Jiří Witzany

Credit Risk Management

Pricing, Measurement, and Modeling

Springer

Credit Risk Management

Jiří Witzany

Credit Risk Management

Pricing, Measurement, and Modeling

Springer

Jiří Witzany
Faculty of Finance and Accounting
University of Economics in Prague
Prague, Czech Republic

ISBN 978-3-319-49799-0 ISBN 978-3-319-49800-3 (eBook)
DOI 10.1007/978-3-319-49800-3

Library of Congress Control Number: 2016962709

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book has been supported by the VŠE institutional grant IP100040 and by the Czech Science Foundation Grant 15-00036S, "Credit Risk Modeling for Financial and Commodity Assets Portfolios."

Chris Sadil

Preface

Credit risk management has been a keystone of prudent traditional banking for thousands of years. Bankers have to consider the risks and expected profits in order to decide to whom and how much to lend. When a loan is granted, there must be monitoring and communication with the borrower, in particular in the case of financial distress or simple unwillingness to repay. These decisions used to be based on experience and expertise, but with the advance of modern mathematical, statistical, and computational tools the discipline has become increasingly sophisticated. Credit risk is not only managed, but also priced and measured using various accurate mathematical and statistical tools and models. The credit risk management decisions must be now based on the results of exact pricing and measurement.

The growth of consumer credit in the last 50 years would hardly have been possible without automatically operated credit scoring. Credit scoring developed using various classification or regression methods, in particular with logistic regression, has become an industry standard accepted as an accurate credit risk assessment tool, often performing better than human credit risk officers. The scoring measures the credit risk not only by using a score discriminating between better and worse loan applications, but also by estimating the probability of default within one year or another given time horizon. The risk can be also priced by calculating the risk premium that should be part of the loan interest rate in order to cover properly the expected credit losses.

Moreover, the complexity of credit risk modeling has increased with the advance of financial derivatives, which essentially always involve counterparty credit risk. Assessment of the counterparty credit risk is essentially impossible to separate from the market risk analysis. Credit risk is directly priced into credit derivatives, whose value depends on one or more reference credit entities. The credit derivatives and related asset backed securities became popular at the end of the nineties and at the beginning of the previous decade. Nevertheless, the global financial crisis has shown that the complexity of their pricing has been seriously underestimated, and the market has become much more cautious with respect to these products since then.

The banking industry has been always regulated to a certain extent, but with the advance of globalization and the experience of many banking and financial crises with negative consequences on vast numbers of bank clients, employment, economic growth, and state budgets, regulation has become global, focusing more and more on risk management issues. The modern regulation approved in Basel by a committee representing the central banks of the most developed nations requires banks to keep sufficient capital to cover unexpected losses related to credit, market, and operational risk. In addition, the financial institutions are required to have sound risk management processes and organization satisfying the outlined regulatory principles. For example, the credit risk management should be able to estimate unexpected credit risk on the bank's loan portfolio independently of the regulatory capital calculation. The latest version of the regulation (Basel III) also requires banks to price the counterparty credit risk (Credit Valuation Adjustment—CVA) and estimate its stressed value in order to calculate a specific additional part of the regulatory capital.

The main goal of this book is to cover the most important historically developed areas of credit risk management, pricing, and measurement as outlined above, but also to focus on the latest developments and research. Each chapter also contains a discussion of relevant regulatory principles and requirements. The book is intended for academic researchers and practitioners in the field of derivatives and risk management and for specialists in banks and financial institutions, as well as for graduate degree students of economics, finance, and financial engineering.

The author refers not only to a long list of literature on the subject of credit risk management, pricing and measurement but also to his experience as a market and credit risk manager in a large Czech bank (Komerční banka) and as a partner of the Quantitative Consulting company participating in or managing many credit risk projects for a number of domestic and international banks. The quality of the English language text has been greatly improved thanks to Chris Sadil. Many thanks belong to K. Sivakumar, who took care of the final text and picture editing. Last but not least, the work could not have been finished without the personal support and patience of my wife Nadia.

Although the materials presented here have been thoroughly checked, some mistakes may possibly remain. Any comments can be sent to jiri.witzany@vse.cz.

Prague, Czech Republic Jiří Witzany

Contents

1 Introduction 1

2 Credit Risk Management 5
2.1 Credit Risk Organization 5
2.2 Trading and Investment Banking 8
2.3 Basel Credit Risk Management Requirements 12

3 Rating and Scoring Systems 19
3.1 Rating Quality Measurement and Validation 19
3.2 Analytical Ratings 34
3.3 Regression Rating Systems 37
3.4 Alternative Automated Rating Systems 74
3.5 Expected Loss, LGD, and EAD Estimations 94
3.6 Basel II Rating Based Approach 110

4 Portfolio Credit Risk 117
4.1 Economic Capital, Expected, and Unexpected Losses 118
4.2 CreditMetrics 120
4.3 CreditRisk+ 133
4.4 CreditPortfolioView 139
4.5 KMV Portfolio Manager 142
4.6 Comparison of the Models 150
4.7 Vasicek's Model and the Basel Capital Calculation 152

5 Credit Derivatives and Counterparty Credit Risk 159
5.1 Credit Derivatives Market 160
5.2 Valuation of Multi-name Credit Derivatives 179
5.3 Advanced Dependence Modeling 190
5.4 Dynamic Intensity of Default Modeling 204
5.5 Basel on Credit Derivatives and Securitization 217
5.6 Counterparty Credit Risk 219

6 Conclusion 241

Index 243

Abbreviations

ABS	Asset Backed Security
AMA	Advanced Measurement Approach
AR	Accuracy Ratio
AUC	Area under the Receiver Operating Characteristic Curve
BCBS	Basel Committee on Banking Supervision
BCVA	Bilateral Credit Valuation Adjustment
BEEL	Best Estimate of the Expected Loss
BVA	Bilateral Valuation Adjustment
CAD	Capital Adequacy Directive
CAPM	Capital Asset Pricing Model
CAR	Capital Adequacy Ratio
CCR	Counterparty Credit Risk
CDO	Collateralized Debt Obligation
CDS	Credit Default Swap
CEBS	Committee of European Banking Supervisors
CEM	Current Exposure Method
CF	Conversion Factor
CFO	Chief Finance Officer
CIR	Cox, Ingersoll, and Ross (Model)
CML	Canonical Maximum Likelihood
CRD	Capital requirements Directive
CRO	Chief Risk Officer
CRR	Capital Requirements Regulation
CVA	Credit Valuation Adjustment
CZEONIA	Czech Overnight Index Average
DD	Distance to Default
DVA	Debit Valuation Adjustment
ECAI	External Credit Assessment Institution
EDF	Expected Default Frequency
EE	Expected Exposure
EIR	Effective Interest Rate
EL	Expected Loss
EMIR	European Market Infrastructure Regulation

ENE	Expected Negative Exposure
EONIA	Euro Overnight Index Average
EPE	Expected Positive Exposure
ETL	Expected Tranche Loss
FBA	Funding Benefit Adjustment
FCA	Funding Cost Adjustment
FVA	Funding Valuation Adjustment
GPL	Generalized Poisson Loss
HJM	Heath, Jarrow, and Morton (Model)
IAS	International Accounting Standards
IFM	Inference for the Margins (Method)
IFRS	International Financial Reporting Standards
IMA	Internal Model Approach
IMM	Internal Market Model (IMM)
IRB	Internal Rating Based (Approach)
IRBA	Internal Rating Based Advanced (Approach)
IRBF	Internal Rating Based Foundation (Approach)
IRS	Interest Rate Swap
ISDA	International Swaps and Derivatives Association
IV	Information Value
KVA	Capital Valuation Adjustment
LGD	Loss Given Default
LHP	Large Homogenous Portfolio
LVA	Liquidity Valuation Adjustment
MA	Maturity Adjustment
MBS	Mortgage Backed Security
MVA	Margin Valuation Adjustment
NI	Net Income
OIS	Over-Night Index Swap (Rate)
OLS	Ordinary Least Squares (Regression)
ONIA	Overnight Index Average
OTC	Over-the-Counter
PD	Probability of Default
PIT	Point in Time
PSE	Private Sector Entity
Q-Q	Quantile-to-Quantile
RBA	Rating Based Approach
RDS	Reference Data Set
RP	Risk Premium
RWA	Risk-weighted Assets
SFA	Supervisory Formula Approach
SIFMA	Securities Industry and Financial Markets Association
SME	Small and Medium Enterprises
SONIA	Sterling Overnight Index Average
SPV	Special Purpose Vehicle

TRS	Total Return Swap
TTC	Through-the-Cycle
UDR	Unexpected Default Rate
VaR	Value at Risk

1 Introduction

Many monographs on credit risk start, and deal, essentially only with different computational methods for measuring and modeling credit risk. As the title indicates, we want to focus not only on various modeling techniques for credit risk pricing and measurement, but also on key issues of credit risk management; i.e., we also want to look at the proper setting of the credit risk organization, credit risk processes, powers, and controls. It may be observed from many failures and mishaps which have occurred in various banks around the world, including the recent systemic crisis, that the key problem often lies, rather, in conflicts of interest, insufficient controls, or the low levels of power vested in those who knew that there was a problem (and who were able to gauge the risks correctly) but were not able to prevent, or limit, the risky transactions. This experience can also be confirmed by the author, who was responsible for overseeing the risks inherent in trading activities, as well as the later credit risks in the classical banking activities of a large Czech bank[1] in the late nineties, and in the first half of the previous decade. On the other hand, in many cases, in particular during the recent crisis, insufficient controls and regulation have been partially connected to the low level of understanding, and consequent underestimation, of the risks involved. Hence, to summarize, one should neither overestimate nor underestimate the importance of credit measurement techniques with respect to the classical credit risk management issues. We shall start discussing the credit risk organization and management issues in Chap. 2.

Nowadays it is impossible to write a textbook on banking credit risk without incorporating a discussion of the Basel II/III Regulation requirements, their impacts and consequences. This regulation is sometimes taken as a law from above, and essentially all risk management effort is focused on the fulfillment of its requirements and standards. Sound risk management should, nevertheless, still be performed in the shareholders' own best interests, even if there were no Basel

[1]Komerční banka, a.s.

J. Witzany, *Credit Risk Management*, DOI 10.1007/978-3-319-49800-3_1

regulations at all. The current regulation has not been created from nothing, but reflects the best practices and experiences of the banking industry before the regulatory framework was first proposed, fine-tuned, and put into practice. It is, on the other hand, far from perfect, and cannot be taken as a dogma, and there is a lot of room for improvement, both in the area of implementation, as well as in the area of possible future changes in the regulation. We will take a critical approach, following the regulation, its requirements, and implementation steps, as well as critically discussing possible improvements and techniques that go beyond the regulation.

The first task faced by anyone considering a new transaction or business relationship with a counterparty, which may or may not fulfill its future obligations, is credit risk assessment. That is some sort of estimate of the adverse possibility of default, which helps one to make the final decision; yes, no, or yes with some additional conditions. In some traditional banks, the bankruptcy of a corporate debtor is considered as ex post facto evidence of wrong initial decision making, along with the corresponding credit assessment. This reasoning might have some logic in the case of the largest debtors or project financing, but does not make any sense in the case of mass segments, such as the small and medium enterprises (SME) sector, and in particular, in the case of retail portfolios with many small debtors. We cannot have a crystal ball which tells us whether each particular client will fail or not. For a large portfolio to experience 1, 10 %, or even more defaults within a year does not mean automatically that the assessment was wrong. If the assessment predicted 1, 10 %, or some other percentage of defaults and the reality matches the expectation, then the assessment was correct. Moreover, if the interest income on the receivables that did not default covers sufficiently not only the cost of funds and administrative costs, but also the losses, with some remaining profit, then there is a sound business case. This common approach of modern banking requires advanced financial analysis and statistical (rating and scoring) techniques that will be described in Chap. 3. The chapter will also discuss alternative classification and data-mining methods as Support Vector Machines, Random Forests, or Neural Network that are the subject of recent research, with the logistic regression still representing an industry standard.

Even if our credit assessment process and pricing methodology are correct, we need to take the large picture into account and draw appropriate conclusions from it. If many things go wrong, do we have enough reserves and capital to survive? The advanced statistical approaches to this question are discussed in Chap. 4. The question is, in fact, the main concern of the regulatory capital requirement. The goal of the regulator is to make sure that the banking system is sufficiently stable, as one bank failure may cause large losses to many other banks, and so, to the whole economy and the tax payers in the end. The Basel II/III Regulation provides a more or less simplified approach, inspired by the economic capital models, and there is still a need for advanced complementary modeling.

Recent decades have seen a tremendous growth in the trading of credit derivative instruments with many positive and, unfortunately also, negative effects such as we could see in the recent global financial crisis. Credit derivatives, like derivatives in general, are derived from more basic underlying instruments or market indicators. The derivatives always involve future cash flows (or exchange of assets) that are subject to the prices of the underlying assets, or values, of the underlying indicators. In the case of the credit derivatives, the underlying assets are typically bonds or loans, and the simplest indicator is the one of default, or no default, of a reference entity whose credit risk is "insured" by the credit derivative. Credit derivative securities, like CDO, treat credit risk like raw milk which is separated into skimmed, medium fat, fat, and creamy milk. The fat and creamy milk can taste good to the markets, but may become sour as we have seen recently. Since the credit derivatives are traded on the financial markets, where prices change almost continuously, the modeling and pricing of the derivatives requires the introduction of stochastic modeling methods in a fashion similar to the classical derivatives. In addition, the financial crisis and the Basel III regulation has drawn more attention to counterparty credit risk management. The Credit Valuation Adjustment (CVA) has become a mandatory component of derivative pricing. The most important types of credit derivatives and the main approaches to their pricing, as well as CVA calculation principles will be introduced in the last chapter.

Credit Risk Management

2

If a person or an organizational unit within a bank or corporation performs credit assessment, then there are two essential questions: Do they have the necessary skills and techniques to assess the credit risk properly, and, secondly, will the assessment really be independent and unbiased? For example, if the assessment is entrusted to a salesperson remunerated according to the number and volume of loans granted, then there is a clear danger of underestimating the risks; i.e., being too optimistic when looking at the applicants' financial situation in order to maximize the business target. The situation is even worse if there is a relationship, potentially even corrupt, between the salesperson and the applicant. Such a situation is, unfortunately, not impossible. Hence a bank may have excellent credit modeling software, and many qualified mathematicians, but if those simple issues (one could say operational risks) are omitted, then there is a big problem. Therefore, we need to discuss, firstly, the appropriate or recommended models of credit risk organization, as well as the separation of powers, both in the case of classical banking (or corporate business) activities, as well as in the case of trading and investment activities. Those recommendations, in fact, go hand in hand with the Basel risk management process standards.

2.1 Credit Risk Organization

The first and the most important requirement for a sound banking credit risk organization, is the separation of powers between the risk management and the business side. A possible organizational structure of a universal bank involved in commercial corporate and retail banking, as well as in investment banking is shown in Fig. 2.1. The risk management division of the bank is headed by the Chief Risk Officer (CRO), who should be a member of the management board. The business divisions, investment banking, and marketing should report to other members of the board. Risk Management is close to the Finance and Accounting sections of the bank, which are usually directed by the Chief Finance Officer (CFO). The

J. Witzany, *Credit Risk Management*, DOI 10.1007/978-3-319-49800-3_2

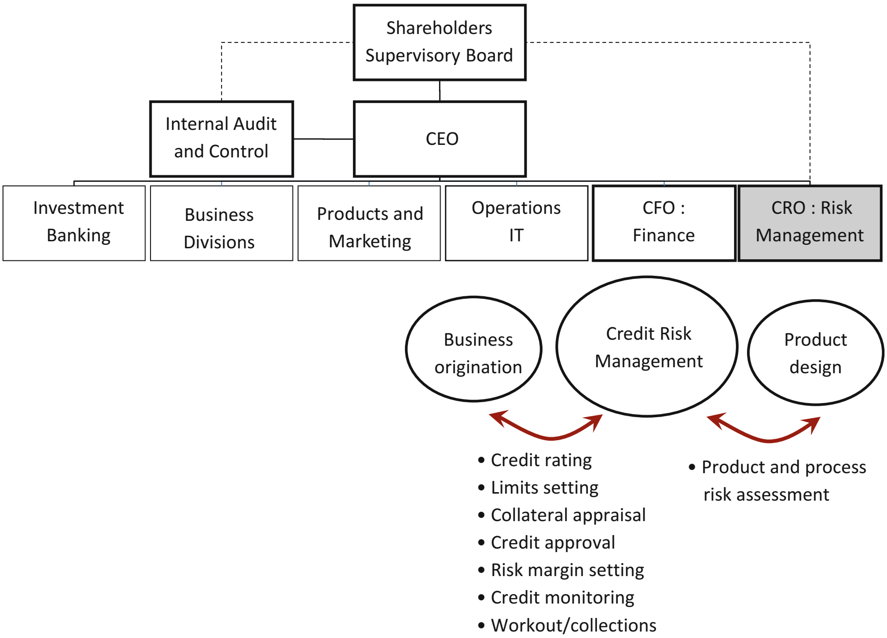


Fig. 2.1 A universal bank organization structure

responsibilities of CRO and CFO can possibly be delegated to just one member of the board. Another controlling unit close to risk management is the Internal Audit and Control section, which stands above all the other departments and divisions in terms of control and auditing supervision. The independence of the Internal Audit and Risk Management sections is often strengthened by direct reporting to the Supervisory Board, which acts on behalf of the shareholders. There is also a compliance department responsible for adherence to legal and regulatory requirements. If the bank belongs to a large banking group, then the CRO is usually nominated by the superior risk management of the group and is only formally appointed by the CEO (Chief Executive Officer).

Credit analysis, the setting of exposure limits, estimation of expected losses entering the risk margin calculation, and final transaction approval should be optimally done independently of those officials who take care of the clients' business. Smaller transactions, in particular retail, may be processed automatically using a software system. The system design and maintenance responsibility must be sufficiently independent of the business and marketing side. This part of the bank is motivated by the maximization of volumes, market share, and/or number of clients. An important component of the approval process is the assessment of collateral; in particular of real estate, which must also be controlled by the risk. Credit risk management may, and usually does, delegate certain limited approval responsibility to the business units. The exact division of underwriting powers is always a

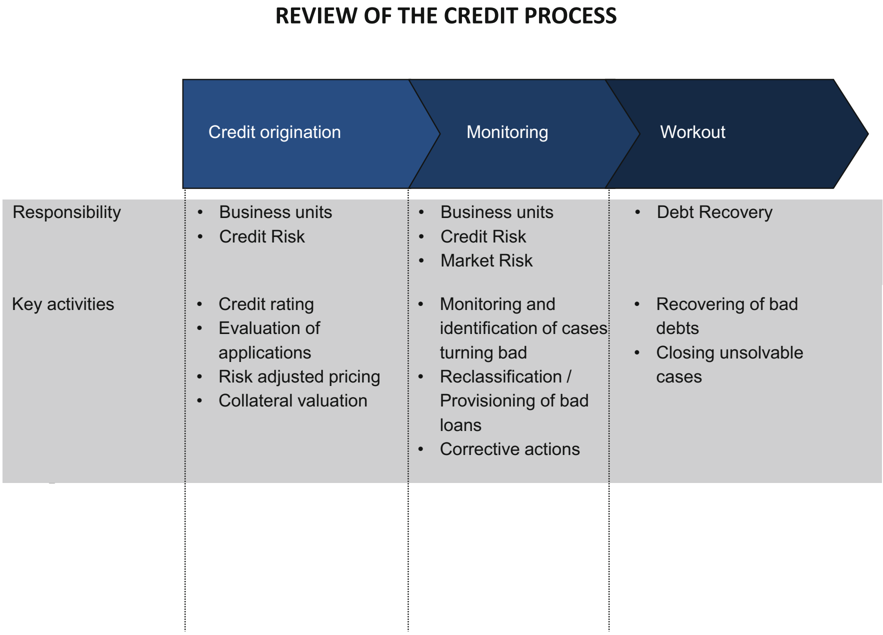


Fig. 2.2 A review of the credit process

compromise between the efficiency of the process, and the conservativeness of the risk management standards.

The key role of risk management does not end with the approval and granting of a loan to the client. The overall credit process and the split of responsibilities can be schematically illustrated by Fig. 2.2. Corporate clients should be obliged to submit regular financial reports, which need to be monitored by the client relationship officials, as well as (depending on the size of the exposure) by credit analysts. If the reports or other information indicate a worsening of the client's situation, then corrective actions should be taken. The action may be the stopping of any new withdrawals, negotiation with the client on the initiation of early repayments based on existing contractual covenants, recommendation to change the current business plan, etc. The last phase of the credit process already belongs to the workout phase when the exposure turns bad. Since the client relationship officials tend to underestimate the problems, the exposure should be transferred to a different specialized unit which may be under the responsibility of risk. The workout activities could also be under the business side management, but at all events, there should be a change of personal and organizational responsibility in the case of a bad client that is going through bankruptcy or restructuring.

The monitoring of retail loans is more mechanical, and usually based on the so-called behavioral scoring. The process of collecting retail loans with delayed repayments normally starts very early with the so-called soft collection method, whereby the clients are contacted by phone. It may end by the selling or outsourcing of the defaulted loans to specialized collection companies that initiate possible legal steps, the sale of collateral, or executions.

Provisioning is a very important part of the monitoring process. Provisions reduce the receivable net accounting value in order to reflect expected losses. The provisioning process is governed by the IAS/IFRS standards, but should be, in principle, consistent with Basel II/III or the economic valuation approach. Provisioning is usually connected to a classification system, classifying performing loans as Standard and Watch, while non-performing or impaired are classed as Substandard, Doubtful, and Loss. Specific provisions are created on the impaired loans. For example, the exposure of loans classified as Loss not covered by valuable collateral is usually provisioned at 100 %. Early and correct provisioning which has an impact on the P/L statement of the bank, as well as on the responsible business unit, is the best way to initiate a corrective action, as well as being a wake-up call. Many banks which tried to hide provisions in order to improve financial results temporarily, eventually ended up with huge losses. Therefore, provisioning should be again primarily under the control of the credit risk management part of the bank.

Last, but not least, we should mention the importance of data for good credit risk management. First of all, the bank should be able to filter out clients where there is serious negative information, including the bank's own experience from the past. For example, if a client engaged in a fraudulent activity when applying for a consumer loan, his/her application should be automatically rejected if the client applies for a small business loan and vice versa. This may seem trivial, but requires well organized and interconnected databases covering all available sources of information. Banks often build large data-warehouses with different areas, including a credit risk data mart. The information is used, not only as a black list, but generally to obtain a rating utilizing all available information. Moreover, there are external banking and non-banking registers (credit bureaus) where the negative and even positive credit information is shared by banking and non-banking institutions. Thus, credit risk must closely cooperate with the bank's IT and Finance in the building and maintaining of those databases which form the key foundation of a sound credit risk management process.

2.2 Trading and Investment Banking

Trading and investment activities should be, in principle, separated from commercial banking. There should be a "Chinese wall" preventing the investment part of the bank from using insider information obtained from clients in the classical credit process. Moreover, the monitoring of financial market instruments exposures is, in general, very complex and goes hand in hand with the ability to price and quantify the risks associated with the instruments. A relatively simple example may be given by an FX forward contract (e.g. selling 10 million EUR for 260 million CZK with settlement in 1 month). The initial market value of an FX forward contract between the bank and counterparty A, if entered into under market conditions, should be

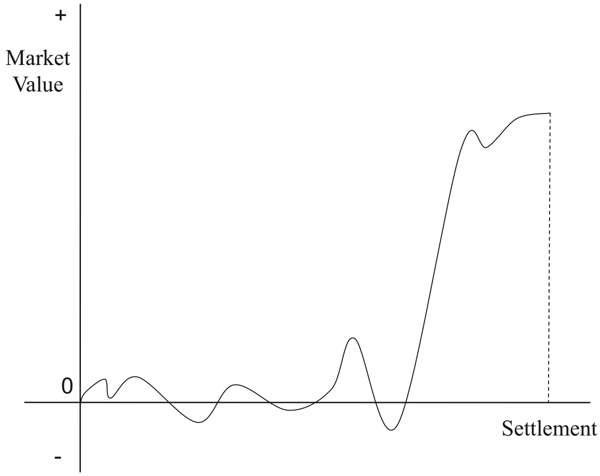


Fig. 2.3 A possible development of the FX forward market value

equal approximately to zero, but can later fluctuate to positive and negative values due to the exchange rate movement as shown in Fig. 2.3.

There are two types of credit risk that should be considered. They are the settlement risk and the counterparty risk. The settlement risk is the risk that the bank pays the notional (10 million EUR in our example) on the settlement day, while the counterparty fails to pay the full corresponding amount (260 million CZK in our example). In the case of bankruptcy, most of the notional amount would be usually lost for the bank. The realization of settlement risk is rare, but it does happen, because financial institutions, from time to time, go bankrupt. The settlement risk can be eliminated, or minimized, through the use of an appropriate settlement procedure, e.g. delivery versus payment.

The more important and complex type of credit risk that needs to be managed is the counterparty risk; i.e. the possibility that the counterparty fails before, or at, maturity, and the transaction must be canceled before the final settlement. If this happens, and the market value from the bank's perspective is positive, then there is a loss, since the value accounted in the P/L must be written-off or provisioned. The difference compared to the settlement risk is that there is no payment (in the case of the FX forward) between the bank and the counterparty, which might declare bankruptcy, or is known not to be able to settle the transaction. It may seem that such a loss is only virtual and not real. The reality of the loss can be explained by considering that such transactions are normally matched by offsetting transactions within a trading portfolio. For example, if the transaction from Fig. 2.3 was simply mirrored by an identical, but opposite, transaction with another counterparty B, then the bank must still settle the transaction and realize the loss with respect to counterparty B, but it gets no profit from the transaction with counterparty A where the settlement is not realized.

Now the question is, what credit exposure of the FX forward transaction should be recorded with respect to the credit limit at the origin, and during the life, of the deal? The market value at the beginning is zero, but there is a credit risk as the market value can easily go, for instance, up to 10 % of the domestic currency nominal amount, and at the same time, we do not know the future. On the other hand, it would be clearly much too conservative to consider the full notional amount as the credit exposure. Also, during the life of the transaction we need to consider the actual market value, if positive, and the possibility of its future increase. This reasoning leads to the classical counterparty risk equivalent exposure formula:

$\text{Exposure} = \max(\text{market value}, 0) + x\% \times \text{Notional amount}.$

The percentage x% corresponds to the estimated volatility of the instrument's market value and to its remaining maturity. Note that the exposure is not static but dynamic depending on its volatile market value. The total exposure with respect to a single counterparty, due to many transactions, can be partially reduced by so-called netting agreements, allowing us to net positive and negative market values in the case of bankruptcy.

The complexity of managing counterparty credit risk with respect to a given set of limits can be seen during the financial market crisis, when market prices swing up and down, and a number of financial institutions go bankrupt. The effect is twofold: first of all, many of the market values go up in significant excess of the initial buffer (x%·Notional Amount), and moreover, an unexpected number of counterparties fail, so that overall losses are much larger than ever expected. There might be a domino effect such as we could see during the recent financial crisis. The importance of sound counterparty risk management is underlined by an exponential growth of derivatives trading in terms of OTC (over-the-counter derivatives not settled in organized exchanges) outstanding notional (Fig. 2.4), and deepening credit relationships of the market participants.

The Counterparty Credit Risk (CCR) measurement has become even more complex with the advance of the Credit Valuation Adjustment (CVA) concept

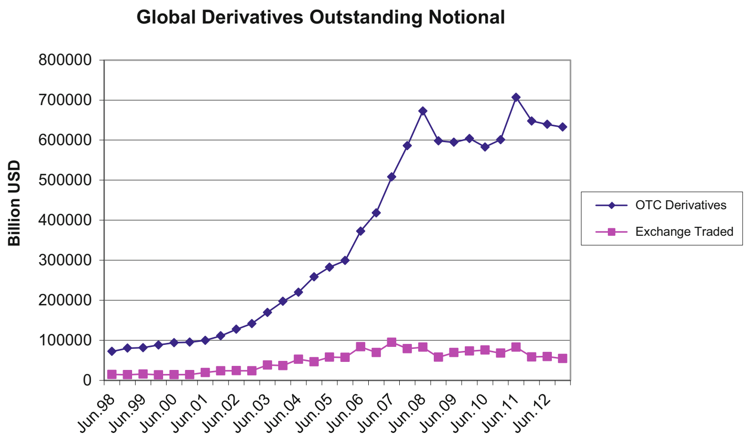


Fig. 2.4 Global OTC derivatives outstanding notional development (Source: www.bis.org)

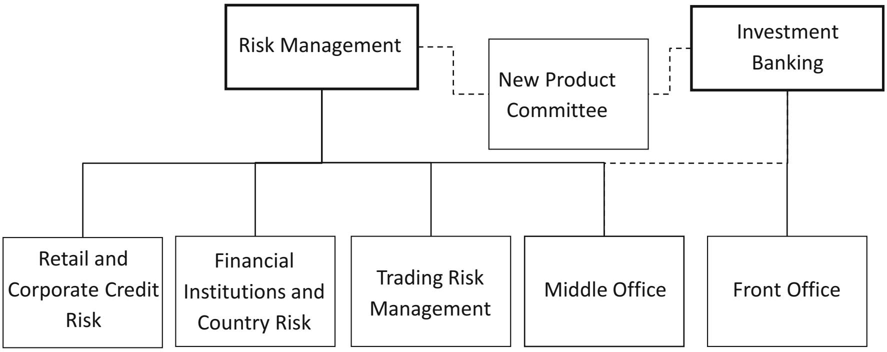


Fig. 2.5 Organization of credit risk management of trading and investment banking

and the new Basel III requirements. The CVA is defined as the expected present value of loss on a derivative position due to the counterparty's default. Although the definition leads to a complex modeling of the derivative price and counterparty default process, there are various simplified methods that allow one to calculate CVA in practice (see Sect. 5.6). The CVA has become a standard component of derivatives pricing. Moreover, variation of CVA due to changes of counterparties' credit quality creates a new source of market losses. The Basel III document (BCBS 2010) in its introduction points out that during the financial crisis, the CVA revaluation was a greater source of losses than those arising from outright defaults. Therefore, the regulation introduces a new capital charge to cover the CVA market losses, meaning that banks not only need to calculate CVA but also to model its future variation.

Due to the complexity and interrelationships of the credit and market risk of the financial market products, credit risk management is usually integrated into a trading risk department. The counterparty risk limits of corporate clients eligible to enter into financial market transactions should, however, be set within the standard credit assessment process. Special attention needs to be paid to credit assessment; i.e., determination of limits and ratings on financial institutions and countries. The country and financial institutions risk assessment is usually based on external ratings and separated from classical corporate and retail credit risk management. A typical organizational structure is shown in Fig. 2.5.

Besides the Trading Risk Management department and the Front Office, i.e. the trading room, the diagram shows the Middle Office and the New Product Committee. The role of the Middle Office is to maintain an up-to-date database of existing transactions, enter new transactions into the trading system, perform reconciliation with the Back Office (reporting to Finance; not shown in the figure, though also very important for sound risk management), preparing various reports, and monitoring market and credit exposure with respect to valid limits on a daily basis. Thus, the Middle Office is primarily an executive branch of the risk management, as well as being the unit supporting the day-to-day business of the trading room.

The responsibility of the New Product Committee is to negotiate and approve new financial markets products. As there are newer and newer financial instruments on the financial markets, it is important to set the appropriate rules, evaluation, and risk assessment methodology (approved by risk management), before any new product transaction is entered into by the dealers. There are many examples where this simple principle has been bypassed with serious consequences. For example, in the nineties, a large Czech bank invested in securities called Credit Linked Notes (CLN) without using the proper credit risk methodology. The securities were issued by a London bank, and so the exposure was naively drawn only with respect to the London bank and British limits that were sufficiently large. Nevertheless, the repayments of the coupons and principal were linked to Russian government bonds. Later, during the Asian crisis, the Russian bonds defaulted and the Czech bank suffered huge losses, afterwards discovering that the exposure had, in fact, been drawn against the Russian government, and not the London bank. The limits set for the Russian government would not have allowed the investment, but the missing methodology, as well as the complexity of the relatively new instrument allowed them to be bypassed.

The importance and difficulty of investment banking credit risk management is underlined by the bankruptcies of, and huge losses suffered by, the leading market players during the recent financial crises; even those that had been assumed to represent the best practice on the markets like JP Morgan or Goldman Sachs. For example, the latter is given as a benchmark of the best credit risk culture in the work of Caouette, Altman, and other respected practitioners and academics published at the beginning of 2008. That publication cites Craig Broderick, responsible for risk management in Goldman Sachs in 2007, saying with self-confidence: *"Our risk culture has always been good in my view, but it is stronger than ever today. We have evolved from a firm where you could take a little bit of market risk and no credit risk, to a place that takes quite a bit of market and credit risk in many of our activities. However, we do so with the clear proviso that we will take only the risks that we understand, can control and are being compensated for."* In spite of that, Goldman Sachs suffered huge losses at the end of 2008 and had to be transformed (together with Morgan Stanley) from an investment bank into a bank holding company eligible for help from the Federal Reserve.

2.3 Basel Credit Risk Management Requirements

The credit risk management organization principles are not only recommendations based on the author's experience, but also follow from the Basel banking regulation. Before we point out the most important parts of the regulation related to the subject, let us firstly review the development and overall structure of the current regulatory framework.

The goal of the Basel II regulatory framework (BCBS 2004) is to set higher risk management and internal control standards for banks all over the world, as well as to introduce a new more risk-sensitive approach to the regulatory capital calculation. A brief overview of the historical development illustrates the fact that the regulation has undergone a long and complex process, and represents, essentially, a compromise between the varying points of view and interests. The first Capital Accord (BCBS 1988) was officially published by the Basel Committee on Banking Supervision, representing the regulators of the most developed countries, in 1988 with an implementation deadline in 1992. The Accord was amended to incorporate market risks in 1996. At that time, discussions on a new regulatory framework had already started, and the first Consultative Paper was published in 1999. After receiving many comments from banks and national supervisors, a second Consultative Paper was issued in 2001. The New Capital Accord (also called Basel II) was finally published, after long and numerous discussions, as well as a number of Quantitative Studies, in June 2004. The document was updated to incorporate certain relatively minor issues in 2005. A comprehensive version incorporating the market risk part (BCBS 1996) was published in 2006. In order to put the New Accord into practice, it had to be implemented into national legislation and regulatory frameworks. Specifically, the European Union published the Implementation Directive (CAD 2006) in June 2006. The Capital Adequacy Directive had to be further incorporated into the national legislations within the EU. The Czech National Bank published the corresponding Provision (CNB 2007) in June 2007, and so the first year in which the new approach was fully applied by banks in the Czech Republic, and similarly in other European countries, was 2008. The financial crises which began in mid-2007 have driven the Basel Committee to propose new amendments and modifications to the Accord. The final versions of amendments to the Basel II Framework, in particular in the area of securitization, as well as Revisions to the Market Risk Framework (BCBS 2009a, b), were published in July 2009. Finally, in December 2010, BCBS approved another reform of the framework called Basel III (BCBS 2010). Basel III does not change the approach to the calculation of capital requirements for classical credit exposures. It does, however, enhance capital requirements for market risk and introduces a new CVA capital requirement. For banks, the most challenging part of the new regulation lies in the strengthened capital requirements for the quality of capital and in the introduction of new capital conservative and countercyclical measures. Other new important requirements are relatively simple leverage and liquidity limits (see Fig. 2.6 for a more detailed overview). There is also an important shift in the EU implementation: the capital and liquidity requirements now take the form of a regulation (CRR 2013) directly applicable as a "single rulebook" to member countries without national discretion, while the supervisory assessment principles remain in the form of a directive (CRD 2013).

Basel II/III has not changed significantly the BCBS (1996) methodology of market risk regulatory capital calculation. A completely new element, brought in by Basel II, is the operational risk capital requirement. The credit risk capital requirement is significantly extended and elaborated by the new regulation, compared to the 1988 Capital Accord, allowing, optionally, a standardized (SA) or internal rating based approach (IRBA), in addition to providing a credit-risk securitization framework.

