk allocation defines the appropriate capital charge of subportfolios or facilities. Unlike risks, the capital allocations should be additive and sum up to total economic capital. Risk-adjusted performances of subportfolios depend on capital and are relevant and fair only to the extent that the capital allocation is.

Within an existing portfolio, the risk contributions are the allocations of the current portfolio risk to subportfolios or facilities. Marginal risk contributions measure the incremental risk "with and without" an additional facility or subportfolio. This chapter explains how such risk allocations can be defined.

Contents

27.1 RISK AND CAPITAL ALLOCATIONS

The risk, and the capital, allocation issue results from to the subadditive property of individual risks due to diversification effects. The capital allocation issue arises for economic capital, not for regulatory capital.

For regulatory capital, the capital charges are added across facilities to obtain the total regulatory capital. The regulations allow for a standard diversification effect when defining individual charges. But they do not address the actual diversification effect that is specific to a portfolio. Banks willing to measure their economic capital, given the specifics of their portfolio, have to define reasonable allocation rules.

The usage of capital should reflect the risk of a segment or of a facility once the diversification effect is accounted for. The standalone risk of a transaction portfolio is intrinsic to a facility, independently of diversification effects within a portfolio. The retained risk is the fraction of risk not diversified away by the rest of the portfolio and measures its risk contribution to the portfolio.

Risk contributions and allocation of capital allow measuring how the overall risk of the portfolio should be distributed to individual facilities or subsets of the portfolio. For the allocation to be relevant, it should have desirable properties. The risk contribution should be lower than the standalone risk, or the risk measured independently of the portfolio context, since a fraction of the risk is diversified away by the portfolio. The allocation should be full, or such that the sum of risk contributions of all subportfolios is equal to the risk capital of the portfolio.

Since risk differs across portfolio segments or facilities, the performances are not comparable. The purpose of risk-adjusting performance measures is to allow comparisons by making the performance of high-risk facilities comparable to the performance of low-risk facilities. The risk adjustment is based on capital allocation as a risk metric, in the return on risk-based capital.

For an existing portfolio, the risk is supposed to be known and is distributed across subportfolios. When an expansion of the portfolio is considered, the decision for new exposures depends on how they change the portfolio risk. The relevant risk measure is the incremental risk of the portfolio from the new transactions. Risk-sensitive pricing imposes that this incremental risk be properly assessed for making sure that the new transactions will effectively compensate capital. The corresponding measure is the incremental, or marginal, contribution to risk or capital. In subsequent definitions, these two situations are differentiated. The term risk contribution refers to the allocation of risk within an existing portfolio and of which the capital is known. Marginal risk contribution, or marginal capital, refers to the incremental effect of additional facilities or subportfolios within a portfolio.

The literature on capital, or risk, allocation raises numerous challenges. Capital allocations should be based on risk metrics that have desirable properties, notably being coherent, and they should satisfy the principle of full allocation.[1]

The risk contributions detailed hereafter, and the examples, refer to the decomposition of the portfolio loss volatility into the contributions of assets or subportfolios. Some risk contributions

1 The Euler family of risk contributions are defined with risk measures $F(x)$ such that $F(hx) = hF(x)$, where x represents a vector of positions x_k and h a constant. In practice, F is the volatility or a quantile of the portfolio loss distribution and the x_k represent the random losses of assets. The functions can be decomposed linearly in risk contributions proportional to the exposures that are the first derivatives of the function, following: $F(x) = \sum_k x_k \partial F(x)/\partial x_k$. For literature on theoretical properties of different risk contributions with other metrics, see Tasche, D. (2008), Capital allocation to business units and subportfolios: The Euler principle, [129].

apply when the portfolio loss volatility, and it economic capital, are known. In the case of an expansion, the marginal risk contributions should be used for pricing purposes.

27.2 RISK CONTRIBUTIONS TO AN EXISTING PORTFOLIO

The risk contributions for an existing portfolio measure how much risk should be assigned to portfolio segments or facilities. One way of defining such risk contributions consists of decomposing the variance of the portfolio losses into additive elements. The distribution of portfolio losses is obtained from credit models. Loss volatility, expected loss and loss quantiles derive from this distribution. The variance of losses is subadditive, but the quantiles are not. However, once a loss distribution is generated for a known portfolio, the capital can be expressed as a multiple of loss volatility. This multiple can be used to move from a decomposition of variance to the allocation of capital.

27.2.1 Risk Contributions to Portfolio Loss Volatility

The decomposition of variance of an existing portfolio can be used to measure how any subset of the portfolio contributes to the total variance, given its size and its correlation with the rest of the portfolio.

The portfolio loss is the sum of the random losses of each individual segment or facility, L_i, for facility i. The variance of the total loss is the variance of a sum of individual losses, which can be decomposed into a sum of covariances of each individual loss with the portfolio loss:

$$\sigma_P^2 = \text{Cov}(L_P, L_P) = \text{Cov}(\Sigma_i L_i, L_P) = \Sigma_i \text{Cov}(L_i, L_P)$$

Each term of the summation in the right-hand side of this identity is the contribution of a particular facility to the variance of the portfolio loss. A similar decomposition can be applied to the volatility of portfolio loss. The portfolio loss volatility is, by definition, the portfolio loss variance divided by the portfolio loss volatility:

$$\sigma_p = \Sigma_i \text{Cov}(L_i, L_P)/\sigma_p$$

The terms of the sum are the covariances between the individual loss and the portfolio loss, divided by the portfolio loss volatility. They measure the risk contributions to the loss volatility of the portfolio. The risk contribution of a segment or a facility, noted RC_i^P, can be defined as: $\text{RC}_i^P = \text{Cov}(L_i, L_P)/\sigma_P$. Since the losses are proportional to the size of exposures, the risk contribution is a linear function of the size of the transactions.[2]

These risk contributions have desirable properties making them relevant. The risk contributions are additive and they have the property of full completion since they sum up to the portfolio loss volatility:

$$\Sigma_i \text{RC}_i^P = \Sigma_i[\text{Cov}(L_i, L_P)]/\sigma_P = \sigma_P$$

2 Given this property, the risk allocation according to the decomposition of volatility belongs to the Euler family of allocations.

An alternate form of the same risk contributions is helpful to highlight some of their properties. Each individual loss is correlated with the portfolio loss with the coefficient ρ_{iP}. The standalone loss volatility, σ_i, is the volatility of losses measured at the level of the facility, and independently of the portfolio context. Writing that $\text{Cov}(L_i, L_P) = \rho_{iP}\sigma_i\sigma_P$ and dividing both terms by σ_P, the risk contribution is:

$$\text{RC}_i^P = \text{Cov}(L_i, L_P)/\sigma_P = \rho_{iP}\sigma_i$$

This new formulation shows that the risk contribution is proportional to its correlation with the rest of the portfolio and to the standalone risk of the facility.

Since all correlation coefficients are lower or equal to 1, this also shows that the risk contribution is always lower than the standalone risk: $\text{RC}_{iP} \leq \sigma_i$. But, because the risk contribution is proportional to standalone risk, it also increases, if correlation is positive, with its magnitude. In other words, risk contributions are higher when they are more correlated with the portfolio and when the standalone risk is higher.

The retained risk, RR_i, of an individual facility is the ratio of the risk contribution, RC_i^P, to standalone risk (σ_i). It is identical to the correlation with the portfolio: $\text{RR}_i = \rho_{iP}$. Since correlations are lower than 1, the retained risk is always lower than the standalone risk of a facility. Moreover, the retained risk varies proportionally with the correlation between individual risks and the portfolio risk. This makes sense intuitively. If a facility contributes significantly to the risk of the portfolio, its retained risk should be higher. If a facility diversifies further from the portfolio risk, its retained risk should be lower, eventually negative. This could happen with credit instruments offsetting the risk of others, like credit derivatives.

These formulas are general. They simplify in some special cases. For example, for portfolios with the same uniform correlation across all pairs of facilities, all risk contributions become proportional to the standalone loss volatilities, the coefficient being the uniform correlation.

If individual losses are independent, all covariance terms are zero. The variance of the portfolio loss collapses to the arithmetic summation of the individual standalone loss variances σ_i^2: $\sigma_P^2 = \text{Cov}(L_P, L_P) = \text{Cov}(\Sigma_i L_i, L_P) = \Sigma_i \sigma_i^2$. The portfolio loss volatility is the square root of this summation. Therefore, the risk contribution becomes the ratio of the standalone loss variance of the facility to the loss volatility of the portfolio: $\text{RC}_i^P = \sigma_i^2/\sigma_P$.

The risk contributions sum up to the portfolio loss volatility, but do not add up to capital, which is a quantile of portfolio losses. In general, quantiles are not subadditive, whether the risk contributions to volatility are. In order to convert these into capital allocations, the multiple of capital, for a given confidence level, to loss volatility is known and can be used as a scaling factor. If K_α is the capital, at the confidence level α, there is always a multiple m_α such that $K_\alpha = m_\alpha LV_P$, whatever the shape of the loss distribution. Once the loss volatility of the portfolio is decomposed into the risk contributions, the capital allocations are obtained by multiplication of these risk contributions by the multiple m_α: $K_\alpha = m_\alpha LV_P = \Sigma_i m_\alpha \text{RC}_i$.

27.3 CALCULATIONS OF RISK CONTRIBUTIONS

An example illustrates the decomposition of the portfolio volatility. The calculations are detailed with a two-obligor portfolio and a default correlation of 10%. Table 27.1 provides the unconditional default probabilities and the exposures of each obligor, A or B. The loss given default is 100% of exposure.

TABLE 27.1 Portfolio of two obligors

	Default probability	*Exposure*
A	7.00%	100
B	5.00%	50
ρ_{AB}	10.00%	

There are four possible values of losses, depending on whether A and B survive, or both default, or only A or B defaults. The joint default probability of A and B results from the general formula:[3] 0.9061%.

Table 27.2 cross-tabulates the possible events. The margins of the matrix are the unconditional survival and default probabilities of A and B. The probabilities of other scenarios derive from the joint default probability and these unconditional probabilities. For example, the probability that B survives conditional on A's default is the unconditional default probability of A (7%) minus the joint probability of default, or 6.094%. The probabilities for other cells are calculated in the same way. The table should be reorganized as a cumulative loss distribution for finding the quantiles (Table 27.3).

With this simple portfolio, it is possible to illustrate that quantiles are not subadditive, which makes contributions to loss volatility more appropriate. The $q\%$-quantile is the value x of the

TABLE 27.2 Joint probability matrix: default correlation of 10%

		A		
		Default	*No default*	*Total*
B	Default	0.906%	4.094%	5%
	No default	6.094%	88.906%	95%
	Total	7%	93%	100%

TABLE 27.3 Cumulative loss distribution

Loss	*Probability*	*Cumulative probability*
0	88.906%	88.906%
50	4.094%	93.000%
100	6.094%	99.094%
150	0.906%	100%
Total	100%	100%

3 See Chapter 23.

loss such that $F(x) \leq q\%$, where F is the cumulative loss distribution. For a discrete distribution, the determination of quantiles should be done carefully as some quantiles do not necessarily match any of the discrete values. For example, there is no value of the loss matching the 95%-quantile. This quantile is such that there are at least 95% of values below and no more than 5% above, which is true for the loss value 100. To see that quantiles are not subadditive, consider, for example, the 1%-quantiles for A and B: these are equal to their full exposures, 100 and 50. The same quantile for the portfolio, as derived from the loss distribution, is also 150. This shows that, for this quantile, there is no diversification effect.

The definitions of individual losses are straightforward under default mode. The loss under default, L_i, with $i = \text{A, B}$, is the product of exposure with the percentage loss under default. The loss for a facility is zero if no default or the loss under default otherwise. Default is a Bernoulli variable d_i taking the value 1 in case of default, with probability d_i and zero with probability $(1 - d_i)$. The random loss is $L_i = L_i d_i$. Its expectation, variance and volatility follow:

$$E(L_i) = L_i d_i$$

$$V(L_i) = L_i^2 d_i(1 - d_i)$$

$$\sigma(L_i) = L_i \sqrt{d_i(1 - d_i)}$$

The random loss of the portfolio is L_P, which sums up the random losses L_i of each obligor. The expected losses of A and B are the default probability times the exposure or $100 \times 7\% = 7$ and $50 \times 5\% = 2.5$, respectively. The expected loss of the portfolio is the sum of standalone expected losses, 9.5. It is independent of the correlation. The loss volatility, σ_P, for the entire portfolio is the volatility of a sum of the random individual losses. It depends on the correlation between individual losses and their standalone volatilities.

The standalone loss volatilities of A and B are determined from the unconditional default probabilities:

$$\sigma_{\text{A}} = 100\sqrt{7\%(1 - 7\%)} = 25.515$$

$$\sigma_{\text{B}} = 50\sqrt{5\%(1 - 5\%)} = 10.897$$

The portfolio loss volatility is lower than the arithmetic summation of these standalone volatilities. It is:[4] $\sigma_P = 28.729$. The diversification effect can be measured by the difference between the sum of the volatilities of A and B and the loss volatility of the portfolio (A + B):

$$\sigma_{\text{A+B}} = 28.729 < \sigma_{\text{A}} + \sigma_{\text{B}} = 25.515 + 10.897 = 36.412$$

The risk contributions to the portfolio loss volatility are, with i being either A or B:

$$\text{RC}_i^P = \frac{\text{cov}(L_i, L_{\text{A}}) + \text{cov}(L_i, L_{\text{B}})}{\sigma_P}$$

4 The calculation of the variance is: $\sigma_P^2 = (150 - 9.50)^2 \times 0.906\% + (100 - 9.50)^2 \times 6.094\% + (50 - 9.50)^2 \times 4.094\% + (0 - 9.50)^2 \times 88.906\% = 825.36$. The loss volatility is the square root.