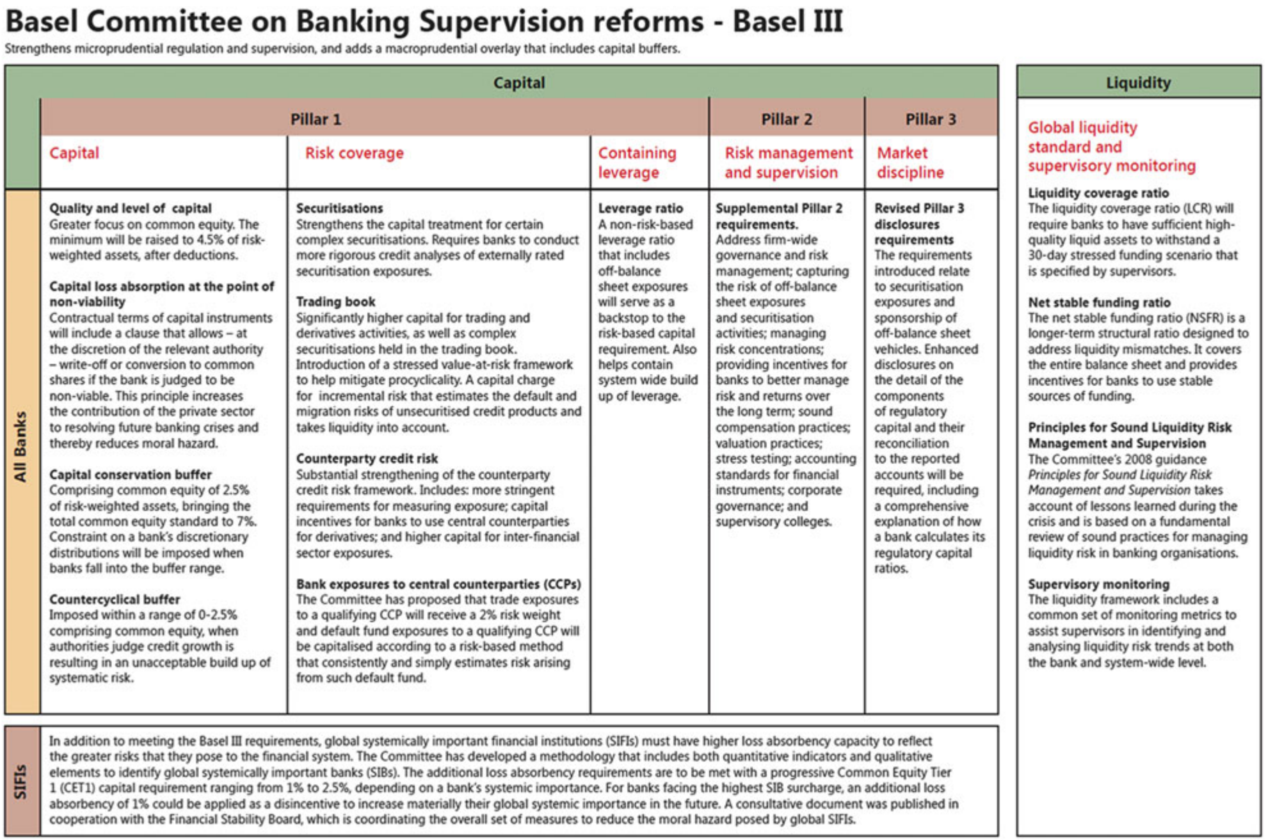


Fig. 2.6 An overview of the new Basel III requirement (BCBS 2010)

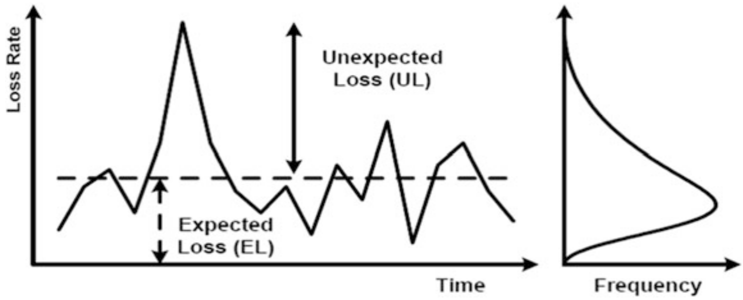


Fig. 2.7 Expected versus unexpected loss (Source: BCBS 2005a)

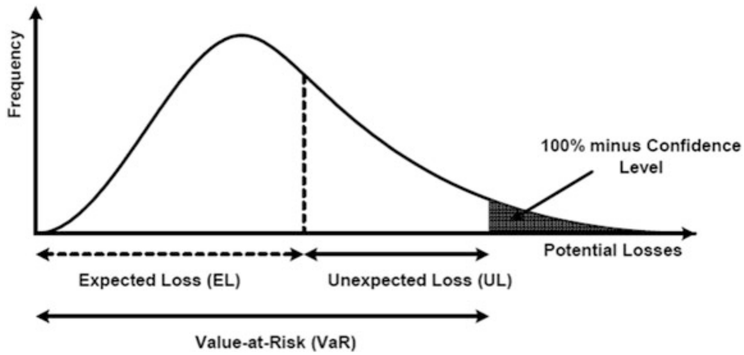


Fig. 2.8 Expected and unexpected losses (Source: BCBS 2005a)

The main guiding principle of the Basel regulation can be seen in Fig. 2.7. Banks regularly suffer (annual) credit, operational, and market expected, or unexpected, losses. The expected part of the losses, based on a long term history, should be covered from annual revenues. The unexpected part, if not covered by the revenues, must be charged against the capital. In good years, the actual losses may be lower than expected, but in bad years, the bank needs a sufficient capital buffer. Hence the goal of the regulation is to set up a procedure estimating the potential unexpected loss on a regulatory probability level. The task is mathematically depicted in Fig. 2.8.

The Value at Risk approach had already been incorporated into the 1996 Market Risk Amendment. Banks may either follow the standardized approach, approximating the market Value at Risk (VaR) through a precisely prescribed calculation procedure, or may use the Internal Model Approach (IMA), where the market VaR is calculated using their own model subject to regulatory approval. Many banks already used some form of internal market VaR model before the implementation of the Amendment and have recently obtained or applied for regulatory approval for the Internal Model Approach.

The new operational risk requirement allows a simplified approach (basic indicator or standardized), or even an internal operational Value at Risk (VaR) model called the Advanced Measurement Approach (AMA). To obtain AMA regulatory approval, banks must collect sufficient historical operational loss data (5 years) that are complete and consistent. The banks must demonstrate to the regulator that potentially severe "tail" loss events at the 99.9 % probability level are covered by the outcome of the internal model. The expected losses are in this case included in the capital requirement due to the nature of operational risk. Currently, most banks apply just the basic indicator or the standardized approach, as the loss data have not been usually stored in the required form in the past, and the operational VaR models are still being investigated and discussed by researchers and practitioners.

The credit risk regulatory capital calculation has undergone significant changes compared to the 1988 Accord. In the original approach, each on-balance sheet, or adjusted off-balance sheet, exposure has been classified into four broad risk buckets and multiplied by a 0, 20, 50, or 100 % coefficient to obtain the risk weighted assets (RWA). In particular, all corporate or retail assets fall into the 100 % bucket without any differentiation. The capital adequacy ratio (CAR), required to be at least 8 %, has been then calculated as the total capital divided by the total risk weighted assets:

$$CAR = \frac{\text{available capital}}{RWA}.$$

Equally, the regulatory capital requirement could be calculated as 8 % times the RWA calculated separately for each individual exposure. The New Accord preserves this portfolio invariant approach (where the capital requirement for an exposure does not depend on the portfolio which it belongs to), but significantly refines risk differentiation in the risk weight determination.

The Standardized Approach uses in principle five buckets (0, 20, 50, 100, 150 %), but more importantly allows the use of ratings obtained from external agencies. The corporate assets may thus fall into four different categories (20, 50, 100, and 150 %) according to their risk rating. Moreover the risk weight for retail assets may be reduced to 75 %, or even 35 %, in the case of residential exposures subject to regulatory approval.

The New Accord does not allow banks to use a full scope credit risk internal model. The advanced approach is, rather, based on a regulatory formula, or a set of regulatory formulas, where certain input parameters are estimated by internal models. The outcomes of the formulas are risk weights, or equivalently capital charges calculated separately per each exposure. The only parameter to be estimated by banks in the Foundation Internal Rating Based Approach (IRBF) is the probability of default (PD). The PD parameter is to be derived from an internal credit rating system with assigned expected probabilities of default calibrated to the historical data. Other key parameters are the Loss Given Default (LGD), Exposure at Default (EAD), or the closely related Conversion Factor (CF), and Effective Maturity (M). The remaining parameters used by the formulas as correlations, or

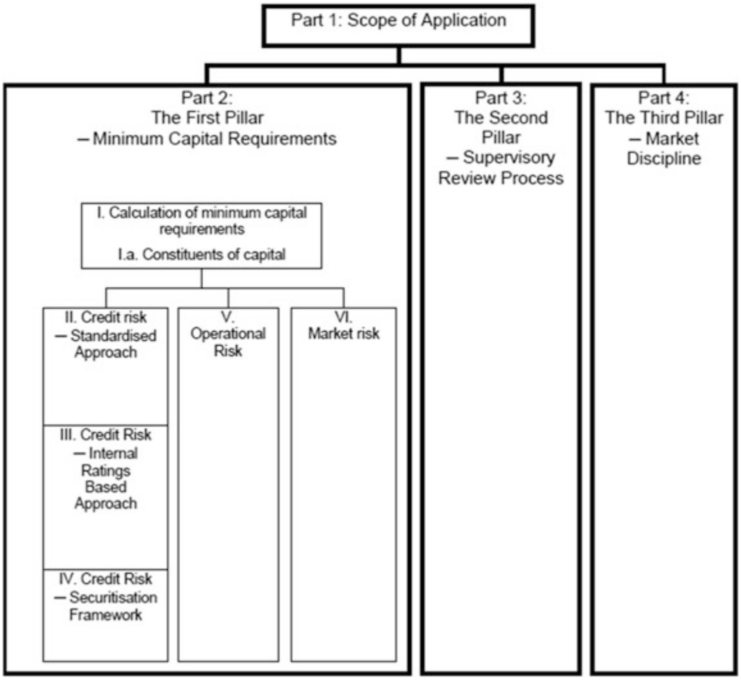


Fig. 2.9 Structure of the Basel II regulation (Source: BCBS 2005a)

maturity adjustments are set directly by the regulation. The LGD and CF parameters are defined by the regulation in the Foundation Approach, and may be estimated by an internal model in the Advanced Internal Rating Based Model (IRBA)). Corporate exposures admit both the IRBF and IRBA approaches, while retail exposures admit only IRBA if the internal rating based approach is to be used. Therefore, application of IRB to retail exposures means one's own estimation of all the three parameters; PD, LGD, and EAD. The internal parameter estimation models must satisfy a number of requirements (minimum standards) set by the regulator in terms of data quality, length of the observation period, structure, methodology etc., and must obtain regulatory approval.

The estimation of the PD, LGD, and EAD parameters will be discussed in more detail in Chap. 3. Let us now focus on the qualitative risk management requirements laid down recently by the BIS regulation. The overall structure of the regulation is shown in Fig. 2.9. The qualitative credit risk management requirements have been formulated in particular as the minimum requirements for the IRB approach, and within the second Pillar for the supervisory review process. A detailed document on the internal banking control system BCBS (1998) has been, to a certain extent, incorporated into the Revised Framework.

Regarding the independence of the credit risk function, the regulation (BCBS 2006a) clearly says:

441. *Banks must have independent credit risk control units that are responsible for the design or selection, implementation and performance of their internal rating systems. The unit(s) must be functionally independent from the personnel and management functions responsible for originating exposures. Areas of responsibility must include:*

- *Testing and monitoring internal grades;*
- *Production and analysis of summary reports from the bank's rating system, to include historical default data sorted by rating at the time of default and one year prior to default, grade migration analyses, and monitoring of trends in key rating criteria;*
- *Implementing procedures to verify that rating definitions are consistently applied across departments and geographic areas;*
- *Reviewing and documenting any changes to the rating process, including the reasons for the changes; and*
- *Reviewing the rating criteria to evaluate if they remain predictive of risk. Changes to the rating process, criteria or individual rating parameters must be documented and retained for supervisors to review.*

The responsibility of the management board is defined in detail in the Supervisory Pillar, for example:

730. *The bank's board of directors has responsibility for setting the bank's tolerance for risks. It should also ensure that management establishes a framework for assessing the various risks, develops a system to relate risk to the bank's capital level, and establishes a method for monitoring compliance with internal policies. It is likewise important that the board of directors adopts and supports strong internal controls and written policies and procedures and ensures that management effectively communicates these throughout the organization.*

We could continue citing many paragraphs of the regulation related to the previous sections. Other qualitative requirements will be explained when we discuss the quantitative part of the Basel document in the following chapter.

3 Rating and Scoring Systems

The main goal of the credit assessment process is to approve acceptable loan applications, reject clients that will probably default in the future, and, moreover, set up loan pricing so that the credit losses are covered by collected credit margins. This does not have to be necessarily achieved through a rating system. Nevertheless, the approach of assigning a credit grade on a finite scale to each client, and/or exposure, has become a general keystone of the modern credit risk management process. The rating can be obtained in many different ways. There are external agencies like Standard & Poor's and Moody's which have been assigning ratings to bond issuers, large corporations, and countries for more than 100 years. Ratings can be also produced by banks internally in many different ways—by experienced credit analysts, by statistical, or even artificial intelligence methods, or by a combination of human and machine assessment. Before we start describing in detail any particular methods, we should, first of all, define our expectation of a rating system, and the methods used to measure how well the expectations are met.

3.1 Rating Quality Measurement and Validation

A general rating scheme is shown in Fig. 3.1. It starts with a client or facility information (e.g. bond issue, or a new transaction under consideration) and produces a rating/scoring on a scale. Well-known examples of scales are those of the external rating agencies, e.g., Standard & Poor's, starting with the worst non-default rating: **C**, and going up to the best rating **AAA**, or the very similar one of Moody's, starting with **C** and going up to the best **Aaa**. Some banks denote rating classes by numbers, often starting with 1 denoting the best grade, and higher numbers denoting worse grades. In the following we will use the reversed numbering as a standard scale, which is also used in practice, with 1 denoting the worst grade, and higher numbers the better credit grades. The number of grades may vary. Generally, rating scales do not have many grades—typically somewhere between 7 and 25. Ratings assigned in the standard numbering are alternatively called rating scores. As we shall see,

J. Witzany, *Credit Risk Management*, DOI 10.1007/978-3-319-49800-3_3

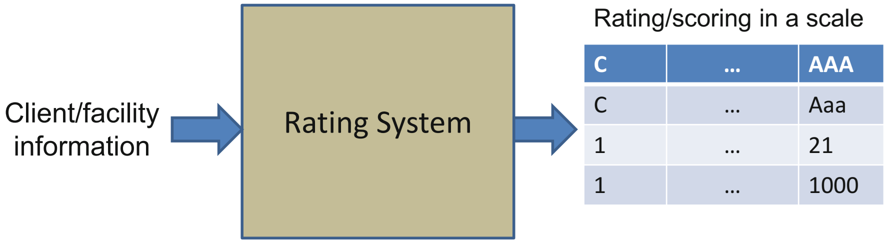


Fig. 3.1 General rating process diagram

statistical rating systems also typically produce a score on a much finer scale, e.g. 1–1000. The scoring itself could be used as a rating, but it is usually transformed into a rating scale with a much smaller number of grades.

The "Rating System" may include any of the methods mentioned, or it may even be an oracle. Yet our question is: "Does the rating perform well or not?" To answer the question we firstly have to specify our definitions and expectations:

- Do we seek an assessment of credit risk in a short time period (e.g. a 1-year horizon) or over a long term horizon?
- Does the rating assess the possibility of default on any of the client's obligations, or does it deal with only one credit facility (e.g. a bond issue)?
- Is the rating based on the client's existing debt, or is it conditional on a new loan the client is applying for?
- Do the rating grades represent any specific probabilities of default or not?
- How do we define default (legally proclaimed bankruptcy, any delay in payments, delay in payments exceeding 90 days, etc.)?

It is clear that as we do not know the future, we can judge the quality of a "black box" rating system based only on its real, observable performance. We want to measure the performance of a given rating system, and to validate, i.e. verify, that it has satisfied our expectations. Let us assume, for example, that the risk horizon is 1 year. Then we need to collect ratings produced by the systems, on a set of non-defaulted clients at time T, and wait until time $T+1$ to record defaulted and non-defaulted clients. The same observation could be made at other times (see Fig. 3.2) forming so called cohorts.

Then we put all debtors, or exposures rated in the cohorts, into the set of initially rated observations, and put the final state into the set of observed defaults and non-defaults, and analyze the number of observed defaults per individual rating class. Our expectation is that there should be relatively fewer defaults on good credit ratings than on worse credit ratings. For each rating: $s = 1, \ldots, N$, we calculate the corresponding observed default rate simply as:

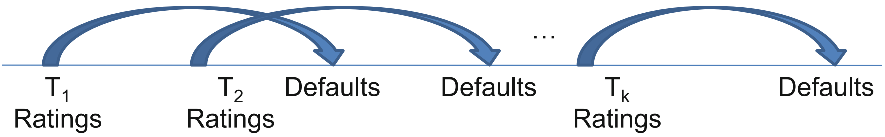


Fig. 3.2 Historical observations of ratings and defaults

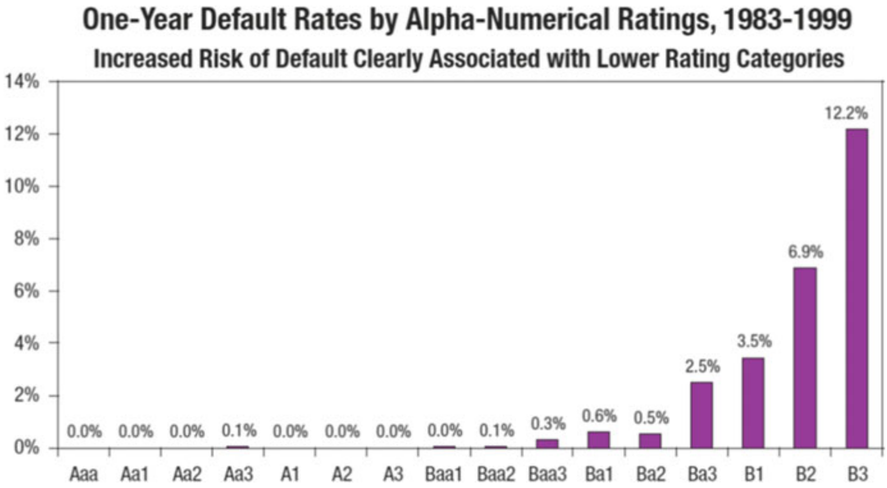


Fig. 3.3 Historical default rates by Moody's ratings Aaa-B3 (Source: moodys.com)

$$p_s = \frac{d_s}{n_s} \tag{3.1}$$

where n_s is the number of observations with rating s, and d_s is the number of defaults recorded within the rating class. For example, Fig. 3.3 shows historical 1-year default rates for Moody's rating grades Aaa-B3, and we can use it to assess visually the performance of Moody's rating system. To calculate the default rate, the study defines cohorts for each grade and for each year in the period as the set of corporate issuers with the given grade at the beginning of the year, and the defaults (based on legal bankruptcy and/or missed payments) are counted at the end of the year, hence there is no overlap, as in Fig. 3.2. The default rate for a grade over the whole period is defined as the total number of defaults divided by the sum of counts of observed issuers in the cohorts according to (3.1).

Looking at Fig. 3.3 we might be quite satisfied, as better ratings generally show much lower observed default rates than worse ratings. Nevertheless, there are a few signs of imperfection; in particular, the rating Aa3 shows a higher default rate than the better ratings A1, A2, A3. Likewise, Ba1 shows a higher rate than Ba2. Intuitively we tend to say that those are just statistical errors, but how do we measure the quality of the system more precisely? If there were explicit probabilities of default assigned to the ratings, we could calculate some kind of average deviation between the expected and observed rate of default. Since the probabilities of default are not usually explicitly assigned to the ratings of external agencies, let us look first at the more difficult case of measuring just the discrimination power of a rating system.

Discrimination Power of Rating Systems

There is a variety of goodness-of-fit measures described in literature. We shall focus on the two generally used in banking practice; i.e., the Accuracy Ratio (AR), also called Somers' D or Gini's Coefficient, which is obtained from the Cumulative Accuracy Profile (CAP), and AUC—Area under the Receiver Operating Characteristic Curve (ROC). There is extensive literature on the subject, including the signal detection theory and medical literature, but we shall mostly follow the study related to the validation of rating systems by Engelman et al. (2003). The two indicators can be defined and, naturally, interpreted in the language of probabilities, where we use ratings to discriminate ex-ante between bad and good debtors, defaulting and not defaulting in the rating time horizon.

Given a good debtor X and a bad debtor Y (the information becoming known ex-post) we say that the rating successfully discriminated (gave a correct signal) between X and Y, if the ex-ante rating(X) > rating(Y). On the other hand we say that the rating was not successful (gave a wrong signal) if rating(X) < rating(Y), and finally we say that the rating did not discriminate (gave no signal) if rating (X) = rating(Y). Let p_1 denote the probability of the successful discrimination of an arbitrary good X, and an arbitrary bad Y; p_2 the probability of wrong discrimination, and p_3 the probability of no discrimination, i.e.

$$\begin{aligned} p_1 &= \Pr\left[\text{rating}(X) > \text{rating}(Y) \middle| X \text{ is good},\ Y \text{ is bad}\right], \\ p_2 &= \Pr\left[\text{rating}(X) < \text{rating}(Y) \middle| X \text{ is good},\ Y \text{ is bad}\right], \\ p_3 &= \Pr\left[\text{rating}(X) = \text{rating}(Y) \middle| X \text{ is good},\ Y \text{ is bad}\right]. \end{aligned}$$

Then, we can define theoretically the accuracy ratio AR as the probability of good discrimination minus the probability of wrong discrimination

$$\text{AR} = p_1 - p_2$$

assigning "negative points" to wrong decisions. The definition above is, in fact, equivalent to the Somers' D-concordance statistic (generally defined for two ordinal valued random variables) and it can be shown to be equal to the Accuracy Ratio as we prove later. Similarly, we can define AUC as the probability of good discrimination, plus one half of the probability of no discrimination:

$$\mathrm{AUC} = p_1 + \frac{p_3}{2},$$

as if our strategy would be to toss a coin if the rating gave us no hint. Thus, AUC measures the ratio of good decisions from all attempts, including the coin tossing strategy. Since $p_1 + p_2 + p_3 = 1$, it is straightforward to verify that $\mathrm{AUC} = \frac{1}{2}(1 + \mathrm{AR})$, so the AUC is just a linear transformation of AR, and vice versa. Note that AUC is always a value between 0 and 1, while the Accuracy Ratio generally takes values between -1 and 1. If the discrimination by rating were always correct, then $\mathrm{AR} = 1 = \mathrm{AUC}$. If it were always incorrect, then $\mathrm{AR} = -1$ while $\mathrm{AUC} = 0$. And if it did not discriminate at all, assigning the same rating to all debtors or making the same number of correct and incorrect decisions, then $\mathrm{AR} = 0$ while $\mathrm{AUC} = 0.5$.

The two measures are usually primarily defined geometrically and some effort is necessary to show that the probabilistic definitions are equivalent to the geometrical ones. The accuracy ratio can be defined from the Cumulative Accuracy Profile (CAP) curve, closely related to the Lorenz curve used to measure income disparity. Let us assume that our rating scale is $\{1,\ldots,N\}$. The CAP curve (Fig. 3.4) is defined as the line in the rectangle $[0, 1] \times [0, 1]$ starting from the origin and linearly connecting the points (x_s, y_s), $s = 1, \ldots, N$ where

$$x_s = F(s) = \Pr\left[\mathrm{rating}(X) \le s \middle| X \text{ is a rated debtor}\right]$$

denotes the cumulative proportion of all debtors with a rating not better than s, while

$$y_s = F\left(s \middle| B\right) = \Pr\left[\mathrm{rating}(Y) \le s \middle| Y \text{ is a rated bad debtor}\right]$$

is the cumulative proportion of bad debtors with a rating not better than s.

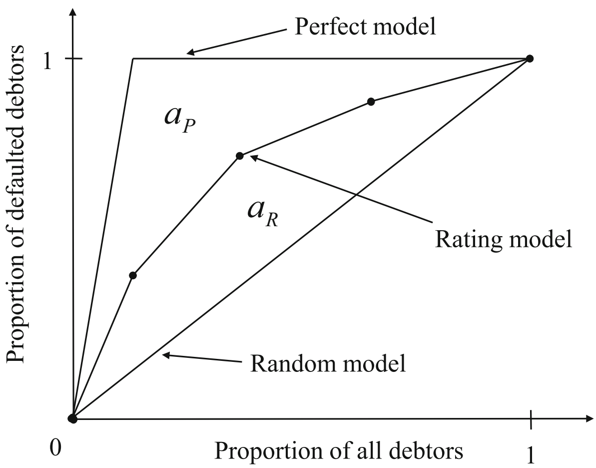


Fig. 3.4 Cumulative accuracy profile

The “random model” in Fig. 3.4 corresponds to the situation where the rating does not discriminate between bad and good clients at all, as if we always just tossed a coin. On the other hand the “perfect model” would assign the lowest rating score to all bad debtors, while all the remaining good debtors would be assigned the next rating score, i.e. $x_1 = \pi, y_1 = 1$, where π is the overall probability of default. Since the goal is to get as close as possible to the perfect model, it is natural to measure the quality of the rating system by the proportion $\frac{a_R}{a_P}$, where a_R is the area between the random model curve (diagonal) and the rating model curve while $a_P = \frac{1}{2} - \frac{\pi}{2}$ is the area between the random model curve and the perfect model curve. We will show at the end of this subsection that this definition is equivalent to the probabilistic definition of the Accuracy Ratio; i.e. $\mathrm{AR} = \frac{a_R}{a_P}$.

A similar curve frequently used in rating validation studies is the ROC (Receiver Operating Characteristic curve used in the signal detection theory). An example of the ROC curve is shown on Fig. 3.5. In the context of rating, the “false alarm rate” is the relative proportion of clients with a low score among all good clients; i.e. for a given score the value is

$$x_s = F(s|G) = \Pr[\text{rating}(X) \le s|X \text{ is good}],$$

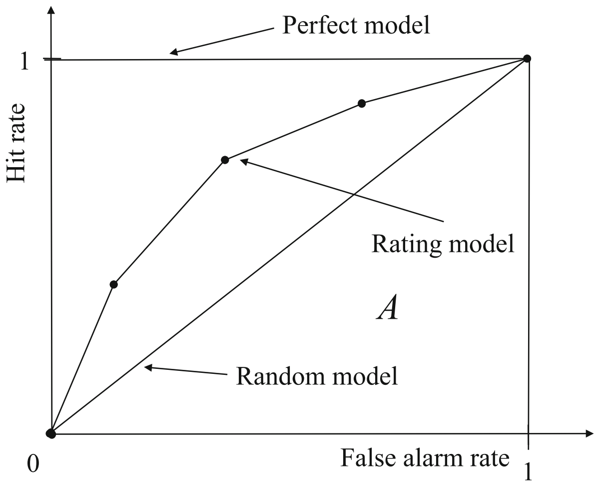
while the “hit rate” is the relative proportion of clients with a low score among all bad clients, i.e.

$$y_s = F(s|B) = \Pr[\text{rating}(Y) \le s|Y \text{ is bad}].$$

The ROC curve then connects linearly the origin and the points (x_s, y_s), $s = 1, \ldots, N$.

The random model ROC curve is again the diagonal, while the perfect model ROC curve connects the origin, the point (0, 1), and the point (1, 1). Thus, it is also

Fig. 3.5 Receiver operating characteristic curve



natural to measure the quality of the rating by the Area under ROC, denoted A in Fig. 3.5, and we will show that this is the equivalent to the probabilistic definition of AUC. Recalling that $\mathrm{AR} = 2\left(\mathrm{AUC} - \frac{1}{2}\right)$ we see that the Accuracy Ratio also equals twice the area between the rating model and the random model ROC curves.

The following *proof of equivalence of the probabilistic and geometric definitions* for AR and AUC can be omitted by the less mathematically inclined reader. Let us start with the area A shown in Fig. 3.5. It can be expressed (defining $x_0 = y_0 = 0$ and $r(X) = \mathrm{rating}(X)$) as

$$\begin{aligned}
A &= \sum_{s=1}^{N} \frac{1}{2}(y_{s-1} + y_s)(x_s - x_{s-1}) = \\
&= \sum_{s=1}^{N} \frac{1}{2}\left(\Pr\left[r(Y) \le s-1 \middle| Y \text{ is bad}\right] + \Pr\left[r(Y) \le s \middle| Y \text{ is bad}\right]\right) \\
&\quad \cdot \Pr\left[r(X) = s \middle| X \text{ is good}\right] \\
&= \sum_{s=1}^{N} \left(\Pr\left[r(Y) \le s-1 \middle| Y \text{ is bad}\right] + \frac{1}{2}\Pr\left[r(Y) = s \middle| Y \text{ is bad}\right]\right) \\
&\quad \cdot \Pr\left[r(X) = s \middle| X \text{ is good}\right] \\
&= \sum_{s=1}^{N} \Pr\left[r(Y) \le s-1,\ \ r(Y) = s \middle| Y \text{ is bad},\ \ X \text{ is good}\right] \\
&\quad + \frac{1}{2}\sum_{s=1}^{N} \Pr\left[r(Y) = s,\ \ r(X) = s \middle| Y \text{ is bad},\ \ X \text{ is good}\right] \\
&= p_1 + \frac{1}{2} p_3 = AUC.
\end{aligned}$$

Regarding AR, and the definition based on the CAP curve, we can proceed similarly but we also need to use the fact that

$$\Pr[r(Y) = s] = (1-\pi)\Pr\left[r(X) = s \middle| X \text{ is good}\right] + \pi\Pr\left[r(Y) = s \middle| Y \text{ is bad}\right]$$

as $\pi = \Pr[X \text{ is bad}]$ is the ex-ante probability of default (i.e., being bad). Looking at Fig. 3.4 we see that

$$a_R + \frac{1}{2} = \sum_{s=1}^{N} \frac{1}{2}(y_{s-1} + y_s)(x_s - x_{s-1}) =$$

$$= \sum_{s=1}^{N} \frac{1}{2}\left(\Pr\left[r(Y) \le s-1 \middle| Y \text{ is bad}\right] + \Pr\left[r(Y) \le s \middle| Y \text{ is bad}\right]\right) \cdot \Pr[r(X) = s]$$

$$= (1-\pi)\sum_{s=1}^{N} \frac{1}{2}\left(\Pr\left[r(Y) \le s-1 \middle| Y \text{ is bad}\right] + \Pr\left[r(Y) \le s \middle| Y \text{ is bad}\right]\right)$$

$$\cdot \Pr\left[r(X) = s \middle| X \text{ is good}\right] + \pi \sum_{s=1}^{N} \frac{1}{2}\left(\Pr\left[r(Y) \le s-1 \middle| Y \text{ is bad}\right] + \Pr\left[r(Y) \le s \middle| Y \text{ is bad}\right]\right)$$

$$\cdot \Pr\left[r(Y) = s \middle| Y \text{ is bad}\right]$$

$$= (1-\pi)AUC + \pi \sum_{s=1}^{N} \frac{1}{2}\left(\Pr\left[r(Y) \le s \middle| Y \text{ is bad}\right] + \Pr\left[r(Y) \le s-1 \middle| Y \text{ is bad}\right]\right)$$

$$\cdot \left(\Pr\left[r(Y) \le s \middle| Y \text{ is bad}\right] - \Pr\left[r(Y) \le s-1 \middle| Y \text{ is bad}\right]\right)$$

$$= (1-\pi)AUC + \pi \sum_{s=1}^{N} \frac{1}{2}\left(\left(\Pr\left[r(Y) \le s \middle| Y \text{ is bad}\right]\right)^2 - \left(\Pr\left[r(Y) \le s-1 \middle| Y \text{ is bad}\right]\right)^2\right)$$

$$= (1-\pi)AUC + \frac{1}{2}\pi.$$

Since $a_P = \frac{1}{2} - \frac{\pi}{2}$, and we already know that $\text{AUC} = \frac{1}{2}(1 + \text{AR})$, using just the probabilistic definitions, it is now easy to verify that indeed: $\text{AR} = \frac{a_R}{a_P}$ *(end of proof)*.

Finally, let us mention some other measures of rating systems' discrimination power, like the Kolmogorov-Smirnov, Kruskall-Wallis, and Kendall's τ statistics. Of those, the most frequently used, the Kolmogorov-Smirnov statistic (KS), which can be defined as $KS = \max_s |x_s - y_s|$ where $(x_s, y_s) = \left(F\left(s \middle| G\right), F\left(s \middle| B\right)\right)$, $s = 1, \ldots, N$ are the points on the ROC curves; i.e. the coordinates are relative proportions of good clients and bad clients. It is easy to see that $|x_s - y_s|$ is in fact the distance of the point (x_s, y_s) from the diagonal (see Fig. 3.5), multiplied by $\sqrt{2}$, and so $KS/\sqrt{2}$ is the maximum distance of the ROC curve from the diagonal. Hence, the KS statistic is similar but less comprehensive, compared to the AR and AUC statistics. Therefore, if computational power allows us to calculate AR or AUC, we should rather use one of those.

Empirical Estimations of AR and AUC

The analyzed ideal definitions of AR and AUC must be in practice numerically approximated, as we have only a finite sample of rated borrowers with observed defaults (validation sample) and not the full population sample. The estimates AR or AUC may be calculated using the geometric or probabilistic definitions, and one has to realize that the estimate generally differs from the ideal value and certainly depends on the validation sample used. If the rating system is developed on a historical sample, then its Gini coefficient (AR) measured on the training sample will normally be higher than the coefficient calculated on another sample. Generally, we prefer *out-of-sample* validation (calculating the measurements on a sample different from the training sample). Optimally, the default observations of the testing sample should really be done in the period following the rating assignment. For example, this is the case of Gini's coefficient reported in Standard & Poor's Annual Global Corporate Default Study (Apr 2014), and shown in Fig. 3.6, indicating a relatively very good ex post performance of the rating decisions. The high Gini, at around 90%, means that the rating agency downgrades the issuers already 1 year before default in the majority of cases.

Given the estimate of the Gini coefficient AR, we should also look at the confidence interval for the real coefficient AR on the confidence level α, using an estimator given, e.g., in Engelman et al. (2003), as well as also being automatically calculated by many statistical packages. The confidence interval could be quite large if the validation sample is too small with just a few defaults. The S&P report, unfortunately, does not explicitly show any confidence intervals, but the time series in Fig. 3.6 indicates that the interval can be relatively large; e.g., somewhere between 80 and 95%, depending on the confidence level. Confidence intervals could also be used for a simple comparison of real Gini coefficients (AR_1 and AR_2) based on estimates AR_1 and AR_2. If $AR_1 < AR_2$, and the two intervals with confidence level α do not intersect each other, then we can conclude that $AR_1 < AR_2$;

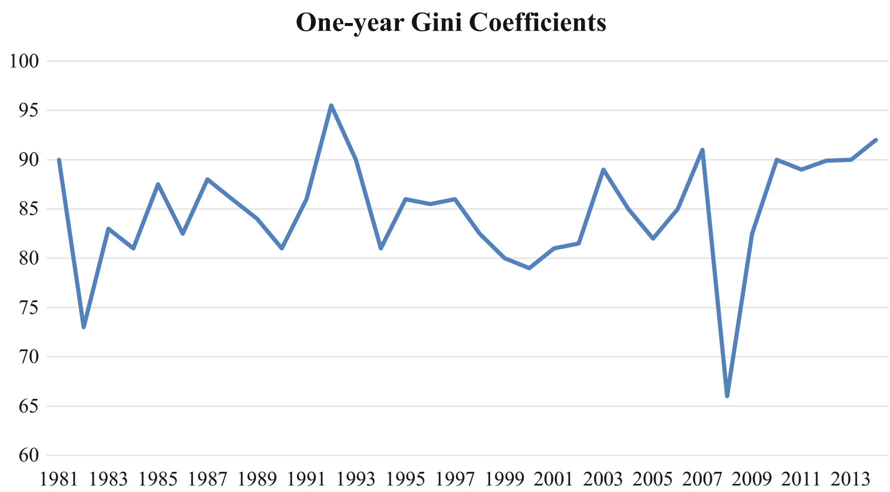


Fig. 3.6 Performance of S&P corporate ratings measured by the Gini coefficient (Based on data reported by Standard & Poor's CreditPro®)

or more precisely, we reject the hypothesis $AR_1 \geq AR_2$, at least on probability level α. In this case we do not take into account the correlation of the estimators, which is usually positive, in particular, if the two coefficients are calculated on the same validation sample. In this case, we can use a better T-statistic in the form:

$$T = \frac{(AR_2 - AR_1)^2}{\hat{\sigma}_1^2 + \hat{\sigma}_2^2 - 2cov_{1,2}}, \tag{3.2}$$

where the variances and covariance are estimated by a formula given in the aforementioned paper (for AUC, and equally for AR), or by advanced statistical packages. The statistic is asymptotically χ^2-distributed with one degree of freedom, provided the null hypothesis $H_0 : AR_2 - AR_1 = 0$ holds.

Example Table 3.1 shows Gini coefficients of two rating systems calculated on the same validation sample. Such output for Gini coefficients, or AUC, is provided by many statistical packages (Stata, SPSS, SAS, etc.).

The 95% confidence intervals are calculated using the reported standard errors, and with the asymptotic normality assumption. The coefficients are relatively close, the confidence intervals overlap, and so we cannot compare the quality of the two rating systems, even on the 95% probability level. The package, however, reports also the T-statistic (3.2) denoted as $\chi^2(1)$. The reported value corresponds to the shown standard errors $\hat{\sigma}_1$ and $\hat{\sigma}_2$, and the correlation 0.8, i.e.

$$T = (71\% - 69\%)^2 / \left(1.2\%^2 + 1.3\%^2 - 2 \times 0.8 \times 1.2\% \times 1.3\%\right) \doteq 9.86.$$

Since the probability of values larger than 9.86 for a χ^2-distribution with one degree of freedom is 0.17% we can reject the null hypothesis H$_0$: $AR_1 - AR_2 = 0$ on the probability level 99% or even higher. Similarly for any $a > 0$, we can reject the hypothesis H$_0$: $AR_1 - AR_2 = a$. In that sense, we can reject $AR_1 \geq AR_2$ on the 99% probability level, and conclude that the second rating has a better performance than the first one on that probability level, though the difference between the Gini coefficients is relatively small.□

Table 3.1 Gini coefficients for two rating systems

	Gini (%)	St. error (%)	95% Confidence interval (%)	
Rating1	69	1.2	66.65	71.35
Rating2	71.50	1.30	68.95	74.05

H_0: Gini(Rating1) = Gini(Rating2)
$\chi^2(1) = 9.89$
Pr[$> \chi^2(1)$] = 0.17%
Source: Stata

Measures of the Correctness of Categorical Predictions

The rating systems are in practice, especially for the retail segment, used in connection with a cut-off rating score s_c to accept or reject loan applications: all applications with the rating score $\leq s_c$ are rejected, while those with the rating score $> s_s$ should be approved. The cut-off score could be determined by a marketing strategy, or by optimizing the expected overall profit on the product. For example, if the bank offers credit cards with a competitive market interest rate, then the cut-off s_c should correspond to the marginal probability of default (e.g. 10%) that still allows the cards to be given to debtors, with ratings around s_c remaining profitable. If the cut-off s_c is given, then, rather than in AR or AUC, we are interested in certain error measurements, or the overall cost of the errors incurred by the approval process, when some bad debtors might be approved, while some good debtors could be rejected.

In fact the problem can be formulated as one of classification: the goal is, given ex-ante applications, to predict good (approve) and bad (reject) applications as accurately as possible. Given a testing sample, the *classification accuracy* (ACC) can be simply defined as the number of accurate predictions (good predicted as good and bad as bad) divided by the number of all cases. The classification can be done even without any rating or scoring system (see for example classification trees in Sect. 3.3), and in that case the accuracy is a straightforward and intuitive performance measure. In the case of rating and scoring systems it is quite ambiguous, since it depends on the selected cut-off. Another disadvantage of ACC is that it might depend on the testing sample proportion of good and bad. For example, a "silly" classification system that predicts all cases as good will have $ACC = n_G/n$ where n_G is the number of good and n the number of all cases.

The overall situation can be characterized in more detail by the so called *confusion matrix* (Thomas 2009), where we split the total number of applications into actual goods and bads $n = n_G + n_B$; the actual goods into those that are predicted as good (approved), and those that are predicted as bad (rejected) $n_G = g_G + b_G$, and the actual bads into those that are predicted as good (approved) and those that are predicted as bad (rejected) $n_B = g_B + b_B$, see Table 3.2.

We are interested in the Type I error when the hypothesis is true (the borrower is good), and the system predicts otherwise (the borrower is rejected). The probability of the Type I error assessing actual goods based on the confusion matrix is b_G/n_G; alternatively g_G/n_G is called the *positive predictive value* or *precision*. Similarly, a Type II error is when the hypothesis is false (the applicant is bad), but the system predicts that the hypothesis is true (the loan is granted). The probability of the Type

Table 3.2 Confusion matrix

	Actual goods	Actual bads	Total
Approved	g_G	g_B	g
Rejected	b_G	b_B	b
Actual numbers	n_G	n_B	n

Table 3.3 Confusion matrix in terms of the underlying probabilities

	Actual goods	Actual bads	Total
Approved	$\pi_G F^c(s_c\|G)$	$\pi_B F^c(s_c\|B)$	$F^c(s_c)$
Rejected	$\pi_G F(s_c\|G)$	$\pi_B F(s_c\|B)$	$F(s_c)$
Total actual	π_G	π_B	1

II error is g_B/n_B; alternatively n_B/n_B is called the *negative predictive value*. Given the confusion matrix, the accuracy can be calculated as $(g_G + b_B)/n$.

The confusion matrix can also be represented in terms of the underlying cumulative proportions of good and bad debtors, $F(s_c|G)$ and $F(s_c|B)$, used to build the ROC, and the ex-ante probabilities of borrowers being good and bad, π_G and π_B—see Table 3.3. The proportion of good predicted as bad is, in fact, equal to the false alarm rate $F(s_c|G)$, while the proportion of bad predicted as good equals one minus hit rate, i.e. $F^c(s_c|B) = 1 - F(s_c|B)$. The relative total number of Type II and Type I errors out of all cases is then $ER = \pi_B F^c(s_c|B) + \pi_G F(s_c|G)$ and the classification accuracy can be expressed as $ACC = \pi_G F^c(s_c|G) + \pi_B F(s_c|B)$.

This approach is more practical, since normally we do not know exactly the actual numbers of goods and bads of rejected applicants in a validation sample and we also do not have the information on defaults on the rejected applications. Nevertheless, we do know the cumulative distributions functions $F(s|G)$ and $F(s|B)$, and we may estimate the overall probabilities π_G and π_B; e.g. based on the scores that should be available for all applicants in the sample.

Example Let us assume that we have a scorecard assigning scores in the range from 0 to 100, and that the proposed cut-off score is 50. The validation sample of historical applications has 10,000 observations, out of which 9000 have been approved, and the others rejected. Our scorecard has been developed on another (training) sample, and we want to estimate the error rates on the validation sample representing a hypothetical set of new applications. The cumulative proportion of goods and bads can be calculated on the 10000 validation sample cases, where we have the information on defaults. Let us say that the cut-off score false alarm rate is $F(50|G) = 14\%$, while the hit rate is $F(50|B) = 76\%$, and so the probability of approving a bad borrower is 24%. Moreover, based on the scores assigned to all applicants in the validation sample, we estimate that the overall proportion of bads is $\pi_B = 10\%$, and the proportion of goods is $\pi_G = 90\%$. Consequently, the total rate of error can be calculated as

$$ER = 10\% \times 24\% + 90\% \times 14\% = 15\%. \square$$

The example above indicates that the Type II error (approval of a bad applicant) has generally a lower weight than the Type I error (rejection of a good applicant). The situation is, in practice, even a little more complicated, since the loss caused by an approved bad application generally differs from the loss caused by rejection of a good application. Here we mean the opportunity cost with respect to a perfect approval system, which accepts all the good applications, and rejects all the bad ones. Let us say that the net loss on a bad loan is l and the potential net profit on a good loan is q, assuming that all the loans are of the same unit size. Then the weighted cost of errors can be expressed as

$$WCE = l\pi_B F^c(s_c|B) + q\pi_G F(s_c|G) = q\pi_G(C \cdot F^c(s_c|B) + F(s_c|G)),$$

where: $C = \frac{l\pi_B}{q\pi_G}$. If the parameters l and q do not depend on the cut-off score (in practice they could, since a change of cut-off should lead to a change in pricing policy), we may try to minimize the weighted cost of errors, i.e. effectively the function

$$w(s_c) = C \cdot F^c(s_c|B) + F(s_c|G) = C - C \cdot F(s_c|B) + F(s_c|G).$$

Note that if $C = 1$, we in fact need to maximize $F(s_c|G) - F(s_c|B)$, which leads exactly to the maximization of the KS statistic. Generally, we need to find the score s, where $w'(s) = 0$; i.e. the point $(x_s, y_s) = (F(s|G), F(s|B))$ on the ROC curve, and where the slope of the tangent line equals exactly $1/C$. Figure 3.7 shows the shift along the ROC, when the profit is maximized just as the tangent line has the required slope, hence, not exactly at the point with the maximum distance from the diagonal, as in case of the Kolmogorov-Smirnov statistic. If the goal is to maximize the volume granted, which can be expressed as $1 - (\pi_B F(s_c|B) + \pi_G F(s_c|G))$, out of the total number of applications, then we just need to move down to the left to the origin A where all applications are approved. However, if we want to minimize future losses on defaulted loans that can be estimated as $l\pi_B F^c(s_c|B)$ out of the total number of applications, we need to move up, i.e. in the direction to point C, where all the applications are, in fact, rejected.

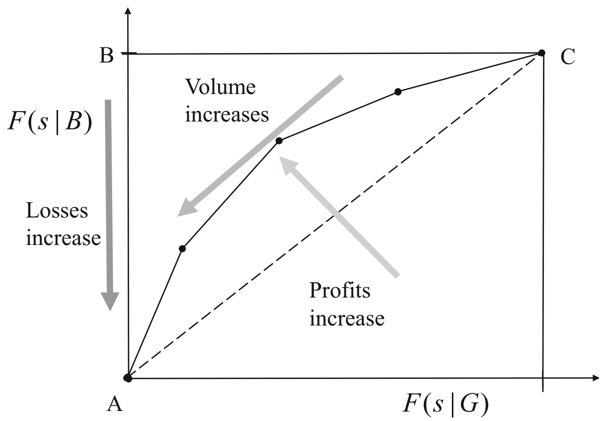


Fig. 3.7 Profit, volume, and losses when moving along the ROC curve

Measurement of Accuracy and Validation of PD Estimates
Let us now consider the case when the rating system produces not only the rating scores on the scale 1,. . .,*N,* but also forecasted probabilities of default $PD_1, \ldots, PD_N$ assigned explicitly to the rating grades. Given a validation sample, we calculate the realized default rates $p_s = d_s/n_s,\ \ s = 1, \ldots, N,$ where n_s is the number of observations, and d_s the number of defaults with the rating *s*. The forecasted and realized probabilities can be compared visually, but again, we want to use some statistics to measure more exactly the quality of the rating system PD estimates.

The most useful and frequently used is the *Hosmer-Lemeshow (HL) Test*, calculating a weighted sum of normalized squared differences between the forecasted probabilities and the realized default rates:

$$S_N^{\chi^2} = \sum_{s=1}^{N} \frac{n_s(PD_s - p_s)^2}{PD_s(1 - PD_s)} = \sum_{s=1}^{N} \frac{(n_s PD_s - d_s)^2}{n_s PD_s(1 - PD_s)} \tag{3.3}$$

The null hypothesis H_0 is that the default probability in grade *s* is PD_s for all *s*. Moreover, If we assume that the defaults are independent, then the statistic asymptotically converge (with all $n_s \to \infty$) to the χ^2-distribution, with $N - 2$ degrees of freedom (Hosmer and Lemeshow 2000). Therefore, we need the Hosmer-Lemeshow (CS) statistic to be a small value with a large corresponding p-value in order not to reject H_0.

Example Table 3.4 shows an example of the Hosmer-Lemeshow Test calculation. There are seven rating grades with predicted PDs going from 50% for the worst rating grade 1, down to 1% for the best rating grade 7. There are 7450 observations with a total of 352 defaults distributed among the rating grades according to the table. The sum of all contributions to the statistic, according to (3.3), is relatively high: $S_7^{\chi^2} \doteq 11,17$ with corresponding p-value; i.e., the probability that $\chi^2 > 11,17$ at 4.8%. The rating grade contributions show that the most significant differences appear for the grades 2, 3, and 5.

The result can be interpreted as a rejection of the hypothesis that the predicted PDs are correct on the 95% level. On the other hand, the correctness hypothesis cannot be rejected on the 99% level.□

Table 3.4 Hosmer-Lemeshow test calculation example

Rating	1	2	3	4	5	6	7
Predicted PD (%)	50	30	15	8	4	2	1
Observations	50	100	300	1000	3000	2000	1000
Defaults	28	37	36	90	102	46	13
Observed PD	56.0	37.0	12.0	9.0	3.4	2.3	1.3
HL test contribution	0.72	2.33	2.12	1.36	2.81	0.92	0.91
HL test total	**11.17**						
p-value	**0.048**						

The construction of the HL test is based on the assumption of the independence of defaults, while the empirical evidence often shows that there is a positive default correlation. In that case, with valid H_0, the deviations of default rates from theoretical PDs would be larger than in the case of independent defaults. Thus, the statistic would be too conservative in rejecting H_0 more frequently than necessary (a false positive type I error). A possible solution proposed by Engelmann and Rauhmeier (2006), is to estimate the default correlation, and determine the statistic's distribution using the Monte Carlo simulation. That means simulating defaults for different rating grades with n_s debtors, probabilities PD_s, and with a given correlation,[1] or correlation matrix. The value (3.3) would be calculated for each scenario, obtaining an empirical distribution of the statistic. The distribution is then used to determine the p-value based on the real validation sample statistic.

Another widely used and a simple method is the *Binomial Test*. Its disadvantage is that it can be used only for a single rating grade s with the forecast probability of default PD_s. We can apply a one-tailed or two-tailed test. If the null hypothesis is that PD is the default probability, and we assume that the events of default are independent, then the probability of j defaults out of n observations is $\binom{n}{j}PD^j(1-PD)^{n-j}$. Therefore, the probability of observing d or more defaults is

$$\begin{aligned} B(\geq d; n_s, PD_s) &= \sum_{j=d}^{n_s} \binom{n_s}{j} PD_s^{\,j}(1-PD_s)^{n_s-j}, \\ B(\leq d; n_s, PD_s) &= 1 - B(\geq d+1; n_s, PD_s). \end{aligned} \tag{3.4}$$

So, if we observe $d > n_s \times PD_s$ defaults and $B(\geq d; N_s, PD_s) \leq 1-\alpha$, i.e. the probability of getting this many defaults is small, then the null hypothesis is rejected on the confidence level α using the right-tailed test. Otherwise, we may say that we fail to reject the hypothesis H_0 on the given confidence level. Similarly, if we observe too few defaults $d < n_s \times PD_s$ and $B(\leq d; N_s, PD_s) \leq 1-\alpha$, then the null hypothesis is rejected on the confidence level α using the left-tailed test.

Example Let us consider the 7th rating grade in Table 3.4 with the forecast probability of default $PD_7 = 1\%$. Out of 1000 non-defaulting debtors at the beginning of a year, we observe 13 defaults during the year. Our expectation of $1\% \times 1000 = 10$ was exceeded by three defaults. Is it enough to reject the forecast probability as too low, or is it still within a tolerance interval on the 95% confidence level? The tail binomial distribution probability (3.4), with the given parameters can be easily calculated, for example, using the Excel function BINOMDIST. Since the result $B(\geq 13; 1000, 1\%) = 20.75\%$ is safely larger than $1-\alpha = 5\%$, we may conclude that the number of defaults is not large enough to reject $H_0 : PD_7 = 1\%$ as a conservative probability of default estimate. The minimum number of observed defaults that would reject H_0 would be in general

[1] Default correlation can be defined simply as the correlation between corresponding binary 0–1 variables. For another treatment of default correlation see also Section 4.2.

$$d_\alpha = \min\{d : B(\geq d; N_s, PD_s) \leq 1 - \alpha\}.$$

With the given parameters, the minimum number of defaults given our confidence level is $d_{95\%} = 16$ as $B(\geq 16; 1000, 1\%) = 4.79\%$.

On the other hand, for the rating grade 5 we observe in Table 3.4 too few defaults, 102, while the expectation is $3000 \times 0.04 = 120$. In this case, the null hypothesis can be rejected on the 95% confidence level, since $B(\leq 102; 3000, 4\%) = 4.83\%$ using the left-tailed test. □

The binomial distribution can be approximated with a normal distribution, if the number of observations is large. However, the approximation is not usually needed with modern computational tools. The test again overestimates the significance of deviations in the realized default rate from the forecast default rate if there is a positive default correlation. This means that the test becomes too conservative. In order to improve the precision of the test, the Monte Carlo simulation generating the empirical distribution with a given correlation can be used. There is also an analytic formula based on the Gaussian copula (see Chap. 4), as proposed by Blochwitz et al. in Engelmann and Rauhmeier (2006).

3.2 Analytical Ratings

Corporate ratings are, even today, when there are so many sophisticated statistical scoring techniques in existence, usually obtained by a rating process that is, at most, only partially automated. The reason is that the key rating factors are not only quantitative, but also qualitative. Management quality, ownership structure, technology, competitive position, legal and regulatory issues can be translated into certain scores, but the assessment must, in any case, be done by an experienced analyst. Figure 3.8 shows a list of possible business and financial factors in a rating process, corresponding to the methodology of Standard & Poor's (Trueck and Rachev 2009).

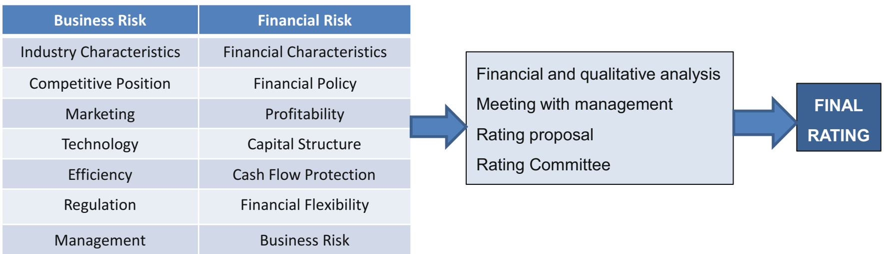


Fig. 3.8 Corporate credit analysis factors and the rating process

Rating agencies generally provide issue-specific ratings and issue credit ratings. Most corporations approach rating agencies to request a rating prior to the sale or registration of a debt issue. Some rating agencies, like S&P, publish ratings for all public corporate debt issues over a certain threshold (USD50 million in the case of S&P). For the purpose of basic research, analysts require financial information about the company, 5 years of audited annual financial statements, the last several interim financial statements, and a description of the company's operations, products, management policies, and plans. Meetings with management and visits to facilities are usually an important part of the rating process. The goal is to review, in detail, the company's key operating and financial plans, management policies, and other factors that may impact the rating. The financial analysis focuses on a number of financial ratios that are compared to the industry averages, and to their past values. Table 3.5 shows a sample of commonly used financial ratios out of the many possible ones.

When the analyst completes the analysis and a rating proposal, a rating committee with a presentation is called. Once the rating is determined, the company is notified of the rating and the major considerations supporting it. Usually, the rating agencies allow the issuer to respond to the rating decision prior to its publication. Such an appeal must be processed very quickly so that the final rating can be published in the media or released to the company. Generally, all of the major rating agencies agree that a rating is, in the end, a credit opinion with interpretation, as shown in Table 3.6. The rating symbols shown in the table may be further modified with + or – (S&P, Fitch), or with 1, 2, 3 (Moody's).

Table 3.5 Commonly used financial ratios (see, e.g., Caouette et al. 2008)

Category	Ratio
Operating performance	Return on equity (ROE) = Net income/equity Return on assets (ROA) = Net income/assets Operating margin = EBITDA/sales Net Profit Margin = Net income/sales Effective tax rate Sales/sales last year Productivity = (sales-material costs)/personal costs
Debt service coverage	EBITDA/interest Capital expenditure/interest
Financial leverage	Leverage = assets/equity Liabilities/equity Bank debt/assets Liabilities/liabilities last year
Liquidity	Current ratio = current assets/current liabilities Quick ratio = quick assets/current liabilities Inventory/Sales
Receivables	Aging of receivable (30,60,90,90+ past due) Average collection period

Table 3.6 Long-term senior debt rating symbols (Source: S&P, Fitch/Moody's)

Rating	Interpretation
Investment grade ratings	
AAA/Aaa	Highest quality; extremely strong; highly unlikely to be affected by foreseeable events
AA/Aa	Very High quality; capacity for repayments is not significantly vulnerable to foreseeable events
A/A	Strong payment capacity; more likely to be affected by changes in economic circumstances
BBB/Ba	Adequate payment capacity; a negative change in environment may affect capacity for repayment
Below investment grade ratings	
BB/Ba	Considered speculative with possibility of developing credit risks
B/B	Considered very speculative with significant credit risk
CCC/Caa	Considered highly speculative with substantial credit risk
CC/Ca	Maybe in default, or wildly speculative
C/C/D	In bankruptcy, or default

Therefore, it is difficult to provide a unique, or reliable, mapping of ratings to default probabilities. There are a number of studies that do so, but the numbers, due to cyclical effects, must be treated very carefully.

The rating agencies have become tremendously influential, covering over $34 trillion in securities on the international financial markets. Many portfolio managers have investment limits on various securities tied to the ratings, and so rating decisions often have a tremendous impact on the prices of securities and the ability of companies to finance themselves. This importance has even been strengthened by the Basel II regulation, allowing banks to calculate capital requirements according to external ratings in the Standardized Approach. One of the more widely discussed weaknesses of the external rating process is the institutional conflict of interest, caused by the fact that ratings are mostly paid for by the issuers, and not by the investors. This has been particularly pronounced during the recent subprime mortgage crisis. It turned out that the agencies assigned positive ratings to many structured bonds (CDOs), which later turned sour, and there is a general suspicion that the high ratings were related to the high fees paid to the rating agencies. The bonds then qualified as acceptable securities for many institutional investors, which suffered huge losses during the crisis.

Credit underwriting is a fundamental function of banks, and so banks should have an internal credit assessment capacity, even in cases where the companies have an external rating. The process and methodology in detail vary with different banks, but are, in principal, similar to the process described above. The process also depends on the type of lending provided: asset-based, project, or unsecured general corporate lending. In the case of asset-based lending, the emphasis is put on the valuation of the borrower's assets, while in the other two types, it is on the borrower's ability to generate future cash flows. Banking credit analysts are turned into experts over the course of their careers, gaining additional authority as they acquire experience and demonstrate their skills. Consequently, banks that lack the expertise tend to develop the automated rating systems discussed in the next section. On the other hand, banks with traditional analytical expertise will hesitate before replacing a human expert with a machine. Financial data based scorings, or other automated ratings, will be used only as an input for the analytical expert assessment.

Regarding the retail, small business, and SME segments, the application of automated rating systems has become almost an industry standard. Nevertheless, in the past, retail loan applications have also been assessed by analysts, and decided by loan officers. For example, in the case of households, the key factors analyzed would be debt to salary ratio, household costs, employment status, age, marital status, education, etc. Regarding small businesses, it usually turns out that accounting numbers have a very low explanatory power, and an expert judgment is very important. Even with a modern automated rating system, for any of the segments, the credit process usually allows for overriding by a competent credit officer. In any case, as pointed out in the introduction to the chapter, any system needs a careful periodical monitoring of performance, and an unbiased comparison with other possible rating methodologies using the methods described in the previous section.

3.3 Regression Rating Systems

Before we turn our attention to the logistic regression model as the most widely used technique in the financial sector, let us list and discuss a variety of possible statistical and other automated rating models:

1. **Econometric models** include classification trees, random forests, linear and multiple discriminant analysis, linear, logit (logistic), and probit regression. The models estimate the probability of default as the target (explained) variable by a set of explanatory variables including financial ratios, categorical variables, and also qualitative assessment indicators. The models' parameters are estimated based on historical data with information on defaults and the explanatory variables values. The structure of the explanatory variables is proposed based on expert judgment and optimized with statistical methods.
2. **Shadow rating** models are not based on the history of defaults, but on a database of the rating decisions of an external credit agency. The models try to mimic, using a regression technique, the expertise of the rating agencies, in particular for segments where there are not so many historical defaults, especially countries, financial institutions, or large corporations.
3. **Artificial neural networks** try to mimic the functioning of the human brain by emulating a network of interconnected neurons. The networks are also trained on historical data. The neural networks are, in a sense, similar to the econometric approach, but the final model is more difficult to interpret, being often characterized as a black box.

4. **Linear programing and Support Vector Machines** aim to separate sets of good and bad observations represented by the vectors of explanatory variables, possibly after a transformation, by a hyperplane in an optimal way.
5. **Rule-based, or expert systems** mimic, in a structured way, the process used by an experienced analyst to arrive at credit decisions. The systems are built on the expertise provided by experienced analysts, and transformed with an applied logic methodology into a set of decision rules within a knowledge-base. Expert systems are used in practice, for instance for small business clients, where econometric techniques do not work well and "human" analysts become too expensive relative to the loan size.
6. **Structural models** are typically based on the option valuation theory, considering default as an event where the assets fall below the debt level. The best known example of this approach is Moody's KMV model, using the information on stock price development. The output of the KMV methodology is called EDF (Expected Default Frequency), and has become quite successful on the U.S. market (see Sect. 4.5). It is, however, not very applicable in emerging countries, where the stock markets are not sufficiently developed. There are also attempts to apply the structural approach to retail clients with a behavior score playing the role of stochastic asset value (Thomas 2009).

In this section we will focus on the "mainstream" regression models, in particular on the logistic regression scorecard that has become a banking industry standard. Other alternative approaches to building an automated rating system will be discussed in the following section. The KMV structural approach will be explained in Chap. 4 as a part of the portfolio modeling approach.

Altman's Z-Score

A linear scoring function assigns to a company a linear combination of financial ratios and other numerical explanatory variables x_i with the goal to discriminate between bad and good borrowers:

$$z = \sum_{i=1}^{k} \beta_i x_i. \tag{3.5}$$

Altman's Z-Score (Altman 1968), is the first successful model of this form for the corporate segment, where the particular coefficients and financial ratios are:

$$Z = 1.2x_1 + 1.4x_2 + 2.3x_3 + 0.6x_4 + 0.999x_5,$$

where the variables are given in Table 3.7

Table 3.7 Z-score model variable group means and F-ratios

Variable	Ratio	Bankrupt group mean	Non-bankrupt group mean	F-ratio
x_1	Working capital/total assets	−6.1%	41.4%	32.60
x_2	Retained earnings/total assets	−62.6%	35.5%	58.86
x_3	EBIT/total assets	−31.8%	15.4%	25.56
x_4	Market value of equity/Book value of liabilities	40.1%	247.7%	33.26
x_5	Sales/total assets	1.5	1.9	2.84

Source: Altman (1968)

The idea of the discriminant analysis method is, given a sample of historical observations, to select the appropriate financial ratios and find their optimal combination so that the score (3.5) maximizes the variance between the group of bankrupt and non-bankrupt companies and, at the same time, minimizes intra group variances. The development performed by Altman on a group of 33 bankrupt and 33 non-bankrupt companies was an iterative process including and excluding variables, correlation testing, and involving a lot of expert judgment which cannot be defined as a straightforward statistical optimization. At the end, when the variables were selected, Altman found the coefficients maximizing the F-value calculated as

$$\frac{N_1(\bar{z}_1-\bar{z})^2+N_2(\bar{z}_2-\bar{z})^2}{\sum_{i=1}^{N_1}(z_{1,i}-\bar{z}_1)^2+\sum_{i=1}^{N_2}(z_{2,i}-\bar{z}_2)^2}, \tag{3.6}$$

where $z_{g,i}$ is the score of the i-th company, N_g the number of companies, $\bar{z}_g$ the group averages for the two groups $g=1,2$, and $\bar{z}$ is the overall average. Altman also estimated an optimal upper boundary of 1.81 (fails), and a lower boundary of 2.99 (non-fail). Any score between the two boundaries is treated as being in the zone of ignorance.

Despite the criticism mentioned in many studies (see; e.g., Duffie and Singleton 2003) which point out that the model may have substantial sample selection bias, the original Z-Score model and its modifications (ZETA Model, private companies, emerging market companies, etc.) have endured as a practical analytical tool to this day.

Linear Regression

The discriminant analysis is closely related to the classical regression model trying to establish a linear relationship between the default variable: $y_i \in \{0,1\}$; and the borrowers' i column vector of characteristics $\mathbf{x}_i$ observed in the period before default:

$$y_i = \boldsymbol{\beta}' \cdot \mathbf{x}_i + u_i.$$

Here, we assume as usual that the first element of $\mathbf{x}_i$ is the constant $x_{i,0} = 1$ corresponding to the absolute term β_0, i.e. $\boldsymbol{\beta}' \cdot \mathbf{x}_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k}$. So, if there are k explanatory variables, the dimension of the vector $\mathbf{x}_i$ is $k+1$. The variables can be continuous, categorical (with a finite range of qualitative values) that are encoded by sets of binary dummy variables (indicating by $\{0,1\}$ each of the possible categorical values), or ordinal (in fact categorical, attaining a finite range of integer values which do not have to be necessarily encoded by dummy variables). The standard ordinary least squares (OLS) regression will yield an estimate **b** of the vector $\boldsymbol{\beta}$, that can be used to calculate the scores for any given set of input parameters. To see that the method is, in fact, equivalent to the discriminant analysis, note that the F-value (3.6) does not change when the coefficients in (3.5) are multiplied by an arbitrary scaling factor, nor when we add an absolute term either. So, without loss of generality, we may assume that $\bar{z}_1 = 1$, and $\bar{z}_2 = 0$. Maximization of (3.6) with these assumptions is equivalent to minimization of

$$\sum_{i=1}^{N_1} (z_{1,i} - \bar{z}_1)^2 + \sum_{i=1}^{N_2} (z_{2,i} - \bar{z}_2)^2$$

which is the same as in the OLS regression. The main drawbacks of the linear regression, as well as of the discriminant approach, follow from the classical regression analysis. Since the variables $y_i \in \{0,1\}$ take only two values, the residuals u_i are heteroscedastic; i.e. the variance depends on i, and so the estimation of $\boldsymbol{\beta}$ is inefficient. The problem can be partially solved with the weighted least squares (WLS) estimator, but the standard errors of the estimated coefficients **b** remain biased. Another problem is that the values of the score $z_i = \boldsymbol{\beta}' \cdot \mathbf{x}_i$ may be negative, or larger than 1, and so are difficult to interpret as the probability of default.

Logistic Regression

In this situation, it is more appropriate to apply an econometric model designed specifically for analyzing binary dependent variables, in particular the Logit or Probit model (Greene 2003). Of those, the Logit, or Logistic Regression model, is the most popular in banking practice, as well as in academic literature. In both approaches, a possible interpretation is to model a latent future credit score variable

$$y_i^* = \boldsymbol{\beta}' \cdot \mathbf{x}_i + u_i,$$

where the score is decomposed into a known (expected) value $\boldsymbol{\beta}' \cdot \mathbf{x}_i$ (the actual score) and an unknown future change u_i of the score that is assumed to have zero mean and a known distribution. The value of the variable y_i^* triggers the default event $y_i = 1$ if and only if $y_i^* \leq 0$. Hence the probability of default conditional on the vector of explanatory variables $\mathbf{x}_i$ is

$$p_i = \Pr[y_i = 1 | \mathbf{x}_i] = \Pr[u_i + \boldsymbol{\beta}' \cdot \mathbf{x}_i \leq 0] = F_i(-\boldsymbol{\beta}' \cdot \mathbf{x}_i), \tag{3.7}$$

where F_i is the cumulative distribution function (cdf) of the random variable u_i. In other words, the link function F_i consistently transforms the score $z_i = \boldsymbol{\beta}' \cdot \mathbf{x}_i$ that takes values in the interval $(-\infty, +\infty)$ to the corresponding probability of default values in (0, 1). A high score implies a low probability of default and a low score means a higher probability of default. The relationship can be also interpreted in a structural model spirit: the score $z_i = \boldsymbol{\beta}' \cdot \mathbf{x}_i$ characterizes the debtor's credit capacity today (at the beginning of the observation period), while u_i denotes the future unknown change of the credit capacity. If the total future score breaches a threshold, then the default takes place.

If the distribution of u_i is normal, then we can use the standard normal cumulative distribution function $F_i(x) = \Phi(x)$ leading to the Probit model. If instead, the residuals are assumed to follow the logistic distribution, then:

$$F_i(x) = \Lambda(x) = \frac{e^x}{1 + e^x} = \frac{1}{e^{-x} + 1}.$$

The mean of the logistic distribution is 0, but note that the variance equals $s^2 = \pi^2/3 > 1$. In order to compare it visually with the standard normal distribution (Fig. 3.9) we need to normalize the logit probability density function (pdf)

$$f(\text{logit}, \text{var} = 1) = \frac{e^{-x/s}}{s(1 + e^{-x/s})^2} \text{ where } s = \sqrt{3}/\pi.$$

Both distributions are quite similar, except for the tails, which are heavier for the logistic distribution (have higher kurtosis).

For a single borrower the probability of default p_i is not observable. However, if we were able to split our sample into a number of groups with common characteristics $\mathbf{x}_i$, and observed default rates p_i for those groups, then the OLS or WLS regression could be applied to the dependent variable $F^{-1}(p_i)$ and the explanatory variables $\mathbf{x}_i$ [see (3.7)]. Such a grouping, however, usually involves considerable problems due to the limited number of observations. A better way to estimate logit and probit models, which does not require any grouping, is the Maximum Likelihood Estimation (MLE) Method. Given an estimate $\mathbf{b}$ of the unknown parameters $\boldsymbol{\beta}$, the probability of default conditional on $\mathbf{x}_i$ is $F(-\mathbf{b}' \cdot \mathbf{x}_i)$. So, if we observe $y_i = 1$, then the likelihood of the observation is $F(-\mathbf{b}' \cdot \mathbf{x}_i)$, and if we observe $y_i = 0$, then the likelihood of the observation is the complementary

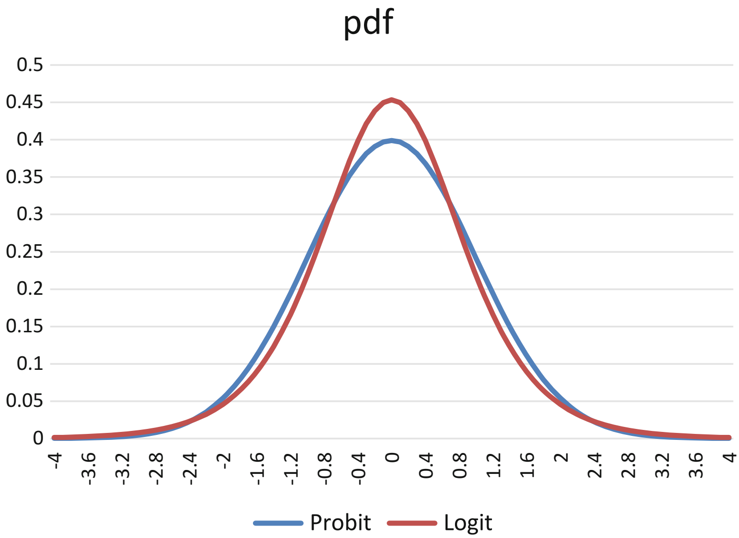


Fig. 3.9 Comparison of the probability density functions of the standard normal (Probit) and the normalized logistic distribution (Logit)

value $1 - F(-\mathbf{b}' \cdot \mathbf{x}_i)$. Assuming that the observations are independent, the total likelihood of the observed dataset given the coefficient vector **b** can be expressed as

$$L(\mathbf{b}) = \prod_i F(-\mathbf{b}' \cdot \mathbf{x}_i)^{y_i} (1 - F(-\mathbf{b}' \cdot \mathbf{x}_i))^{1-y_i}.$$

The estimation of the vector **b** is then based on the maximization of the likelihood function, or rather, for numerical purposes on the maximization of the log-likelihood function

$$\ln L(\mathbf{b}) = l(\mathbf{b}) = \sum_i [y_i \ln F(-\mathbf{b}' \cdot \mathbf{x}_i) + (1 - y_i) \ln(1 - F(-\mathbf{b}' \cdot \mathbf{x}_i))]. \tag{3.8}$$

Usually, the choice of the link function is not theoretically driven. In fact, the differences between the probit and logit models are often negligible (see Fig. 3.9). One argument for the logit distribution is that the fatter tails of the logit distribution give larger weights to extreme events corresponding better to the real world datasets. The logit model also turns out to be numerically more efficient, and its coefficients can be relatively easily interpreted.

Notice that the equation $p_i = \Lambda(\boldsymbol{\beta}' \cdot \mathbf{x}_i)$ can be rewritten as

$$\frac{1-p_i}{p_i} = e^{\boldsymbol{\beta}' \cdot \mathbf{x}_i}, \text{ or equivalently } \ln\left(\frac{1-p_i}{p_i}\right) = \boldsymbol{\beta}' \cdot \mathbf{x}_i. \tag{3.9}$$

The left hand side of the first Eq. (3.9) is called the *good-bad odds* i.e., the probability of good divided by the probability of bad. The right hand side means that one unit increase of the k-th variable has a multiplicative impact of e^{β_k} on the odds, or equally, looking at the second equation, an additive effect of β_k on the *log good-bad odds*.

Henceforth, we will work with the logistic regression model. If $F_i = \Lambda$, then the partial derivatives of (3.8) take the following relatively simple form:

$$\frac{\partial l}{\partial b_j} = \sum_i (y_i - \Lambda(-\mathbf{b}' \cdot \mathbf{x}_i))x_{i,j} = 0 \text{ for } \; j = 0, \ldots, k. \tag{3.10}$$

It can be shown that the *Hessian* (the matrix of the second order derivatives) is negatively definite, consequently, the log likelihood function is strictly concave, and so the solution exists and is unique. The solution can be found using Newton-Raphson's algorithm usually after just a few iterations. Since $x_{i,0} = 1$ corresponding to the intercept coefficient b_0, we have a useful observation that the average of the predicted probabilities (on the training sample) equals the overall default rate in the sample. The estimated parameters **b** are asymptotic normal with the mean $\boldsymbol{\beta}$, and their standard errors can be obtained by inverting the Hessian (Greene 2003). The standard errors s. e. (b_j) are reported automatically by most statistical packages, and it is not necessary to put down the formulas. The significance of the individual parameters is then easily tested with the Wald statistic $W = \frac{b_j}{\text{s.e.}(b_j)}$ which has the asymptotic normal $N(0, 1)$ distribution. This means that we reject the hypothesis H_0: $\beta_j = 0$ on the confidence level α if $|W| \geq \Phi^{-1}(1-\alpha/2)$. Similarly, we may obtain a confidence interval for the true coefficient β_j given the estimate b_j and its standard error.