TABLE 27.4 Risk contributions to the portfolio loss volatility

	A	*B*	*Portfolio*
Loss volatilities	25.515	10.897	36.412
Allocation – standalone risks	70.07%	29.93%	100%
Risk contributions to loss volatility	23.628	5.101	28.729
Allocation – risk contributions	82.24%	17.76%	100%

The covariances depend on standalone volatilities and the correlation coefficient according to: $\mathrm{Cov}(L_\mathrm{A}, L_\mathrm{B}) = \rho_\mathrm{AB}\sigma_\mathrm{A}\sigma_\mathrm{B}$. The risk contributions of A and B are expanded using $\rho_\mathrm{BA} = \rho_\mathrm{AB} = 10\%$, plus $\sigma_\mathrm{A} = 25.51$ and $\sigma_\mathrm{B} = 10.90$:

$$\mathrm{RC}_\mathrm{A}^P = (25.51^2 + 10\% \times 25.51 \times 10.90)/28.73 = 23.628$$

$$\mathrm{RC}_\mathrm{B}^P = (21.79^2 + 10\% \times 25.51 \times 10.90)/28.73 = 5.101$$

By definition, the risk contributions sum up to the loss volatility of the portfolio: $\mathrm{RC}_\mathrm{A}^P + \mathrm{RC}_\mathrm{B}^P = 28.729$. Table 27.4 provides the standalone loss volatilities, the portfolio loss volatility and the risk contributions to loss volatility and capital. For comparison purposes, the allocations of portfolio loss volatility based on the standalone risks of A and B are included.

The allocation based on risk contributions differs from an allocation according to the standalone loss volatilities. This allocation would be based on the ratios of individual standalone risks to their sum. The standalone loss volatilities of A and B are, respectively, 25.515 and 10.897, summing up to 36.412. The corresponding percentage allocations are 70.07% and 29.93%, respectively. When the risk contributions are used instead, the percentage allocations to A and B become 82.24% and 17.76%. The difference is that the allocations based on standalone risks do not embed the dependencies between A and B. Instead, the allocations based on risk contributions do, and should be preferred.

The risk contributions to volatility serve as the basis of capital allocations. The ratio of capital to loss volatility depends on the confidence level. With a confidence level of 0.906%, the loss quantile of the portfolio is 150 and the capital is the loss quantile in excess of the expected loss, 150 – 9.5 = 140.5. Since the portfolio loss volatility is 28.729, the ratio of capital to loss volatility is: 140.5/28.73 = 4.891.

The calculation of the risk contributions to the portfolio loss volatility can be generalized to a large number of facilities, using matrix formulas (see Appendix in section 27.5).

27.4 MARGINAL RISK CONTRIBUTIONS

Marginal risk contributions are the incremental risks due to a new facility or a new subportfolio added to an existing portfolio. Chapter 28, on risk-adjusted performance, shows that the marginal risk contributions are the correct references for risk-based pricing because they guarantee that the return remains in line with the required overall return on capital.

The above standard risk contributions are defined with an existing portfolio, not for an expansion of the portfolio. Marginal contributions can be defined as the incremental variance, the volatility of the portfolio loss or the incremental capital. Some properties of marginal risk contributions are illustrated using the same two-obligor portfolio.

27.4.1 Calculations of Marginal Risk Contributions

Determining the marginal contributions of either B or A to the portfolio loss volatility or to capital is simple. The final portfolio, A + B, is always the starting point. The marginal risk contributions are differences between the risk of A + B and the risks of A or B. The marginal risk contributions are noted MRC_A or MRC_B. The marginal risk contribution of B is relevant when B is the second to enter into the portfolio A.

The portfolio loss volatility with both facilities is already calculated (28.729). The initial loss volatilities are the standalone loss volatilities. When the initial portfolio is A, the marginal contribution of B is: $MRC_B = 28.729 - 25.515 = 3.214$. This shows that the marginal risk contribution of B is lower than its risk contribution once in the portfolio and is also lower than its standalone risk:

$$MRC_B < RC_B < \text{Standalone risk (B)}$$

$$3.214 < 5.101 < 10.897$$

The calculations can be repeated when shifting the entrance order, using B as initial portfolio, and B + A as final portfolio. The marginal risk contribution of A is: $MRC_A = 28.729 - 5.101 = 17.832$, and a similar ranking is observed.

$$MRC_A < RC_A < \text{Standalone risk (A)}$$

$$17.832 < 23.628 < 25.515$$

The sum of the marginal contributions of A and B is 21.046, lower than the portfolio loss volatility. Therefore, they cannot be used as allocations of the risk in the final state. In general, the marginal risk of a facility depends on which facilities are already in a portfolio, and, for this reason, it depends on the order of entrance in the portfolio.

The above ranking of risk contributions within the portfolio does not apply to capital because the ratio of capital to loss volatility changes considerably when adding the new facility in this example with only two large facilities. This can be seen with the 0.906%-quantile of loss as a measure of capital. With this level, the standalone loss quantiles of A and B are 100 and 150 and they add up exactly to the loss quantile of the portfolio when combined. This is true for some other values of the confidence level (98% or 99%, for example).

27.4.2 General Properties of Risk Contributions to the Portfolio Loss Volatility

The example highlights empirically some comparisons of risk contributions to marginal contributions and some properties of marginal risk contributions, which are general. An overview of these properties is useful before getting into the details.

The marginal risk contributions to the portfolio loss volatility are lower than the risk contributions after inclusion in the portfolio, and the latter are also lower than standalone loss volatilities. Accordingly, marginal risk contributions to portfolio loss volatility add to less than portfolio loss volatility. The marginal risk contributions of the same facility as first entrant and second entrant differ because they depend on the existing facilities before inclusion. Such properties are apparent here because individual exposures of the example are large compared to total exposure.

27.4.3 Marginal Contributions to Volatility

The marginal risk contribution of a facility X to the volatility of a portfolio is MRC_X. It is the difference between the loss volatilities of the portfolio with and without the facility X. The initial portfolio, before the new facility is added,[5] is P. The portfolio becomes the final portfolio $P + X$ once the new facility is added. In subsequent notations, the superscript, P or $P + X$, designates the portfolio used as reference. By definition, the marginal risk contribution of X, or MRC_X, is:

$$\text{MRC}_X = \sigma_{P+X} - \sigma_P$$

All (non-marginal) risk contributions within the portfolio $P + X$ are additive and sum up to the loss volatility of the final portfolio ($P + X$):

$$\sigma_{P+X} = \text{RC}_P^{P+X} + \text{RC}_X^{P+X}$$

For comparing the marginal contribution of X to the risk contribution of X once in the portfolio, the loss volatility of $P + X$, σ_{P+X}, is replaced by the sum of the risk contributions of P and X once in the final portfolio $P + X$:

$$\text{MRC}_X = \text{RC}_X^{P+X} + (\text{RC}_P^{P+X} - \sigma_P)$$

Adding the new facility X to the initial portfolio P triggers two effects. The new facility increases the exposure and the risk of the portfolio. However, simultaneously, it diversifies away some of the risk of the initial portfolio. In other words, a facility in the initial portfolio has a lower risk contribution within the larger final portfolio. Therefore, the risk contribution of the initial portfolio, P, to the final larger portfolio, $P + X$, is lower than its standalone risk. This implies that the second term, within parentheses, of the marginal risk contribution is negative. The retained risk (RC_P^{P+X}) allocated to the initial portfolio within the larger final portfolio is lower than its standalone risk σ_P because the new facility X diversifies away some of the risk of the initial portfolio P.

The implication is that the marginal risk contribution of the new facility X, given the initial portfolio P, is lower than its risk contribution once in the final portfolio $P + X$. Mathematically, $\text{RC}_P^{P+X} \leq \sigma_P$, which implies that $\text{MRC}_X < \text{RC}_X^{P+X}$. Moreover, the standalone loss volatility of facility X is always larger than its risk contribution within the portfolio $P + X$, which measures the risk retained after diversification. Combining both results, we find the general property that

5 Unlike when dealing with non-marginal risk contributions.

the marginal risk contribution to loss volatility is lower than its risk contribution to the final portfolio, which is also lower than the standalone loss volatility:

$$\mathrm{MRC}_X^P < \mathrm{RC}_X^{P+X} < \sigma_X$$

Such inequalities would apply to capital allocations only to the extent that capital is proportional to loss volatility before and after inclusion of a new facility. This is acceptable for incremental facilities much smaller than the portfolio in terms of size and risk. For larger subportfolios, the allocation of capital should be conducted once from the final capital of the portfolio is known and the ranking of marginal versus non-marginal capital allocations is not mechanical anymore.

27.5 APPENDIX: MATRIX FORMULA FOR RISK CONTRIBUTIONS

The risk contributions derive from the simple matrix formulas of the variance of the portfolio loss. This matrix extends to any number of obligors. The matrix formulas are illustrated with the same simple example of a two-obligor portfolio. The variance of the portfolio loss distribution is:

$$\text{Portfolio loss variance } \sigma_P^2 = \mathrm{X}\,\Sigma\,\mathrm{X}^{\mathrm{T}}$$

X: The row vector of exposures.
X^{T}: The column vector of exposures, the transpose of the above.
Σ: The variance–covariance matrix of the portfolio, where each term is measured per unit of exposures.
ρ_{AB}: The default correlation between A and B.

The only difference with the calculations in the text is that the covariances and volatilities should be expressed per unit of exposure. The loss volatility per unit exposure, with default probability d_i, is: $\sqrt{d_i(1-d_i)}$, with d_i equal to 7% and 5%. The covariance is: $\rho_{\mathrm{AB}}\sigma_{\mathrm{A}}^1\sigma_{\mathrm{B}}^1 = 10\% \times 25.515\% \times 21.794\% = 0.5561\%$, where the superscripts refer to unit exposures (Table 27.5).

The variance calculation with matrix format is shown in Table 27.6.

The decomposition of variance is obtained by multiplying 6.788 and 2.931 by the corresponding exposures, 100 and 50, or 679 for A and 147 for B. The risk contributions are obtained by dividing this decomposition of variance by the portfolio loss volatility, 28.729, and are 23.628 and 5.101 for A and B, respectively. The variance is obtained after multiplication of the row vector for the decomposition of variance by the transposed vector X^{T}, and its square root is the portfolio loss volatility.

TABLE 27.5 Variance–covariances matrix

6.5100%	0.5561%
0.5561%	4.7500%

TABLE 27.6 Calculations of risk contributions

				X^T
		$\sum$ (unit exposure)		
		6.5100%	0.5561%	100
		0.5561%	4.7500%	50
X vector				
		σ_{AP}	σ_{BP}	Variance
100	50	6.788	2.931	825.358

28

Risk-adjusted Performance Measures

The standard measure of performance of the bank is the return on economic capital or regulatory capital. When a target return on capital is fixed, the issue is to find out whether transactions are contributing to the global return of the bank. Since risk varies across transactions, the measure of performance has to be risk adjusted.

Performances adjusted for the risk of transactions are the return on risk-adjusted capital (RoRaC[1]) and shareholder value added (SVA). The metric for risk adjustment is the capital allocation to transactions or subportfolios. The measure of income, in the banking portfolio, requires that the fund transfer pricing system, for allocating income, be implemented together with capital allocation.

This chapter defines the RoRaC and SVA measures of performance and shows that the relevant capital allocations for risk-based pricing are the marginal risk contributions.

Contents

1 The RoRaC acronym is preferred to RaRoC for "risk-adjusted return on capital" because the latter suggests that the adjustment applies primarily to return, whether or not it applies primarily to capital.

28.1 RISK-ADJUSTED MEASURES OF PERFORMANCE

In risk-adjusted measures of performance, the risk metrics used are the capital allocated to a transaction and the expected loss. The two traditional measures of risk-adjusted performances are the RoRaC and SVA. The funds transfer-pricing system is used for allocating income and the risk allocation system for risks.

The RoRaC is the ratio of earnings to allocated capital used as the risk-adjustment metric:

$$\text{RoRaC} = \frac{\text{Earnings} - \text{Expected Loss} - \text{Operating costs}}{\text{Allocated Capital}}$$

Earnings include interest income and fees in the banking book, and are net of expected credit loss. They can be pre-tax or after tax, and after or before, allocated operating costs. For comparison purposes with the required return on capital of the bank, it should be calculated in the same way.

The SVA is the earnings net of the expected loss and the cost of allocated capital. The calculation involves explicitly the target return on capital for the bank. Such target is the required return on capital by shareholders, or cost of equity capital, k. In the SVA, the cost of capital deducted from earnings is in monetary units, and is the product of the percentage cost of equity and the amount of capital allocated:

$$\text{SVA} = \text{Earnings} - \text{Expected loss} - k\ \text{allocated capital}$$

The two measures have different usages. From a purely theoretical standpoint, the return measure is relevant for portfolio optimization. From a practical standpoint, they complement each other. The RoRaC ratio does not provide information either on the size of the transaction, or on the size of earnings generated. The SVA figure depends on size, but it does not reveal whether a high value, for example, results from a low return on capital combined with a large transaction size or the opposite.