The rating score of a debtor a with a vector of characteristics $\mathbf{x}(a)$ can be defined either directly as the probability of being good $p_G(a) = \Lambda(-\mathbf{b}' \cdot \mathbf{x}(a))$ scaled to an interval, e.g. 0–1000, or as the original linear function value $s(a) = \mathbf{b}' \cdot \mathbf{x}(a)$ that has the additive property of log-odds scores (a unit change in an explanatory variable has a constant impact on the score). Note that the log-odds score $s(a)$ is usually nonnegative, since the predicted probability of default $p_B(a) = \frac{1}{1+e^{s(a)}}$ should be normally less than 50%. For example $PD = 0.1\%$ corresponds to the log-odds score $\ln(0.999/0.001) = 6.9$, while $PD = 20\%$ corresponds to $\ln(0.8/0.2) = 1.39$. Thus, for example, if the log-odds score $s(a) = \mathbf{b}' \cdot \mathbf{x}(a)$ is multiplied by the constant 150, we get a scaled log-odds score that is approximately in the range 0–1000.

We have intentionally used the notation a for a new debtor that does not belong to the development sample with observations indexed by $i = 1, \ldots, N$. Our goal is not just to maximize the likelihood on the development sample, where we already know whether a default happened or not, but rather to achieve good prediction capabilities on a set of debtors where we do not know the future. This requires, in particular, a careful selection of the explanatory variables.

To test robustness of the resulting function, the given data are usually split into two parts: a development (training or in) sample (e.g. 70% of the data, 70% of bad debtors and 70% of good debtors), and a testing (out or validation) sample (30%). The coefficients of selected variables are estimated only on the development sample, but how well the scorecard fits the data is then measured not only on the in-sample, but also on the out-sample. If the values drop significantly, comparing in-sample and out-sample goodness of fit measures, then the model should not be accepted. Given a set of selected variables, the procedure can be repeated randomly, dividing the initial dataset into the development and the testing part. The resulting empirical distribution of the estimated coefficients and goodness of fit measures give us an even better picture of the scoring model's robustness.

Selection of Variables

It is obvious that on a fixed development sample the best value of the likelihood function, and the closely related Gini coefficient, is maximized when all the available explanatory variables are used. If the number of possible explanatory variables is large (over 20), then the usual result is that many of the coefficients are insignificant; i.e., we cannot reject the null hypothesis H_0: $\beta_j = 0$ on a reasonable confidence level (at least 90%). Then, we cannot even be sure about the sign of the coefficient; if our estimate b_j is positive, the "true" coefficient β_j can be, in fact, negative, and vice versa. It may also happen that the hypothesis $\beta_j = 0$ is not statistically rejected, yet the sign of the estimated coefficient b_j contradicts economic reasoning or experience. Thus, even though the likelihood, and Gini, has been maximized on the development sample, the wrong coefficients might significantly decrease the predictive power on the testing sample and on future sets of loan applications.

The conclusion is that the coefficients of the final scoring function should be significant, with not too wide confidence intervals, and should not contradict our expertise. This can be achieved by limiting the number of explanatory variables to a number optimally between 7 and 15, or even less, eliminating correlated variables and variables with low explanatory power. This goal can be achieved in a number of ways, for instance with *the forward selection procedure*, adding one by one the variables with the best explanatory power, or with *the backward selection procedure,* removing one by one the variables with the worst explanatory power. This is typically done based on the coefficients' p-values or using a test comparing the likelihood of the restricted model with fewer explanatory variables and the likelihood of the larger model.

It can be shown (Hosmer and Lemeshow 2000) that the statistic

$$G = -\ln\left(\frac{\text{likelihood of the restricted model}}{\text{likelihood of the larger model}}\right)$$

has asymptotically the $\chi^2(k-l)$ distribution, provided the number of variables of the larger model is k, of the restricted model l, and the null hypothesis is that the true coefficients β_j of the new variables are all zero. So, in the forward selection procedure we may add, one by one, new variables, maximizing the $\chi^2(1)$ statistics. The null hypothesis must also be rejected at some specified minimal level, and if there is no such new variable, the procedure stops. *The forward stepwise selection procedure*, which can be characterized as a combination of the forward and backward procedures, in addition eliminates at each step already included variables that have become insignificant when new variables have been added.

It is not recommended to apply the automated selection procedure to too long a list of potential explanatory variables. The procedure then becomes quite time consuming, in particular for large datasets, and its outcome might be also influenced by undesirable noise in the data, e.g. by correlated variables that will surely exist if the list of variables is too long, etc. Therefore, the general recommendation is to create *a short list of variables*, around 20–30, that have an acceptable explanatory power based on the univariate analysis discussed below. The selection of the variables can be based on the univariate Gini coefficient (e.g., requiring at least 10% value) and/or on the Information Value described below (typically, at least 4%). In addition, the variables should not be correlated—if there are two highly correlated variables, then the one with the lower explanatory power is to be eliminated from the short list. The correlation cut off is set expertly typically at about 30–70%. The automated selection procedure applied to a carefully designed short list usually produces a reasonable outcome. Nevertheless, additional expert modifications, like forcing some variables to stay in the final list or eliminating others, are often applied. One also has to distinguish the corporate segment where the variables are usually continuous (financial ratios) and the retail segment with explanatory variables being mostly categorical.

A practical approach applicable to the corporate segment (Hayden 2003) or to the retail segment may be the following:

1. Perform a **univariate analysis** on pre-selected financial ratios, testing the explanatory power and linearity assumption of log odds according to (3.9), using only one single explanatory variable without the remaining ones. In this case, the sample needs to be split into a number of bins according to the variable values. The number of bins depends on the number of defaults in the sample, requiring that rates of default for the bins can be reasonably estimated. The corresponding average bin variable values, and log odds, can be plotted on a chart with an OLS interpolation. Figure 3.10 shows an example of the result for two variables where the number of bins used is 50. The first chart confirms a satisfactory linear dependence of the log odds on the first tested variable (Current Liabilities/Total Assets). The second variable (Sales Growth), however, shows a non-linear, even non-monotone, impact on the log-odds. The

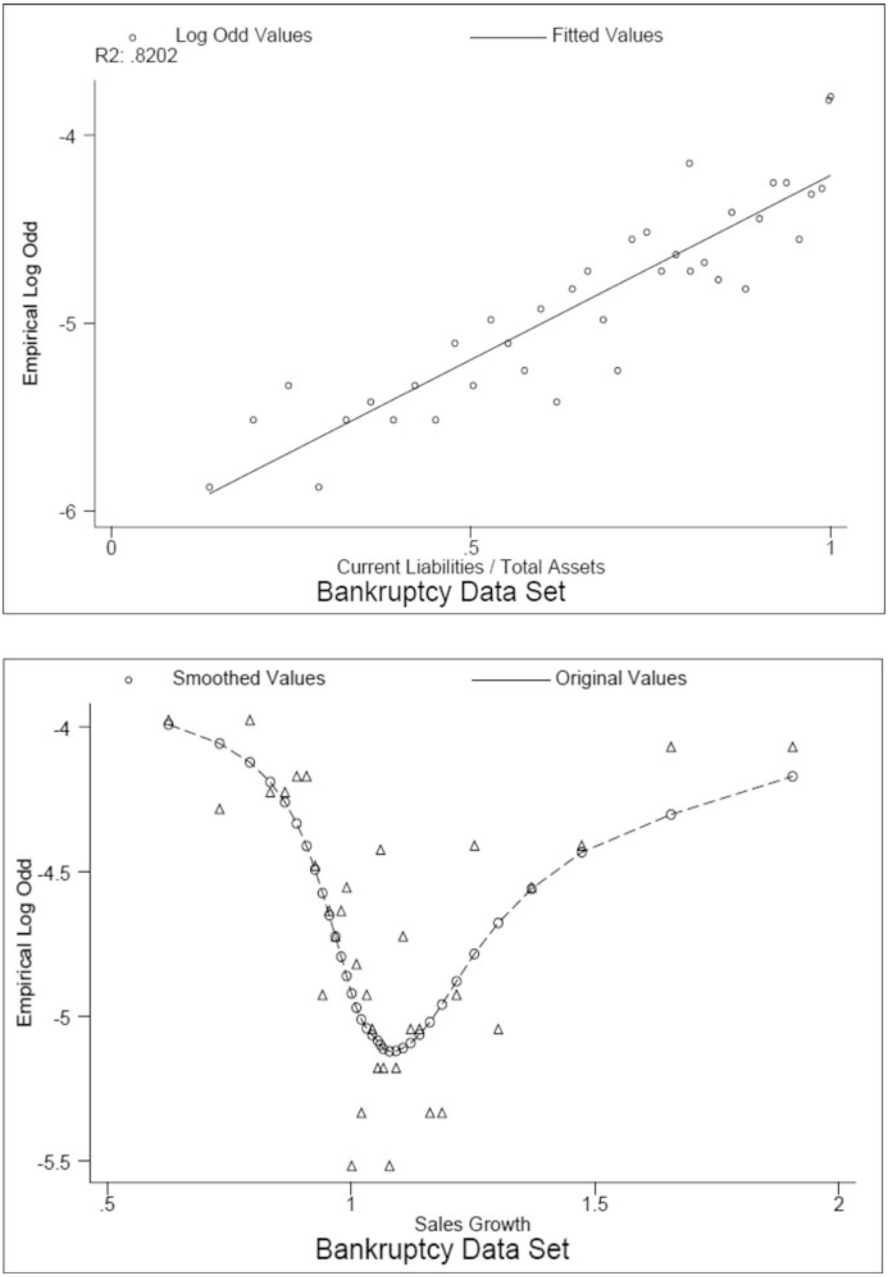


Fig. 3.10 An example of univariate analysis (Source: Hayden 2003)

solution is either to reject the variable and find an alternative with linear behavior, or to design a transformation of the original variable so that the resulting dependence is linear. The variables that show no, or too low, explanatory power (e.g., measured simply by R^2 or by the Gini coefficient) in the univariate analysis are rejected.
2. Calculate the **correlation matrix** on the set of variables selected in the previous step. It is known that if highly correlated variables are included in the model, then the coefficients will be significantly and systematically biased. The rule of thumb is to eliminate correlations above 50% by grouping the variables according to the correlations, and selecting those with the best explanatory power from each group.
3. Run the forward (or backward, or stepwise) selection **logistic regression** on the set of variables from step 2. The automated selection procedure might be combined with expert judgement, forcing some variables into the regression, or vice versa removing variables with a coefficient sign contradicting expert opinion.
4. **Validate the result** by calculating in- and out-sample Gini's coefficients and the Hosmer-Lemeshow statistics. If the values are below an acceptable level, try to find alternative explanatory variables. Acceptable values of Gini's coefficient start at about 30%. The Hosmer-Lemeshow statistic p-value should not be too small; e.g. not less than 10% as explained in Sect. 3.1.

The procedure described above is more appropriate for the corporate segment, where the explanatory variables like financial ratios are naturally continuous. Categorical variables are used more often in the retail segment, and the analysis is usually based on slightly different techniques. Examples of categorical variables are: marital status, education, employment status, driver's license, residential status, phone line, etc. But even ordinal variables like age, or number of children, are usually represented by categorical variables with just a few possibilities; for example: age 18–24, 24–32,. . ., 65 and older. A categorical variable attaining C values $c = 1, \ldots, C$ is, as usual in a regression analysis, encoded by $C - 1$ dummy variables. It means that there are $C - 1$ beta coefficients to be estimated and it is, therefore, desirable not to have more categories than necessary.

The explanatory power of a categorical variable can be measured not only by the univariate logistic regression Gini's coefficient, but can also be, as a first approximation, easily investigated graphically by looking at the average default rates, or log-odds, for subsamples obtained when we split the development sample into C parts according to the categorical values. If all the default rates are approximately the same, then the variable does not look promising. Categories showing similar default rates, or with a low number of good or bad debtors, should be merged. Figure 3.11 shows, as an example, percentages of bad borrowers according to marital status taken from the crosstab in Table 3.8. The chart indicates that all categories have an impact on the default rate, with "Married male" being not surprisingly very close to "Married female", and "Divorced" being close to "Partner". Moreover, the categories "Widowed" and "Partner" have a low number of

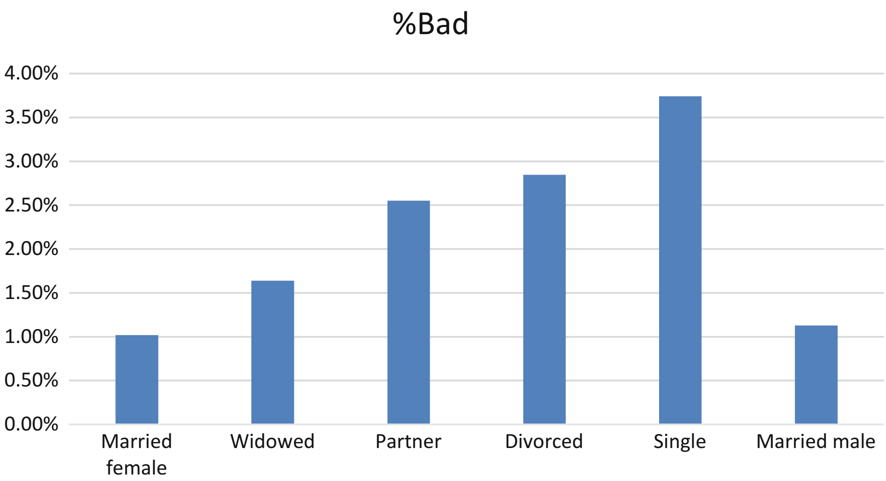


Fig. 3.11 Percentage of bad borrowers according to "Marital Status" values

Table 3.8 Crosstab for the variable "marital status"

	Borrower Marital Status						
	Married female	Widowed	Partner	Divorced	Single	Married male	Total
Good	3 300	120	191	1 092	643	5 430	**10 776**
Bad	34	2	5	32	25	62	**160**
Total	**3 334**	**122**	**196**	**1 124**	**668**	**5 492**	**10 936**
%Bad	1.02%	1.64%	2.55%	2.85%	3.74%	1.13%	**1.46%**

observations; a rule of thumb is that there should be at least 5% of all observations in each class (i.e. over 500 in this case). Therefore, the categories "Widowed" and "Partner" need to be merged with a relatively close category which is "Widowed". So, in addition, after merging the categories "Married male" and "Married female" into "Married" we are left with only three categorical values: "Married", "Single", and "Other—Widowed, Partner, Divorced" that have, at least visually, an acceptable discriminative power between bad and good loans.

In a more quantitative approach we should look at the CS statistic (3.3) and its p-values. The CS statistic of the distribution of PDs into five original bins, according to the categorical values in Table 3.8, turns out to be 49.44, while the CS statistic, after the proposed merging into three bins, is just a little bit less, 48.12. The Chi-squared distribution p-values are, in both cases, very low (<0.001), and the p-value in the second case is even lower than in the first since the number of degrees of freedom has been reduced from four to two. Note, however, that the CS statistic does not capture the issue of a low number of observations or defaults in a bin. In this particular case, the primary reason for merging the categorical values was the low number of defaults in the categories "Widowed" and "Partner", and the CS statistic confirmed that we lost almost no discriminatory power. If we tried to create just two bins; e.g., "Married" and all the other categories in the second bin, then the CS statistic would be reduced more significantly to 44.71, but still remaining relatively high.

Other widely used and sophisticated quantitative measures of discrimination power applied to categorical variables are *the Weight of Evidence* (*WoE*) and *the Information Value* (IV) [Porath in Engelmann and Rauhmeier (2006); Thomas (2009)]. The Weight of Evidence of a categorical value c can be defined as the change between the overall log odds ratio and the log odds ratio given the value c. It follows from the Bayes theorem that it can be defined as:

$$WoE(c) = \ln\Pr\left[c\middle|Good\right] - \ln\Pr\left[c\middle|Bad\right].$$

Note that the value $\Pr\left[c\middle|Good\right]$ is calculated as the relative frequency of c among all good loans and, analogously, $\ln\Pr\left[c\middle|Bad\right]$. For example, we can calculate based on Table 3.8

$$\begin{aligned} &WoE(\text{Single}) = \ln\Pr\left[\text{Single}\middle|Good\right] - \ln\Pr\left[\text{Single}\middle|Bad\right] = \\ &= \ln 643/10776 - \ln 25/160 = -0.96. \end{aligned}$$

To be more precise, as usual, what we are calculating above is not the theoretical population WoE but an estimate based on the given sample. In order to explain the definition and importance of the Weight of Evidence, let us decompose the good-bad odds according to the Bayes theorem

$$\frac{\Pr\left[Good\middle|c\right]}{\Pr\left[Bad\middle|c\right]} = \frac{\Pr\left[c\middle|Good\right]}{\Pr\left[c\middle|Bad\right]} \times \frac{\Pr[Good]}{\Pr[Bad]}.$$

Applying ln to both sides of the equation we obtain

$$\ln o(c) = \ln o_{Pop} + WoE(c), \tag{3.11}$$

where $o_{Pop} = \frac{\Pr[Good]}{\Pr[Bad]}$ denotes the population good-bad odds, and $o(c) = \frac{\Pr\left[Good\middle|c\right]}{\Pr\left[Bad\middle|c\right]}$ the odds conditional on the information given by c. Therefore, instead of inspecting the log-odds of a categorical variable we may look at WoE. One of the advantages is that we can directly identify categorical variables that have a positive impact on the logodds, i.e. implying lower predicted probability of bad, and those that have a negative impact. For example, Fig. 3.12 shows that only the "Marital Status" categories "Married male" and "Married female" have positive WoE, while all the other categories have negative WoE values. This may better explain the decision to merge the category "Widowed" with "Partner" and "Single" and not with "Married". Table 3.9 compares the log-odds and WoE values and verifies that the difference indeed always equals the dataset population WoE.

Fig. 3.12 Weight of evidence according to the "Marital Status" values

Table 3.9 Log-odds, WoE, and their difference for the variable "Marital Status"

	Borrower Marital Status						
	Married female	**Widowed**	**Partner**	**Divorced**	**Single**	**Married male**	**Total**
log G/B	4.58	4.09	3.64	3.53	3.25	4.47	**4.21**
WoE	0.37	- 0.12	- 0.57	- 0.68	- 0.96	0.26	-
Diff.	4.21	4.21	4.21	4.21	4.21	4.21	4.21

In fact, the identity (3.11) could be used to build a naïve Bayes scorecard. Let us assume that there are n categorical explanatory variables $c_1, \ldots, c_n$. If we were able to calculate the weight of evidence for each combination $\mathbf{c} = \langle c_1, \ldots, c_n \rangle$ of possible categorical values:

$$WoE(\mathbf{c}) = \ln\Pr[\mathbf{c}|Good] - \ln\Pr[\mathbf{c}|Bad],$$

then we could define the score, conditional on the information $\mathbf{c}$ according to (3.11) as:

$$s(\mathbf{c}) = s_{Pop} + WoE(\mathbf{c}), \tag{3.12}$$

where the population log odds score is $s_{Pop} = \ln o_{Pop}$. Note that the equation naturally separates the population log-odds from the information implied by the information $\mathbf{c}$, which does not depend on the level of the overall population odds. A positive value of $WoE(\mathbf{c})$ indicates that the group is better than the average, and the negative value that it is worse.

To get a better intuition of the order of the weight of evidence values which could be interpreted as significant, notice that $WoE = \ln 2 = 0.69$ doubles the good-bad-odds, while, for example, $WoE = \ln 1.2 = 0.18$ increases the good-bad-odds only by 20%. The probability of default given by the log-odds score can be expressed, by definition, as

$$\Pr\left[Bad\middle|\mathbf{c}\right] = \frac{1}{1 + e^{-s(\mathbf{c})}}.$$

The score card defined by (3.12) would be perfect if we were able to estimate precisely $WoE(\mathbf{c})$ for all possible combinations of the categorical values. This is, however, in practice virtually impossible, since for a larger number of categorical variables there is a huge number of their combinations, with just a few, or no, observations in the corresponding bins. One possible approach is to accept the strong assumption that the categorical variables are independent. Then

$$\Pr\left[\mathbf{c}\middle|Good\right] = \Pr\left[c_1\middle|Good\right]\cdots\Pr\left[c_2\middle|Good\right] \text{ and}$$

$$\Pr\left[\mathbf{c}\middle|Bad\right] = \Pr\left[c_1\middle|Bad\right]\cdots\Pr\left[c_2\middle|Bad\right].$$

Consequently,

$$WoE(\mathbf{c}) = WoE(c_1) + \cdots + WoE(c_n), \text{ and so}$$

$$s(\mathbf{c}) = s_{Pop} + WoE(c_1) + \cdots + WoE(c_n). \tag{3.13}$$

Therefore, under the assumption of the independence of the explanatory categorical variables, the weight of evidence of a single categorical value presents its contribution to the naïve log-odds score. Nevertheless, in reality the independence assumption is too strong, and the naïve Bayes score given by (3.13) is just an approximation of a more precise score obtained by logistic regression handling possible dependencies between the variables. In any case, the decomposition (3.13) illustrates well the importance of the concept of *WoE*.

A comprehensive discrimination measure of a categorical variable can be calculated as a probability weighted average of the Weights of Evidence for $c = 1, \ldots, C$ called the *Information Value* (IV). The formal definition is:

$$IV = \sum_{c=1}^{C} WoE(c) \cdot \left(\Pr\left[c\middle|Good\right] - \Pr\left[c\middle|Bad\right]\right).$$

Note that the value is always nonnegative since the sign of $WoE(c)$ is always the same as the sign of $\Pr\left[c\middle|Good\right] - \Pr\left[c\middle|Bad\right]$ but, on the other hand, there is no upper bound. The information value can be indeed interpreted as the average *WoE,* weighted by the probability distribution of categorical values in the population of goods, minus the average *WoE,* weighted by the probability distribution of categorical values in the population of bads. Since *WoE* should be positive for good debtors and negative for bad debtors, the resulting Information Value reflects the

Table 3.10 WoE and IV for "Marital Status" after coarse classification

	Borrower Marital Status			
	Single	Div., Part., Wid.	Married	Total
Good	643	1 403	8 730	**10 776**
Bad	25	39	96	**160**
Total	668	1 442	8 826	**10 936**
WoE	- 0.96	- 0.63	0.30	
IV	0.09	0.07	0.06	**0.23**

discrimination power between bad and good, given by the information contained in the categorical variable values.

Therefore, the measures *WoE* and *IV* are good indicators to assess the relative discrimination power of the categorical values and variables. The disadvantage is that there is no appropriate asymptotic probability distribution for the measures, so we cannot tell from the absolute values whether the discriminatory value is satisfactory or not. Regarding the Information Value, the rule of thumb is that the minimum required to keep a variable on the short list is 0.04–0.10 (while for the univariate Gini coefficient it is 0.10–0.15). In the coarse classification process one should control the decrease of the two values (IV and/or univariate Gini) in order to avoid unnecessary loss of information.

For example, Table 3.10 shows WoE and the Information value of the "Marital Status" after coarse classification. The Information Value before coarse classification has been 0.33 and so there is only a small and acceptable decrease of the value, which remains high at 0.32.

To summarize, a standard scoring function development process on a general dataset with categorical and numerical potential explanatory variables should include the following steps:

1. **Preselect variables** from a long list in order to get at most 20–30 variables where a more detailed univariate analysis and course classification will be performed. The selection is usually based on statistical and expert criteria. For each variable, numerical or categorical, the univariate Gini coefficient can be automatically calculated as the (in-sample) Gini coefficient of the logistic regression function with the single variable. A typical cut-off value for the univariate Gini is then around 10%, depending on the number of available variables. For categorical variables, or binned numerical variables, the Information Values should also be calculated, requiring, for example, 4% as a minimum. The selection does not have to be purely mechanical; some variables might be preselected on an expert basis in spite of their lower explanatory power, or, vice versa, some variables with relatively high statistics might, for some reason, be expertly eliminated.
2. Perform a **univariate analysis** of the preselected variables that need to be more closely inspected in terms of log-odds dependence on their values. This is

straightforward for categorical variables where the log-odds can be easily calculated for each categorical value, i.e. as $\ln G/B(c) = \ln \frac{1-\hat{p}(c)}{\hat{p}(c)}$, where $\hat{p}(c) = \frac{d(c)}{n(c)}$, $n(c)$ is the number of observations where the categorical variable takes the value c and $d(c)$ is the number of observed defaults, again conditional on the categorical value c. For a numerical variable, the sample needs to be split into a number of equally sized bins so that there are sufficient numbers of observations and defaults to calculate the empirical log-odds for individual bins.

3. Based on the univariate analysis and on our approach to the scoring function development, various **transformations of the preselected variables** can be applied. For the retail segment, it is quite common to use categorical variables only. In this case, numerical variables are transformed into categorical by binning. On the other hand, for corporate segments it is more usual to use numerical variables only, so if there is a categorical variable, it needs to be transformed into a numerical one. The standard transformation uses the dummy variables. However, it is also possible to replace a single categorical variable by a single numerical variable. This can be achieved by replacing the categorical values by the corresponding Weight of Evidence (WoE) values defined above. This transformation is also called *"WoEization"* and is sometimes applied to all categorical variables, even in the retail segment, to transform categorical variables into numerical ones. This idea can be also used to define easily a transformation of a numerical variable with strongly non-linear dependence of the log-odds like the second variable in Fig. 3.10. The variable has, in fact, already been transformed into a categorical one by binning, and it can be transformed back into a numerical one by assigning the WoE values to the corresponding bins. The dependence of the log-odds on the explanatory variable value will then be perfectly linear according to (3.11), since the difference between the bin log-odds and WoE equals the population log-odds that are independent of the bins.
4. The categorical variables that have too many values need to be **coarse classified** by merging categories with similar default rates (or WoE) and/or with too low numbers of observations.
5. **Correlation analysis** should be performed to eliminate highly correlated variables, typically with correlation larger than 50%. It is straightforward to calculate the correlations between continuous variables. For categorical ones, one may calculate the correlation between dummy variables. However, the correlation analysis may become too complicated, having a number of dummies for each single categorical variable. A possible solution is to replace the categorical variables by their WoE numerical equivalents, at least for the purpose of the correlation analysis. An alternative and useful correlation statistic is the Variance Inflation Factor (VIF)[2] that measures how much collinearity exists in

[2]The Variance Inflation Factor is defined as $VIF = \frac{1}{1-R_j^2}$ where R_j^2 is the coefficient of regression of the explanatory factor j on all the other explanatory factors.

the regression analysis. A typical cut-off corresponding to highly correlated variables would be around 5–10.

6. Perform a manual expert opinion based **selection of variables, and/or an automated selection of variables**, e.g. using the stepwise regression. The final list usually does not have more than 10 significant variables.
7. Estimate the regression coefficients on the in-sample dataset and calculate **in- and out-sample performance**, i.e. the Gini coefficient (AUC) and other statistics. It is advisable to repeat the cross-validation for different in/out sample splits. If the out-sample Gini coefficient does not decrease too much (by, at most, a few percentage points) compared to the in-sample Gini, the selection of variables and their transformations are accepted. The final regression coefficients can be estimated on the full data set. Note that the partially expert-based univariate analysis and selection of variables described above works with the full sample. Nevertheless, in order to better test the robustness of the development, the univariate analysis and selection of variables can also be performed only in-sample. In this case it is advisable to have a sufficiently large dataset (in particular enough bad observations), and the coarse classification and selection procedures should be automated enough so that different in-out-sample splits can be tested. If the "WoEization" approach is applied, then one has to keep in mind that the WoE values in fact play the role of estimated parameters and so should be estimated in-sample (not on the full sample) in the validation procedure.

Case Study

We will illustrate the development of a logistic regression scorecard on a consumer loan dataset obtained from a large Czech bank. The dataset was also used in Witzany (2009b, d–f) to analyze the sensitivity of quality of a scoring function, depending on the definition of defaults. In this case study, however, we use only one standard definition of default; i.e., 90 days-past-due with a minimum past due balance of 100 CZK. The dataset contains application data on 10,936 accounts observed in the period 1/2002–10/2008. For each of the accounts we have the information of default, or no default, in the 12-month horizon after origination. The dataset contains just 160 defaults, and 10,776 non-defaulted observations, i.e. the experienced default rate is only 1.48%. The explanatory variables labeled in the second column of Table 3.11 are mostly categorical (sex, age, marital status, etc.). Monthly and other incomes are used as numerical variables and the number of dependents may be used as numerical or categorical.

Looking at Table 3.11, it is clear that there are too many categorical values, which would probably cause over-fitting and low robustness of the model if we did

Table 3.11 Scoring dataset description

Scoring dataset	Observations	10,936
31 Jul 2009 14:56	Variables	21
Variable name	Variable description	# of categorical values
id_deal	*Account number*	*N/A*
def	*Default (90 days, 100 CZK)*	*2*
mesprij	*Monthly income*	*N/A*
pocvyz	*Number of dependents*	*8*
ostprij	*Other income*	*N/A*
k1pohlavi	*Sex*	*2*
k2vek	*Age*	*15*
k3stav	*Marital status*	*6*
k4vzdel	*Education*	*8*
k5stabil	*Employment stability*	*10*
k6platce	*Employer type*	*9*
k7forby	*Type of housing*	*6*
k8forspl	*Type of repayment*	*6*
k27kk	*Credit card*	*2*
k28soczar	*Social status*	*10*
k29bydtel	*Home phone lines*	*3*
k30zamtel	*Employment phone lines*	*4*

not perform any coarse classification and selection of variables. There would be 76 dummy variables encoding the categorical values, and it is clear that for some variables like Education, Age, Employment Stability, Employer Type etc., the number of categories would be too high. Let us, nevertheless, first look at the results of the logistic regression without any selection of variables and without any coarse classification performed using the SAS software package. In order to compare in-sample and out-sample results, the dataset is split into a training sample with 65% of goods and bads, and 35% of goods and bads in the remaining validation sample. Table 3.12 lists only part of the list of coefficients (as there are over 79 lines in the full table), and their significance from the unrestricted model. The significance of all the variables, with the exception of “Other income”, is unacceptably low. Poor robustness of the model is illustrated by Table 3.13, where we generate different random in-sample/out-sample selections, apply the logistic model, and calculate the in-sample and out-sample Gini coefficients. We could be satisfied with the in-sample Gini, which is around 70%, but the out-sample coefficient turns out to be just around the range of 40–50% with quite high variability. The difference between the average in-sample and out-sample Gini is almost 28%. This poor performance, or even worse, should be expected if the scorecard was implemented in practice and used to score new applications.

Thus, in order to improve robustness, i.e. the out-sample performance of the model, we need to coarse classify the categorical variables and then run a variable selection procedure in the way described above. First, let us look at the univariate

Table 3.12 SAS logistic regression coefficients sample without coarse classifications and variable selection

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	24.2156	291.9	0.0069	0.9339
MESPRIJ		1	-5.40E-06	0.000011	0.2467	0.6194
OSTPRIJ		1	0.000058	0.000024	5.7435	0.0166
POCVYZ	**0**	1	-3.9438	173.6	0.0005	0.9819
POCVYZ	**1**	1	-3.3988	173.6	0.0004	0.9844
POCVYZ	**2**	1	-3.5687	173.6	0.0004	0.9836
POCVYZ	**3**	1	-3.0501	173.6	0.0003	0.986
POCVYZ	**4**	1	-3.2527	173.6	0.0004	0.985
POCVYZ	**5**	1	6.161	382.8	0.0003	0.9872
POCVYZ	**6**	1	5.6907	778.8	0.0001	0.9942
K1POHLAVI	**M**	1	-0.1789	0.1568	1.3024	0.2538
K2VEK	**1**	1	8.6268	313.1	0.0008	0.978
K2VEK	**2**	1	-2.6363	39.7479	0.0044	0.9471
K2VEK	**3**	1	-2.511	39.7473	0.004	0.9496
K2VEK	**4**	1	-2.3	39.7467	0.0033	0.9539
K2VEK	**5**	1	-1.1525	39.7509	0.0008	0.9769
…	**….**					

Table 3.13 In-sample and out-sample Gini coefficients for the unrestricted model

Full model Gini coefficients		
Run	In-sample (%)	Out-sample (%)
1	67.9	47.6
2	72.1	35.7
3	71.9	26.2
4	64.2	57.7
5	70.2	40.0
Ave	**69.3**	**41.4**
Std	3.3	11.9

Gini coefficients and Information values of the individual explanatory variables (of course, not including the Account Number variable) in Table 3.14 in order to eliminate the variables with low discrimination power. The Gini and IV of Sex, Credit Card, and Employment Phone lines are too low and we will not consider these variables any more. The Gini of monthly and Other Income is also quite low (we do not calculate IV since those are numerical variables), but it makes sense to define Total Income as the sum of Monthly and Other Income. The Gini above 18% now becomes more interesting, and we can include this variable in the short list. The relationship between Total Income and log-odds can be further investigated by

Table 3.14 Univariate Gini and Information Value before and after coarse classification and/or "Woeization"

		Full sample			After course classification			Woe 1	WoE 2
variable name	**variable description**	**# cat.**	**Gini**	**IV**	**# cat.**	**Gini**	**IV**	**Gini**	**Gini**
mesprij	*Monthly Income*	N/A	6.80%	N/A	N/A	N/A	N/A	N/A	N/A
pocvyz	*Number of Dependents*	8	26.30%	30.78%	3	22.80%	26.68%	26.00%	22.80%
ostprij	*Other Income*	N/A	8.50%	N/A	N/A	N/A	N/A	N/A	N/A
k1pohlavi	*Sex*	2	3.20%	0.44%	N/A	N/A	N/A	N/A	N/A
k2vek	*Age*	15	31.40%	36.56%	3	28.70%	31.73%	30.80%	28.70%
k3stav	*Marital Status*	6	21.70%	23.27%	3	21.60%	22.73%	21.70%	21.60%
k4vzdel	*Education*	8	20.30%	15.94%	4	19.80%	15.52%	20.30%	19.80%
k5stabil	*Employment Stability*	10	36.60%	51.44%	4	35.50%	48.83%	36.30%	35.50%
k6platce	*Employer Type*	9	26.90%	25.36%	3	23.90%	24.18%	26.40%	23.90%
k7forby	*Type of Housing*	6	18.90%	13.39%	2	18.00%	12.34%	18.40%	18.00%
k8forspl	*Type of Repayment*	6	14.00%	15.33%	2	12.80%	13.13%	13.90%	12.80%
k27kk	*Credit Card*	2	4.80%	7.38%	N/A	N/A	N/A	N/A	N/A
k28soczar	*Social Status*	10	9.80%	7.48%	4	8.60%	5.11%	8.90%	8.60%
k29bydtel	*Home Phone Lines*	3	21.80%	20.25%	2	21.40%	18.68%	21.80%	21.40%
k30zamtel	*Employment Phone Lines*	4	5.80%	4.80%	N/A	N/A	N/A	N/A	N/A
prij	*Total Income*	N/A	18.50%	N/A	N/A	N/A	N/A	N/A	N/A
prij_bin	*Binned Total Income (10)*	10	25.40%	20.79%	3	20.30%	15.65%	25.40%	20.30%

Table 3.15 Coarse classification of the education variable

	Basic educ.	Skilled	Voc. educ.	Voc. grad. educ.	Full sec. educ.	Full sec. gen. educ.	High voc. educ.	University	Total
Good	704	1814	2530	202	3475	550	155	1346	10776
Bad	18	37	45	3	42	7	1	7	160
Total	**722**	**1 851**	**2 575**	**205**	**3 517**	**557**	**156**	**1353**	**10 936**
WoE	- 0.54	- 0.32	- 0.18	- 0.00	0.21	0.15	0.83	1.05	
recoding	**A**	**A**	**B**	**B**	**C**	**C**	**D**	**D**	

categorizing the variable into 10 equally sized bins. Table 3.14 shows that Gini of the Binned Total Income increases to 25.4%, and IV is quite high at 20.79%, and so this variable is included into the short list.

Now we can proceed with the coarse classification. We have already shown (Fig. 3.12 and Table 3.10) how to reduce the number of categories for Marital Status from 6 to 3 while keeping the univariate Gini still above 21% and IV above 22%. Similarly, we can reduce the number of categories for the other variables, for example 8 to 4 for the Education variable (Table 3.15) keeping IV still at 15.5%.