The RoRaC ratio is defined as long as the capital is not zero, which happens only when the risk is zero. For risk-free transactions, the ratio is undefined. Any earnings are then measured as the gap between revenues and operating costs. SVA is not subject to infinite or undefined ratios because it is a simple difference.

The expected loss is netted from revenues. For an individual transaction, the expected loss looks theoretical since the loss given default can either occur or does not exist. However, it makes sense to set aside a fraction of income, since, once aggregated, the provision should offset the statistical loss of the portfolio.

The revenues for a loan include the net interest income plus any upfront and recurring fees. The upfront fee increases earnings in the early stage of a transaction and not after, which creates a distortion of the ratio. It is common to use all-in-revenue as the annualized value of interest income and fees calculated over the life of the transaction. The all-in-spread is the spread over the cost of debt.

These measures are primarily used for the banking portfolio, and apply to transactions, clients' portfolios or business units. For the market portfolio, the earnings would be the profit and loss (P&L) of trading desks and the capital and allocation of value-at-risk (VaR).

The cost of equity capital is the required return on equity by shareholders.[2] It is defined by the well-known CAPM (Capital Asset Pricing Model) model, which adds to the risk-free rate a premium that depends on the risk of the stock measured by the β of the stock. It serves as hurdle rate for the RoRaC ratio, the return that subportfolios or facilities should beat.

Once a hurdle rate is defined, various references for the level of capital can be used: the available capital, the regulatory capital and the economic capital. Regulatory capital is risk adjusted, but is based on a portfolio that regulators consider as representative for the industry. Economic capital is supposed to be closer to the actual risks given the specifics of a bank's portfolio. The actual capital is the reference for shareholders and for determining their required return.

If available capital is higher than economic capital, and if the required return on available capital is 20%, the return on economic capital is higher than 20%. For instance, with an actual capital of 100, the target profit is $20\% \times 100 = 20$. If economic capital is 80, the effective return on economic capital is higher and equal to $20/80 = 25\%$. The return on economic capital is scaled by the ratio of available to economic capital:

$$\text{Return of economic capital} = 20\% \times (\text{Available capital}/\text{Economic capital}) = 25\%$$

The correct reference for measuring risk is the economic capital since it is the closest image of the true risk of the firm. However, the actual capital or regulatory capital is a constraint imposed on the firm. Suppose the firm complies with the requirements of shareholders of a minimum return of 20% on actual capital. The higher return on economic capital is an incentive to take more risks since the available capital overstates the portfolio true risk. The bank should increase the risk of its portfolio or call back some of the excess capital. If so, the economic capital should tend towards available capital.

28.2 RISK-BASED PRICING AND MARGINAL RISK CONTRIBUTIONS

The goal of risk-based pricing is to determine the minimum return of new transactions, in line with the risk and with the target bank's return. When considering a new facility, its pricing should be such that the return on the capital, once the facility is included in the portfolio, is at least in line with the target return on capital.

For example, a portfolio has a capital of 100. The hurdle rate is 20% pre-tax and before operating expenses. Assuming that an initial portfolio has a return of 20%, it generates earnings of: $20\% \times 100 = 20$. The marginal capital assigned of a new transaction is 15. The target net income when the new transaction is included in the portfolio is $20\% \times 115 = 23$. This is an increase of $23 - 20 = 3$. This increase is exactly $20\% \times (115 - 100) = 3$, or the hurdle rate times the incremental capital. Pricing on a different base than marginal capital would not maintain the portfolio minimum return on capital.

2 In a non-financial corporation, the cost of capital is the weighted cost of equity and debt, using the weights of equity and debt. In a bank, income is net of the cost of debt, and the relevant cost of capital is the cost of equity.

This mechanism raises an apparent paradox. Although the pricing rationale seems straightforward, the marginal rule does not imply that the return of a facility, once included in a portfolio, remains in line with the target. Once in the portfolio, the risk contribution of the facility becomes higher than its marginal risk contribution before inclusion in the portfolio, according to the general property of risk contributions. This implies that the return after inclusion be lower than before. Assume, for example, that, once in the portfolio, the capital allocated to the facility is 17, a capital higher than marginal capital. The revenue of 3 does not provide 20% on 17, but 3/17 or 17.65%. Then, one can raise the issue of whether, and how, the earnings of the transaction are sufficient to keep the overall portfolio return at 20%.

Suppose that the capital before including the new facility is 100. After inclusion of the new facility, it becomes 115, since 15 is the incremental capital. If the new facility has a capital allocation of 17 once included in the initial portfolio, it implies that the sum of the capital allocations of the facilities existing before the inclusion of the new one dropped from 100 to 98. This is the only way for the final capital to be 115, since 98 + 17 = 115. This is effectively possible since the new facility increased the diversification of the existing facilities. The additional diversification implies that the capital allocated to existing facilities declines by 17 − 15 = 2. The new facility increases the overall risk incrementally and, simultaneously, diversifies the existing risks. This resolves the pricing paradox.

Once the facility is in the portfolio, its return on allocated capital drops to 17.65% because the allocated capital is 17, higher than incremental capital 15. This decline is offset by a higher return on allocated capital for other facilities, since the capital allocated to these facilities becomes lower. Pricing according to the risk contribution of the new facility once in the final portfolio would ignore the diversification effect on risk contributions of existing facilities. On the other hand, pricing on marginal risk contribution captures both the incremental risk and the increased diversification.

The argument can be generalized as follows. The existing portfolio is P. The required return, pre-tax and net of operating expense, is k, equal to the bank's cost of equity. If the current capital is K, the target operating income is kK. The initial portfolio becomes $P + X$, X being the new transaction. By definition, the marginal capital of the new facility to the portfolio is: $K_{P+X} - K_P$. The rule for pricing is that the required return, r, of the new facility on incremental capital be at least equal to the hurdle rate, k. This implies that the additional revenue from the new facility be at least: $k(K_{P+X} - K_P)$. The return of the portfolio $P + X$ sums up the income on the existing portfolio plus the income on the new facility:

$$kK_P + r(K_{P+X} - K_P) \geq kK_{P+X}$$

This condition implies that $r \geq k$, or that the risk-based return is higher or equal to the hurdle rate.

28.3 RISK PREMIUM AND RISK-BASED PRICING

It is useful to show how the RoRaC ratio is expanded as a function of its various components, and the all-in-spread, r, charged to the customer under risk-based pricing. It is convenient to express all items as percentages of exposure because the required customers' rate is expressed in percentage of the exposure. Calculations can be conducted before tax, because pricing is.

In the following notations, small letters refer to percentages of exposure, except for the required return on capital, k, which is a percentage of capital. The notations are:

Exposure: X.
Asset all-in revenue: r (%).
Risk-free rate: i_f (%)
Cost of debt: i (%), inclusive of the credit spread applicable to the bank.
Allocated debt: D.
Operating costs: oc (%).
Expected loss: el (%).
Allocated capital RC (K), in monetary units.
Required return on capital: k, as a percentage of capital, not exposure.

The target revenues should absorb the cost of debt, the operating expenses and generate a minimum required return k on capital. The RoRaC formula is:

$$\text{RoRaC} = \frac{rX - iD - \text{el}\,X - \text{oc}\,X}{K}$$

In risk-based pricing, the RoRaC ratio should be above the required return on equity capital. A sample numerical calculation of the target rate charged to the customer matching the required return on capital can be found in Chapter 10. The calculation is expanded again below, with general formulas.

The earnings before tax are:

$$\text{EBT} = rX - iD - \text{Operating costs (oc)} - \text{Expected loss (el)} \geq k \text{ capital}$$

The capital charge can be considered as substituted to debt, and the debt is then lower than exposure by the same amount. Substituting in earnings and expressing operating costs and expected loss as percentages of exposure:

$$rX - i(X - K) - (\text{oc} + \text{el})X \geq kK$$

The risk premium for credit risk is the cost of the capital charge. This additional charge for credit risk is the differential between the required return on equity capital and the cost of bank's debt multiplied by capital: $(k - i)K$. The differential $r - i$ represents the all-in-spread in percentage of exposure, which is the difference between the all-in-rate charged to the borrower and the cost of bank's debt:

$$r - i \geq (\text{oc} + \text{el}) + (k - i)K/X$$

An alternate calculation considers that the capital should be invested in risk-free assets; otherwise the capital would be at risk. In this case, the debt becomes equal to exposure and the revenue from investing capital at the risk-free rate should be added to the client's revenues. It can be shown that the required all-in-spread for matching the required return on equity capital is: $(r - i) \geq (\text{oc} + \text{el}) + (k - i_f)K/X$. The risk premium is slightly different because it is calculated over the risk-free rate instead of the interest rate applicable to the bank, including its spread above the risk-free rate.

The risk premium is an excess return required for compensating the capital charge for credit risk. It is proportional to the ratio of capital to exposure and to the excess of cost of equity capital over the risk-free rate. When there is no credit risk, there is no capital charge and expected loss is also zero. The minimum spread over the bank's cost of debt should only offset operating costs.

28.4 SHAREHOLDER VALUE ADDED MEASURES

The SVA is a measure of performance in excess of the cost of equity capital expressed in monetary units.[3] A positive SVA means that the transaction creates value, and is the condition for accepting new facilities. SVA adjusts the revenues with expected loss, then with operating costs, and finally with the cost of the risk-based capital, equal to the hurdle rate k times allocated capital. The SVA definition, when capital is substituted to debt, is:

$$\text{SVA} = (r - i)X - (\text{el} + \text{oc})X - (k - i)K$$

A RoRaC above the hurdle rate implies that the SVA is positive. This can be seen by dividing the SVA by exposure:

$$\text{SVA}/X = (r - i) - (\text{el} + \text{oc}) - (k - i)K/X \geq 0$$

When capital drops to zero, the required earnings are at least equal to the operating costs plus the cost of debt. But very high RoRaC or infinite ratios create confusion in reports. SVA does not have this flaw. If risk is zero, the SVA equals earnings minus operating and debt costs, and remains meaningful.

Negative measures occur when competition does not allow to charge the full risk-based client's rate, or when some facilities are priced at market rates, for large clients, for example. Market spreads of large corporates might not compensate the bank's risk, although that depends on country and geographic areas. Typical spreads to large corporates of high-grade quality in the bond market are lower than 100 basis points. With such low spreads, negative RoRaC might appear. A negative ratio is meaningful, but not very usable. The negative SVA remains meaningful and measures the destruction of value for shareholders.

28.5 RISK-BASED PERFORMANCE, PRICING AND CAPITAL ALLOCATION

The calculations of RoRaC and SVA for the same two-obligor portfolio used in Chapter 27 illustrate the performance measures before and after inclusion in the portfolio.

The two exposures of A and B are respectively 100 and 50, with default probabilities 7% and 5%. It is now assumed that the required return on capital before tax is 20%. For simplification

3 For the definition of SVA, see Uyemura, D. G., Kantor C. C., Pettit J. M. (1996), EVA® for banks: Value creation, risk management, and profitability measurement, [131].

TABLE 28.1 Initial and final portfolio characteristics

	A	*B*	*A + B*
Marginal allocations	25.515	3.214	28.729
Expected loss	7.0	2.5	9.5
Required return on capital	*20%*	*20%*	
Required spread	12.103	3.143	15.246

purposes, the operating costs are ignored. With this portfolio, the effect of pricing based on marginal contributions is shown to maintain the required return on capital, but the return on capital of the facility declines once the allocation of risk is based on the final portfolio.

Since quantiles would not show risk diversification in the example, it is assumed instead that the capital is proportional to loss volatility for illustrating the effect on risk-adjusted performance of a new facility. It is convenient to use contributions instead of capital, since all calculations remain valid if capital is a multiple of risk contributions.

The initial portfolio is A, and the new facility added to the portfolio is A. For the initial portfolio, A alone, the standalone risk is its loss volatility, 25.515. The RoRaC should be 20%. The corresponding target spread for A, S_A, sums up the expected loss, 7, plus the return on its capital: $S_A = 7 + 20\% \times 25.515 = 12.103$. The SVA of A with this spread is zero since the return is exactly equal to the required return on capital: $25.60 - 7 - 20\% \times 93 = 0$.

When adding B, the pricing should be in line with its marginal risk contribution to the loss volatility of the portfolio, 3.214.[4] The required spread for B, S_B, is the sum of expected loss, 2.5, plus 20% times the marginal risk contribution of B, 3.214:

$$S_B = 2.5 + 20\% \times 3.214 = 3.143$$

These calculations are shown in Table 28.1, which also shows the risk contributions and the expected loss of A + B. The loss volatility is determined from the loss distribution, and is also equal to the sum of the initial loss volatility of A and the marginal contribution of B.

The sum of the two spreads is 15.246. For the portfolio A + B, it is easy to check that the RoRaC is exactly 20%:

$$\text{RoRaC}_{A+B} = \frac{15.246 - 9.5}{28.729} = 20\%$$

Once in the portfolio, the risk allocations change but the spreads and expected losses of A and B remain the same since they depend only on their exposures. Table 28.2 recalculates the return on capital of both facilities and the portfolio.

4 From the calculations in Chapter 27.

TABLE 28.2 Return on capital with an incremental facility

	A	*B*	*A* + *B*
Allocation of capital A + B	23.628	5.101	28.729
Spread	12.103	3.143	15.246
Expected loss	7.0	2.5	9.5
Return on capital A + B	21.6%	12.6%	20%

The table shows that the ex post returns on capital of A and B diverge from the marginal returns, but that the return on capital of the portfolio remains the same.

29

Credit Derivatives

Credit derivatives are products used for trading the credit risk of assets independently of the asset themselves.[1] As other derivatives, they serve for hedging and for trading purposes. The most commonly used credit derivatives are credit default swaps (CDSs), which provide protection against the default of an underlying asset. Basket products serve for insuring the risk of portfolios, and they can be used for trading credit risk correlations. From a banking perspective, credit derivatives are very flexible instruments for hedging their risk concentrations or diversifying their portfolio by selling credit protections.

This chapter explains the nature of credit derivatives and their usages for hedging, for taking new exposures or for trading. It also shows how credit derivatives can serve for managing the credit risk of loan portfolios, or credit portfolio management. An example shows how derivative trades can potentially enhance the bank's return on capital.

Contents

1 Useful references include Das, S. (ed.) (2005), Credit Derivatives: CDOs and Structured Credit Products, [51] and Duffie, D., Singleton, K. J. (2003), Credit Risk: Pricing, Measurement and Management, [58].

29.1 DEFINITIONS OF CREDIT DERIVATIVES

Credit derivatives are instruments serving to trade credit risk by isolating the credit risk from the underlying transactions. The products are used for trading the credit risk of assets independently of the reference asset. For banks, the products do not necessitate entering directly into loan arrangements or purchases of bonds, nor do they necessitate prior agreement of the issuer of the asset protected by the instrument. Hence, they are flexible instruments that can be used to rebalance the credit of their portfolios.

CDSs provide a specified payment under default events, and are by far the most commonly used instruments. The underlying asset of which default triggers payment is the underlying asset, typically a bond. When a default event occurs for the underlying asset, the seller of the derivative provides a payment to the party suffering the loss. Credit derivatives are similar to insurance on credit risk, except that they are traded at prices that reflect the credit standing of the underlying asset.

The terminology deserves some attention. Credit derivatives involve a protection seller, a protection buyer and the underlying risk. The seller of a credit derivative is the seller of protection, or insurance, and the buyer of credit risk. The buyer of a credit derivative is the buyer of protection, or insurance, and the seller of credit risk.

The intermediate player making the market is the protection seller. A firm buying of protection does not need to hold the asset in its portfolio. A protection on the default of a loan can be a credit derivative on a traded asset of which credit risk is correlated to the credit exposures to which the buyer is exposed.

29.1.1 Credit Default Swaps

Figure 29.1 summarizes the exchanges of payment for a CDS. The buyer of a CDS pays a recurring premium and receives a payment equal to loss under default if the underlying asset defaults, or is subject to a credit event that qualifies for triggering the CDS. The derivative has a paying leg and a receiving leg, like a swap. For the buyer of a CDS, the paying leg is the premium and the receiving leg is the payment contingent upon default of the underlying asset.

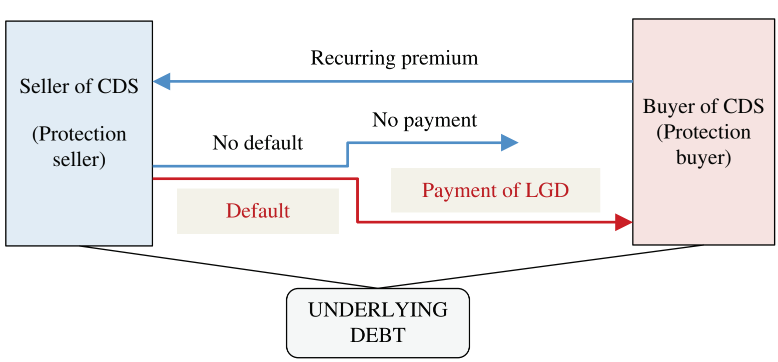


FIGURE 29.1 CDS, protection seller and buyer, and reference asset

CDSs are quoted with the premium in basis points. For example, a CDS on US debt at 64 bps means that the buyer pays $64,000 per year for insuring $10 million of US debt maturing in five years.

CDSs are triggered by default events. In general, credit products can be triggered by credit events that entail a loss of value, a downgrade or a default. For the reference asset, credit events potentially include:

- Payment obligation default;
- Bankruptcy or insolvency event;
- Restructuring or equivalent;
- Rating downgrades beyond a specified threshold;
- Change of credit spread exceeding a specified level.

Termination payments are settlement payments if the credit risk of the underlying actually materializes. When assets are tradable with a reasonable liquidity, prices are a sound basis for calculating the payment under default. The post-default price defines the loss under default, a standard practice of rating agencies for assessing the loss given default (LGD). An alternate option is to use a fixed LGD amount as a percentage of face value of the bond. Delivery is in cash, or with securities considered as acceptable substitutes of the defaulted asset.

There are instruments other than CDSs, such as basket derivatives, which provide protection against the defaults within a basket or portfolio.

29.1.2 Correlation Products

A “first-to-default” derivative provides a payment if any single asset within the basket is the first to default. First-to-default products are considered as instruments for trading the correlation between portfolios of credit products. Diversification reduces the risk of the basket but it does not reduce the risk of first-to-default derivatives and, in fact, the reverse happens. Moreover, the probability of triggering a basket derivative is higher than for a single asset. The rationale is based on joint survival probabilities.

The probability of zero default within the basket is the joint probability that all survive. This joint survival probability is higher when correlation is positive. The probability that one asset, or more, defaults is the complement of the joint survival probability. If correlation increases, the basket is less diversified, the joint survival probability increases and the probability that at least one default is lower. If the correlation is lower, the basket is more diversified, its joint survival probability is lower and the probability that at least one defaults is higher. The credit risk of the basket varies inversely with the joint survival probability, and hence inversely with correlation.

For example, with two independent assets of default probabilities 1% and 2%, the joint default probability is the product of the two default probabilities. The survival probability of both is the joint survival probability, or $99\% \times 98\% = 97.02\%$. The probability that any one of the two assets defaults is simply $1 - 97.02\% = 2.98\%$. This is higher than any one of the two single default probabilities, which shows that the risk of first-to-default baskets is greater than for single assets. If the credit correlation is positive, the joint survival probability increases above 97.02% and the probability that at least one defaults becomes lower than 2.98%. The risk of the product is lower when the correlation of the portfolio increases.

Because the risk of default products is inversely related to correlation, the seller of such products can be seen as selling default correlation, or being "short" correlation. Conversely, the buyer is "long" correlation.

Other N-to-default products provide payments when at least N assets default. They are triggered when several assets default, such as two, three or more assets of a portfolio. When N defaults are reached, the seller pays the loss to the buyer.

Basket products have benefits for both protection buyer and seller. From the buyer's standpoint, there might be very large exposures of high-grade obligors to hedge. Insuring them together in a portfolio makes sense. From the protection seller's standpoint, there is some comfort in finding high-grade exposures in the basket. When pooled together into a first-to-default derivative the credit quality of the derivative is lower than that of any one high-grade asset. Nevertheless, it remains an eligible investment in low-risk assets with an enhanced return compared to the returns of individual high-grade assets.

29.2 USAGE OF CREDIT DERIVATIVES

Credit derivatives are used for transferring or customizing credit risks, taking synthetic exposures or replicating risky assets.

Banks are protection seekers. A bank having a concentration in an industry has a number of choices: capping exposures and diversifying the loan portfolio in other industries are standard practices. Credit derivatives offer the additional option of buying a direct protection for some of the assets held.

Credit derivatives serve to customize exposures. Users can leverage the actual exposure to an underlying asset by using a notional different from the actual asset value. Lenders can reduce their excess exposures. Investors in credit derivatives can leverage their investments and customize their expected return.

A bank can also diversify some the risk of its portfolio by selling CDSs. CDSs sold replicate assets and are synthetic exposures. Synthetic exposures are unfunded, since there is no need to invest cash to get a credit exposure. However, a sold CDS has a capital charge for credit risk.

Investors look at credit derivatives as providing return enhancements when they perceive credit events as having a low probability. Credit derivatives also allow taking exposures to which they have no access. Replicating a risky cash asset is achieved by investing in a risk-free asset and selling a CDS. The recurring fee from the CDS provides the compensation for credit risk, on top of the risk-free rate.

Credit derivatives are used for trading different expectations on the same risks. Credit spreads embed expectations of default probabilities and recoveries. In theory, the spread should be the loss rate, under risk-neutral probabilities. If different parties have different views on probabilities of default or on future recoveries, they can trade these views through CDSs.

29.3 CREDIT PORTFOLIO MANAGEMENT

Credit portfolio management is closely related to the business model of banks. Under the "buy and hold" business model, the bank originates loans and keeps them in its balance sheet until maturity (or default). Under the "originate and distribute" business model, the bank originates loans but sells their risk, thereby freeing up some capital for originating new business. The main techniques for managing the credit portfolio are securitizations and credit risk trading. Securitizations have been used for a long time for managing credit portfolios. Traditional securitizations involve the sale of the assets to a vehicle financed in the market. Securitizations are discussed in Chapter 30. Credit derivatives further facilitated this task by allowing the sale of the credit risk of portfolios without selling the assets.

Traditional banking builds up on relationship management, building a relationship with clients and selling customized solutions. Relationship banking creates a customer base, but it also tends to create concentration and specialization. By contrast, an active credit portfolio management refers to actions for reshaping and enhancing the risk–return profile of the credit portfolio of the bank.

There are many incentives for developing active portfolio management. A primary motivation is developing new business by offloading credit risk onto the markets, and freeing capital for expanding activities. This was the main purpose of traditional securitizations. Other transactions can be used to take advantage of price discrepancies of the same credit risk, on the balance sheet of a bank and in the market.

Credit portfolio management can be seen as an equivalent of asset-liability management for credit risk when loans are transferred from the originating units to a credit portfolio unit. Transfers within the bank should be done at mark-to-market values, since the portfolio management unit operates in capital markets. The portfolio management unit can then be a profit center and manages the portfolio with its own goals.

29.4 TRADING CREDIT RISK AND THE RETURN ON CAPITAL OF A BANK

By trading credit risk, banks also alter their credit exposure, their capital charge and their revenues or costs. If properly adapted, the economics of CDS transactions might enhance the return on capital. The capital base depends on exposures hedged or gained from the CDSs, and the return depends on CDS premium paid for hedging, or the premium received from selling credit protection. Regulatory arbitrage is a classic example.

The purpose of regulatory arbitrage is to take advantage of price discrepancies for the same credit risk, on balance sheet and in the market. In general, the capital-based price of credit risk on the balance sheet and the market price of the same credit risk do not match. By taking advantage of such differences, it becomes feasible to enhance the return on capital. A similar type of arbitrage applies to securitizations, an example of which is provided in Chapter 30. The mechanism is illustrated for CDS with a numerical example detailing the economics of the transaction.

When a bank hedges a loan by buying a CDS[2] from another bank, its capital charge is lower if the risk weight of the loan is higher than the risk weight of the bank selling the CDS. The discrepancy can be used to enhance the return on capital. Assume that the risk weight of corporate borrowers is 100% and only 20% for a banking counterparty of good credit quality. The capital charge of a corporate loan is 8% and becomes 20% of 8%, or 1.6%, if the loan is hedged with the bank.

2 The CDS does not have the loan as underlying asset but a traded asset, which is correlated with the risk of the corporate borrower.

The lending bank buys protection if the transaction enhances its return on capital. The pre-tax earnings decrease by the fee paid for the credit protection bought and there is also an incremental increase of the cost of debt due to the capital saving from the CDS. The net effect on the return on capital depends on the relative magnitudes of earning and capital variations.

With a capital ratio of 8%, the funding of the loan is 92% debt and the rest is capital. Suppose that the spread on loans over funding cost is 100 bps, with a portfolio yield of 6% and bank's debt costing 5%. The earnings from the existing portfolio depend on how much debt and capital back the loans. If capital is 8% of the loans,[3] the debt size is 92 and its cost is $5\% \times 92 = 4.6$. Before any derivative transaction, the earnings pre-tax are: $6 - 4.6 = 1.40$. The return on capital is: $1.40/8 = 17.50\%$.

When the bank buys protection, it benefits from the decline of the capital. Using the low risk weight of the seller of the CDS of 20%, the capital charge declines to 1.6. If the premium paid for hedging the loan with a CDS is 60 bps, the direct effect on earnings is a decline of 0.6 for a notional of the CDS of 100. But earnings also decline further because the debt backing the loans increases by the amount of capital saved. The additional debt is 6.4, the difference between the original capital and the capital after hedging. The cost of incremental debt is: $5\% \times 6.4 = 0.32$. With the CDS hedge, the earnings always decline because of the premium paid on a CDS and on the incremental cost of debt.

The new earnings become the revenue from loans, unchanged, minus the premium paid and minus the additional cost of debt. They become: $6 - 5 - 0.32 - 0.6 = 0.480$. The new return of capital after hedging is: $0.48/1.60 = 30\%$. Hedging has leveraged the return on capital from 17.50% to 30%.

The condition for a positive leverage effect can be made explicit, using the following notations:

r: Return of portfolio.
i: Cost of bank's debt.
k_1: Capital before hedging, in percentage of portfolio size.
k_2: Capital after hedging with CDS, in percentage of portfolio size.
p: CDS premium, in % of notional hedged.