It is also interesting to look at Fig. 3.13 showing that the dependence of WoE on the ten equally sized binned categories of Total income is not monotonic as one would expect. Low income below 15,000 has, apparently, a negative WoE, but the Woe corresponding to income below 11,000 is close to zero. Similarly, The WoE for income above 15,000 is positive, but close to zero for incomes between 21,000 and 26,000. Since there is no fundamental reason for the non-monotonicity we have decided to coarse classify the first four bins (monthly income below 15,000) into the "Low income" bin, the following 5 bins (income between 15,000 and 291,000) into "Medium income", and finally the last bin (income above 291,000) as "High

Bin	1	2	3	4	5	6	7	8	9	10
Income (x1 000)	11.1	12.5	13.7	14.8	16.0	17.5	19.2	21.8	26.4	291.7
WoE	-0.09	-0.77	-0.38	-0.32	0.35	0.69	0.42	0.05	0.13	0.69

Fig. 3.13 Dependence of WoE on binned income

income". The resulting IV dropped to 15.6%, but we have achieved a much better robustness and consistency of the categorical variable.

The overall outcome of the course classification, i.e. the numbers of categorical values, Gini, and IV, is shown in Table 3.14. To proceed with the selection of variables we have several options: we can use the categorical variables or we can replace the categorical values with the corresponding WoE values and perform regression with the numerical explanatory variables. The "WoEization" can be done before coarse classification, saving us some work, or after coarse classification, achieving better robustness of the estimated WoE values. The Gini coefficients for the two options are shown in the last two columns of Table 3.14. The univariate Gini coefficients after "WoEization" without coarse classification are, by definition, the same as for the original categorical values[3] while the Gini after "WoEization" with coarse classification are the same as the categorical Gini after coarse classification. The transformed numerical variables also allow us to perform a comprehensible correlation analysis (Table 3.16). The mutual correlations are at an acceptable level, not above 50%, and so we do not eliminate any additional variables from the short list. The highest mutual correlation of 50% can be observed, not surprisingly, between the Number of Dependents and Marital Status.

Table 3.17 shows the results of the regression with the categorical variables after coarse classification. The stepwise selection procedure applied to the full dataset has selected nine variables that remain significant in the final model: Number of

[3]Some small differences might be caused by approximate WoE values for categories with no good or bad observation.

Table 3.16 Correlation analysis of the explanatory variables after coarse classification and "WoEization"

Pearson Correlation Coefficients, N = 10936											
	PRIJ_BIN1	POCVYZ1	K2VEK1	K3STAV1	K4VZDEL1	K5STABIL1	K6PLATCE1	K7FORBY1	K8FORSPL1	K28SOCZAR1	K29BYDTEL1
PRIJ_BIN1	100%	17%	2%	14%	19%	8%	1%	4%	-1%	1%	7%
POCVYZ1	17%	100%	21%	50%	7%	11%	2%	12%	5%	2%	7%
K2VEK1	2%	21%	100%	25%	2%	24%	6%	17%	7%	-3%	9%
K3STAV1	14%	50%	25%	100%	4%	13%	4%	18%	3%	-4%	12%
K4VZDEL1	19%	7%	2%	4%	100%	11%	8%	3%	7%	8%	15%
K5STABIL1	8%	11%	24%	13%	11%	100%	25%	5%	3%	32%	7%
K6PLATCE1	1%	2%	6%	4%	8%	25%	100%	-1%	2%	6%	4%
K7FORBY1	4%	12%	17%	18%	3%	5%	-1%	100%	3%	-5%	11%
K8FORSPL1	-1%	5%	7%	3%	7%	3%	2%	3%	100%	2%	3%
K28SOCZAR1	1%	2%	-3%	-4%	8%	32%	6%	-5%	2%	100%	-4%
K29BYDTEL1	7%	7%	9%	12%	15%	7%	4%	11%	3%	-4%	100%

Table 3.17 Stepwise selection of categorical variables and the validation performance

Summary of stepwise selection							
	Effect			Number	Score	Wald	
Step	Entered	Removed	DF	In	Chi-square	Chi-square	Pr > ChiSq
1	POCVYZ1		2	1	68.0450		<.0001
2	K5STABIL1		3	2	56.5916		<.0001
3	K29BYDTEL1		1	3	21.1297		<.0001
4	K8FORSPL1		1	4	19.4712		<.0001
5	K6PLATCE1		2	5	16.2111		0.0003
6	K2VEK1		2	6	11.5255		0.0031
7	PRIJ_BIN1		2	7	9.4163		0.0090
8	K28SOCZAR1		3	8	9.7886		0.0205
9	K7FORBY1		2	9	6.7938		0.0335

Categorical variables selection		
Run	In-sample (%)	Out-sample (%)
1	59.1	61.4
2	60.6	53.2
3	59.9	56.6
4	60.2	51.2
5	59.4	57.6
Ave	59.8	56.0
std	0.6	4.0

Dependents, Employment Stability, Home Phone Line, Type of Repayment, Employer Type, Age, Income, Social Status, and Type of Housing. Note that some apparently important explanatory variables like Education did not enter the model, while seemingly irrelevant one like Home Phone Line did. The out-sample

Table 3.18 Performance of the model after "WoEization" and without coarse classification

	Analysis of maximum likelihood estimates					
	Parameter	DF	Estimate	Standard Error	Wald Chi-square	Pr > ChiSq
	Intercept	1	4.2436	0.0947	2006.6314	<.0001
1	PRIJ_BIN1	1	0.7017	0.1829	14.7250	0.0001
2	POCVYZ1	1	0.4855	0.1425	11.6116	0.0007
3	K2VEK1	1	0.4339	0.1458	8.8638	0.0029
4	K4VZDEL1	1	0.5198	0.2363	4.8393	0.0278
5	K5STABIL1	1	0.6811	0.1374	24.5883	<.0001
6	K6PLATCE1	1	0.7859	0.1617	23.6186	<.0001
7	K7FORBY1	1	0.4447	0.2227	3.9870	0.0459
8	K8FORSPL1	1	0.7795	0.1849	17.7723	<.0001
9	K29BYDTEL1	1	0.6725	0.1888	12.6855	0.0004

WoE1 variables selection

Run	In-sample (%)	Out-sample (%)
1	63.3	59.3
2	61.7	62.1
3	66.0	52.1
4	65.7	52.1
5	61.5	59.5
Ave	63.6	57.0
std	2.1	4.6

performance, where we fix the list of nine variables, estimate coefficients in-sample, and calculate Gini out-sample, is shown in the second part of Table 3.17 with average Gini around 56%, and is now much better than in Table 3.13. The difference between the in-sample and out-sample performance is now much tighter—below 4%, as expected.

Tables 3.18 and 3.19 show the results of the stepwise selection after "WoEization" with and without coarse classification. Note that the lists of selected variables slightly differ. The first WoE regression uses nine variables as the categorical regression, but Education now replaces Social Status, while the second WoE regression uses ten variables including both Education and Social Status. The out-sample performance of the WoE regression without coarse classification appears to be the best, but the slightly lower Gini of the WoE regression with coarse classification has a much lower variation and indicates better stability.

To conclude, any of the three approaches above would yield acceptable and more-or-less similar results. The alternatives involving coarse classification promise better robustness and a more transparent scoring function, while the approach based on "WoEization" without coarse classification saves considerable development time devoted to the transformation of variables.

The case study demonstrates that the development of a scoring function is not a fully automated process—it involves a number of expert decisions based on

Table 3.19 Performance of the model after "WoEization" and coarse classification

	Analysis of maximum likelihood estimates					
	Parameter	DF	Estimate	Standard Error	Wald Chi-square	Pr > ChiSq
	Intercept	1	4.1271	0.0955	1869.4498	<.0001
1	PRIJ_BIN1_woe	1	0.5007	0.2197	5.1950	0.0227
2	POCVYZ1_woe	1	0.4253	0.1248	11.6230	0.0007
3	K2VEK1_woe	1	0.4806	0.1570	9.3758	0.0022
4	K4VZDEL1_woe	1	0.4836	0.2417	4.0052	0.0454
5	K5STABIL1_woe	1	0.6367	0.1437	19.6249	<.0001
6	K6PLATCE1_woe	1	0.7539	0.1791	17.7248	<.0001
7	K7FORBY1_woe	1	0.5666	0.2363	5.7513	0.0165
8	K8FORSPL1_woe	1	0.7415	0.1938	14.6359	0.0001
9	K28SOCZAR1_woe	1	0.8378	0.3491	5.7610	0.0164
10	K29BYDTEL1_woe	1	0.6921	0.1941	12.7171	0.0004

WoE2 variables selection		
Run	In-sample (%)	Out-sample (%)
1	62.8	52.6
2	62.6	53.1
3	60.7	58.6
4	62.5	54.9
5	59.8	57.2
Ave	61.7	55.3
std	1.3	2.6

experience or a deeper understanding of the explanatory variables and underlying data. It is important to go carefully through the basic steps, starting from univariate analysis, continuing with variable selection, and ending with a thorough validation. If the final result is not satisfactory then one needs to go back iteratively and possibly change previously taken decisions until the results are considered good enough.

Reject Inference

A typical scoring function development dataset has a major problem called "reject bias" caused by the fact that financial institutions observe defaults only on exposures that have been previously approved. In other words, there are no good-bad observations for borrowers that have been rejected. For example, a very simple approval process might incorporate various "KO" (Knock-Out) criteria, as in an extremely conservative approach rejecting all applicants with income below 25,000. Based on Fig. 3.13, the case study set of observations would be reduced by more than 80%, and we could make hardly any statistical inference on incomes below 25,000. In a less extreme case, low income could be harshly penalized in the original scoring, causing a very low number of low income approved applications. Therefore the income variable would be probably eliminated in the new scoring function due to low discrimination caused mainly by the small number of low income observations and not by the fact that income is not a discriminating factor per se.

A number of techniques known as *reject inference* have been proposed to overcome the reject bias. The cleanest way to solve the problem is to construct a sample in which no case was rejected. This can be achieved by accepting all applicants over a limited testing period. Another possibility mentioned by Thomas et al. (2002) is to randomly accept applicants that would normally be rejected (with scores below the cut-off) in a proportion that inversely depends on the estimated probability of default. Specifically, if $p(s)$ is the probability of default depending on the estimated score s, then the approval rate $a(s)$ should satisfy the inequality $p(s)a(s) \leq DR_{\max}$ controlling the expected population default rate below a predetermined maximum limit $DR_{\max}$, i.e. we can set $a(s) = DR_{\max}/p(s)$. The approved cases then need to be reweighted by $w(s) = 1/a(s)$ in order to obtain a representative good-bad population sample based on accepted and rejected applications.[4]

Among a number of reject inference methods proposed in literature we should mention two basic approaches: *reweighting* and *augmentation* (see Crook and Banasik 2004; Anderson et al. 2009). The reweighting reject inference scheme assigns new weights to the approved cases in order to better represent the full sample distribution. The first step is to build an accept-reject scorecard (AR) using the dataset with all available explanatory factors and the accept/reject information. Then, for accepted observations with the AR score s (or within a score band), we calculate the empirical approval rate $a(s)$ and reweight the approved cases by $w(s) = 1/a(s)$ in order to represent the rejected cases where the good-bad observations are missing. This procedure implicitly assumes (see also the original paper of Hsia 1978) that the probability of default conditional on the AR score s is the same for the approved and rejected cases, i.e.

$$\Pr\left[Bad \middle| s, A\right] = \Pr\left[Bad \middle| s, R\right]. \tag{3.14}$$

The augmentation method extends the dataset with the rejected observations. Let us assume, combining the prior and new models, that p_j is the estimated probability of default of a rejected case j. In principle, there are two ways how to assign the rejected cases with an unknown outcome to the good-bad categories. The first, also called the *fuzzy method*, splits the observation j into two: j' assigned bad with the weight p_j and j'' assigned good with the complementary weight $1 - p_j$. In this way we more or less perfectly represent the available knowledge of the reject cases in

[4]The logistic regression can be easily generalized to include positive weights of the observations. The weights are applied as multiplicative factors in the log-likelihood function (3.8). Standard statistical SW packages allow one to use the observation weights.

the augmented dataset. Another approach is to classify each reject j so that the estimated default probabilities are somehow respected. One possibility is to classify j as randomly bad or good with the probabilities p_j and $1 - p_j$. In order to avoid additional sampling error, it might be preferable to split the reject observations into bands according to the estimated default probabilities, and sample good and bad from the bands in appropriate proportions. Finally, we can also apply the *hard cut-off method* where all rejected cases j with p_j below a certain cut-off are classified as bad and others (above the cut-off) as good. The cut-off is usually set as the average expected default probability on the rejected subsample.

The default probabilities on the rejected cases can be estimated using the old scoring function, in particular if the previous model worked quite well and we just want to improve it using the new observations. Another possibility is to use a pure *extrapolation approach* where a new model is developed on the approved subsample and the new scoring function is used to estimate the default probabilities on the reject cases. Finally, we may develop the accept-reject scorecard, use the assumption (3.14), and estimate

$$\Pr[Bad|s,A] = \sum_{\mathbf{x},\, s_{\mathrm{AR}}(\mathbf{x})=s} \Pr[Bad|\mathbf{x},A]\Pr[\mathbf{x}|s_{\mathrm{AR}}(\mathbf{x}) = s]$$

where $\Pr[Bad|x,A]$ is the default probability estimated using the new model and $\Pr[x|s_{\mathrm{AR}}(x) = s]$ is the relative frequency of the characteristic vector $\mathbf{x}$ with the accept-reject score. In other words, $\Pr[Bad|s,A]$ is estimated as the average probability of default given by the new model weighted by the conditional distribution of the characteristic vectors in the full sample. This method may be useful, for example, if the old model is not known and we have only the information on the accept-reject decisions.

It is obvious that none of the augmentation methods is perfect, and so we should control the proportion between the approved cases, where real outcomes are known, and the augmented rejected cases, where the classification is artificial. If the proportion of the rejected cases is too large, then an additional reweighting multiplier can be applied so that the overall weight of the reject subsample is below a reasonable limit, e.g. 30 or 50%. Various studies (e.g., Crook and Banasik 2004) show that reject inference does not bring an automatic improvement of the final scoring function and that the outcome depends on the specific situation and good expert decisions.

Combination of Qualitative and Quantitative Assessment

As shown above, scoring function explanatory variables can be quantitative as well as qualitative. However, in the case of corporate borrowers' assessment, it appears better to separate the automatic financial rating based on the quantitative financial ratios from the qualitative rating based more on the expert's judgment. The

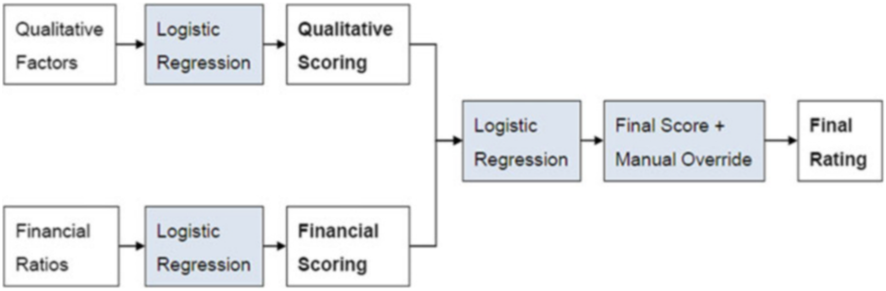


Fig. 3.14 Qualitative and quantitative corporate rating process

qualitative rating could be purely traditional, or more mechanical, assessing various qualitative factors such as management experience, business relationships, market position, etc., in a questionnaire. The answers would be numerically encoded and translated into a qualitative rating score. The final qualitative rating could be manually adjusted under certain conditions. The qualitative and quantitative scores could be then combined by a logistic regression model with two explanatory variables, producing the single final score, and a rating grade (Fig. 3.14). The decomposition of the process provides a better interpretation, and understanding, of why a borrower receives a high or low rating. This is important for the corporate segment, where the automated rating system usually has just a decision-making support role.

Rating Calibration and PD Estimation

Even though the logistic regression tries to predict directly the probabilities of default, the process of scoring function development, and the assignment of the probabilities of default to scoring values, are usually separated. There might be several reasons for this. The development sample might use a proportion of good and bad, e.g. 50:50, not corresponding to the real observed default rate. Therefore, the PDs estimated by the logistic regression might be biased due to the dataset construction. Another more fundamental issue is related to changes in the overall default probabilities (for a specific segment and product) over time. It appears that the probability of default of a borrower depends not only on the borrower's specific factors, but also on a set of systematic factors, such as GDP growth, unemployment, interest rates etc. A scoring function based on borrower specific information, and on a set of historical data, should keep its discriminatory power, but will not predict the PDs correctly if the systematic factors (usually not used as explanatory variables) change. Therefore, scoring functions developed some time ago do need recalibration; i.e., a new function assigning probabilities to the scores. The scoring function development process is time consuming, and even a freshly developed scoring based on a large dataset spanning a number of years needs recalibration. If the dataset covers good and bad years, i.e. goes *through the cycle* (TTC), then the resulting logit function gives estimates of some sort of average TTC probabilities of default. That is desirable from the regulatory point of view, but business requires rather *Point in Time* (PIT) estimates of the probabilities of default; i.e., the best available predictions for the coming period, typically 1 year, depending on the product maturity. Therefore, the TTC scoring function needs to be recalibrated using up-to-date default information to produce PIT PD predictions. Another situation is when there is only a relatively short history of a product and a small number of defaults defined in the desired time horizon. The scoring function can be developed based on a modified definition of default observed in a short horizon, for instance, default on the first payment or just 60 DPD (days past due), and then recalibrated to the target definition of default.

The recalibration process may be set up as follows. Let $s(a)$ be a linear transformation of the log odds score $-\boldsymbol{\beta}' \cdot \mathbf{x}(a)$, where the coefficients come from the logistic regression. The values are typically scaled into a conventional range such as 0–100, or 0–1000. For new prospective borrowers we employ the single variable Logit model, with the score $s(a)$ being the only explanatory variable:

$$p(a) = \Lambda(\alpha + \beta s(a)), \text{ i.e. } \ln \frac{p(a)}{1 - p(a)} = \alpha + \beta s(a).$$

To estimate the simple model, we use the latest observations covering the desired default time horizon. For example, if the prediction horizon is 1 year, we use the set of borrowers that were not in default 1 year ago and observe the defaults which have occurred during the past 12 months. In this way, we utilize both the discriminatory power of the scoring function and the latest information on defaults. The intercept α and the slope β could possibly be estimated by an OLS regression on the log odds explained by the score splitting the sample into a number of bins according to the scores. The simple approach may result in a prediction bias, the difference of the average predicted PD, and the overall frequency of default in the calibration sample. The problem can be solved simply by running the single variable logistic regression on the calibration sample. According to (3.10) there will be no prediction bias.

Figure 3.15 gives a typical example of the development of actual default rates versus the predicted PD. The PD predicted by a scoring function developed at the end of 2003, without any subsequent calibration, shows a more-or-less stable prediction of PDs at a level around 4% on a given product portfolio (e.g. consumer loans). The small fluctuations are caused by changes in portfolio composition (in terms of explanatory variable distribution). The actual experienced portfolio rates of default are, however, much more volatile, and diverge from the original scorecard predictions, going up, notably, to 7% in 2007 and 2008. Thus the non-calibrated PDs, provided by the original scorecard, would significantly underestimate the risk. If the bank wanted to keep the same original scoring function, then a solution would be given by quarterly recalibrations. The predicted PD, based

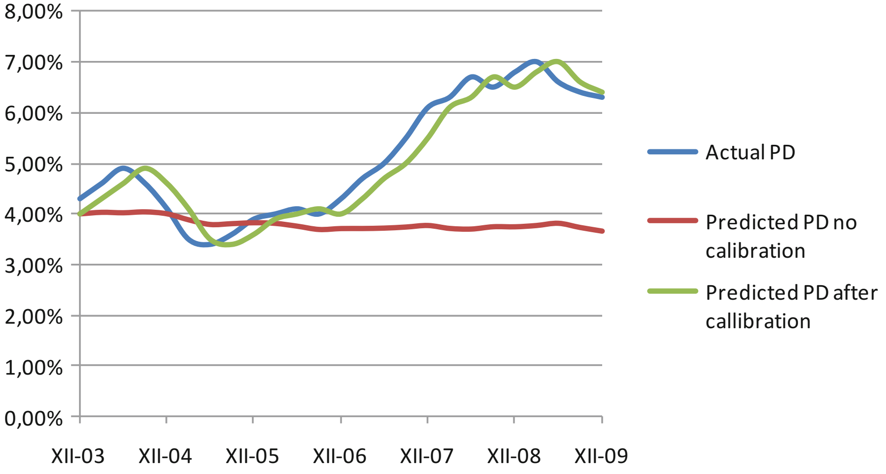


Fig. 3.15 Comparison of actual and predicted rates of default with and without calibration

on the most recent data, still has certain lag compared to the reality, but performs much better than the non-calibrated scoring function.

The scores that take many possible values are usually translated into rating grades on a smaller scale. From the business perspective the scale might be very simple: 1—“yes”, 2—“maybe”, 3—“no”, corresponding to the approval process. Applications rated as “yes” are automatically approved, applications rated as “no” declined, and applications rated as “maybe” are sent for further expert assessment. If the rating is used to differentiate the credit margin according to the risk, then there should be more grades in the approval zone. Better differentiation may also be useful in the grey zone, but the number of grades on a reasonable scale for the retail segment usually does not exceed 10, and 25 for the corporate segment. There are different philosophies as to how to translate the scores into rating grades. The simplest approach is to fix the score value intervals for each rating grade, with a targeted distribution of borrowers. For example, if the score value range is [0, 1000], and the rating scale is the Moody’s one, then the range for “Aaa” could be [981, 1000] for “Aa1”: [961, 980], and so on up to “Ca” with [0, 100]. In this approach, the ratings are a kind of transformed rounded values of the scores. The rating grades will express TTC rather than PIT probabilities of default, depending on the cyclicality of the explanatory variables and on the calibration approach. An alternative philosophy is to assign the rating grades according to fixed bands for PIT PD values. The rating grade 1 may be defined by the PD interval [0, 0.5%), the grade 2 by [0.5%, 1.5%), and so on, with the last grade reserved for the highest PD values. In that case, the assignment of scoring values to the rating grades depends on the PD recalibration. While for the first approach, for a single rating class, historical observed frequencies of default fluctuate with the cycle, for the second approach, the observed values should be within, or not too far from, the PD interval defining the rating grade.

Shadow Rating

The shadow rating is typically employed when default data are rare, but there are external ratings from respected rating agencies. This may be the case of governments, regions, municipalities, financial institutions, or large corporations. The bank often does not have, or does not want to build, a specialized credit analysis department, but at the same time, it is difficult to develop a scoring system using an ordinary statistical approach, as there are too few historical defaults. If a significant part (but not all) of the segment is rated by external agencies, then the idea is to mimic the external ratings by an internally developed scoring function that uses the existing external ratings and related PDs as a development dataset. The first step requiring expertise is the identification of the explanatory factors potentially used by the rating agencies. The factors can include macroeconomic factors, and financial or country risk indicators. Secondly, the external ratings have to be calibrated, i.e. probabilities of default have to be attached to them. In this case one can use the frequency of default over an observation period preceding the rating decision (PIT philosophy), or over a fixed, long time horizon going through the cycles (TTC philosophy). Then we apply the logistic regression model of the form:

$$p_i = \Lambda(\boldsymbol{\beta}' \cdot \mathbf{x}_i), \text{ i.e. } \ln \frac{p_i}{1 - p_i} = \boldsymbol{\beta}' \cdot \mathbf{x}_i,$$

where $\mathbf{x}_i$ is the vector of the explanatory variables of an externally rated borrower, and p_i is the predicted probability of default (given by the external ratings in the development sample). The coefficients $\boldsymbol{\beta}$ can be estimated by a simple OLS regression of log-odds, explained by the factors selected by a method as described above. The performance of the system does not have to be very high, and it should be considered, rather, as a supporting decision tool for an analyst who has the power to override the final decision. For details and a case study see Erlenmaier in Engelmann and Rauhmeier (2006).

Survival Analysis

As we can see, the classical logistic regression approach does not capture well the time scale of the default dynamics. Sometimes we need to predict default in a short time horizon; in other situation, over a long time horizon. There is also empirical evidence that for most products the probability of default depends on aging; i.e., time on the books. For example, unsecured consumer loans have a large probability of default shortly after being granted, since some borrowers might not be willing to repay from the very beginning, while the probability declines later for loans that have "survived" the initial period (see Fig. 3.16). In the case of mortgages the typical pattern is different: initially the probability of default is low, but later, after a few years, it may go up due to the loss of repayment capacity (e.g. unemployment, divorce, jump in interest rates after a reset, etc.). Therefore, to model the behavior

Fig. 3.16 An example of monthly empirical default rates depending on time after the first drawing for unsecured consumer loans and mortgages

of the loan over its life cycle, we need more than a prediction of PD in a single time horizon.

Survival analysis is appropriate in situations where we observe a population of objects that stay in a certain state (survive) for some time until an exit (death or failure) happens. This is, literally, our situation; interpreting default as an exit. Survival analysis also allows the use of censored observations; i.e., observations of objects which are known to have survived until a certain time, but on which no more information is available. This is another useful feature, incorporating the most recent data when we have many newly granted loans (within the last year), where the default/non-default observations are not normally used in the logistic regression. A number of studies (see. e.g., Andreeva 2006; Thomas 2009) have confirmed that the method may give better results in terms of discrimination power measures than the classical logistic regression.

Generally, the goal is to study the time until failure, and the probability of survival or failure in a given time period. The key survival analysis concepts (Greene 2003) are the *survival function*, and the *hazard rate*. Let T be the random variable representing the time of exit of an object, $f(t),\ t \geq 0$, its continuous probability density function, and $F(t)$, its cumulative distribution function. Then, $F(t)$ is the probability of exit (failure) of an individual until the time t, while the survival function $S(t) = 1 - F(t)$ gives the probability of survival until t. The hazard rate is defined as

$$\lambda(t) = \frac{f(t)}{S(t)}.$$

This gives the rate at which objects that have survived until the time t, exit exactly at time t; specifically, $\lambda(t)\delta t$ is approximately the probability of exit during the time interval $(t, t + \delta t]$, provided the object is still alive at t. In the case of credit risk survival modeling, the hazard function is usually called *the intensity of default*. The probability of exit $F(t)$ in this case, corresponds to the cumulative probability of default in the time horizon t. It is also useful to define the *cumulative hazard function*

$$\Lambda(t) = \int_0^t \lambda(s)ds.$$

It can be seen that the cumulative survival function then is $S(t) = e^{-\Lambda(t)}$, and the cumulative exit (default) rate $F(t) = 1 - e^{-\Lambda(t)}$.

Figure 3.17 shows an example of default hazard functions for consumer loans (Pazdera et al. 2009), depending on borrower's education level, with a typical patter where the hazard is higher during an initial period after the credit origination. This is also confirmed by Andreeva (2006) where the credit card data from several European countries show the general pattern of significantly higher hazard during the first year of the exposures.

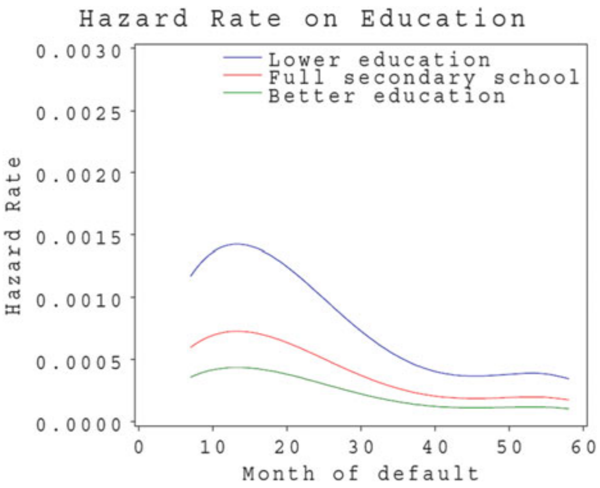


Fig. 3.17 Consumer loan default hazard functions (Source: Pazdera et al. 2009)

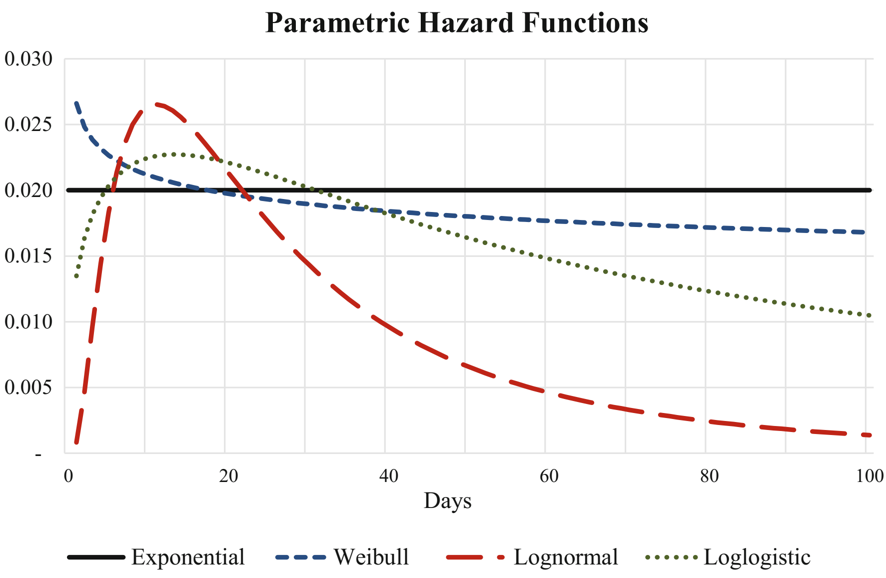


Fig. 3.18 Parametric hazard functions

The models are specified through the hazard function, given in parametric or semi-parametric forms. The parameters are, moreover, allowed to depend on explanatory variables characterizing the objects under observation. The simplest model is the exponential one with the constant hazard function $\lambda_{\text{Exponential}}(t) = \lambda$. Other, more general, parametric models allow different shapes of the hazard function (Fig. 3.18).

For example, the parametric Weibull model is specified by:

$$\lambda_{\text{Weibull}}(t) = \lambda p(\lambda t)^{p-1}, \ \ S_{\text{Weibull}}(t) = e^{-(\lambda t)^p},$$

while the Lognormal and Loglogistic models have the form:

$$f_{\text{Lognormal}}(t) = (p/t)\phi(p\ln(\lambda t)), S_{Lognormal}(t) = \Phi(-p\ln(\lambda t))$$

$$\lambda_{\text{Loglog}}(t) = \lambda p(\lambda t)^{p-1}/[1 + (\lambda t)^p], \ \ S_{\text{Loglog}}(t) = \Lambda(-p\ln(\lambda t)) = \frac{1}{1+(\lambda t)^p}.$$

In fact, the Lognormal and Loglogistic models are better characterized by defining the log of default time, $\ln T$, cumulative distribution; standard normal[5] with mean (*intercept*) $\mu = -\ln \lambda$ and standard deviation (*scale*) $\sigma = 1/p$; and the logistic distribution with mean $\mu = -\ln\lambda$ and variance $\sigma^2 = \pi^2/(3p^2)$. It turns out that the Lognormal model is a parsimonious choice fitting most credit products relatively well.

The coefficient $\lambda = e^{-\mathbf{x}'\boldsymbol{\beta}}$ in all cases depends on the vector of covariates $\mathbf{x}$ (without the constant 1). The coefficients $\boldsymbol{\theta}=(\boldsymbol{\beta}, \mathbf{p})$ are estimated using a maximum likelihood method, maximizing the total log-likelihood function:

$$\ln L(\boldsymbol{\theta}) = \sum_{\substack{\text{uncensored} \\ \text{observations}}} \ln \lambda(t|\boldsymbol{\theta}) + \sum_{\text{all observations}} \ln S(t|\boldsymbol{\theta}). \tag{3.15}$$

The lognormal and loglogistic parameterizations can also be formulated as *accelerated failure time models* where $\ln T = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$ and ε has the specified distribution.

The parametric models are attractive for their simplicity, but may impose too much restriction on the structure of data. Fewer restrictions are imposed by the Cox (1972) proportional hazard model, which we shall focus on. The proposed hazard function has a semi-parametric form:

$$\lambda(t, \mathbf{x}) = \lambda_0(t)\exp(\mathbf{x}'\boldsymbol{\beta}),$$

where $\lambda_0(t)$ is called the baseline hazard function independent of the explanatory variables $\mathbf{x}$ while $\exp(\mathbf{x}'\boldsymbol{\beta})$ determines the risk level.

The baseline hazard is a step function estimated on a discrete set of points where exits or censoring take place. The corresponding survival function is in the form

[5] Here, Φ and ϕ denote the cumulative distribution and the probability density function of the standard normal distribution.

$$S(t,\mathbf{x}) = \exp\left(-\int_0^t \lambda_0(s)\exp(\mathbf{x}'\boldsymbol{\beta})\right) = S_0(t)^{\exp(\mathbf{x}'\boldsymbol{\beta})}, \text{ where } S_0(t)$$
$$= \exp\left(-\int_0^t \lambda_0(s)\right). \tag{3.16}$$

The coefficient vector $\boldsymbol{\beta}$ is estimated using the partial likelihood: if an object i, with covariates $\mathbf{x}_i$ exits at time t_i, if we assume that there is only one exit at that time, and if A_i is the set of objects alive at t_i, then the partial likelihood that just $i \in A_i$ exits is:

$$L_i(\boldsymbol{\beta}) = \frac{\lambda(t_i,\mathbf{x}_i)}{\sum_{j\in A_i}\lambda(t_i,\mathbf{x}_j)} = \frac{\exp(-\mathbf{x}_i'\boldsymbol{\beta})}{\sum_{j\in A_i}\exp(-\mathbf{x}_j'\boldsymbol{\beta})}. \tag{3.17}$$

The coefficients $\boldsymbol{\beta}$ are then obtained by maximizing $\ln L = \sum_{i=1}^{K} \ln L_i$ numerically using the Newton–Raphson algorithm. Generally, we need to handle ties; i.e., multiple exits at the same time. This happens, typically, when we have only monthly data. The partial likelihood function (3.17) can be generalized in a straightforward manner for the case of d_i ties (frequency weights) at the same time t_i. However, due to computational complexity,[6] the exact partial likelihood function is usually approximated by an estimate due to Breslow (1974), or due to Efron (see Kalbfleisch and Prentice 2002). Given $\boldsymbol{\beta}$, the baseline hazard function values are estimated separately for each of the unit time intervals, where the function is assumed to be picewise constant maximizing the likelihood function:

$$L_t = \prod_{i=1}^{n}\left[\lambda_0(t)\exp(\mathbf{x}_i'\boldsymbol{\beta})\right]^{dN_i(t)}\exp\left(\sum_{u=0}^{t} -\lambda_0(u)\exp(\mathbf{x}_i'\boldsymbol{\beta})Y_i(u)\right).$$

Here $dN_i(t)$ is an indicator of the fact that subject i died in the time interval $(t-1, t]$, and $Y_i(t)$ is an indicator of the fact that subject i is at the time $t-1$ still alive. By differentiating the log likelihood function with respect to $\lambda(t)$ the maximum likelihood estimator can be derived in the Breslow-Crowley form:

[6]The summation in the denominator of (**3.17**) must, in general, be done over all subsets of A_i with the size d_i.

$$\hat{\lambda}_0(t) = \frac{\sum_{i=1}^{n} dN_i(t)}{\sum_{i=1}^{n} \exp(\mathbf{x}_i'\boldsymbol{\beta}) Y_i(t)}. \tag{3.18}$$

If there are no explanatory variables; i.e. $\exp(\mathbf{x}_i'\boldsymbol{\beta}) = 1$, then the estimator gives us the Kaplan-Meier hazard rate function, and the corresponding Kaplan-Meier survival function.

So, to apply the survival analysis approach, we need the survival-time dataset with characteristics vectors $\mathbf{x}_i$, and times of default or censoring t_i indicated by $c_i \in \{0, 1\}$, for borrowers $i = 1, \ldots, N$. The time is measured from the granting of the loan until default, or until the end of the observation period. The first step, provided by the main statistical packages, would be to draw the Kaplan-Meier survival function and the corresponding hazard function. Then we may decide for a parametric or non-parametric model. It is advisable to select the best explanatory variables through the classical logistic regression procedure with a standard default horizon. After that we can compare, relatively easily, the different survival models. The non-parametric Cox models better fit the specific shape of the hazard function. On the other hand, the parametric models may give better regression results and are useful if we have only short time horizon data and want to perform a longer time horizon analysis. The best survival models show similar performance compared to the Logistic Regression models, but their main advantage is that they provide consistent PD estimates for various time horizons. For a comparison of the models using a discrimination power measure we need, however, to revert to a fixed time horizon. For a more detailed discussion and illustration of survival modeling, see also Stepanova and Thomas (2002).

In practice, the survival analysis is often combined with the standard scoring or rating outcomes. The logistic regression allows us to perform a thorough univariate analysis and selection of explanatory variables, as explained and illustrated in the previous subsection, which is not so straightforward in the survival analysis set-up. Nevertheless, the survival analysis may use the score as a numerical variable or the rating as a categorical explanatory variable.

For example, Fig. 3.19 shows the estimated Kaplan-Meier (empirical), exponential, Weibull, and lognormal hazard function estimated with a categorical variable for ratings A, B, C on a real-life large mortgage portfolio. The duration (months on book) of the observations ranges from 1 to 73 months. The number of observations of course declines, and the empirical hazard rate becomes noisier as we approach the maximum observation span. If we wanted to use the empirical hazard rate, in fact the Cox model, then there would be no estimations beyond 73 months. Therefore, it is more appropriate to choose a parametric model that smoothes out the noisy empirical hazard functions and extrapolates its values beyond the observation horizon of 73 months. The exponential, Weibull, and lognormal models have

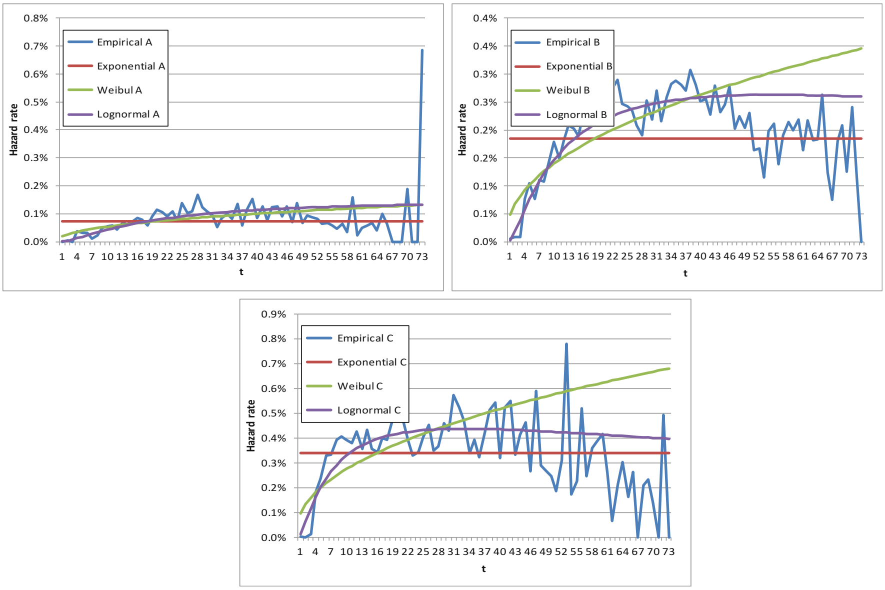


Fig. 3.19 Comparison of different survival models conditional on ratings on a mortgage portfolio

been compared in terms of log-likelihood with the lognormal model performing the best as indicated visually by Fig. 3.19. The intercept (mean) of the lognormal model depends on the rating category, while the scale is the same for all the rating classes. Note that the rates in Fig. 3.19 are measured on a monthly basis, for example, the hazard 0.3% on a monthly basis corresponds to the annualized default probability 3.5%, etc. The hazard rates and the corresponding survival probabilities can be used to estimate credit risk margins consistently in the range 0.2–0.8% p.a. as described in Sect. 3.5 (based on a relatively low loss given default for a mortgage product that is below 20%).

3.4 Alternative Automated Rating Systems

Although logistic regression provides generally very good predictive results, there is ongoing research into alternative methods that might lead to better performance of the classification system in specific situations. For example, the goal of collection scoring is to predict the impact of various collection actions (phone calls, mails, legal actions, etc.) where there is no need to explain the outcome to clients or internally to sales persons in the bank. What matters most is the prediction power, and in this case the neural networks, support vector machines, or random forests might be preferred over the classical logistic regression.

Classification Trees

We will start the overview of alternative systems with classification trees, which present, in a sense, a simpler and, to a certain extent, more transparent method than logistic regression. However, the performance of classification trees is usually worse compared to logistic regression, and they are used rather as a complementary analytical tool, for example, in order to identify the most important variables, potential segmentation, etc.

Given a set of characteristics for a set of borrowers, it is natural to try to separate the better ones from the worse ones using a classification tree (or recursive partitioning algorithm—see Breiman et al. 1984), where we divide the borrowers firstly according to the most relevant variable, then those subgroups according to another variable, or, with respect to the first variable, with a limited range of values, and so on. The rules of division must be developed on a training sample for which we have the initial characteristics and indicators of default after a given time period. The division rules need to be chosen in order to maximize the homogeneity of the two subgroups or to provide a divergence measure of the difference in default risk between the two subsets. For an ordinal, or continuous, variable, the tree algorithm looks for possible split thresholds, while for a categorical variable it even needs to consider all the possible subsets of the categorical values. The splitting is repeated until no group can be split into two subgroups based on a chosen criterion when the homogeneity or discrimination measure improvement is not significant, or the group becomes too small, or the tree is too deep, etc. The ratio of bad in the nodes can be possibly used as a PD predictor or as a score. The terminal nodes are then classified as good or bad based on a cut-off.

Then, each new applicant can be classified according to the rules pertaining effectively to a rating, and possibly to a PD estimate. As usual, the performance of the classification tree (Gini coefficient, KS statistic, classification accuracy, etc.) can be assessed on a validation (testing) dataset.

The validation dataset is sometimes used for so called *pruning*. The idea is to make the tree “grown” on the training set more robust by pruning some of its branches in order to improve its performance on the validation dataset. There are many possible approaches to pruning, and it is often rather an expert process. In this case, there should be a third testing set to measure the final performance of the pruned tree, since the validation set has, in fact, been used in the second phase of the tree training (pruning).

Popular splitting measures include the Kolmogorov-Smirnov statistic, the Gini index, Entropy index, or the chi-squared statistic.

The Kolmogorov-Smirnov statistic is calculated as described in Sect. 3.2, but for a simple two-value classification of the group into the *left* (l) and *right* (r) subgroup. Therefore, the statistic is simply:

$$KS = \left|\Pr\left[l\middle|B\right] - \Pr\left[l\middle|G\right]\right|.$$

The Gini index, which should not be confused with the Gini coefficient, is one of the impurity index measures that aim to measure how impure each node v of the tree is (a node is optimally pure either if all its members are good or all are bad). If $i(v)$ is such an index, then the strategy is to split the node into the left and right subgroups so that their weighted impurity decreases as much as possible. Therefore, we maximize the change in impurity

$$I = i(v) - i(l)\Pr[l] - i(r)\Pr[r].$$

One can propose as a possible basic impurity index the minimum proportion of good or bad in a node, $i(v) = \min\left(\Pr\left[G\middle|v\right], \Pr\left[B\middle|v\right]\right)$. Although this looks useful, this index does not work well in practice: for example, if the parent node v and the child nodes l and r all contain a minority of bads, then the change of impurity will be always equal to zero and the measure does not help to decide which split is the best. Instead of being a linear proportion, the Gini index is a quadratic function of the proportion of good (or bad) and gives relatively more weight to purer nodes (see Fig. 3.20):

$$i_G(v) = \Pr\left[G\middle|v\right] \times \Pr\left[B\middle|v\right] = \Pr\left[G\middle|v\right] \times \left(1 - \Pr\left[G\middle|v\right]\right).$$

The entropy index is another nonlinear impurity measure defined as follows:

$$i_E(v) = -\Pr\left[G\middle|v\right]\ln\left(\Pr\left[G\middle|v\right]\right) - \Pr\left[B\middle|v\right]\ln\left(\Pr\left[B\middle|v\right]\right).$$

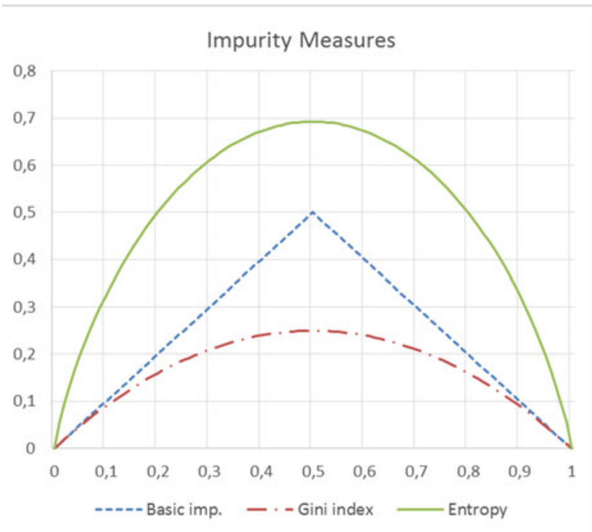


Fig. 3.20 Comparison of alternative impurity measures (basic impurity, Gini index, and entropy index)

This measure is related to the information statistic and can be intuitively defined as the level of disorder. In other words, it is a measure reflecting the average amount of information of a good-bad split in the given proportion.

Finally, let us mention the *chi-squared (CS) statistic*. Generally, given K groups with $n_1, \ldots, n_K$ observations, and $d_1, \ldots, d_K$ defaults, the statistic aims to test the hypothesis that the distribution of defaults is the same for all the groups. So, it is natural to set the average expected probability of default $PD = \sum_{k=1}^{K} n_k / \sum_{k=1}^{K} d_k$, and, similarly to the Hosmer-Lemeshow statistic (Sect. 3.2), calculate the sum of normalized squared differences between the expected and observed numbers of default:

$$CS = \sum_{k=1}^{K} \frac{(n_k PD - d_k)^2}{n_k PD(1 - PD)}. \tag{3.19}$$

Note the variance of the variable defined as the number of defaults on n_k debtors (provided that the events of default are independent, and the probability of default is PD) indeed is $n_k PD(1 - PD)$. The statistic has (asymptotically) the chi-squared distribution with $K - 1$ degrees of freedom, so we can use the p-values to compare partitioning with different numbers of groups. In the case of the (binary) classification trees we have only two groups, $K = 2$, and we try to maximize the CS statistic as a divergence measure between the rates of default in the left and in the right group.

The Case Study Continued

Figure 3.21 shows an example of a classification tree built in the SAS Enterprise Miner application on the same data set as the one used in the Sect. 3.3 case study. The dataset where categorical variables have been already coarse classified as in Sect. 3.3 has been split into training and validation sets in the proportion 70:30 with approximately the same proportion of bads. The tree shown in Fig. 3.12 has depth limited to 3 in order to have only a few nodes on the display. The algorithm using the Gini impurity index as the branching criterion identified the length of employment as the most important, then age, number of dependents, etc. Note that the base default rate of 1.45% (on the root with 7,653 observations) splits into 2.47% in the left subgroup of borrowers with length of (current) employment less than 7 and into a much lower default rate of 0.72% in the right group of borrowers with stable employment for 7 or more years. Then, following the right branch, the number of dependents shows a high discrimination power with a high default rate of 3.91% for childless borrowers and a low default rate of 0.58% for those having at least one child. Following the right subgroup again, it is interesting to note that borrowers with very long employment duration (9 years or more) have a much higher default

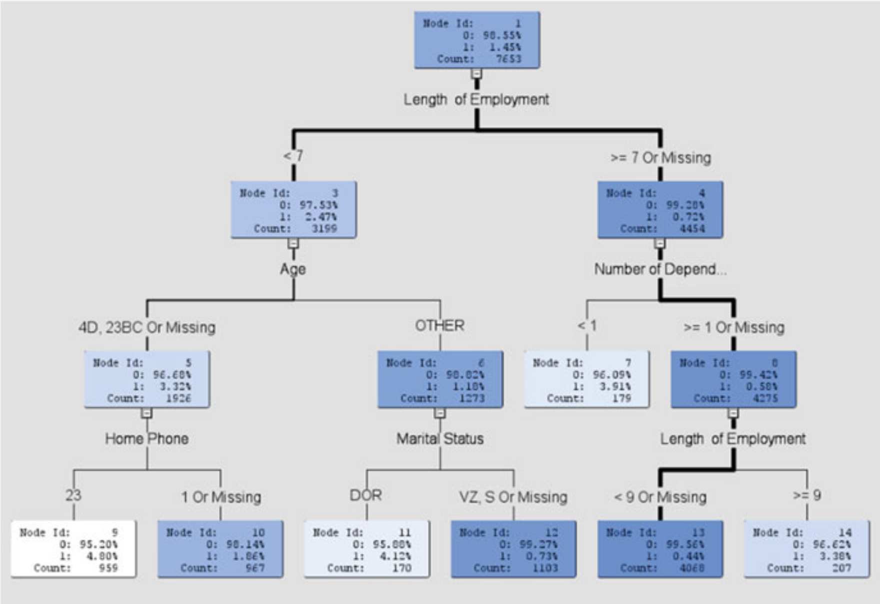


Fig. 3.21 Classification tree with maximum depth set 3 built in SAS Enterprise Miner 13.2

Table 3.20 A comparison of the logistic regression, classification tree, and neural network models

Data set	Training Sample		Testing Sample	
Performance	**KS statistic**	**Gini coeff.**	**KS statistic**	**Gini coeff.**
Logistic Regr.	0.44	0.62	0.41	0.53
Class. Tree	0.55	0.70	0.39	0.47
Neural Network	0.54	0.69	0.37	0.52

rate than those employed for between 7 and 8 years. This dependence might be considered inconsistent, and the two branches could possibly be pruned on an expert basis.

Therefore we can see that the classification tree can provide a relatively nice interpretation and insight into the importance of the various variables and their combinations.

The full tree was built automatically using the Gini impurity measure and setting 150 as the minimum leaf size. The depth of the tree is 7 and it would be difficult to display in a legible way. Table 3.20 and Fig. 3.22 show a comparison of the prediction performance of the classification tree with the logistic regression and a neural network model (with three neurons in one hidden layer as described below), all automatically developed in the SAS Enterprise Miner. It is interesting to compare the training and testing sample results. While the classification tree and neural network models perform better compared to the logistic regression on the

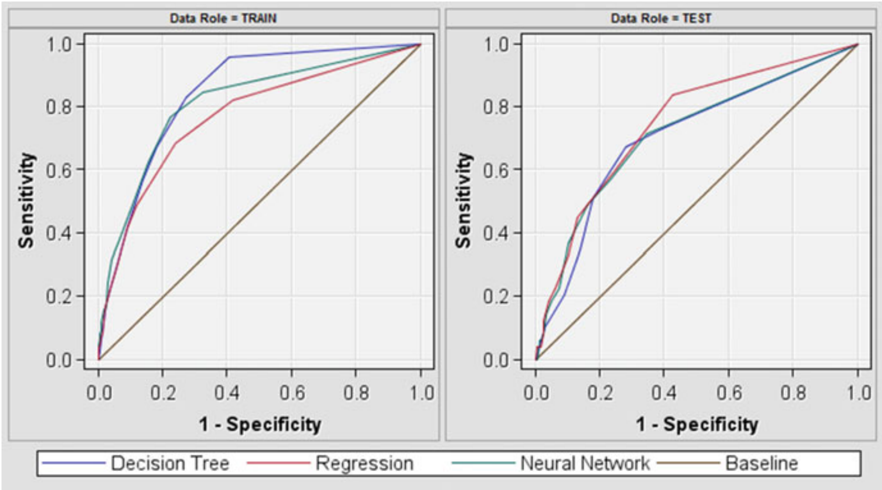


Fig. 3.22 A comparison of the ROC for the logistic regression, classification tree, and neural network models

training model, the logistic regression outperforms the other models on the testing sample. The robustness of the logistic regression is related to a more careful (stepwise) selection of variables. In order to obtain a fair comparison we have used the dataset after the coarse classification that was done as part of the logistic storing function development. However, the coarse classification is not usually done before applying the tree or neural network models, making the risk of overfitting even more serious. As mentioned above, the robustness of the classification tree can be improved by automated or interactive pruning, but the general experience is that it is difficult to achieve performance of the logistic regression obtained by the standard development procedure.

Artificial Neural Networks

Neural networks have been inspired by the architecture of the human brain, where a large number of dendrites carry signals neurons which send converted signals to other neurons, etc. The artificial neural networks approach is classified as a non-statistical method for scorecard development (Thomas et al. 2002). Nevertheless, an artificial network model can be on a high level described mathematically as a nonlinear predictive function $f(\mathbf{x};\mathbf{w})$, whose predictive power is assessed by standard statistical methods and whose vector of parameters $\mathbf{w}$ is estimated by various numerical and, to a certain extent, statistical methods. The outcome is a numerical or classification value depending on the vector of explanatory variables $\mathbf{x}$, and the form of the function is given by the neural network architecture. The simplest architecture is a single-layer neural network shown in Fig. 3.23.

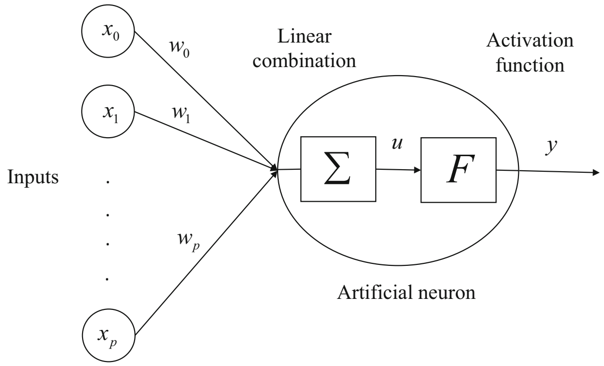


Fig. 3.23 A single-layer neural network

Algebraically, the output of the network is $y = F(u)$ where $u = \sum_{j=0}^{p} w_j x_j$ is a linear combination of the explanatory variables (as usual with $x_0 = 1$ so that w_0 represents the constant) and F is a nonlinear *activation function*. A popular one is our favorite logistic function $F(u) = \frac{1}{1+e^{-au}}$ which transforms the linear combination u onto a value in the interval $(0, 1)$. Other alternative logistic functions are the simple threshold function ($F(u) = 1$ if $u \geq 0$ and $F(u) = 0$ if $u < 0$), or the tangent hyperbolic function, $F(u) = \tanh(u)$, etc.

Note that the single-layer neural network with the logistic activation function is equivalent to the logit model. However, the simplest network is not able to handle many classification problems like the well-known XOR problem (unless we can use nonlinear transformations and combinations of the inputs). It is obvious that the neural networks provide much more flexibility if we allow more neurons and layers processing inputs from previous layers. Figure 3.24 shows an example of a two-layer neural network with three *hidden neurons*. The inputs are firstly combined and transformed by the three hidden neurons and their outputs are finally processed by the output neuron. In this case we have to index the coefficients and intermediate inputs/outputs in a slightly more complicated way, where the superscript stands for the layer (not for a power). The neural networks can have more output neurons and any number of hidden neurons and layers. The number of parameters is, of course, rapidly increasing, with the number of hidden neurons making the calibration computationally more difficult and causing possible overparametrization. In spite of that, due to increased computational power, *deep neural networks* with many hidden layers have become quite popular and successful in many artificial intelligence areas such as automatic vision or speech recognition.

A neural network $y = f(\mathbf{x}; \mathbf{w})$ is trained, i.e. its coefficients calibrated, using similar numerical techniques to those used for the statistical models based on a training dataset $\langle y_t, \mathbf{x}_t \rangle$. Here we consider, for the sake of simplicity, only one-dimensional output y_t and index the dataset by $t = 1, 2, \ldots$ in order to

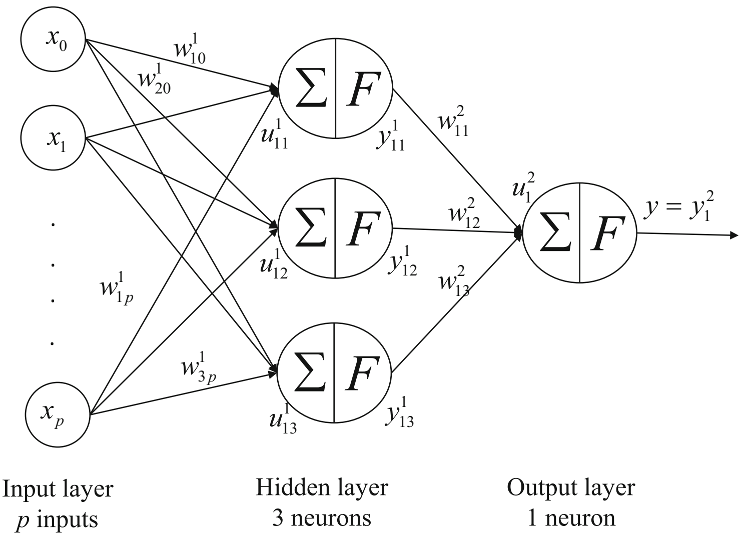


Fig. 3.24 A multi-layer neural network

emphasize that additional training cases may come over time and the training procedure can be incremental. Firstly, for an observation $\langle y_t, \mathbf{x}_t \rangle$ and an estimation $\hat{y}_t = f(\mathbf{x}_t; \mathbf{w})$ given a set of parameters $\mathbf{w}$, we need to define an error function, e.g. the usual squared error

$$E(t) = \frac{1}{2}(y_t - \hat{y}_t)^2.$$

If we consider a block of observations or the full dataset, then we use the average value

$$E_{\text{mean}} = \frac{1}{N}\sum_{t=1}^{T} E(t).$$

If want to calibrate the network based on the full dataset, then the goal is simply to find the vector of parameters $\mathbf{w}$, minimizing the error function $E_{\text{mean}}(\mathbf{w})$. If we want to apply incremental learning, then we simply need to adjust the weights $\mathbf{w}$ to reflect appropriately the information contained in the training case t. The *back-propagation algorithm* is based in both cases on the gradient-descent method where we move in each iteration in the direction of the negative gradient $-\nabla E = -\left\langle \frac{\partial E}{\partial w_{ij}^c} \right\rangle$, i.e. $\Delta w_{ij}^c = -\eta \frac{\partial E}{\partial w_{ij}^c}$ where η is a positive training rate coefficient to make the changes smaller or larger. The partial derivatives can be calculated analytically using the chain rule, provided the activation function has an analytical derivative (which is the case, e.g., for the logistic function). For example, in the case of a single layer neural network $\hat{y} = F(u)$, $u = \sum_{j=0}^{p} w_j x_j$, and $E = 0.5(y - \hat{y})^2$ for a training case $\langle y, \mathbf{x} \rangle$. Hence,

$$\frac{\partial E}{\partial w_j} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial u} \cdot \frac{\partial u}{\partial w_j} = -(y - \hat{y}) \cdot F'(u) \cdot x_j,$$

and so we change w_j by a positive multiple of $(y - \hat{y}) \cdot F'(u) \cdot x_j$. In the case of a multiple-layer neuron network, the chain rule must be applied more times, e.g. for the two-layer network and for an input layer coefficient w_{ij}^1 shown in Fig. 3.24 where $\hat{y} = y_1^2 = F(u_1^2)$, $u_1^2 = \sum w_{1k}^2 y_{1k}^1$ and so on, the partial derivative can be expressed as follows:

$$\frac{\partial E}{\partial w_{ij}^1} = \frac{\partial E}{\partial y_1^2} \cdot \frac{\partial y_1^2}{\partial u_1^2} \cdot \frac{\partial u_1^2}{\partial y_{1i}^1} \cdot \frac{\partial y_{1i}^1}{\partial u_{1i}^1} \cdot \frac{\partial u_{1i}^1}{\partial w_{ij}^1} = -(y - \hat{y}) \cdot F'(u_1^2) \cdot w_{1i}^2 \cdot F'(u_{1i}^1) \cdot x_j.$$

Generally, the partial derivatives can be expressed analytically recursively starting from the output layer and going back to the deeper hidden layer neurons. The gradient-descent algorithm unfortunately, unlike the Newton-Rapson algorithm, does not know which training constant η is optimal, even if the optimization is done on the full dataset. A too small η may cause the algorithm to be very small and get stuck in a local minimum, while a too large η may cause oscillations and divergence of the algorithm. This is one of the reasons why sequential back-propagation is preferred. Regarding termination of the training, the theoretical condition would be $\nabla E = 0$. However, the practical condition to stop the back-propagation is to require that the absolute value of the change $\Delta E(t) = E(t) - E(t-1)$ should be sufficiently small. There are many possible numerical improvements of the basic back-propagation algorithm. One possibility, for example, is the momentum method setting

$$\Delta w_{ij}^c(t) = \alpha \Delta w_{ij}^c(t-1) - \eta \frac{\partial E(t)}{\partial w_{ij}^c}$$

where $0 < \alpha < 1$ reducing the oscillation of the weights.

Table 3.20 and Fig. 3.22 show that the neural network (in this case with one hidden layer) can be quite competitive compared to the logistic regression. By increasing the number of neurons we certainly increase the in-sample performance, but not necessarily the out-sample performance due to overfitting. Nevertheless, even with a high out-sample performance, credit analysts usually remain sceptical due to the difficult interpretability of the relatively high number of estimated coefficients and the black-box nature of the model.

Nearest-Neighbor Approach

The idea of the nearest neighbor approach is very simple: given a historical dataset of observations with known outcomes of the target variable $\langle y_i, \mathbf{x}_i \rangle$ and given a new case with a vector of explanatory variables $\mathbf{x}$ and unknown outcome y, let us find k "similar" or "near" historical cases $\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_k}$ and estimate y based on the known outcomes $y_{i_1}, \ldots, y_{i_k}$, e.g. calculating the average value $\hat{y} = \frac{1}{k}\sum_{j=1}^{k} y_{i_j}$. In the case of a binary variable, the average would be a score, and an appropriate cut-off value can be used to obtain a binary prediction.

In order to make the idea work, we have to define the "similarity" or "nearness" more exactly. The first proposal is to use the ordinary Euclidean metrics $d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \cdot (\mathbf{x} - \mathbf{y})}$ assuming that all variables are numerical (categorical being transformed to the dummy variables). Of course, there are many more general possibilities, for example, defining the distance by $d_A(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T A(\mathbf{x} - \mathbf{y})}$ where A is a positive definite matrix. Henley and Hand (1996), in their extensive study of nearest-neighbor applications in credit scoring, proposed a mixture of the Euclidean distance and the distance in the direction $\mathbf{w}$ that best separates good and bad, which can be obtained from discriminant or linear regression analysis, i.e. $d_A(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T (I + D\mathbf{w}^T \cdot \mathbf{w})(\mathbf{x} - \mathbf{y})}$. They show, on a training sample, that the nearest-neighbor approach gives competitive classification accuracy compared to linear, logistic regression, or other methods. According to the empirical study, the best choice of the weight D is somewhere between 1.4 and 1.8, and the results stabilized for k ranging between 100 and 3000. They argue that it is optimal if the training set has an equal mix of goods and bads. Although the nearest-neighbor approach is not frequently used for credit scoring, it has some attractive features. In particular, the system can be dynamically updated by adding new observations and possibly dropping the oldest ones. Individual decisions can be analyzed based on the sets of the k nearest neighbors, especially if k is not too large. On the other hand, the approach can be viewed as even more "black box" than neural networks since it has basically no clear internal structure. In the case of the Henley and Hand (1996) approach, one still has to run at least a regression on the explanatory variables. It might then be suggested that it is better to use a regression scorecard directly without the search for nearest neighbors.

Linear Programming and Support Vector Machines

The goal of a score card $s(\mathbf{x}_i) = \sum_{k=1}^{n} \beta_k x_{ik}$ is to separate good and bad cases. Optimally, we would like to have a cut-off value c so that $s(\mathbf{x}_i) \leq c$ for all bad cases $\mathbf{x}_i \in A_B$ and $s(\mathbf{x}_i) > c$ for all good cases $\mathbf{x}_i \in A_G$ (on the training and/or validation sample). The separation is then used on new observations in order to obtain ex ante predictions. Note that we do not have to consider the intercept β_0 that is replaced by the generally non-zero cut-off value c. If the explanatory factors are all numerical (with categorical factors replaced by the dummy variables) then the problem can be formulated geometrically as a linear separability problem of the two sets of points A_G and A_B in the space R^n. Since, usually, it is not possible to obtain a perfect division of goods and bads, we need to allow for possible errors $s(\mathbf{x}_i) \leq c + \varepsilon_i$ for $\mathbf{x}_i \in A_B$ and $s(\mathbf{x}_i) \geq c - \varepsilon_i$ for $\mathbf{x}_i \in A_G$ where $\varepsilon_i \geq 0$ for all i. This leads to a *linear program* where the objective is to minimize the total sum of errors $\sum_i \varepsilon_i$ that can be solved by classical linear programing methods. Alternatively, we may simplify the program by assuming a constant error term $\varepsilon_i = \varepsilon$ and minimizing just ε over possible values of the coefficient vector $\boldsymbol{\beta}$ and the cut-off c. The advantage of the linear programing is that we can add additional expert constraints, e.g. requiring that the coefficient of one variable is larger than the coefficient of another variable, e.g. $\beta_1 \geq \beta_2$. However, the linear programing formulation above has one pitfall: if one sets all $\beta_i = 0$ and $c = 0$, then the minimization problem is trivially solved with all errors equal to zero, $\varepsilon_i = 0$. A straightforward solution is to require that there is a gap between the two sets, i.e. requiring that $s(\mathbf{x}_i) \geq c + a - \varepsilon_i$ for the good cases and a very small predetermined positive constant a. Then one has to decide how to choose the constant a and also how to deal with points within the gap. An elegant solution has been proposed by Glover (1990) requiring that the distance between the mean of bad scores and the mean of good scores should be equal to 1. The linear program solution cannot then be a trivial one. On the other hand, since the coefficient can be multiplied by any nonzero constant, the condition can always be satisfied, provided that not all the scores equal to zero. Therefore, there is no loss of generality by adding Glover's condition that ensures that the solution is non-zero.

The separation problem can also be formulated as maximization of the distance between two hyperplanes dividing the bad and good points (see Fig. 3.25).

After a normalization of the coefficient vector, the two hyperplanes can be described by the following two vector equations:

$$\mathbf{w} \cdot \mathbf{x} - c = -1 \text{ and } \mathbf{w} \cdot \mathbf{x} - c = 1.$$

Therefore, the *hard-margin* requirement is that

$$\mathbf{w} \cdot \mathbf{x}_i - c \leq -1 \text{ for all } \mathbf{x}_i \in A_B \text{ and } \mathbf{w} \cdot \mathbf{x}_i - c \geq 1 \text{ for all } \mathbf{x}_i \in A_G.$$

Since, geometrically, the distance between the two planes equals $2/\|\mathbf{w}\|$, the maximum distance is achieved by minimizing $\|\mathbf{w}\|^2 = \sum w_k^2$ subject to the constraints

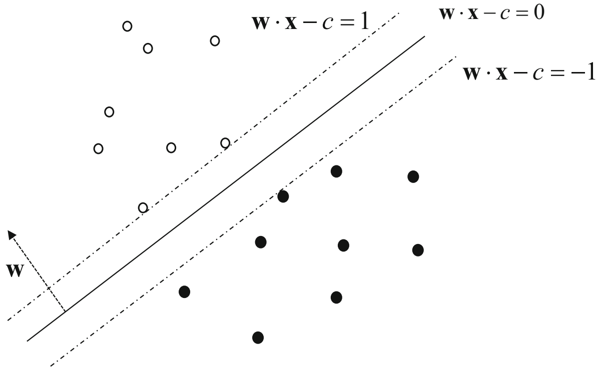


Fig. 3.25 Hyperplanes separating the good and bad points

above. This optimization problem is not a linear program any more, but can be solved by the techniques of quadratic programming. The idea is a basis of the method, in machine learning, called *support vector machines* (Cortes and Vapnik 1995). To explain the terminology at least partially, the support vectors in Fig. 3.25 are the vectors $\mathbf{x}_i$ that are nearest to the mid separation hyperplane $\mathbf{w}\cdot\mathbf{x}-c=0$ and so determine (support) the two max-margin hyperplanes.

As in the case of linear programing, the hard-margin problem is only exceptionally solvable, and one has to introduce soft margins:

$$\mathbf{w}\cdot\mathbf{x}_i-c\leq-1+\varepsilon_i \text{ for all } \mathbf{x}_i\in A_B \text{ and } \mathbf{w}\cdot\mathbf{x}_i-c\geq 1-\varepsilon_i \text{ for all } \mathbf{x}_i\in A_G,$$

with $\varepsilon_i\geq 0$. Equivalently, we can directly set $\varepsilon_i=\max(0,1-(\mathbf{w}\cdot\mathbf{x}_i-c))$ for $\mathbf{x}_i\in A_G$ and $\varepsilon_i=\max(0,1+(\mathbf{w}\cdot\mathbf{x}_i-c))$ for $\mathbf{x}_i\in A_B$. Then we wish to minimize the average error and the squared norm $\|\mathbf{w}\|^2$, which can be achieved, e.g. by minimizing $\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i+\lambda\|\mathbf{w}\|^2$ for a predetermined positive constant λ. Choosing a sufficiently small λ ensures that the soft-margin SVM solution will be the same as the hard-margin SVM solution if the data are linearly separable, but still yields a solution if they are not.

The linear SVM algorithm can be generalized by a general nonlinear transformation of the points $\mathbf{x}_i\in R^n$ to $\varphi(\mathbf{x}_i)$ in a higher-dimensional (or even infinite dimensional) space where the originally linearly non-separable sets might become separable. Then, given a new vector of explanatory factors $\mathbf{x}\in R^n$, the separating planes can be used for a prediction based on $\varphi(\mathbf{x})$. In order to keep the *nonlinear SVM* classification algorithm efficient, the kernel trick is usually applied. The idea is to specify the dot product $\varphi(\mathbf{x})\cdot\varphi(\mathbf{y})$ as a function $k(\mathbf{x},\mathbf{y})$ instead of the function φ directly.

Example Let $\mathbf{x}, \mathbf{y} \in R^2$ and

$$k(\mathbf{x},\mathbf{y}) = (\mathbf{x}\cdot\mathbf{y})^2 = x_1^2y_1^2 + 2x_1y_1x_2y_2 + x_2^2y_2^2$$
$$= \left\langle x_1^2, \sqrt{2}x_1x_2, x_2^2 \right\rangle \cdot \left\langle y_1^2, \sqrt{2}y_1y_2, y_2^2 \right\rangle.$$

Therefore, the implicitly determined transformation from R^2 to R^3 is $\varphi(\mathbf{x}) = \langle x_1^2, \sqrt{2}x_1x_2, x_2^2 \rangle$.

The most popular kernels include: homogenous and non-homogenous polynomial $k(\mathbf{x},\mathbf{y}) = (\mathbf{x}\cdot\mathbf{y})^d$ or $k(\mathbf{x},\mathbf{y}) = (\mathbf{x}\cdot\mathbf{y}+1)^d$, the Gaussian radial basis $k(\mathbf{x},\mathbf{y}) = \exp\left(-\gamma\|\mathbf{x}-\mathbf{y}\|^2\right)$, etc. It follows from the example above that the non-linear transformations implicitly allow us to process various interactions between the explanatory variables.

Another trick that makes the algorithm more efficient is to realize that even in a very high dimensional space it is sufficient to search for the separating hyperplanes in the space spanned by the transformed vectors $\varphi(\mathbf{x}_i)$. Therefore, we look for the perpendicular vector in the form $\mathbf{w} = \sum \alpha_i\varphi(\mathbf{x}_i)$, i.e. minimize $\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i + \lambda\|\mathbf{w}\|^2$ over all $\boldsymbol{\alpha}$ in R^N where N is the number of observations. Since $\mathbf{w}\cdot\varphi(\mathbf{x}_j) = \sum \alpha_i k(\mathbf{x}_i,\mathbf{x}_j)$ the matrix of the products $k(\mathbf{x}_i,\mathbf{x}_j)$ is the only input (plus the good/bad information and the constant λ) of the quadratic optimization problem that can be solved by relatively efficient methods described in the extensive SVM literature. It can be shown that $\alpha_i = 0$ provided $\varphi(\mathbf{x}_i)$ lies on the correct side of the boundary and the vector $\mathbf{w}$ can be written as a linear combination of the support vectors. Therefore, the number of nonzero parameters is not usually as large as N. □

Regarding the application of this method to credit scoring, there are a number of empirical studies comparing logistic regression, discriminant analysis, and other methods with SVMs. Generally, SVMs perform very well on a level comparable with the best models, namely the logistic regression or the neural networks. In spite of good empirical results, SVMs have not been extensively used in banking credit risk management practice, again due to the low interpretability of the estimated parameters and a perceived risk of overfitting.

Ensemble Models and Random Forests

Ensemble models are motivated by the proverb according to which more heads are better than one. That is, given several models, let us collect their answers and use an average or majority vote to produce the final ensemble model answer, assuming that collective wisdom is better than the individual models. To formulate the argument more exactly, let us assume that we have a number of models $M_1,\ldots,M_n$ that produce unbiased estimations $\hat{y}_{M_i} = y + \varepsilon_i$ of the true target variable y conditional on explanatory factors $\mathbf{x}$ with errors ε_i that all have mean zero and standard deviation equal to some σ. Here, we assume, for the sake of simplicity, that y is a continuous variable or the objective probability conditional on $\mathbf{x}$ in the case of a binary outcome. If the model errors were independent, then the average estimation

$$\hat{y}_{Ens} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_{M_i} = y + \overline{\varepsilon}$$

error again has mean equal to zero, but the standard deviation is reduced to $\sigma/\sqrt{n}$ due to the model error "diversification" effect. In practice, of course, the individual model errors do not have to be independent and the standard deviations would not be the same, but as long as the correlations are lower than one, there is a chance that the ensemble model will perform better than the individual models.