The initial earnings in percentage of the size of the loan portfolio are simply: $r - (1 - k_1)i$. The final earnings in percentage of the size of the portfolio are: $r - (1 - k_2)i - p$. The returns on capital, before hedging and after hedging, are the ratios of these earnings to k_1 and k_2, respectively. The new return on capital is higher than the initial return on capital when:[4]

$$\frac{k_2}{k_1} < 1 - \frac{p}{r - i}$$

In this example, the ratio of final capital to initial capital is: $k_2/k_1 = 1.6\%/8\% = 0.2$. This ratio is lower than the above expression, as: $1 - p/(r - i) = 1 - 0.6\%/(6\% - 5\%) = 0.4$.

The bank can also sell a CDS for minimizing its hedging cost. The capital charge for a sold credit derivative is identical to the capital charge of a direct exposure to a corporation. The seller earns the recurring fee from the CDS sold but has an additional exposure, and capital

3 That is on the risk-weighted loans, using a risk weight of 100% and LGD 100%.

4 This inequality is obtained by writing that: $[r - (1 - k_1)i]/k_1 < [r - (1 - k_2)\mathrm{i} - p]/k_2$.

charge, for credit risk. By entering into such trades, the bank might enhance its risk-adjusted return.

Such synthetic exposure is unfunded, hence there is no impact on bank's debt, but there is the additional capital charge. Let us assume that the capital charge is k_3 and the premium on CDS sold is p. The return on capital from CDS sold is simply: p/k_3. For example, assume a capital charge of 8% of notional and a premium of CDS sold of $p = 1\%$. The return on capital is: $1\%/8\% = 12.5\%$. This is less than the original return on capital. However, by combining CDSs purchased and sold, the bank mitigates its cost and could enhance its initial return. This calculation does not consider any diversification effect from the new synthetic exposure.

Credit derivatives can also serve for other arbitrage between returns based on ratings and economic returns. A bank can assemble a portfolio of credit derivatives, sell the risk to investors and earn the corresponding market premium. The investors buy the risk based on the assessment of the same risk by rating agencies. The bank can potentially earn a positive differential between the market price of the risks sold in the market and the rating-based spread paid to investors. The mechanism is similar in synthetic securitizations, as explained in Chapter 30.

30

Securitizations

Securitization is a mechanism that allows assets to be sold on the capital markets, using the cash flows from assets for compensating investors in asset-backed notes. The securitized assets are removed from the balance sheets of banks, which frees up banks' resources and capital for expanding credit. Securitizations have been a core process of the "originate and distribute" business model, whereby banks originate credit and distribute the financing of their loans across investors in the capital markets.

This chapter details the motivations and principles of securitizations. An example is used for explaining the economics of the transaction from the standpoint of a bank selling its credit risk into the market.

Contents

30.1 THE MOTIVATIONS OF SECURITIZATIONS

Historically, and until the 2008 financial crisis, banks progressively shifted from the "originate-and-hold" business model to the "originate-and-distribute" business model.

Traditional lending consists of lending and keeping the loans in the banking portfolio until maturity. One drawback of the classical practice is that assets held on balance sheet require a financing that could potentially be used elsewhere if they could be sold. Loans withheld in the balance sheet also freeze the capital required by regulators, thereby excluding origination of new loans if the bank has capital constraints. Securitizations were a major source of financing for banks and a way of saving capital.

The securitizations allowed banks to turn around the constraints of the “originate-and-hold” policy and to keep developing their loan business. The mechanism facilitated the financing of growing economies as the same capital base could be reused for new loans and because securitizations became a major source of funding for banks.

In a securitization, a pool of assets, such as a pool of loans, is sold to a special purpose entity (SPE), which issues bonds sold to investors for financing the acquisition. Investors look for credit spreads of issued bonds and base their investment decisions on the ratings assigned by rating agencies to such loan-backed bonds. Investors care primarily about the credit quality of bonds, measured by credit ratings, rather than the composition of the pool of loans, over which they have no direct information, making the role of rating agencies critical in the process. An additional benefit when investing in SPE securities is that the credit spreads of notes issued by securitization vehicles have been more attractive than those of comparable corporate bonds with the same rating.

Securitizations can take advantage of discrepancies between the cost of financing on-balance sheet and the cost of financing the same assets in the market. Such discrepancies arise, for example, from regulatory capital charges standardized across banks, which have no reason to match the specific risk of the banks’ portfolios. Discrepancies between the costs of financing on-balance sheet and in the market can also arise from differing perceptions of the same risk by banks and by credit rating agencies.

The securitization process was disrupted at the time of the financial crisis as investors perceived the securitization vehicles as disseminating toxic risks throughout the financial system.[1] In theory, the credit ratings of rating agencies should provide reliable information about this risk. The practice is that rating agencies were not well equipped to quantify such risks. Once it became obvious to all that the risk was higher than expected, it was too late. All that rating agencies could do was accelerate downgrades, which confirmed the fears of investors and created a high risk aversion for the perceived risk of these products. As downgrades extended, the market value of bonds issued by securitization vehicles lost value and triggered losses to investors.

The consequence is that they disappeared as the crisis unfolded. Their virtual disappearance dried up a major source of funds and left banks with the pressure of capital constraints. Securitizations have virtually stopped since the crisis. Still, the securitization mechanism relies on sound economics and should reappear perhaps in new forms, and under more conservative approaches for assessing the risk of asset-backed bonds.

30.2 THE PRINCIPLES OF SECURITIZATIONS

The principle of securitization is simple. Financial assets generating cash flows, interest and principal repayments, can be sold to investors as long as the cash flows from assets are sufficient to provide them with an adequate compensation. The assets are sold to special purpose vehicles who issue bonds backed by the assets and acquired by capital market investors. Often the seller is the bank who originated the loans. Investors are compensated by

1 See Longstaff, F. A. (2010), The subprime credit crisis and contagion in financial markets, [94].

the cash flows from assets. The SPE issues a series of notes of various risks, or tranches, which are acquired by investors. Usually, several notes are issued by the same vehicle. The proceeds of these issues serve for acquiring the securitized assets. The pool of loans makes up the asset side of the SPE. In a non-recourse sale of loans to the SPE, the investors have no claim on the seller of loans. The credit risk is not borne by the seller of assets but by the investors in the SPE. Such effective transfer of credit risk allows the bank to benefit from a capital relief and frees up banks' capital for other usages.

30.2.1 The Mechanism of Securitizations

A securitization is feasible only if the investors in the notes issued by the SPE receive a return in line with the risk of their investments. Rating agencies provide the required risk assessment of these notes, which allows the rated notes to become tradable securities. The cost of financing the assets in the securitization vehicle is the weighted average return of SPE notes required by investors. The return from securitized assets should, at least, match this weighted cost of financing.

Any asset, or pool of assets, that complies with such economics can be securitized. Originally, pools of mortgages were the first assets to be securitized. Subsequently, a wide variety of other assets followed. They include credit cards receivables, leasing receivables, bonds of various credit standing and industrial or commercial corporate loans. Depending on the nature of assets, different names were used such as CDO (collateralized debt obligation), CBO (collateralized bond obligation) and CLO (collateralized loan obligation).

The pool of assets securitized can be static, replenished periodically or managed. Long-term loans, such as mortgages, form static pools. Short-term assets, such as credit card receivables, need to be periodically replenished to ensure a longer life of the securitization vehicle. Traded assets form a portfolio that can be actively managed during the life of the transaction, subject to the terms of contract governing the SPE. In theory, any asset generating cash flows can be sold to investors through an SPE. For example, an office building generates rents, which can be used to compensate market investors, and it can be securitized.

Cash securitizations are transactions where assets are effectively sold to an SPE. Synthetic securitizations use credit derivatives for selling only the credit risk of a pool of assets without selling the underlying assets. In this second class of transactions, the SPE is the seller of the credit derivative, hence the buyer of the risk, and the bank is the buyer of protection and seller of the risk. The proceeds from investors are invested in risk-free assets. The bank pays an insurance premium to the SPE, which is added to the risk-free return from the SPE investment in Treasuries, and allows the SPE to reconstruct a credit risky return.

30.2.2 Organization of Securitizations

Securitizations involve several players. The bank sells the assets to the securitization vehicle, which is bankruptcy remote. The SPE issues a series of notes, differing by their risk, of which credit standing is assessed by rating agencies for making them eligible investments for investors. The "arranger", typically an investment bank, structures the transaction, and is usually an investment bank. The "servicer", usually the bank who sold the loans, takes care of

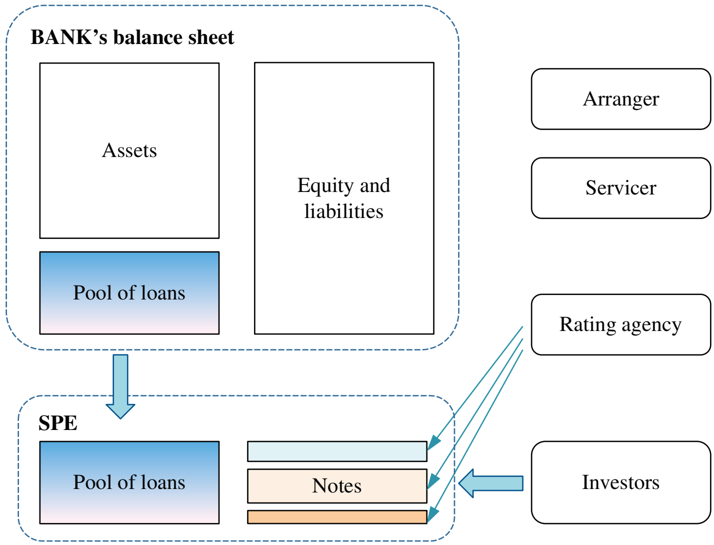


FIGURE 30.1 Structuring a securitization

day-to-day relations with the borrowers whose loans have been securitized, during the life of the transaction. Figure 30.1 provides a view of the financial players involved in a typical transaction.

30.2.3 The Structuring of Securitizations

There are many mechanisms used for backing the notes issued by a vehicle with the assets of the vehicle. In a pass-through vehicle, the cash flows of the assets are passed on to the notes issued by the vehicle. If several notes are issued by the same vehicle, each class of notes can be backed by a dedicated subpool of assets. In a pure pass-through mechanism, the risk of each note is identical to the risk of the pool of assets.

Other securitizations use a credit enhancement mechanism for mitigating the risk of notes issued. Investors can be protected from the first losses of the portfolio securitized by oversizing the size of the pool relative to what is needed to service the claims of investors. Since only a fraction of the pool is sufficient to compensate these investors, the overcollateralization of notes provides a safety cushion against adverse deviations of the flows generated by the pool. For instance, an oversized pool of assets would generate a cash flow of 100, when only 80 are necessary to compensate investors. The risk of loss for investors materializes only when the actual cash flow goes under 80.

Modern securitizations relied on a structuring of the notes issued by the SPE. The SPE does not issue a single class of bond that would bear the entire risk of the portfolio of assets securitized. Instead, it issues several types of notes of different seniority levels. A specific feature of a securitization is that the default of one class of notes, such as the riskiest notes, does not trigger the default of other more senior notes of the SPE. Bonds backed by assets are called "notes" or "structured notes".

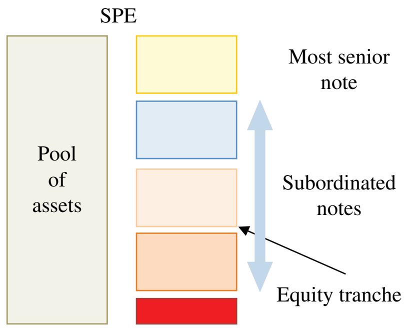


FIGURE 30.2 Structure of notes issued by and SPE

The variety of notes issued allows customizing the risk of these various issues to the profiles of potential investors. Risk-averse investors buy the notes with the lowest credit risk, while some investors are attracted by the higher spreads of riskier notes. Structuring consists of defining how many classes of notes should be issued, the size and the risk of each one.

The mechanism has a number of benefits. It differentiates the risks of the various classes of notes issued by the SPE, while simultaneously ensuring a protection for the more senior notes. The credit risk protection of senior investors results from the loss protection provided by subordinated notes. The process also implies that the risk of the issued notes is disconnected from the risk of the securitized portfolio.

The difference between the expected return of assets and the cost of financing the vehicle is the "excess spread". The cost of financing the SPE is the weighted average of the required returns by investors, the weights being those of each note. By definition, the excess spread does not belong to the investors. It can be released to the seller of assets over the life of the transactions or at maturity.

The structuring of the transaction consists of defining the sizes of the various notes issued and their seniority levels (Figure 30.2). The subordination level of a note is defined by its attachment point and its thickness, or size. The attachment point of a particular note defines the size of all subordinated notes ranking below. The higher this "safety cushion", the lower the risk of the note protected by more subordinated notes. The most subordinated tranche is the equity tranche, which concentrates the first losses of the portfolio and has no rating.

Either the seller of assets, or a third party acting as a "credit enhancer", holds the last tranche. If the seller keeps the equity tranche, it is held on the asset side for a value of zero since the portfolio losses will entirely absorb this tranche. On the liability side, a cash account is dedicated to the tranche and debited when losses hit the equity tranche. If the equity tranche is sold, the investor benefits from an upfront premium and a higher spread than other tranches, as a compensation for holding this highest risk. The return depends on how long the equity tranche survives, and hence on the timing of defaults of the pool of assets.