Figure 3.26 shows an ensemble model diagram built from the logistic regression, a classification tree, and a neural network on the case study dataset in SAS Enterprise Miner. The model simply averages the posterior probability of the default estimates. The discrimination power of the model (Table 3.21) is indeed slightly higher compared with the individual models (Table 3.20).

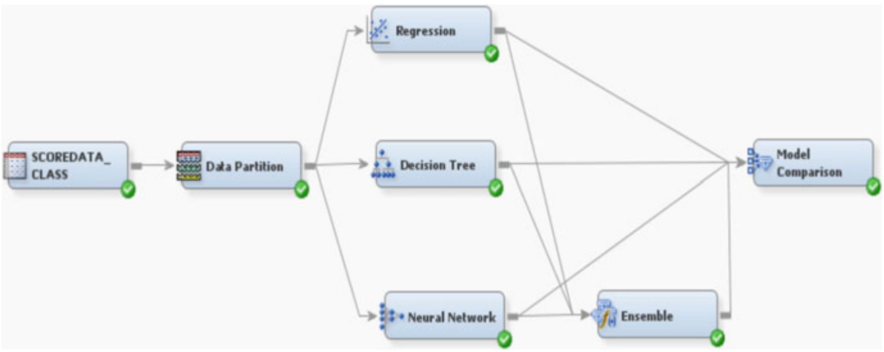


Fig. 3.26 An ensemble model diagram

Table 3.21 Performance of two ensemble models and a random forest

Data set	Training Sample		Testing Sample	
Performance	**KS statistic**	**Gini coeff.**	**KS statistic**	**Gini coeff.**
Ensemble (logistic+tree+NN)	0.56	0.71	0.41	0.534
Ensemble (3 trees)	0.65	0.65	0.40	0.49
Random Forest	0.535	0.68	0.327	0.512

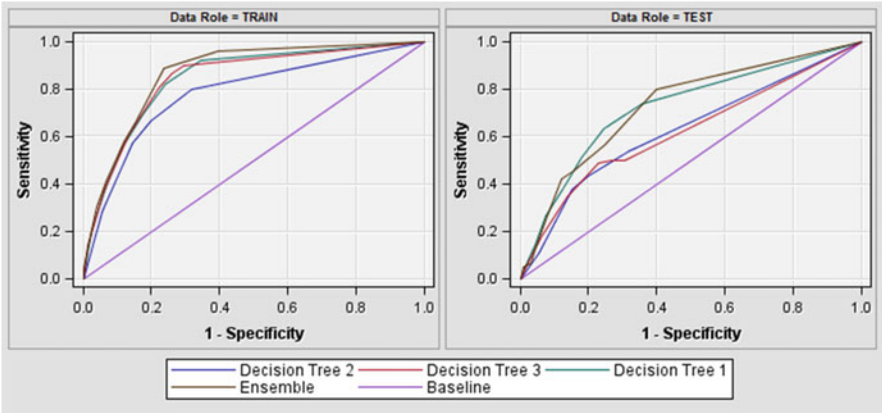


Fig. 3.27 ROC for three decision trees and the corresponding ensemble model

Similarly, we can build an ensemble model from several classification trees. However, we have to use trees that are, to a certain extent, independent. This can be done manually, for example starting with the different variables used for splitting at the root of the tree. Figure 3.27 shows the ROC of three different classification trees and the apparently better ROC of the corresponding ensemble trees. The trees are based on different random training subsamples of the dataset, and the root splitting variable was modified in an interactive mode. While the individual trees' Gini coefficient is around 40%, the ensemble tree's Gini is much better at 49% (Table 3.20).

This result can be improved by "growing" a *Random Forest* with many trees generated randomly in an automated way (Breiman 2001). In order to achieve the relative independence of the trees, the algorithm firstly selects a random subsample (*bagging*), or bootstraps with resampling, and in addition chooses a random subset of the explanatory variables that will be used to build the tree. Since the individual classification trees are weaker (due to selection only of some explanatory variables) it is important to generate a larger number of trees, i.e. a forest. The SAS Enterprise Miner application, in fact, allows one to "grow" a random forest easily. Table 3.20 shows that a random forest with 1000 classification trees has the Gini coefficient comparable to the best models.

To conclude, let us mention the results of some of the many published empirical studies comparing different classification techniques. For example, Baesens et al. (2003) compare various state-of-the-art credit scoring algorithms such as logistic or linear regression, neural networks, or support-vector machines on eight real-life credit scoring datasets. They conclude that least squares, support-vector-machines and neural network outperform the other methods, but the simple classifiers, such as the logistic and linear regression, also perform very well for credit scoring. Other studies like Haltuf (2015) or Kesely (2015) demonstrate, rather, the superiority of the classical logistic regression. Lessmann et al. (2008) compare 22 classifiers on ten public domain datasets: Random forests are the winners in terms of AUC for five datasets, neural networks for two datasets, SVM also for two datasets, and, finally, the classical statistical inference methods like the logistic or linear regression outperform other methods only in the case of one dataset. The dominance of ensemble methods is confirmed in an updated credit scoring study of Lessmann et al. (2015). The set of investigated methods from Baesens et al. (2003) is extended by a large number of homogenous and heterogeneous ensemble methods. Homogenous ensembles (e.g., random forests) use one base model and generate randomly a large number of classifiers. Heterogeneous ensembles create the classifiers using different classification algorithms. Unlike the simple heterogeneous ensemble of the three models shown in Fig. 3.26, these models must involve a selection procedure with a random element in order to generate automatically a large number of classifiers. Among 41 methods benchmarked using different alternative performance measures over eight credit scoring datasets, the heterogeneous ensembles occupy the first 11 places, followed more or less by the homogenous ensembles, and then by the individual classifiers. The random forest ranks 12th, artificial neural networks 14th, and logistic regression 16th.

Ensemble models suffer from the same “black box” criticism as the neural networks and the other alternative models. Nevertheless, the most recent empirical tests show that these models can outperform the logistic regression industry standard in a systematic way and so should be considered at least in areas where the interpretability of the model and its parameters is not of the highest priority (such as fraud or collection scoring, etc.).

Markov Chain Models

An alternative, and in a sense simpler, way to capture default and loss dynamics, is to use the Markov Chain modeling approach. The idea is that there are a finite number of credit behavior states, including the default state, that loans migrate between the states, and we may estimate the migration probabilities and make further predictions. The states could be rating grades, but are more often simply states according to the number of days-past-due (e.g. 0–30, 31–60, 61–90, 90+, i.e. default), or some other elementary characteristics. There is a basic time period, for instance, one month, or one quarter, and the goal is to model the behavior over a longer time horizon including more periods. The Markov assumption is that the process is memoryless, i.e. that transition to a state depends only on the previous state, not on the history of the state transitions before. Specifically, if there are J states $j = 1, \ldots, J$, and the states at times $t = 0, 1, 2, \ldots$ are represented by random variables $X_0, X_1, X_1, \ldots$, then the Markov property says that the probability of transition from state i at time $t - 1$ to state j taken at time t does not depend on any possible path of previous states $k_0, \ldots, k_{t-2}$:

$$\Pr\left[X_t = j \middle| X_0 = k_0, \ldots, X_{t-2} = k_{t-2}, X_{t-1} = i\right] = \Pr\left[X_t = j \middle| X_{t-1} = i\right].$$

The probability $p_t(i,j) = \Pr\left[X_t = j \middle| X_{t-1} = i\right]$ is called the transition probability, and we can define the transition matrix $P_t = (p_t(i,j))_{i,j=1}^{J}$. It is obvious that all elements of the matrix are non-negative and the sum of each row is one. The key observation is that the multiplication of matrices corresponds to migration over multiple periods. For example:

$$\Pr\left[X_2 = j \middle| X_0 = i\right] = \sum_{k=1}^{J} p_1(i,k) \cdot p_2(k,j) = (P_1 P_2)(i,j)$$

where $P_1 P_2$ denotes ordinary matrix multiplication. Similarly, the T-stage transition probabilities are given by the matrix $P_1 P_2 \ldots P_T$. Therefore, if J is the default state, then $P_1 P_2 \ldots P_T(i,J)$ is the probability of default of a borrower starting from state i at time 0 in the T-stage time horizon. So, if we estimate, based on historical data, the transition matrices P_t, then we have a dynamic model of default probabilities in different time horizons. The transition matrices may indeed depend on "aging"; i.e., time on books, as can be seen from Fig. 3.27. The transition probabilities in P_t must be then calculated, starting with loans which were granted t periods ago. Nevertheless, notice that we do not need a long period to estimate the matrices P_t for different values of t (as t measures time from the origination of an exposure, not the absolute time). It is sufficient to have just a one-period observation window where we need to count the number of cases $n_{t-1}(i)$ with age $t-1$ and in the state i at the beginning of the period, and then the number of cases $n_t(i,j)$ that were in state i at time $t-1$ and subsequently moved into state j at time t. Then the estimated transition probability is

$$\hat{p}_t(i,j) = n_t(i,j)/n_{t-1}(i). \tag{3.20}$$

The estimation can be simplified if we assume that the Markov chain is homogenous, that is: $P_t = P$ does not depend on t. This would be the case, rather, of classical corporate exposures. Then the power matrix P^T fully characterizes the distribution of the variable X_T. The default state J is often defined as an absorbing one; loans that enter the state never leave it; i.e., $p(J,J) = 1$ and $p(J,j) = 0$ for $j \neq J$. Other examples of absorbing states are repayment or "not rated".

Table 3.22 shows, as an example, the S&P rating transition probability matrix that is usually published in rating agencies' statistical reports. The matrix has, in fact, two absorbing (or persistent) states; "D" denoting default and "N.R." meaning not rated. So, to get a 9×9 matrix, the table could be amended with two rows for the states "D" and "N.R." having 1.00 on the diagonal, and 0.00 elsewhere. The powers of the matrix unfortunately do not give the full default rate estimation, as part of the ratings will end with a positive probability in the state "N.R.", where we do not know exactly the proportion of defaulters. The problem may solved by an adjustment of the matrix, assuming that there is an average distribution of rating grades in the "N.R." class.

Table 3.22 S&P global average 1-year transition rates 1981–2004 (%)

From/To	AAA	AA	A	BBB	BB	B	CCC/C	D	N.R.
AAA	87.44	7.37	0.46	0.09	0.06	0.00	0.00	0.00	4.59
AA	0.60	8.65	7.78	0.58	0.06	0.11	0.02	0.01	4.21
A	0.05	2.05	8.96	5.50	0.43	0.16	0.03	0.04	4.79
BBB	0.02	0.21	3.85	84.13	4.39	0.77	0.19	0.29	6.14
BB	0.04	0.08	0.33	5.27	75.73	7.36	0.94	1.20	9.06
B	0.00	0.07	0.20	0.28	5.21	72.95	4.23	5.71	11.36
CCC/C	0.08	0.00	0.31	0.39	1.31	9.74	46.83	28.83	12.52

An Application of MC Models: Estimating Provisions

A simple matrix algebra can be used to estimate loan loss provisions (expected losses) of impaired (defaulted) exposures over an indefinite horizon conditional on a current credit behavior status (see also Prášková and Lachout 2012). Let us assume that the states $j = 1, \ldots, J$ corresponding to DPD bands, or other characteristics of principal exposures where repayment improvement or worsening is still possible, are transitional, and that there are three absorbing states; "repaid", "cured", and "written-off", which have the special codes R, C, and L. In order to estimate provisions on a defaulted exposure in state k we need to estimate the proportion of the principal exposure that is going to be written-off in an infinite horizon, i.e.

$$p_\infty(k, L) = \Pr\left[X_\tau = L \middle| X_t = k\right] \qquad (3.21)$$

where τ denotes the (random) time when an absorbing state (repaid, cured, or written-off) is reached. Let $P = (p(i,j))_{i,j=1}^{J}$ denote the time-homogenous transition matrix between the transitional states and U the $3 \times n$ matrix rows transition probabilities $p_\infty(k, L), p_\infty(k, C), p_\infty(k, R)$, where $p_\infty(k, C)$ and $p_\infty(k, R)$ are defined analogously to (3.21). Note that it is implicitly assumed in (3.21) that the probabilities of migration from a state at time t into an absorbing state in the infinite horizon do not depend on t. Over a single period, an exposure in the state k might migrate directly to an absorbing state, or to another transitional state j from which it can again migrate to an absorbing state in the infinite horizon. Therefore, we obtain the following recurrent equation:

$$p_\infty(k, L) = p(k, L) + \sum_{j=1}^{J} p(k,j) p_\infty(j, L),$$

and similarly for $p_\infty(k, C)$ and $p_\infty(k, R)$. The three recurrent equations can be expressed in the matrix form as:

$$U = Q + PU \tag{3.22}$$

where Q is the one period transition probability matrix from the transitive to the absorbing states, i.e. with rows $p(k,L)$, $p(k,C)$, $p(k,R)$. The matrices P and Q may be estimated based on one period observations and the unknown matrix U can be then easily obtained solving (3.22) by the matrix algebra:

$$U = (I - P)^{-1}Q.$$

Markov Property Testing

The literature on applications of Markov chain matrices is quite extensive (see, e.g. Thomas 2009; Trueck and Rachev 2009). However, there are not too many papers dealing with the problem of how to test the Markov property. It is usually implicitly assumed that the Markov property holds, but given a dataset and a definition of the states, it is not so easy to test it statistically. For example, Ait-Sahalia (1996) proposes a test statistic based on the Chapman-Kolmogorov equation, which, for a discrete Markov chain, simply says that by multiplying two consecutive transition matrices we get the corresponding 2-stage transition probabilities:

$$\Pr\left[X_{t+1} = j \middle| X_{t-1} = i\right] = \sum_{k=1}^{J} p_t(i,k) \cdot p_{t+1}(k,j) = (P_t P_{t+1})(i,j).$$

According to Chen and Hong (2012), the test is not complete in the sense that it does not reject some non-Markovian processes; they propose a more complex test examining a growing number of lag dependencies based on the concept of the conditional characteristic function.

A simple test of the Markov property given a finite credit dataset and a definition of states is described in Thomas et al. (2002). Let $n_{t-1}(i)$, $n_t(i,j)$ be defined as in (3.20) and let $n_t(i,j,k)$ denote analogously the number of exposures that were in state i at time $t-2$, then in state j at time $t-1$, and finally in state k at time t. The time-homogenous probability of moving from i to j can be estimated by:

$$\hat{p}\,(i,j) = \sum_{t=1}^{T} n_t(i,j) / \sum_{t=1}^{T} n_{t-1}(i),$$

and the probability of moving from to k conditional on previous states i and j:

$$\hat{p}\,(i,j,k) = \sum_{t=2}^{T} n_t(i,j,k) / \sum_{t=2}^{T} n_{t-1}(i,j).$$

These are estimators of probabilities $p(i,j,k)$ and $p(i,j)$. The Markov property implies that $p(i,j,k) = p(j,k)$ for all i and so a χ^2 test can be applied. According to Thomas et al. (2002)

$$S = \sum_{i,k=1}^{J} \frac{n(i,j)(\hat{p}\,(i,j,k) - \hat{p}\,(j,k))^2}{\hat{p}\,(j,k)}$$

has a χ^2 distribution with $(J-1)^2$ degrees of freedom.

Continuous Time Markov Processes

Observing credit state migrations that can take place essentially any time, it is often more appropriate to consider a continuous time process. The Markov chain variable $X_t \in \{1,..,J\}$ is now indexed by continuous time $t \geq 0$. It is, in fact, a discrete state stochastic process whose evolution is described by transition matrices $P(s,t)$ whose elements are transition probabilities $p_{ij}(s,t)$ between states i and j from time s to t. The Markov property implies the familiar matrix identity:

$$P(s,u) = P(s,t)P(t,u) \quad \text{for } s < t < u. \tag{3.23}$$

The continuous time discrete-state processes can be studied in an analogy to survival analysis where we deal only with two states. Given two states $i \neq j$, let us define the transition intensity (hazard rate)

$$\lambda_{ij}(t) = \lim_{\Delta t \to 0} \frac{p_{ij}(t, t+\Delta t)}{\Delta t},$$

where we implicitly assume that the limit always exists. Note that the limit would not exist for $i = j$ since in this case $p_{ij}(t, t+\Delta t)$ approaches one. Therefore, if $\Lambda(t)$ is the matrix with zeros on the diagonal and $\lambda_{ij}(t)$ elsewhere, then we have formally $P(t, t+dt) = I + \Lambda(t)dt$ and so according to (3.23)

$$P(0,T) = \prod_{0}^{T} (I + \Lambda(t)dt).$$

The product should be interpreted as a limit of finite products when $dt = T/K$ with K being large (going to infinity). The exponential survival analysis model with a constant hazard rate corresponds to the assumption of a constant transition intensity matrix $\Lambda = \Lambda(t)$. Then, indeed,

$$P(0, T) = \exp(I + \Lambda T) = I + \Lambda T + \frac{1}{2}\Lambda^2 T^2 + \cdots.$$

The constant intensities can be estimated from historical migration data in a straightforward way. Let us consider two states $i \neq j$ and an exposure that has entered the state i, then after a certain time T_l it may migrate to state j, or it may migrate into another state, or it may stay in state i until the end of the observation window. The observation (of a migration form i to j) is censored in the last two cases and uncensored in the first case. Let $T^* = \sum T_l$ and N be the number of uncensored observations, i.e. of exits.[7] Then we have the following estimate:

$$\hat{\lambda}_{ij} = N/T^*. \tag{3.24}$$

The times T_l, measured in years, could be based on daily data utilizing maximally the available historical information including censored observations.

Remark It is easy to see that (3.24) is a consistent maximum likelihood estimate for the exponential survival model. Since $S(t) = \exp(-\lambda_{ij} t)$ is the survival function, that is the probability of continuously staying in state i without transition to j until time t, and according to (3.15), the log-likelihood function we need to maximize is

$$\ln L(\lambda_{ij}) = -\lambda_{ij} T^* + N \ln \lambda_{ij},$$

where N and T^* are defined as above. Its maximum can be found by setting the first derivative equal to zero, i.e. indeed

$$\frac{\partial \ln L(\lambda_{ij})}{\partial \lambda_{ij}} = -T^* + \frac{N}{\lambda_{ij}} = 0,$$
$$\lambda_{ij} = N/T^*.$$

3.5 Expected Loss, LGD, and EAD Estimations

So far, we have focused just on ratings, and the probabilities of defaults, which play the role of a decision support tool in the process of the acceptance or rejection of loan applications. But the bottom line for the banking business is the final profit/loss of a loan product portfolio, which depends on the appropriate loan pricing, rates of default, and realized losses on the defaulted loans. The interest rate of a loan should reflect the *internal cost of funds*, the *risk premium*, and the *administrative cost*. The

[7] Note that an exposure may be in state i more than once during our observation window.

internal cost of funds may be given by a fixed or floating rate, corresponding to the interest the bank is paying on customer deposits and on the interbank market. It is usually defined as the marginal financing rate close to the offered rates quoted on the interbank market. The internal cost of funds is the minimum that must be covered when granting a loan, otherwise the loan granting business lacks any rationale. Secondly, the interest income on loans, specifically the business margin, must cover the administrative costs related to the loans, and last, but not least, the risk premium must pay for the expected credit losses that do happen on a larger portfolio. After all those costs, the bank still has to account some positive net profit which should bear a certain relation to the bank's capital. It is often emphasized that the business margin, on a stand-alone basis, does not often guarantee sufficient profitability. Clients with loans are expected to have a side-business with the bank which generates additional income, related, for instance, to current account and maintenance fees, asset management, insurance, etc. The cross-selling effect is often taken into account when the business margin is calculated.

Let us now focus on the key element of the credit margin calculation called the *Expected Loss* (*EL*). Let us fix a time horizon T, and, for a loan i, define the absolute loss $L_i^{abs}(T)$ as the amount of provisions charged at the end of the period in the case of default of the loan, and zero otherwise. According to IAS rules, provisions can be created only in the case of credit impairment, which usually coincides with the definition of default. In some cases banks may create provisions earlier, but those should not be considered as incurred losses, rather, as reserves. If a loan is written-off during the period, then $L_i^{abs}(T)$ will be simply the written-off amount. In the case of default, the provisions should be estimated as an economic loss on the exposure at the time of the default, reflecting the expected recovery cash flows, or bad debt market value. Having defined the random variable $L_i^{abs}(T)$, we can define the theoretical *absolute expected loss* $EL_i^{abs} = E\left[L_i^{abs}(T)\right]$. The *relative expected loss* would be then $EL_i = EL_i^{abs}/EAD_i(T)$, where $EAD_i(T)$ is the expected total exposure of the loan at the time of default (usually approximated by the initial exposure balance). Let us for the sake of simplicity assume that a given portfolio of N loans is homogenous in terms of cumulative probability of default PD, in the time horizon T, in terms of the relative expected loss EL, and size A. Now let us calculate the relative risk premium RP that is collected only on loans that do not default during the period, solving the simple equation requiring that expected losses are covered exactly by the collected premiums,

$$N \times (1 - PD) \times A \times RP = N \times A \times EL, \quad \text{i.e.}$$

$$RP = \frac{EL}{1 - PD}. \tag{3.25}$$

The EL (expected loss) and RP (risk premium) rates need to be annualized if calculated over a horizon different from 1 year. We could possibly improve the model, taking into account the randomness of the time of default, exposure, recovery rates, partial installment payments, and, in particular, the time value of money. Nevertheless, the Eq. (3.25) is, in most situations, sufficient to estimate roughly the risk margin that must be added to the internal cost of funds and administration costs as a minimum, in order not to run a money losing business. The risk premium may serve just to set appropriate pricing, or as an internal cost charged to business units as an insurance premium. The central budget in the latter case must cover the credit losses from the created “credit insurance reserves”. In this system, business units can focus on the distribution of loans without fear of unexpected credit losses. The system still needs to have a participation element motivating business places to maintain a satisfactory loan portfolio quality.

To estimate the expected loss key parameter, it is useful to break down the expected loss into the probability of default, and the absolute loss at default:

$$EL^{abs} = E[L] = E\left[L^{abs} \middle| \text{Default}\right] \times \Pr[\text{Default}] = E\left[L \times Exp \middle| \text{Default}\right] \times PD.$$

The variables Exp, and the percentage loss $L = L^{abs}/Exp$, may be considered in most situations to be independent, and so we can further break down the expected conditional loss as

$$E\left[L \times Exp \middle| \text{Default}\right] \cong LGD \times EAD,$$

where:

$$LGD = E\left[L^{abs}/Exp \middle| \text{Default}\right] \text{ and } EAD = E\left[Exp \middle| \text{Default}\right].$$

The estimation of Exposure at Default is non-trivial for revolving exposures like credit cards, overdrafts, lines of credit, etc. For other products, it is simply set equal to the outstanding exposure. Hence, for the absolute loss we have the following elementary decomposition

$$EL^{abs} = PD \times LGD \times EAD. \tag{3.26}$$

In the case when EAD equals the actual exposure, the percentage loss rate can be simply expressed as $EL = PD \times LGD$.

According to (3.25), the risk premium can be calculated per each rating class, or for a whole product pool. Retail products like credit cards, or consumer loans, often have a flat interest rate based more on marketing criteria. In general, the credit risk and marketing management’s task is to find an optimum scoring cut-off level s_0, for the product application approval (only clients with a score of at least s_0 are approved) maximizing the overall profit. Given s_0, we may estimate the average $PD(s_0)$, the expected loss $EL(s_0)$, and the risk premium $RP(s_0)$ given by (3.25). In fact, since there is a one-to-one correspondence between the average PD, and the cut-off score, the optimization can also be based on a target average default rate PD_0. The estimation must use an assumption based on the distribution of the

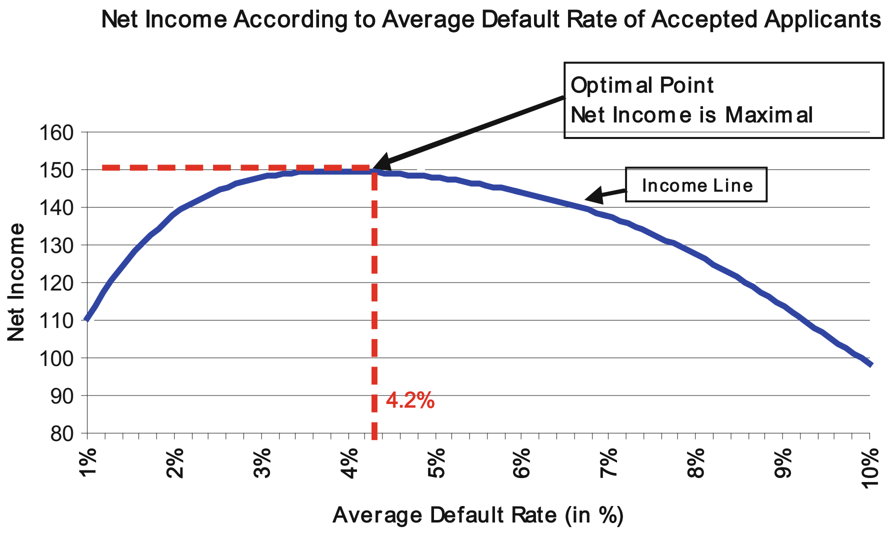


Fig. 3.28 Cut-off optimization example

approved loans across the score values, and on an *LGD* estimation. If the loan fixed interest rate is R, the cost of the funds is R_0, and the administrative costs as a percentage rate is C, then the percentage average net income is $NI^{rel}(s_0) = R - R_0 - RP(s_0) - C$. With a higher cut-off score, the quality of the production improves and the percentage net income goes up. On the other hand, a higher cut-off implies a lower volume of loans granted; i.e., the rejection rate goes up. If V is the volume of all loan applications estimated by the marketing department, then the potentially approved volume $V(s_0) = V \times \Pr[s \geq s_0]$, will decrease with a higher cut-off score s_0. The absolute net income is then simply $NI = NI^{rel}(s_0) \times V(s_0)$. An example of the relationship between the cut-off (or average *PD*), and the overall net income is shown in Fig. 3.28. The conservative cut-off corresponding to $PD_0 = 1\%$, or the too optimistic one corresponding to $PD_0 = 8\%$, clearly do not lead to the best results. In fact, in this simple model, the optimum, around $PD_0 = 4.3\%$, corresponds to the point where the marginal net income; i.e., net income on loans with a score exactly at s_0, equals zero (see also Sect. 3.1). The optimization should also take into account the various levels of the product price, and the cost of risk capital, which will be discussed in Chap. 4.

LGD and Recovery Rate Estimation

In order to estimate the LGD, we first need to specify the notions of realized (ex post), and expected (ex ante), Recovery Rate (RR), and the complementary Loss Given Default (LGD). Realized RR can be observed only on defaulted receivables, while the expected recovery rate is estimated for non-defaulted receivables based on currently available information. The RR and LGD are expressed as percentages of the exposure outstanding at default (EAD), and $LGD = 1 - RR$ is simply the complementary loss rate based on the recovery rate that is usually less than 1. For market instruments like bonds, or other debt securities, we may define the realized market RR as the market value of the principal (plus coupon accrued at default) of the security shortly (typically 1 month) after the default. Applicability of the definition assumes the existence of an efficient, and sufficiently liquid, market for defaulted debt. For other receivables we have to observe the net recovery cash flows CF_t from the receivable generated by a work-out process. The work-out process may be internal or external, where a collection company is paid a fee for collecting the payment on behalf of the receivable owner. The process may also combine the ordinary collection and partially the sale of the receivables to third parties. In any case, the work-out process involves significant costs that must be deducted from the gross recoveries. The net cash flows must be finally discounted, with a discount rate r appropriately reflecting the risk (BCBS 2005a, b, c).

$$RR = \frac{1}{EAD} \sum_{i=1}^{n} \frac{CF_{t_i}}{(1+r)^{t_i}}. \quad (3.27)$$

The work-out recovery rate should, in a sense, mimic the market recovery rates. The relationship between the two ex-ante notions is an analogy between the fundamental value and the market value of a stock. Thus, the discount rate can be based on a measure of the RR systematic risk, and the general price of the risk (see Witzany 2009a). Since the market recovery rate is never negative, and can be hardly larger than 1, normally we assume that RR, as well as $LGD = 1 - RR$, lie in the interval [0, 1]. The calculation of the work-out recovery rates according to (3.27) may, however, in some cases lead to negative values due to high costs and low, or no, recoveries, and, on the other hand, it may lead sometimes to values larger than 1 in the case of large and successfully collected late fees.

Having collected and calculated the realized recovery rates, the next task is to estimate the LGD for non-defaulted accounts. In the case of new loan applications, banks need to estimate not only the probability of default (PD) in the given horizon, but also the LGD in the same horizon.

There are, in principle, two approaches that can be combined: pool level and account level regression based LGD estimations. The basic pool level approach is based on elementary criteria like the collateralization level (e.g., according to LTV—*loan to value ratio*, in the case of mortgages), debtor type (e.g., industry in the case of corporate exposures), loan seniority etc. in order to split historical defaulted exposures and realized LGD observations, as well as non-defaulted exposures, into LGD homogenous pools. Average historical LGDs, and the corresponding confidence intervals, are then computed for every pool. The pools that do not discriminate between the LGD estimates should be merged, and the pool estimates are used as ex ante predictions.

In the regression based approach, we try to estimate a regression model

$$LGD_i = f(\boldsymbol{\beta}'\mathbf{x}_i) + \varepsilon_i, \tag{3.28}$$

where $\mathbf{x}_i$ are LGD explanatory variables (e.g. collateral value, financial ratios, etc.), $\boldsymbol{\beta}$ is the vector of the regression coefficients, and f is an appropriate link function. We need to apply a link function, since the values of an ordinary linear regression term $\boldsymbol{\beta}'\mathbf{x}_i$ lie in $(-\infty, \infty)$, while LGD values are more or less in the interval [0, 1]. The link function could simply be the Logit function, or rather, a function transforming the normal distribution into an appropriate LGD distribution, e.g., beta, mixed beta, or an empirical distribution. Figure 3.29 shows the recovery rate distributions for corporate loans according to seniority, as reported by Moody's. At least one of the classes (Sr. Unsecured), shows a bimodal shape (either rather low, or rather high, recovery rates). This is even more pronounced for consumer loans, where a logistic regression discriminating between low and high recoveries could be applicable.

The performance of the model is, in general, measured using the in-sample or out-sample classical statistic R^2 (see, e.g. Greene 2003). Alternatively, Spearman's rank correlation or Somers'D (Gini coefficient) can be used. The predicted values $LGD(a) = f(\boldsymbol{\beta}'\mathbf{x}(a))$, can be used as account level LGD estimations, or to define

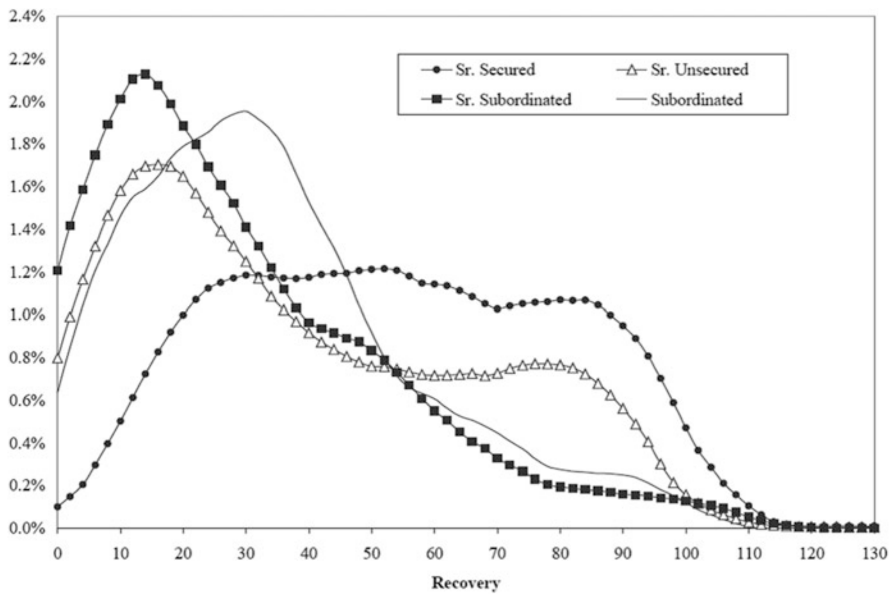


Fig. 3.29 Probability densities of recovery by seniority (Source: Schuermann 2002)

LGD rating and corresponding LGD pools. LGD predictions can be recalibrated on the rating pools analogously to PD calibration. The advantage of the approach is that the regression univariate, and multivariate, analysis helps to analyze the most important explanatory factors, and to define the optimal LGD pools.

There are many issues encountered by banks estimating LGD. Firstly, there are often not enough observations. Some products, such as mortgages, or large corporate debtors, may have rare defaults and, moreover, banks in the past did not pay too much attention to the systematic collection of recovery data. Even if a bank started to collect data a few years ago, the problem is that the standard internal recovery process takes a long time; up to 3 or even more years; and so we do not know the ultimate recovery rates for the most recent defaults. This issue may be overcome by an extension method, which extrapolates partial recovery rates using existing completed recovery observations. Alternatively, there is the possibility of applying a modified Survival Analysis method (see Witzany et al. 2010).

Secondly, a number of studies (see Fig. 3.30) show the cycle dependent variability of recovery rates and LGDs (Schuermann 2002; Altman et al. 2002), and the bank must decide whether the estimations are to be Point-in-Time (PIT), or Through-the-Cycle (TTC). In the case of pool level estimates, this can be achieved by analyzing the monthly, or quarterly LGD time series; i.e., time series of average LGDs on receivables which have defaulted in a given period and that belong to the pool. If the time series is long enough, then the appropriate PIT LGD estimations (prolonging a series based on an econometric model), or TTC LGD estimations,

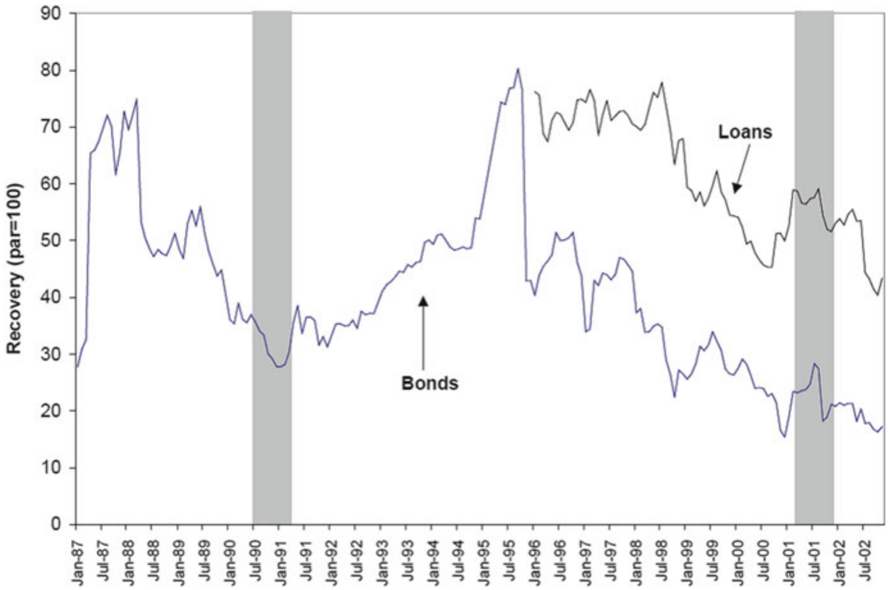


Fig. 3.30 Defaulted bond and bank loan recovery index, U.S. debtors. Shaded regions indicate recession periods (Source: Schuermann 2002)

usually as weighted long run averages, can be produced. In the case of account level predictions, one possible method is to include relevant macroeconomic indicators in the vector of explanatory variables **x**. Then $LGD(a) = f(\boldsymbol{\beta}'\mathbf{x}(a))$ with systematic variables set to the values forecast for the coming period, produces PIT estimations. On the other hand, if the macroeconomic explanatory variables are set equal to the long term average values and other variables are TTC, then the estimations reflect the TTC philosophy.

Another issue is that the work-out methods often change, some receivables being sold, others collected internally, whilst others are collected by external companies with an incentive scheme, etc. The pooling criteria and account level regression then must respect the main identified processes.

It should be pointed out that the LGD parameter is as important as the PD. The methodology and experience with its estimation is, nevertheless, much more limited. There are many new academic papers on the subject, and we shall probably see new interesting and sophisticated approaches to LGD modeling in the near future.

Provisions and Write-Offs

LGD estimations should go hand in hand with the provisioning process. Provisions decrease the value of receivables with respect to problematic credit borrowers. Since the amount impacts directly on the bank's P/L statement, provisioning is a very sensitive process that should not be under the control of the business part of the bank, nor even of the credit analysts who had approved the original loan. Provisioning is one of the most important items scrutinized by external auditors during an annual review.