Agencies rate notes according to their risks. Senior notes can be investment grade because the likelihood that the pool of assets losses exceeds the safety cushion provided by the subordinated notes is near zero for the highest grades. For example, if the size of subordinated notes is 30% of the portfolio, the probability of losing 30% of the portfolio is extremely small. A senior note can therefore be rated AAA. When moving down the scale of seniority, the risk of notes increases as they benefit from a much lower loss franchise.

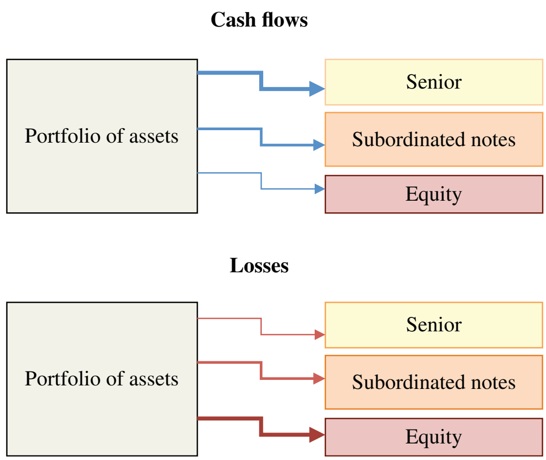


FIGURE 30.3 Waterfall of cash flows and of losses

The risk of a particular tranche, including the most senior tranches, depends on how badly the subordinated tranches are hit by losses. In extreme cases, such as in the financial crisis of 2008, the magnitude of losses absorbed a significant fraction of the subordinated tranches, making the upper tranches riskier. The upper tranches were downgraded and their value declined, triggering losses for the holders. In other words, the investors in senior tranches can be hit by adverse migrations, even if no default impacts the tranche in which they have invested.

The structure routes the cash flows generated by the pool of assets to investors using priority rules based on the seniority level, first to senior notes and last to the equity tranche. The mechanism is the "waterfall" of cash flows, describing how cash flows cascade from one tranche to the next. The "waterfall of losses" follows symmetrical paths. The losses first hit the equity tranche, then the subordinated notes and finally the senior notes. Figure 30.3 shows the waterfalls of cash flows and losses.

30.3 THE ECONOMICS OF SECURITIZATION

This section addresses cash securitizations, the purpose of which is to free up capital for expanding business volume. The main economic issues for the seller of loans are to find out how the bank's earnings and the bank's return on capital are impacted by a securitization. Ideally, a securitization should improve the return on capital of the bank. The capital base is lower to the extent that the risk is transferred to investors, but the securitization also impacts earnings, eventually adversely, and the overall effect on the return on capital is ambiguous.

The expected return on assets of the SPE should be at least equal to the return required by investors in the SPE notes. The difference is the excess spread, which potentially belongs to the seller of the loans. The spread depends on a number of factors:

- The yield of assets (loans) sold by the bank to the SPE;
- The market spreads applicable to the notes issued by the SPE;
- The direct costs of setting up a securitization.

A first criterion for assessing the potential benefit of a securitization for the seller is the differential between the cost of financing through securitization and the on-balance sheet cost of financing. A securitization dissociates the quality of the original flows generated by the pool of assets from the quality of the flows to investors, since the original flows are routed to notes according to priority rules. A consequence is that the cost of financing through the securitization vehicle should be different from the cost of funding on balance sheet.

When assets are held on-balance sheet, their cost of financing is the weighted average of cost of equity and debt. When held as assets of the SPE, it is the weighted average of the costs of the various notes issued by the vehicle. An example details the economics of a standard transaction. Its purpose is to assess the impact on return on capital of a standard securitization.

30.3.1 Analysis of a Securitization Transaction

For detailing the economics of a transaction, a number of – simplified – inputs are required for characterizing the pool of loans securitized and the cost of financing on balance sheet and in the securitization. They are as follows.

The portfolio of loans of a bank has an annual expected return of 10.20%. This return is net of expected loss and net of operating costs for servicing the loans. The risk-free rate is 9%. The debt on the balance sheet of the bank is rated A, corresponding to a spread of 1% over the risk-free rate. The assets have a risk weight of 100% and the capital ratio is 8%. The required return on equity before tax is 20%. The bank considers securitizing an amount of loans, of size $1,000 million, from a total loan portfolio, the size of which is $10,000 million.

In the securitization, the pool of loans is financed with notes sold to investors in the SPE. There are two classes of tranches issued to investors. The subordinated tranche represents 25% of the amount securitized. The credit spread required over the 9% risk-free rate is 2% (200 bps). The senior tranche represents 75% of the pool of loans, is investment grade, is rated AA and its credit spread is 0.50% above the risk-free rate.

The direct costs of the securitization include a legal cost and the servicing cost. The servicing of loans is born by the bank that has the relationship with the clients, and its cost is netted from the return of loans. The legal cost, annualized, is 0.15% of the face value of securitized loans. These costs are senior costs: they should be paid before investors are compensated.

30.3.2 The Costs of Funding On-Balance Sheet and through Securitization

The cost of financing loans on balance sheet is the weighted average cost (WAC) of the bank. With the above assumptions, the calculation is straightforward, given the bank's rating and the corresponding cost of bank's debt. The weights of equity and debt are 8% and 92%, respectively. The required return on equity is given and the cost of debt results from the bank's credit rating. The calculations are shown in Table 30.1.

The bank's WAC is 10.8%. This cost of financing is higher than the net yield of loans. The consequence is that the effective return on capital is lower than the minimum of 20% required. This can be checked directly by calculating the earnings of the bank with the current portfolio. Given expected return on assets, the book return on capital is 12.50%, lower than the required 20% (Table 30.2).

TABLE 30.1 WAC calculations, on-balance sheet

		Cost BS
Risk-free rate		9%
Rating		A
Spread		1.00%
Interest cost		10.00%
	Weights	*Cost*
Debt	92%	10.00%
Equity	8%	20.00%
Weighted cost of capital		10.800%

TABLE 30.2 Effective return on capital of the bank

	Balances	*Yield*	*Rev. & costs*
Revenue from loans	10,000	10.20%	1,020
Bank's debt	9,200	10.00%	920
Bank's capital	800		
Bank's earnings, before securitization			100
Bank's effective return on capital			12.50%

The cost of financing through securitization is the WAC of the notes issued, using the credit spreads matching the ratings of the two classes of notes. The direct operating cost of setting up the securitization vehicle is not included at this stage. The cost of financing loans through securitization is 9.825% (Table 30.3).

TABLE 30.3 WAC of securitization

Cost senior	*Cost junior*
9%	9%
AA	Ba
0.50%	1.80%
9.50%	10.80%
Weights	*Cost*
75.00%	9.50%
25.00%	10.80%
WAC	9.825%

It is lower than on-balance sheet (10.80%). This cost is also lower than the net yield of loans. In general, the cost of financing through an SPE depends on the structuring of notes (the weights assigned to each one of them) and the credit spreads prevailing in the market. The cost of financing through securitization might or not be lower than on-balance sheet, since it depends on market spreads, which can narrow or widen depending on prevailing conditions.

The full cost of financing through securitization includes the direct operating costs of setting up the SPE and the servicing cost. The all-in cost of financing with securitization is the sum of the required return by investors and of direct costs, 0.15% or 9.975%. This all-in cost of financing with securitization is still lower than the cost of financing on-balance sheet. The all-in cost of securitization is also lower than the loan yield. The securitization generates a positive excess spread, calculated as the algebraic difference between the net yield of loans and the all-in cost of securitization. In this example, the excess spread is: $10.100\% - 9.975\% = 0.225\%$. With a securitized portfolio of 1,000, the value of this spread is 2.25 (in million).

This excess revenue does not belong to investors and goes back to the bank as additional earnings. The effective usage of the excess spread depends on the covenants governing the securitization. The excess spread can remain trapped in the SPE as an additional safety cushion under adverse conditions, before being released.

30.3.3 The Bank's Earnings and Return on Capital after Securitization

Once the loan is sold to the SPE, the bank is left with a smaller portfolio, $10,000 - 1,000 = 9,000$. The capital allocated to this portfolio is lower: $8\% \times 9,000 = 720$. It is assumed that the capital freed is allocated to other new activities. The debt financing the portfolio is now $9,000 - 720 = 8,280$.[2] The bank loses the earnings of the securitized portfolio but earns the excess spread. The bank's earnings after securitization and the new return on capital are calculated in Table 30.4, when the balance sheet size declined by 1,000, the size of securitized loans.

TABLE 30.4 New earnings and return on capital after securitization

	Balances	*Yield*	*Rev. & costs*
Excess spread	1,000	0.225%	2.25
Earnings after securitization			
Bank's revenue from loans	9,000	10.20%	918.00
Excess spread			2.25
Bank's debt	8,280	10.00%	828.00
Bank's capital	720		
Bank's earnings after securitization			92.25
Bank's ROC after securitization			12.81%

2 We assume that the capital is not invested risk free, which implies and that debt is allocated as a complement to capital.

The new earnings of the bank are lower than before securitization because the revenues from the excess spread of the SPE do not offset the lost earnings from the portfolio. The bank lost earnings of 10 but gained 2.25 excess spread, with final earnings of 92.25. The earnings are lower than the original 100, when the loans are on-balance sheet, as a net effect of the lower size of the bank's portfolio and of the excess spread.

The variation of earnings can also be derived from the differential cost of financing the securitized portfolio. The initial cost of financing is the cost of bank's debt, weighted 92%. The final cost of financing is the cost of the SPE notes, weighted 100%. The differential earnings, before and after securitization, is the difference between these two costs:

$$(92\% \times 10\% - 100\% \times 9.975\%) \times 1,000 = -7.75$$

The difference matches the variation of bank's earnings before and after securitization, or: $92.25 - 100 = -7.75$.

The new return on capital is 12.81%, higher than before securitization, because the capital base is now lower. The securitization enhances the return on capital in this example. It is interesting to make explicit the conditions of such enhancement.

The return on capital of the bank's portfolio, ignoring excess spread, is exactly 12.5%. As long as yields and weights remain identical, the return on capital is independent of the scale of operations. If the bank securitizes a fraction of its portfolio, the return on capital of the remaining portfolio is the same: 12.5%. But the bank gets additional earnings equal to the positive excess spread from the SPE. The new return on capital can be decomposed into the initial return on capital plus an additional return – on the same capital – resulting from the excess spread. The capital after securitization is 720 and the excess spread 2.25. The incremental return on capital is $2.25/720 = 0.312\%$. This is exactly the difference between the original return on capital, 12.50%, and the final return on capital of 12.812%. This finding shows that a necessary and sufficient condition for the return on capital to be improved is that the excess spread be positive.

If capital is inversely proportional to the amount securitized, and if securitization improves the return on capital, it is tempting to increase the amount securitized. The resulting improvement on return of bank's capital is the "leverage effect of securitization". Leverage is positive as long as the all-in cost of financing through securitization remains fixed. The higher the amount securitized, the higher the final return on capital after securitization. Obviously, there are limits to such leverage effect. It is unlikely that the return on asset remains constant when the proportion of the total portfolio securitized increases too much. If the credit quality of securitized assets declines, their expected yield declines, the junior tranche should be larger, its spread increases and the cost of financing through securitization should also increase. The lower quality of new assets securitized sets up a practical limit to the amount securitized.[3]

30.3.4 The Pricing of Assets Sold to the SPE

An alternative way for the seller to capture upfront the excess spread is to sell assets at a price such that the return on assets acquired by the SPE matches the required return for investors.

3 This might not apply to securitizations of short-term assets, such as credit cards, since the rapid amortization of such loans makes it necessary to periodically replenish the pool of assets with new short-term loans, of which quality is not necessarily downgraded.

The securitized loans can be sold at a higher price than face value, given that they yield more than what is required to absorb senior costs and compensate investors.

Suppose, for example, that the duration of loans is two years. The change of value resulting from a variation of the discount rate is proportional to the duration and to the variation of the discount rate. If the discount rate is exactly 10.20%, the loan yield, the value equals face value. If the discount rate is lower, 9.975%, their value is higher than face value. The new value can be approximated with the duration formula. For a face value of 100%, the new value, V, depends on duration and the difference of discount rates according to:

$$(V - 100\%)/100\% = -\text{Duration}\,(r - y)$$

In this formula, the duration is 2, the original yield, y, is 10.20% and the final yield, r, is 9.975%. The fair value of loans, V, is:

$$V = 100\%[1 - 2(9.975\% - 10.20\%)] = 100.45\%$$

The price generates a capital gain of 0.45% of the face value of loans sold, or 4.5. Selling at this fair price the loans to the SPE is equivalent to collecting the excess spread as this initial upfront capital gain.

30.4 VARIATIONS IN THE SECURITIZATION SCHEME

There are variations around the basic scheme of standard cash securitizations. For example, the bank might retain a fraction of the retained tranche. In other instances, a synthetic securitization can be used, selling the risk of loans without selling the loans, with credit derivatives.

When the bank retains a fraction of the senior tranche, a similar economic rationale applies. The bank initially loses the earnings of the securitized portfolio. It has an additional cost of debt because the remaining portfolio retained by the bank is now financed with 100% bank's debt, instead of the initial mix of debt and equity. The debt allocated to the retained tranche increases only if the risk is effectively transferred to the SPE, which saves the capital charge on the retained tranche. The bank also earns the excess spread from the new SPE.

The size of the new SPE is scaled down by the amount of the retained tranche. The weights of the senior and junior tranches change in the SPE. If the structuring of the SPE is the same, the senior tranche has a lower size and the junior tranche has a higher size. As a consequence, the WAC of capital of the SPE is higher.

The bank's earnings combine several effects: the lost earnings from the portfolio; the additional cost of bank's debt, weighted 100%, for the retained tranche; and the algebraic excess spread from the SPE, recalculated with the new weights of the senior and junior tranches. The resulting variation of earnings can be also seen as a differential cost of financing, without and with securitization. The securitized portfolio is financed 100% with bank's debt for the retained tranche at the bank's cost of debt, and the rest is financed 100% with the SPE's debt at the new SPE's weighted cost of debt.

Such a structure makes sense if the cost of financing with bank's debt is lower than the cost of SPE's debt. In that case, the on-balance sheet spread for the retained tranche is higher than if the tranche is left with the SPE. Retaining a fraction of the senior tranche makes sense for highly rated banks.

Synthetic securitizations are a different scheme: the entire portfolio is retained on-balance sheet. The bank buys from the SPE a credit derivative that ensures the portfolio in exchange for the recurring premium paid by the bank to the SPE. On the SPE side, investors earn the premium, which is added on top of risk-free rate, from investing risk-free the cash from investors, to reconstruct a yield. The capital allocation for the portfolio ensured by the credit derivative is lower. For making the transaction safe for the bank, the risk-free assets of the SPE can be pledged as collateral for the benefit of the bank. In 2000, the former Bistro structure from JP Morgan was designed along such lines.

The revenue from the portfolio on-balance sheet remains with the bank. The bank's debt should be scaled up to match a lower capital allocated to the portfolio. The bank's earnings, after the synthetic securitization, are lower because of the higher cost of debt and because of the cost of the premium paid to the SPE. The incremental effect depends on the capital saved, on the incremental cost of debt and on the premium paid to the SPE.

30.5 THE RISK OF ASSET-BACKED NOTES

The resiliency of a securitization structure designates its ability to sustain adverse conditions without triggering losses for the issued notes. The methodologies for assessing the credit standing of notes include stress scenarios and credit portfolio models.

All factors that have an influence on the loss rates of the portfolios backing the notes directly affect the credit standing of the notes. For loans, the parameters that influence the potential losses of notes include: the delinquency rate (delays in payment), the charge-off rate (losses due to default), the payment rate (both monthly payments of interest and principal), the recovery rate (both percentage amount and timing of recoveries) and the average yield of the portfolio of loans.

A standard technique for assessing the risk of the notes is to stress-test such factors and find out under which scenarios losses affect the notes. One common stress-test consists of determining the multiples of expected charge-offs that trigger losses for investors. The higher the multiple sustainable without any loss, the higher the rating. Minimum required multiples of charge-off rates for senior notes are, for example, five or six times the expected average charge-off rate. The technique relies on the waterfall model showing how losses cascade to the series of notes, and when.

Credit portfolio models can serve the same purpose of assessing the risk of the notes issued by the SPE. But the technique is more involved, since it requires the generation of the loss distribution of portfolio and its allocation to each note issued.

For generating portfolio loss distributions, standard techniques are used. With the asset value model of default, defaults are triggered whenever the asset value, associated to individual obligors, falls under the threshold matching the default probability assigned to an individual loan. A uniform correlation captures dependencies. Alternatively, the loss distribution, and its timing, can also be generated from the simulations of times to default of individual obligors of the portfolio. The waterfall mechanism allocates the portfolio losses to each tranche.

Losses flow to structured notes according to their seniority level. A note benefits from the protection of all subordinated notes, the size of which determines the loss franchise before hitting more senior notes. If a simulated loss is lower than the size of these subordinated tranches, it is allocated to the subordinated notes and none is allocated to the senior note. If the loss exceeds the size of the subordinated tranche, the excess loss is allocated to the senior note.

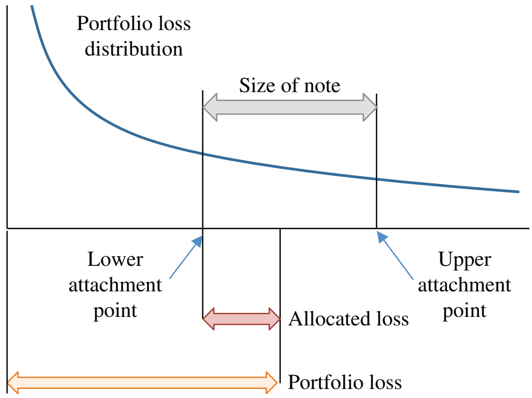


FIGURE 30.4 Loss distributions of the portfolio and of a structured note

The distribution of losses for each note has a lower bound of zero and is capped by its size. The mechanism generates an expected loss and the loss distribution for each tranche. The loss of the note follows the truncated portfolio loss distribution at its lower and higher attachment points. Figure 30.4 shows the truncated portfolio loss distribution relative to some individual tranche.

With securitizations, the probability that some individual tranche defaults is the probability that a minimum number of all assets default. Its risk is comparable to an *N*-to-default basket. Like such baskets, the risk of tranches of securitizations is directly dependent on correlations between defaults. For this reason, they are considered as correlation products.

The rationale can be explained from extreme cases. Consider a portfolio of two assets financed by two tranches, a junior one and a senior one. Moreover, assume the size of each tranche is identical to the size of each asset. For the senior tranche to default, both assets should default. This happens only when the joint probability of default is high. The risk of the senior tranche increases with the correlation.

The junior tranche defaults if only one asset defaults, like a first-to-default basket of two assets. The risk of the junior tranche increases with the probability of a single default. The probability of a single default is one minus the joint survival probability. The joint survival probability increases when the correlation between defaults increases. Accordingly, the risk of the junior tranche declines when the correlation is up.

When the correlation increases, the risk of the senior tranche is up and the risk of the junior tranche is down. Because tranches embed different correlations, they are products whose price is tied to correlation. Such products have prices listed as indexes for standardized tranches. The main advantage of index tranches is that they are standardized. In the ITraxx index, the lowest tranche, known as the equity tranche, absorbs the first 3% of losses on the index due to defaults. The next tranches absorb, respectively, 3% to 6% of the losses, and are insulated from the first 3% of losses.

References

1 Acerbi, C., Tacshe, D. (2002) On the coherence of expected shortfall, *Journal of Banking and Finance*, **26** (7), 1487–1503.
2 Acharya V.V., Viswanathan S. (2011) Leverage, moral hazard, and liquidity, *Journal of Finance*, **66**, 91–138.
3 Acharya, V., Philippon, T., Richardson, M., Roubini, N. (2009) The financial crisis of 2007–2009: Causes and remedies, Financial Markets, *Institutions & Instruments*, **18** (2), 89–137.
4 Adrian, T., Shin, H-S. (2010) Liquidity and leverage, *Journal of Financial Intermediation*, **19**, 418–437.
5 Aebi, V., Sabato, G., Schmid, M. (2012) Risk management, corporate governance, and bank performance in the financial crisis, *Journal of Banking & Finance*, **36**, 3213–3226.
6 Allen, F., Santomero, A.M. (2001) What do financial intermediaries do? *Journal of Banking & Finance*, **25**, 271–294.
7 Allen, N.B., Herring, R.J., Szegö, G.P. (1995) The role of capital in financial institutions, *Journal of Banking & Finance*, **19**, 393–431.
8 Altman, E.I. (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *The Journal of Finance*, **23** (4), 589–609.
9 Altman, E.I., Caouette, J., Narayanan, P. (1998) *Managing Credit Risk: The Next Great Financial Challenge*, John Wiley & Sons.
10 Altman, E.I., Haldeman, R., Narayanan, P. (1977) ZETA analysis: A new model to identify bankruptcy risk of corporation, *Journal of Banking and Finance*, 29–55.
11 Altman, E.I., Sabato, G. (2007) Modelling credit risk for SMEs: Evidence from the US market, *Abacus*, **43** (3), 332–357.
12 Altman, E.I., Saunders, A. (1997) Credit risk measurement: Developments over the last 20 years, *Journal of Banking & Finance*, **21** (11), 1721–1742.
13 Amato, J.D., Remolona, E.M. (2003) The credit spread puzzle, *BIS Quarterly Review*, 51–63.
14 Arnold, B., Borio, C., Ellis, L., Moshirian, F. (2012) Systemic risk, macroprudential policy frameworks, monitoring financial systems and the evolution of capital adequacy, *Journal of Banking & Finance*, **36**, 3125–3132.
15 Artzner, P., Delbaen, F., Eber, J.M., Heath, D. (1997) Thinking coherently, *Risk*, **10**, November, 68–71.
16 Asarnow, E., Edwards, D. (1995) Measuring loss on defaulted bank loans: A 24-year study, *The Journal of Commercial Lending*, **10** (2), 11–23.
17 Basel Committee on Banking Supervision (1988) International convergence of capital measurement and capital standards, Bank for International Settlements, BCBS publications.
18 Basel Committee on Banking Supervision (1994) Risk management guidelines for derivatives, Bank for International Settlements, BCBS publications.

19 Basel Committee on Banking Supervision (1996, updated 2005) Amendment to the capital accord to incorporate market risk, Bank for International Settlements, BCBS publications.

20 Basel Committee on Banking Supervision (2001) Principles for the management and supervision of interest rate risk, Bank for International Settlements, BCBS publications.

21 Basel Committee on Banking Supervision (2006) International convergence of capital measurement and capital standards: A revised framework – Comprehensive version, Bank for International Settlements, BCBS publications.

22 Basel Committee on Banking Supervision (2008) Principles for sound liquidity risk management and supervision, Bank for International Settlements, BCBS publications.

23 Basel Committee on Banking Supervision (2009) Guidelines for computing capital for incremental risk in the trading book, Bank for International Settlements, BCBS publications.

24 Basel Committee on Banking Supervision (2010) Basel III: A global regulatory framework for more resilient banks and banking systems, Bank for International Settlements, BCBS publications.

25 Basel Committee on Banking Supervision (2011) Revisions to the Basel II market risk framework, Bank for International Settlements, BCBS publications.

26 Basel Committee on Banking Supervision (2013) Basel III: The liquidity coverage ratio and liquidity risk monitoring tools, Bank for International Settlements, BCBS publications.

27 Basel Committee on Banking Supervision (2013) Fundamental review of the trading book: A revised market risk framework, Bank for International Settlements, BCBS publications.

28 Basel Committee on Banking Supervision (2014) The standardised approach for measuring counterparty credit risk exposures, Bank for International Settlements, BCBS publications.

29 Berndt, A., Douglas, R., Duffie, D., Ferguson, M., Schranz, D. (2008) *Measuring default risk premia from default swap rates and EDFs*, Tepper School of Business, 49.

30 Bierwag, G. (1987) *Duration Analysis – Managing Interest Rate Risk*, Ballinger Publishing Company, Cambridge, Massachusetts.

31 Black, F., Derman, E., Toy, W. (1990) A one-factor model of interest rates and its application to Treasury bond options, *Financial Analysts Journal*, **46** (1), January–February, 33–39.

32 Black, F., Scholes, M. (1973) The pricing of options and corporate liabilities, *Journal of Political Economy*, **81**, 637–659.

33 Borio, C., Drehmann, M. (2009) Assessing the risk of banking crises – revisited, *BIS Quarterly Review*, 29–46.

34 Borio, C., Drehmann, M., Tsatsaronis, K. (2013) Stress-testing macro stress testing: Does it live up to expectations? *Journal of Financial Stability*, **12**.

35 Brigo, D. (2012) Counterparty risk FAQ, credit VaR, PFE, CVA, DVA, closeout, netting, collateral, re-hypothecation, WWR, Basel, funding, CDS and margin lending, available at SSRN.

36 Brigo, D., Mercurio, F. (2007) *Interest Rate Models – Theory and Practice: With Smile, Inflation and Credit*, Springer.

37 Brigo, D., Morini, M., Pallavicini, A. (2012) *Counterparty Credit Risk, Collateral and Funding*, Wiley Finance.

38 Brigo, D., Morini, M., Pallavicini, A. (2013) *Counterparty Credit Risk, Collateral and Funding: With Pricing Cases for All Asset Classes*, John Wiley & Sons.

39 Brunnermeier, M.K. (2009) Deciphering the liquidity and credit crunch 2007–2008, *Journal of Economic Perspectives*, **23** (1), 77–100.

40 Brunnermeier, M.K., Pedersen, L.H. (2009) Market liquidity and funding liquidity, *Review of Financial Studies*, **22**, 2201–2238.

41 Canabarro, E., Duffie, D. (2003) *Measuring and marking counterparty risk, in Asset/Liability Management for Financial Institutions*, Institutional Investor Books, 122–134.

42 Carty, L.V., Lieberman, D. (1998) Historical default rates of corporate bond issuers, 1920–1996, in S. Das (ed.) (2003) *Credit Derivatives: Trading and Management of Credit and Default Risk*, John Wiley & Sons, 317–348.

43 Castagnolo F., Ferro G. (2014) Models for predicting default: Towards efficient forecasts, *The Journal of Risk Finance*, **15** (1), 52–70.

44 Cherubini, U., Luciano, E., Vecchiato, W. (2004) *Copula Methods in Finance*, John Wiley & Sons.

45 Choudhry, M. (2011) *Bank Asset and Liability Management: Strategy, Trading, Analysis*, John Wiley & Sons.

46 Commission of the European Communities (2008) Proposal for a Regulation of the European Parliament and of the Council on Credit Rating Agencies, Brussels.

47 Cooper, I., Mello, A. (1991) The default risk of swaps, *Journal of Finance*, **46** (2), 597–620.

48 Credit Suisse (1997) *CreditRisk+: A credit risk management framework*, Credit Suisse Financial Products.

49 Crook, J.N., Edelman D.B., Thomas Lyn. C. (2007) Recent developments in consumer credit risk assessment, *European Journal of Operational Research*, **183** (3), 1447–1465.

50 Das, S. (2003) *Swaps and Financial Derivatives: Products, Pricing, Applications and Risk Management*, 3rd edition, John Wiley & Sons, New Jersey.
51 Das, S. (ed.) (2005) *Credit Derivatives: CDOs and Structured Credit Products*, John Wiley & Sons.
52 Delianedis, G., Geske, R.L. (2003) Credit risk and risk neutral default probabilities: Information about rating migrations and defaults, EFA 2003 annual conference paper.
53 Derman, E. (2009) Models, *Financial Analysts Journal*, **65** (1), 28–33.
54 Diamond, D. (1991) Debt maturity structure and liquidity risk, *Quarterly Journal of Economics*, **106**, 709–737.
55 Diamond, D.W. (2007) Banks and liquidity creation: A simple exposition of the Diamond-Dybvig model, *Fed Res Bank Richmond Economic Quarterly*, **93** (2), 189–200.
56 Diamond, D.W., Dybvig, P.H. (1983) Bank runs, deposit insurance, and liquidity, *Journal of Political Economy*, **91** (3), 401–419.
57 Diamond, D., Rajan, R.G. (2011) Fear of fire sales, illiquidity seeking and the credit freeze, *Quarterly Journal of Economics*, **126** (2), 557–591.
58 Duffie, D., Singleton, K.J. (2003) *Credit Risk: Pricing, Measurement and Management*, Princeton University Press, New Jersey.
59 Elizalde, A. (2006) Credit risk models I: Default correlation in intensity models, CEMFI Working Paper 0605, available at www.abelelizalde.com, CEMFI.
60 Elton, E.J., Gruber, M.J., Agrawal, D., Mann, C. (2001) Explaining the rate spread on corporate bonds, *The Journal of Finance*, **56** (1), 247–277.
61 Engelmann, B., Hayden, E., Tasche, D. (2003) *Testing rating accuracy*, *Risk*, www.risknet.
62 Financial Stability Forum (2009) Addressing procyclicality in the financial system.
63 Finger, C., Ta, T. (2002) CreditGrades technical document, RiskMetrics Group.
64 Froot, K.A., Stein, J.C. (1998) Risk management, capital budgeting, and capital structure policy for financial institutions: An integrated approach, *Journal of Financial Economics*, **47** (1), 55–82.
65 Frye, J. (1997) *Principals of risk: Finding VaR through factor-based interest rate scenarios, in VaR: Understanding and Applying Value at Risk*, Risk Publications, London, 275–288.
66 FSA (2009) Strengthening Liquidity Standards, London.
67 Golub, B.W., Crum, C.C. (2010) Risk management lessons worth remembering from the credit crisis of 2007–2009, *Journal of Portfolio Management*, **36** (3), 21–44.
68 Golub, B.W., Tilman L.M. (1997) Measuring yield curve risk using principal components, analysis, value at risk, and key rate durations, *The Journal of Portfolio Management*, **23** (4), 72–84.
69 Gordy, M.B. (2000) A comparative anatomy of credit risk models, *Journal of Banking & Finance*, **24** (1–2), 119–149.
70 Gordy, M.B. (2003) A risk factor model foundation for rating-based bank capital rules, *Journal of Financial Intermediation*, **12**, 199–232.
71 Gordy, M.B., Howells, B. (2006) Procyclicality in Basel II: Can we treat the disease without killing the patient? *Journal of Financial Intermediation*, **15**, 395–417.
72 Gregory, J. (2010) *Counterparty Credit Risk*, John Wiley & Sons.
73 Gregory, J. (2012) *Counterparty Credit Risk and Credit Value Adjustment: A Continuing Challenge for Global Financial Markets*, John Wiley & Sons.
74 Gupton, G.M., Finger C.C., Bhatia, M. (1997) Creditmetrics: Technical Document, JP Morgan & Co.
75 Hayre, L.S. (1990) Understanding option-adjusted spreads and their use, *The Journal of Portfolio Management*, **16** (4), 68–69.
76 Ho, T.S., Saunders, A. (1981) The determinants of bank interest margins: Theory and empirical evidence, *Journal of Financial and Quantitative Analysis*, **16** (4), 581–600.
77 Holton, G.A. (2003) *Value-at-Risk: Theory and Practice*, Academic Press.
78 Hull, J. (2014) *Options, Futures and Other Derivatives*, 9th edition, Prentice Hall.
79 Hull, J., White, A. (1993) One-factor interest-rate models and the valuation of interest-rate derivative securities, *Journal of Financial and Quantitative Analysis*, **28** (2), 235–254.
80 Hull, J., White, A. (2012) CVA and wrong-way risk, *Financial Analysts Journal*, **68** (5), 58–69.
81 Jarrow, E. (1998) *Volatility*, Risk Publications, London.
82 Jarrow, R.A., Van Deventer, D.R. (1998) The arbitrage-free valuation and hedging of demand deposits and credit card loans, *Journal of Banking & Finance*, **22** (3), 249–272.
83 Jorion, P. (1996) Risk2: Measuring the risk in value at risk, *Financial Analysts Journal*, **52** (6), 47–56.
84 Jorion, P. (2000) Risk management lessons from long-term capital management, *European Financial Management*, **6** (3), 277–300.
85 Jorion, P. (2007) *Value-at-Risk*, 3rd edition, McGraw-Hill, New York.

86 Kacperczyk, M., Schnabl, P. (2010) When Safe Proved Risky: Commercial Paper during the Financial Crisis of 2007–2009, *Journal of Economic Perspectives*, **24** (1), Winter, 29–50.
87 Kealhofer, S. (2003) Quantifying credit risk I: Default prediction, *Financial Analysts Journal*, **3**, 30–44.
88 Kealhofer, S. (2003) Quantifying credit risk II: Debt valuation, *Financial Analysts Journal*, **59** (1), 78–92.
89 Laux, C., Leuz, C. (2010) Did fair-value accounting contribute to the financial crisis? *Journal of Economic Perspectives*, **24** (1), Winter, 93–118.
90 Li, D. (2000) On default correlation, a copula approach, *Journal of Fixed Income*, **9** (4), 43–54.
91 Liu, B., Kocagil, A.E., Gupton, G.M. (2007) Fitch equity implied rating and probability of default model, Quantitative Research Special Report.
92 Lo, A.W. (1999) The three P's of total risk management, *Financial Analysts Journal*, **55** (1), 13–26.
93 Longin, F.M. (2000) From value at risk to stress testing: The extreme value approach, *Journal of Banking & Finance*, **24**, 1097–1130.
94 Longstaff, F.A. (2010) The subprime credit crisis and contagion in financial markets, *Journal of Financial Economics*, **97** (3), 436–450.
95 Malz, A.M. (2011) *Financial Risk Management: Models, History, and Institutions*, Wiley.
96 Matten, C. (2000) *Managing Bank Capital*, 2nd edition, John Wiley & Sons.
97 Matz, L., Neu, P. (2006) *Liquidity Risk Measurement and Management: A Practitioner's Guide to Global Best Practices*, John Wiley & Sons.
98 McNeil, A.J. (1999) *Extreme Value Theory for Risk Managers, Internal Models and CAD II*, Risk Books, London.
99 Memmel, C., Schertler, A. (2011) Banks' Management of the Net Interest Margin: Evidence from Germany, available at SSRN: http://ssrn.com/abstract=1944140.
100 Merton, R.C. (1973) Theory of rational option pricing, *Bell Journal of Economics and Management Science*, **4** (1), 141–183.
101 Merton, R.C. (1974) On the pricing of corporate debt: The risk structure of interest rates, *The Journal of Finance*, **29** (2), 449–470.
102 Merton, R., Perold A. (1993) Theory of risk capital in financial firms, *Journal of Applied Corporate Finance*, **6** (3), 16–32.
103 Modigliani, F., Miller, M.H. (1958) The cost of capital, corporation finance, and the theory of investment, *American Economic Review*, **48**, 261–297.
104 Moody's Analytics (2011) Moody's market implied ratings description, methodology, and analytical applications, Moody's.
105 Moody's (2008) European corporate defaults and recovery rates: 1985–2007, London.
106 Moody's Investors Services (2011) Corporate default and recovery rates, 1920–2010.
107 Morgan, J.P. (1995) Risk Metrics Technical Manual, New York, JP Morgan Bank.
108 Neftci, S.N. (2000) *Introduction to the Mathematics of Financial Derivatives*, 2nd edition, Academic Press, New York.
109 Neftci, S.N. (2008) *Principles of Financial Engineering*, Academic Press, New York.
110 Nelson, C.R., Siegel, A.F. (1987) Parsimonious modeling of yield curves, *Journal of Business*, **60**, 473–489.
111 Nijskens, R., Wagner, W. (2011) Credit risk transfer activities and systemic risk: How banks became less risky individually but posed greater risks to the financial system at the same time, *Journal of Banking & Finance*, **35**, 1391–1398.
112 O'Kane, D., Turnbull, S. (2003) Valuation of Credit Default Swaps, Fixed Income Quantitative Credit Research, Lehman Brothers, 28–44.
113 Perold, A. (2005) Capital allocation in financial firms, *Journal of Applied Corporate Finance*, **17** (3), 110–118.
114 Poon, S-H., Granger, C. (2003) Forecasting volatility in financial markets: A review, *Journal of Economic Literature*, **XLI**, 78–539.
115 Poon, S-H., Granger, C. (2003) Practical issues in forecasting volatility, *Financial Analysts Journal*, **61** (1), 45–56.
116 PricewaterhouseCoopers Publications (2009) IFRS Manual of Accounting 2009: Global Guide to IFRS.
117 Pykhtin, M., Zhu, S. (2007) A guide to modelling counterparty credit risk, *Garp Risk Review*, 16–22.
118 Risk Publications (1996) *Vasicek and Beyond*, London.
119 Risk Publications (2003) *Liquidity Black Holes: Understanding, Quantifying and Managing Financial Liquidity*, Risk, London.
120 RiskMetrics Group (2002) CreditGrades: Technical Document, New York.
121 Rosen, D., Saunders, D. (2012) CVA the wrong way, *Journal of Risk Management in Financial Institutions*, **5** (3), 252–272.

122 Satchell, S., Xia, W. (2006) Analytic models of the ROC curve, applications to credit rating model validation, Quantitative Finance Research Center, University of Technology, Sydney.
123 Schönbucher, P., Schubert, D. (2001) Copula-dependent default risk in intensity models, TIK-Report 103.
124 Sharpe, W. (1964) Capital asset prices: A theory of market equilibrium under condition of risk, *Journal of Finance*, **19**, 425–442.
125 Shashidhar, M. (2011) Market-implied risk-neutral probabilities, actual probabilities, credit risk and news, *IIMB Management Review*, **23** (3), 140–150.
126 Singh, M. K. (1997) Value at risk using principal components analysis, *The Journal of Portfolio Management*, **24** (1), 101–112.
127 Standard & Poor's (2003) *Corporate Ratings Criteria*, McGraw-Hill, New York.
128 Stoughton, N.M., Zechner, J. (2007) Optimal capital allocation using RAROC™ and EVA®, *Journal of Financial Intermediation*, **16** (3), 312–342.
129 Tasche, D. (2008) Capital allocation to business units and sub-portfolios: The Euler principle, in Pillar II in the New Basel Accord: The Challenge of Economic Capital, Risk Books.
130 Tasche, D. (2013) The art of probability-of-default curve calibration, *Journal of Credit Risk*, **9** (4), 63–103.
131 Uyemura, D.G., Kantor C.C., Pettit J.M. (1996) EVA® for banks: Value creation, risk management, and profitability measurement, *Journal of Applied Corporate Finance*, **9** (2), 94–109.
132 VanHoose, D. (2007) Theories of bank behavior under capital regulation, *Journal of Banking & Finance*, **31**, 3680–3697.
133 Vasicek, O. (1991) Limiting loan loss probability distribution, Moody's-KMV.
134 Vasicek, O. (2002) Loan portfolio value, *Risk*, 160–162.
135 Wilson, T. (1997) Portfolio credit risk I, *Risk*, **10** (9), 111–117.
136 Wilson, T. (1997) Portfolio credit risk II, *Risk*, **10** (10), 56–61.
137 Wilson, T. (1998) *Credit Portfolio View*, McKinsey & Co.
138 Zhou, C. (2001) An analysis of default correlations and multiple defaults, *The Review of Financial Studies*, **14** (2), 555–576.
139 Zumbach, G. (2007) The RiskMetrics 2006 Methodology, RiskMetrics Group.

Index

Index compiled by Terry Halliday