The process is (in the case of international banks) regulated by the International Accounting Standards (IAS 39 issued by IASB—International Accounting Standards Board and replaced by International Financial Reporting Standards—IFRS 9 published in 2014 and becoming effective from 2018). According to the IAS 39 standards, the provisions reducing the asset value are charged if there is objective evidence of "impairment". The IAS notion of "impairment" might be interpreted a little bit differently compared to the Basel II definition of default. Usually, impaired receivables are always considered as defaulted. However, there might be defaulted receivables where no material loss is expected due to high quality collateral. Since "impairment" is related to a material loss, these receivables are not necessarily considered as impaired. A typical case of "impairment" evidence is a legal event (bankruptcy, restructuring), days past due, etc. The provisions should be equal to the decrease of the asset value (cash flows discounted by the *effective interest rate*, EIR) due to the impairment. Interest accrued after impairment is calculated using the "amortized cost" principle, i.e. as the EIR applied to the net present value. For large impaired exposures, the determination of provisions depends on expert analysis. For smaller exposures, various statistical methods could be applied. In fact, the IAS allows the creation of provisions statistically, on a portfolio basis, if there is an overall deterioration of certain key indicators (e.g., overall payment discipline of credit card borrowers, or property prices in the case of mortgages, regional unemployment in the case of consumer loans, etc.). There is a popular flow rate model based on the Markov Chain which classifies the receivables into various buckets, according to past due days, and calculates the transition probabilities, including the two persistent states of repayment and write-off.

IFRS 9 in addition requires the creation of an impairment allowance based on the expected present value of credit losses from defaults over the next 12 months, unless there is a significant increase of the credit risk of the borrower. The allowance is created from the very origination of a receivable. If the credit risk increases, then the allowance must cover the projected lifetime losses, but interest is still calculated based on the *gross carrying amount*. In the case of default, the concept does not change—provisions must cover the lifetime credit expected loss, in this case conditioned by an incurred (not expected) default, but the interest is based on the "amortized cost" principle. The new accounting standards aim to be forward looking, requiring banks to recognize losses on time, in particular, following the experiences from the recent financial crises. The credit allowances should create an additional buffer against traditional impairment losses and presumably make the banking sector safer. However, the complexity of the expected credit loss estimations, in particular over the lifetime horizon, might bring undesirable flexibility of P/L management to the banks, or on the other hand, a new pro-cyclical source of P/L volatility related to swings in general credit expectations.

EAD and Conversion Factor Estimation

Exposure at default is the third key parameter in the expected loss decomposition (3.26). Its precise estimation is more important for products such as revolving loans and lines of credit, where the drawn exposure at the time of default may differ significantly from the actual on-balance sheet exposure. There is quite limited literature on the subject (Araten and Jacobs 2001; Moral 2006; Jacobs 2008), so we will follow mainly Witzany (2009c).

A popular approach also incorporated into CAD (2006) and CRR (2013) is based on the notion of the *Conversion Factor* (*CF*); estimating the utilization of the undrawn amount upon default. If we know the conversion factor, we may calculate:

$$EAD = \text{Current Exposure} + CF \times \text{Undrawn Limit}.$$

Another approach mentioned in CEBS (2006), is to express the conversion factors from the total credit limits, not only from the undrawn limit. We will call this coefficient the *Credit Conversion Factor* (*CCF*). This method, with the $EAD = CCF \times \text{Limit}$, is also called *the momentum approach*.

Firstly, let us define more precisely the key notions. Ex-post EAD, on a defaulted facility, is defined simply as the gross exposure $Ex(t_d)$ at the time of default t_d, where $Ex(a,t) = Ex(t)$ denotes the on-balance sheet exposure of the facility a at time t. We omit the argument a whenever it is clear from the context.

It is not so straightforward to define the ex-post conversion factor on a defaulted facility, since it requires a retrospective observation point called the *reference date* t_r, where we observe the undrawn amount $L(t_r) - Ex(t_r)$ with $L(t)$ denoting the total credit limit at time t. Since the conversion factor measures the utilization from the undrawn amount, we need $L(t_r) - Ex(t_r) > 0$. Then it makes sense to define the ex-post CF as:

$$CF = CF(a, t_r) = \frac{Ex(t_d) - Ex(t_r)}{L(t_r) - Ex(t_r)}. \tag{3.29}$$

Note that an observed (ex-post) conversion factor may, in practice, be negative if the drawn exposure between the reference date and the default date declines, but it is also larger than 1 if the exposure at default exceeds the limit effective at the reference date. This may happen if there is an increase in the limit, or a breach of the limit, for example caused by interest and late fees. We will admit such observed values, but the estimated ex-ante conversion factor still has to be non-negative (a regulatory requirement), and will usually be expected to be lower than, or equal to, 1 (estimated CF larger than 1 being, exceptionally, acceptable). Notice that the expression (3.29) is very sensitive to the drawn amount if the undrawn amount is small.

Regarding the ex-ante EAD and CF, we will start with a full probabilistic definition and an analysis of the concept. Let T denote the time of default of the non-defaulted facility a from the perspective of time t. Since we do not know the time of default, T is a random variable, and $T < \infty$, as we may assume that any debtor eventually defaults in the infinite time horizon. Let us assume that EAD is defined in the 1-year horizon, so, the theoretical definition is

$$EAD = EAD(a, t) = E\left[Ex(T) \middle| t < T \leq t + 1\right]. \tag{3.30}$$

In order to break down the unknown time of default, and EAD conditional on the time of default, we need to introduce the time to default density function $f_a(s)$; i.e. $f_a(s)\Delta s$ being the unconditional probability that default happens over the time interval $[s, s + \Delta s)$. The time to default density function generally depends on the properties of the facility a including the time from origination. Consequently, EAD can be expressed as the $f_a(s)ds$ weighted average of expected exposure upon default at $T = s$:

$$EAD = EAD(a,t) = \frac{\int_t^{t+1} E[Ex(T)|T=s]f_a(s)ds}{P[t < T \leq t+1]}. \quad (3.31)$$

Thus, according to the analysis, ex-ante *EAD* also depends on the probability distribution (density function) of the time to default. In particular, for short term retail loans we have seen that the time to default density function may be large shortly after drawing, and later it may significantly decline.

The distribution of the time to default depends on a particular product, as well as on the time from the facility origination. There is a significant dependence of *EAD* on the time to default, as confirmed by the study by Araten and Jacobs (2001). Consequently, we will use the definition (3.31), which can be also called the *PD-weighted approach.*

In practice, we need to approximate the integral (3.31) numerically. It can be approximated by a discrete summation: Let us split the 1-year time interval into a sequence of subintervals $(t_0, t_1], \ldots, (t_{n-1}, t_n]$ where $0 = t_0 < t_1 < \cdots < t_n = 1$. Next we estimate EAD_i conditional on time of default T being in the interval $(t_{i-1}, t_i]$, and the probability $\hat{p}_i$, that default happens during this interval for $i = 1, \ldots, n$. Consequently $\hat{p} = \sum_{i=1}^{n} \hat{p}_i$ estimates the probability of default within 1 year. The probabilities can be taken on a portfolio basis or conditional on the facility properties. Then, in line with (3.31) we get the simple approximation:

$$EAD = \frac{1}{\hat{p}} \sum_{i=1}^{n} \hat{p}_i EAD_i \quad (3.32)$$

An empirical study in Witzany (2009c) shows that the estimation may give significantly different results, compared to EAD estimated in a fixed time (usually 1-year) horizon.

The Reference data set (RDS) is a set of ex-post observations used for ex-ante *EAD* estimations. Our notation follows Moral (2006). An observation $o = \left(a, t_r, t_d, \overrightarrow{RD}\right)$ consists of a defaulted facility identification, the reference date, the date of default, and a vector of risk drivers containing at least the information on exposures and limits at the reference and default dates ($Ex(t_r)$, $L(t_r), Ex(t_d), L(t_d)$).

As explained in the definition of ex-post *EAD*, and *CF*, a single observation is not determined only by the facility that defaulted at time t_d, but also by the reference date t_r, at which we measure the retrospective drawn, and undrawn, amount. We do not exclude the possibility of more than one reference date for a given single defaulted facility in order to capture the dependence of *EAD* and *CF* on the time to default. The most common choice (and the most conservative, in line with the

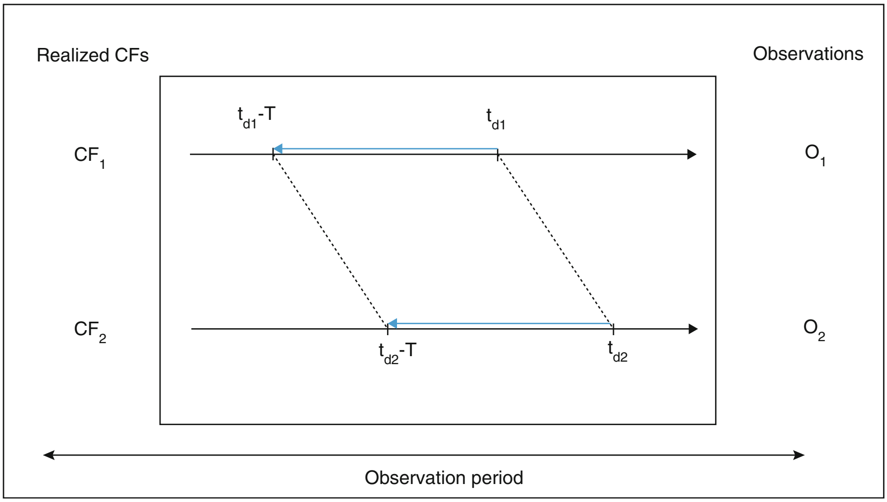


Fig. 3.31 Fixed time horizon approach

analysis above), is the 1-year horizon corresponding to the unexpected credit loss estimation horizon. Generally, there are different alternatives (Moral 2006): *Fixed Time Horizon*, *Cohort Approach*, or *Variable Time Approach*.

The Fixed Time Horizon Approach sets $t_r = t_d - T$ where T is a fixed time horizon (see Fig. 3.31). RDS defined in this way, in fact, leads to an estimation of *EAD* and *CF*, conditional on the time to default being equal exactly to T. Therefore, a number of RDS with different fixed time horizons, and based on the same set of defaulted facilities, may be constructed in the *PD*-weighted approach. Banks often use $T = 1$ year as a standard choice.

The Cohort Method divides the observation period into intervals $(T_0, T_1], \ldots$ $(T_{n-1}, T_n]$ of a fixed length, usually 1 year (see Fig. 3.32). Defaulted facilities are grouped into cohorts according to the default date. The reference date of an observation is defined as the starting point of the corresponding time interval; i.e., if $t_d \in (T_i, T_{i+1}]$ then we set $t_r = T_i$. In this case, the time to default probability distribution is implicitly captured in the data. However the beginnings of intervals may cause an initial seasonal bias (for example T_i some time before Christmas will probably show higher drawing on credit cards, or overdrafts, than during some other months). So, it is advisable to set T_i at "normal" periods of the year with average drawings.

The Variable Time Horizon Approach uses a range of fixed horizon values $T_1, \ldots, T_k$, e.g. 1 to 12 months, or 3, 6, 9, and 12 months (see Fig. 3.33). For each observation we calculate the realized conversion factors for the set of reference dates $t_r = t_d - T_i, \ \ i = 1, \ldots, k$. The difference, compared to the fixed horizon, approach is that we put all the observations $(a, t_d - T_i, t_d, \ldots)$ into one RDS. In the fixed horizon approach, we admit different time horizons only in different reference

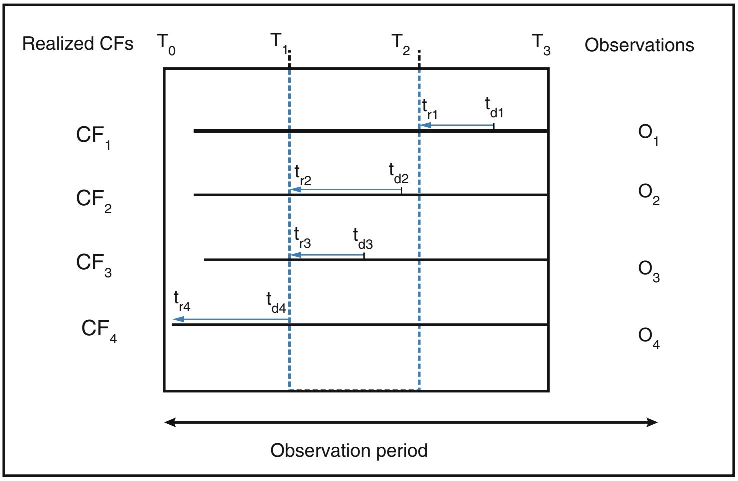


Fig. 3.32 Cohort approach

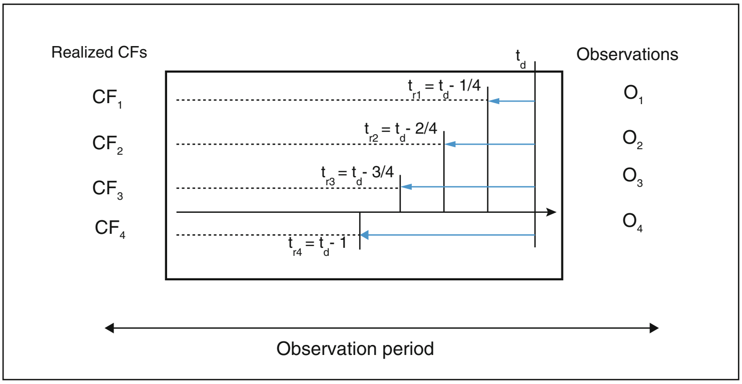


Fig. 3.33 Variable time horizon approach

data sets used for conditional *EAD* estimation. When all the observations are put into one RDS, there might be a problem with homogeneity; for example, the facilities that have been already marked as risky with restrictions on further drawing, should be treated separately. Moreover, there is an issue of the high correlation of the different observations obtained from one defaulted account. The RDS, on the other hand, captures implicitly the possible dependence of *EAD* and *CF* on the time to default, but the distribution of the time of default (appearing flat in the *RDS*), is not realistically reflected. It is not, by definition, suitable for the *PD*-weighted approach.

To summarize, it is recommended to use the fixed-time horizon RDS for the *PD*-weighted approach (different time horizons for different RDS). Otherwise, the cohort method should be preferred, unless the drawings show strong seasonality. In that case, we would recommend the variable-time horizon approach.

When estimating EAD at the *pool level approach*, defaulted and non-defaulted receivables are classified into a number of disjoint pools $l = 1, \ldots, m$ that are homogenous with respect to selected risk drivers, and which contain, at the same time, a sufficient amount of historical data. Pool l dataset $RDS(l)$ is used to obtain an estimation of the conversion factor $CF(l)$. Then, for a non-defaulted facility a belonging to the pool l based on the conversion factor approach we set:

$$EAD(a,t) = Ex(a,t) + (L(a,t) - Ex(a,t)) \times CF(l).$$

There are several approaches to the conversion factor estimation. The simplest approach is to calculate *the sample (default-weighted) mean*:

$$CF(l) = \frac{1}{|RDS(l)|} \sum_{o \in RDS(l)} CF(o)$$

where, given a reference data set with calculated ex-post conversion factors $CF(o), o \in RDS$, the same weight is assigned to each observation, disregarding the magnitude of undrawn amount or the time of the observation. In particular, the observations with very low undrawn amounts may bring a significant random error into the estimation. This problem is generally solved by the *weighted mean approach*:

$$CF(l) = \frac{\sum w_o \times CF(o)}{\sum w_o}, \tag{3.33}$$

where w_o ($o \in RDS(l)$) are appropriate positive weights. The natural candidates for the weights are the undrawn limits $w_o = L(o) - Ex(o)$. Then we get

$$CF(l) = \frac{\sum (EAD(o) - Ex(o))}{\sum (L(o) - Ex(o))}.$$

In order to choose the best approach it is natural to start with the standard goodness-of-fit measure

$$GF = \sum_{o \in RDS(l)} (EAD(o) - EAD(o))^2.$$

In other words, we are looking for estimation methods producing ex-ante *EAD* estimates that minimize the sum of the squared differences between realized *EADs*, and the ex-ante predictions. If we restrict ourselves to estimations of the form

$$EAD(o) = Ex(o) + CF(l) \times (L(o) - Ex(o))$$

then we need to minimize

$$GF = \sum (EAD(o) - E(o) - CF(l) \cdot (L(o) - Ex(o)))^2 \qquad (3.34)$$

which is equivalent to the OLS *regression without constant*:

$$EAD(o) - Ex(o) = \alpha + \beta(L(o) - E(o)) + \varepsilon(o) \text{ with } \alpha = 0 \text{ and } \beta = CF. \qquad (3.35)$$

Consequently:

$$CF(l) = \frac{\sum \left(EAD(o) - Ex(o)\right) \cdot \left(L(o) - Ex(o)\right)}{\sum (L(o) - Ex(o))^2}. \qquad (3.36)$$

Note that this formula corresponds to the weighted mean approach (3.33), with $w_o = (L(o) - Ex(o))^2$. The formula (3.36) could be recommended as the most consistent pool level *CF* estimation approach.

It is also possible to refine the regression (3.35) with $\beta = CF$ expressed in terms of other explanatory variables (macroeconomic, facility, or debtor level risk drivers). CF can be modeled in different parametric forms. The simplest linear form would be $CF = \mathbf{b}'\mathbf{f}$, where $\mathbf{f}$ is a vector of relevant risk drivers and $\mathbf{b}$ is the vector of linear regression coefficients. Alternatively, we may use a link function, for example, the exponential function $CF = e^{-\mathbf{b}'\mathbf{f}}$, where the outcome is always positive, but may be also larger than 1. If the historical data confirm that $CF \in [0, 1]$, then the logit function might be more appropriate:

$$CF = \Lambda(\mathbf{b}'\mathbf{f}) = \frac{e^{\mathbf{b}'\mathbf{f}}}{1 + e^{\mathbf{b}'\mathbf{f}}}.$$

The coefficients are obtained by numerically optimizing, either the sum of squared errors (3.34), or by using the maximum likelihood approach. The account level estimations should, rather, be used to analyze the most relevant risk drivers, and to define the optimal pools where the pool level estimate (3.36) is applied. For more details, see Witzany (2009c).

Risk Premiums Revisited

The concept of survival probabilities together with LGD and EAD estimates can be used to set up much more precise actuarial calculation of the risk premiums. Specifically, let us assume that we have estimated the survival function $S(t)$, possibly conditional on a rating class or even specific on the characteristics of an individual exposure. In addition, assume that $EAD(t)$ is the expected exposure at time t (measured from the exposure origination) and $LGD(t)$ the loss given default rate, possibly depending on time, but usually assumed to be constant. The equivalence principle used in a simplified form in (3.25) can now be applied taking into account the default and survival probabilities over the full life of the product. The life-time discounted expected loss can be approximated as

$$EL_{LT} = \sum_{k=1}^{K} LGD(k\Delta t) \times EAD(k\Delta t) \times e^{-rk\Delta t} \times (S((k-1)\Delta t) - S(k\Delta t)),$$

where the time to maturity $T = K\Delta t$ is split into shorter time intervals, e.g. months corresponding to installment frequency, and r is an appropriate discount rate. Therefore, we are adding up the discounted losses at times $k\Delta t$ weighted by the default probabilities over the time intervals $((k-1)\Delta t, k\Delta t]$ for $k = 1, \ldots, K$. On the other hand, if RP denotes the unknown premium rate paid always at $k\Delta t$ based on the outstanding balance and conditional on survival (that default has not taken place until $k\Delta t$) then the expected life time income can be expressed as

$$EI_{LT} = RP \times \sum_{k=1}^{K} EAD(k\Delta t) \times e^{-rk\Delta t} \times S(k\Delta t).$$

Finally, the actuarial equivalence principle equation $EL_{LT} = EI_{LT}$ can be easily solved for the risk premium as follows:

$$RP = \frac{\sum_{k=1}^{K} LGD \times EAD(k\Delta t) \times e^{-rk\Delta t} \times (S((k-1)\Delta t) - S(k\Delta t))}{\sum_{k=1}^{K} EAD(k\Delta t) \times e^{-rk\Delta t} \times S(k\Delta t)}. \tag{3.37}$$

As mentioned at the end of Sect. 3.3, the estimated real-life mortgage portfolio products survival functions shown in Fig. 3.19 can be used to estimate consistently credit margins for the three considered ratings A, B, and C. The future exposure at default $EAD(t)$ is easily calculated based on the installment calendar that is known for non-revolving products like mortgages. The loss given default parameter assumed to be constant, i.e. independent on time, has been estimated slightly below 20% based on the historical data. It is obvious, looking at Fig. 3.19, that the risk premium based on the simplified formula (3.25) may differ quite significantly from the calculation given by (3.37).

3.6 Basel II Rating Based Approach

The development and the main goals of the Basel II/III regulation have been introduced in Sect. 2.3. As shown in Fig. 2.9, there are three Regulatory Pillars: Minimum Capital Requirements, Supervisory Review Process, and Market Discipline. Let us look closer at the quantitative First Pillar that gives the rules for capital adequacy calculation.

The Capital Adequacy Ratio (CAR) is defined and expressed as the capital divided by the Risk Weighted Assets (RWA):

$$\text{Capital Ratio} = \frac{\text{Total Capital}}{\text{Credit RWA} + \text{Market RWA} + \text{Operational RWA}}. \tag{3.38}$$

The minimum requirement for the Capital Ratio is 8% plus various conservative and countercyclical margins introduced by Basel III (BCBS 2010). Originally (BCBS 1988), the risk weighted assets had been calculated as on-balance, or off-balance, exposures multiplied by a set of coefficients, and there was no Market Risk, or Operational Risk Part. The idea was to have sufficient capital in the case of a stressed scenario, when the losses amount to 8% or more of RWA. In the context of the current Basel II regulation, the principle did not change, but banks are, rather, asked to calculate the regulatory capital related to credit, market, and operational risks, which should cover unexpected losses (see Fig. 2.7), and compare the required regulatory capital with the available shareholders capital. The formula (3.38) could be now rather written as

$$\text{Capital Ratio} = \frac{\text{Total Capital}}{\text{Regulatory Capital}} \times 8\%. \tag{3.39}$$

If the bank's capital is greater than, or equal to, the required regulatory capital, then the bank is fine. Otherwise, there is a problem. The formulas (3.38) and (3.39) are equivalent through the relationship: $\text{RWA} = 12.5 \times \text{Regulatory Capital}$, or equally, $\text{Regulatory Capital} = 8\% \times \text{RWA}$.

Here, we focus on the credit risk regulatory capital (or RWA), which can be calculated using several approaches (see Fig. 3.34): Standardized (SA) and internal rating based (IRB), with two sub-alternative approaches; foundation (IRBF) and advanced (IRBA).

Compared to Basel I regulation, all Basel II approaches incorporate credit ratings: external in the case of SA, and internal in the case of IRB. The empirical and theoretical evidence is that the unexpected credit risk is related to expected losses, and the main idea of the regulation is to differentiate assets according to their credit ratings. The Standardized Approach just refines the differentiation of the risk weights in RWA calculation, while the Internal Rating Based Approach works with PD, LGD, EAD, and other parameters, to estimate the contribution of each single

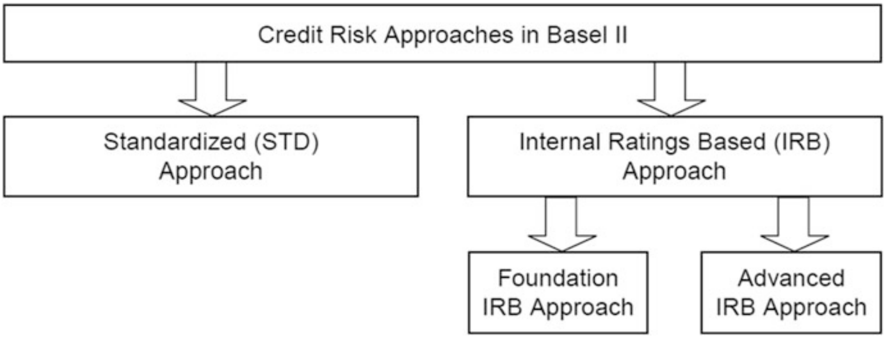


Fig. 3.34 Different Basel II approaches to credit risk measurement (Source: BCBS 2005a)

asset to unexpected loss. While PD is implied by the internal ratings, the other two parameters, LGD and EAD, are either set by the regulator in the Foundation Approach, or estimated by the bank in the Advanced Approach. The regulation works with five broad asset classes: corporate, sovereign, bank, retail, and equity. Within the corporate asset class, and the retail asset class, there is a number of sub-classes. The banks may apply any of the approaches separately within the asset classes. While the SA approach is the minimum approach, the IRBF or IRBA approaches are subject to a regulatory approval.

We have intentionally omitted too frequent citation of Basel II in the previous sections in order to emphasize that the various credit measurement techniques were developed before, and independently of, the regulation. The new regulation uses the already existing concepts, and at the same time, fosters new developments. The motivations for business oriented credit measurement and regulatory credit calculations are close, but not identical. The regulation needs rather conservative and stable (through-the-cycle) estimates, while business needs the most accurate and up-to-date (point-in-time) estimates. Moreover, the regulation sets down a number of qualitative standards which could be too restrictive for the development and application of new credit measurement and management methods. Therefore, it is optimal to have the same rating system, PD, LGD, and EAD estimates for both purposes, but in some areas, differences could exist.

The Standardized Approach

The main difference of the standardized approach, compared to the old Accord, is that the risk weights depend on ratings assigned by eligible External Credit Assessment Institutions (ECAIs); i.e., credit rating agencies like S&P, Moody's, or Fitch. The list of ECAI's is approved by national supervisors based on several criteria: objectivity, independence, international access/transparency, disclosure, resources, and credibility. Table 3.23 shows the weights according to the S&P-like rating

Table 3.23 Risk weights for sovereigns, banks, and corporations

Rating	Sovereign risk weights (%)	Bank risk weights (%)	Rating	Corporate risk weights (%)
AAA to AA−	0	20	AAA to AA−	20
A+ to A−	20	50	A+ to A−	50
BBB+ to BBB−	50	100	BBB+ to BBB−	100
BB+ to B−	100	100	BB+ to BB−	100
Below B−	150	150	Below BB−	150
Unrated	100	100	Unrated	100

Source: BCBS (2006a)

scale. National supervisors are, however, responsible for assigning the ECAI's ratings to the available weights. If there are more ratings available, then the second worst weight applies.

Moreover, retail loans not secured by residential real estate receive a flat 75% weight while mortgages have 35% weight. The weight is increased to 100–150% for past due (90 days and more) retail loans and 50–100% for mortgages. Unrated exposures are assigned a flat weight of 100% corresponding to an average weight of the rated exposures. However, this can be criticized as an incentive for higher risk debtors to avoid obtaining any external rating (note that the ratings B- and worse receive the weight 150%).

Thus, generally the RWA is calculated as $E \times w$; i.e., the Exposure multiplied by the Risk Weight that is determined by the ratings and the regulatory tables. For on-balance sheet items, E is just the outstanding exposure, while for off-balance sheet items, E is calculated as a credit conversion factor (CCF) multiplied by the off-balance sheet exposure. For example, for guarantees, the CCF is 20% if the original maturity is up to 1 year, and 50% if the maturity is over 1 year. The regulator also reflects various credit risk mitigation techniques. The exposure may, in the *simple approach*, be reduced due to collateralization by assets like cash or gold, or the risk-weight may be substituted by a reduced one, according to the guarantee counterparty. In the *comprehensive approach*, the risk weight w is reduced by a formula depending on the collateral quality.

The idea behind the RWA coefficient tables lies in the empirical and theoretical fact that unexpected credit losses are related to the expected losses. For example, if the expected loss rate on a homogenous portfolio was 3%, then the unexpected or stressed loss rate would be three to four times higher, say 11%. While the expected loss should be covered by the margin, the unexpected loss of $11-3\% = 8\%$, needs to be covered by the capital. The 8% required capital cushion corresponds to the $12.5 \times 8\% = 100\%$ risk weight (BB+ to B− weight for all asset classes in Table 3.23). Indeed, for the S&P rating classes BB+ to B−, the average historical default rate is at about 5%, and, considering a normal LGD value of

60%, we are at the expected loss around $3\% = 5\% \times 60\%$. This is a very rough calculation just to illustrate the logic of the risk weights, depending on the ratings, or effectively on the corresponding expected loss rates. The main advantage of this new approach, compared to the old Accord, is that it reasonably differentiates within the most important asset classes (corporations, banks, and sovereigns) according to the risk. Yet, the differentiation is still very approximate and the calculation does not reflect other key factors, such as portfolio diversification and asset correlation.

The Internal Rating Based Approach

In the IRB approach, Table 3.23 is replaced by a formula, or rather a set of formulas, that vary according to the different asset classes and sub-classes. The formulas in general look as follows:

$$\begin{aligned} &RWA = EAD \times w \\ &w = K \times 12.5 \\ &K = (UDR(PD) - PD) \times LGD \times MA, \end{aligned} \tag{3.40}$$

where *UDR* denotes an unexpected default rate calculated by a regulatory formula, with *PD* being the key input parameter, *LGD* is the regulatory, or own, estimate of the loss given default, and *MA* is the maturity adjustment, again given by a specific formula depending on the receivable's maturity. Note that the RWA formula is just mechanically derived from the capital requirement ratio *K*, applied to the exposure at default *EAD*. The maturity adjustment is applied only to non-retail receivables; for retail receivables $MA = 1$. The formulas vary for different asset classes, and differ in particular in the *UDR* calculation, where the asset correlation is either given as a constant, or by a formula depending on *PD* and size of firm in the case of small and medium enterprises (SME).

Figure 3.35 illustrates the dependence of the risk weight *w* in the case of corporate exposures, with LGD set at 45%, and maturity M at 3 years. Note that the weights are relatively close to Table 3.23 when the rating grades are replaced by the average long term default rates.

We have intentionally omitted presenting the full formulas in this section. The formulas and the theory behind them will be outlined in Sect. 4.7. The point is that the formulas (3.40) are, in a sense, continuous and higher dimensional alternatives of Table 3.23, and the key PD parameter is estimated through an internal rating system.

An eligible internal rating system for an asset class must satisfy a number of regulatory conditions and must be approved by the national supervisor. It is difficult to summarize all the qualitative requirements set up by the BCBS (2006a). The guiding principle is that the regulator wants to see higher credit risk measurement standards leading to unbiased, or rather conservative PD estimates. The internal

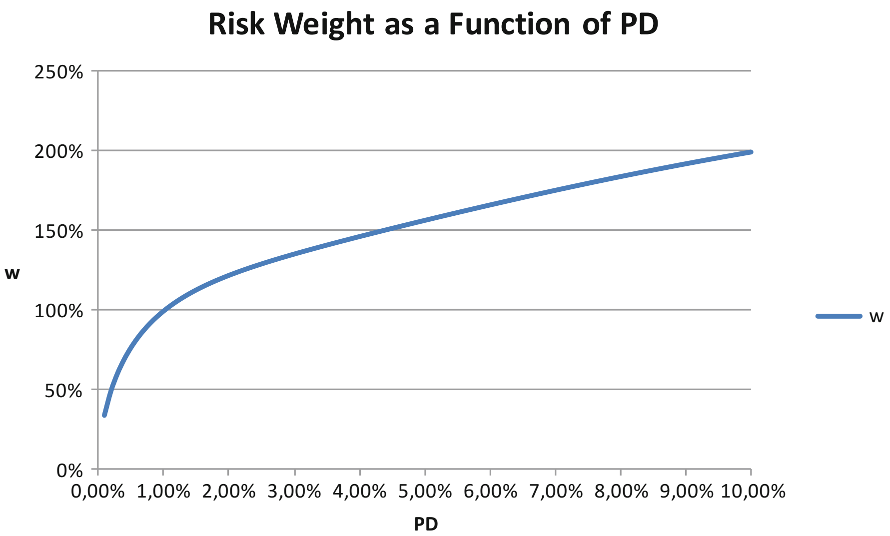


Fig. 3.35 Basel II IRB risk weights for corporates (LGD = 45%, M = 3)

rating system may be expert based, mechanical, or a combination of both. It must differentiate meaningfully across the different risk classes. The minimum number of rating grades is explicitly given in the case of corporate, sovereign, and bank exposures as seven for non-defaulted, plus one for defaulted, receivables. For those asset classes the rating system must have two dimensions, borrower and transaction specific; while in the case of retail exposures, both dimensions could be covered by a one-dimensional rating system. The rating risk horizon must be at least 1 year. According to §415 of the BCBS (2006a): "*The range of economic conditions that are considered when making assessments must be consistent with current conditions, and those that are likely to occur over a business cycle within the respective industry/geographic region.*" Note that this requirement shifts the rating system closer to the TTC (Through-The-Cycle) philosophy, diverging from a more usual PIT (Point-In-Time) business oriented internal rating system.

PD estimates are assigned to the rating grades. The estimates must be based on long-term averages of the 1-year default rates of debtors in the grade, and must incorporate a margin of conservatism that is related to the likely range of estimation error. The length of the underlying historical observation period must be at least 5 years. The bank must use, consistently, a definition of default satisfying the regulatory conditions according to which a debtor that is "unlikely to pay", or 90 days past due, must be classified as defaulted. The rating and PD estimation process must be well documented. Moreover, the bank must regularly (at least once a year) perform a validation (see Sect. 3.1), and report the results to its supervisor.

Note that the PD requirements again emphasize the TTC, and the rather conservative philosophy of the regulatory estimates, which may differ from the PIT business oriented values.

The Foundation (IRBF) Versus Advanced (IRBA) Approach

The LGD and CF parameters are set down explicitly by the regulation in the Foundation Approach, but internally estimated by the bank in the Advanced Approach. Banks may choose between the two approaches only in case of non-retail exposures. For retail exposures only the Advanced Approach is possible in the context of the IRB.

Under the Foundation Approach, senior claims on corporations, sovereigns and banks not secured by recognized collateral are assigned a 45% LGD, while subordinated exposures are assigned 75% LGD. The values may be reduced by recognizable collateral. The EAD is defined as the on-balance sheet exposures, or off-balance sheet exposure times and the CF coefficient is defined as in the SA approach.

In the Advanced Approach (IRBA), banks produce their own LGD and EAD estimates, subject to a number of qualitative requirements similar to the PD estimates. However, there is an important difference between the PD parameter that is stressed through the formula (3.40), and LGD, EAD parameters that are not explicitly stressed. For this reason the regulation does not require only the incorporation of an estimation error, but also other margins of conservatism that should be part of the final estimates. According to §468 of the BCBS (2006a, b): *"A bank must estimate an LGD for each facility that aims to reflect economic downturn conditions where necessary to capture the relevant risks. This LGD cannot be less than the long-run default-weighted average loss rate given default calculated based on the average economic loss of all observed defaults within the data source for that type of facility. In addition, a bank must take into account the potential for the LGD of the facility to be higher than the default-weighted average during a period when credit losses are substantially higher than average."* Moreover, the banks must also take into account the potential dependence of the borrower risk and the collateral or collateral provider, or currency risk, in the case of a currency mismatch between the loan and the collateral value. Overall, the produced LGD, and similarly the EAD, estimates are not to be the expected future values, but rather, certain stressed values. A reasonable interpretation and implementation of those requirements is a challenge for banks and their national supervisors. There is an ongoing discussion among practitioners and academicians on these two puzzling concepts, as well as on the related potential systemic risks (see Sect. 4.7).

The banks also have to estimate LGDs on defaulted assets, again reflecting the possibility that they would have to recognize additional, unexpected losses during the recovery period. For each defaulted asset, the banks must, moreover, obtain their best estimate of the expected loss (BEEL) on that asset, based on current economic circumstances and facility status (§471 of the BCBS 2006a). The specific capital requirement for defaulted assets is then defined as:

$$\begin{aligned} \mathrm{RWA} &= EAD \cdot w \\ w &= K \cdot 12.5 \\ K &= \max(0, LGD - BEEL) \end{aligned}$$

Note, that by definition, LGD should be never less than the BEEL. Thus, the capital requirement on the defaulted exposure is simply defined as the difference between the unexpected losses, and the expected losses. The regulation does not specify the probability level at which the unexpected loss should be estimated. Therefore, a consistent statistical estimation of the parameter is even more challenging than LGD for non-defaulted exposures. The BEEL value should be, at the same time, compared with provisions and write-offs. The former being less than the latter may "attract supervisory scrutiny" and must be justified by the bank. Such a situation may happen when the credit risk management and modeling process is separated from the more traditional accounting based provisioning process. It would be simplest to stipulate that provisions are greater than, or equal to, BEEL, but since the Basel regulation cannot set strict rules on accounting (regulated by the accounting authorities), there is, instead, this soft rule strongly urging the banks to synchronize the two concepts.

Finally, banks applying IRBA must compare the total amount of expected losses $EL = PD \times LGD \times EAD$, with total eligible provisions (§43, §375, and §380 of the BCBS 2006a). If the total expected loss amount exceeds total eligible provisions, banks must deduct the difference from their capital in a specified way. Note that banks currently create provisions, according to IAS 39, on the class of all impaired assets that, generally, is not identical to the class of defaulted assets. The principle of portfolio provisions also allows one to create provisions on individually non-defaulted, yet statistically impaired, exposures. The existing provisions should reflect the incurred losses (on an individual or portfolio basis) caused by the past events, while expected losses capture future, not yet realized events, and should be more or less covered by the margin income. The discrepancy between provisions and expected losses should be partially resolved by the IFRS 9 implementation effective since 2018 (see Sect. 3.5).

Portfolio Credit Risk

4

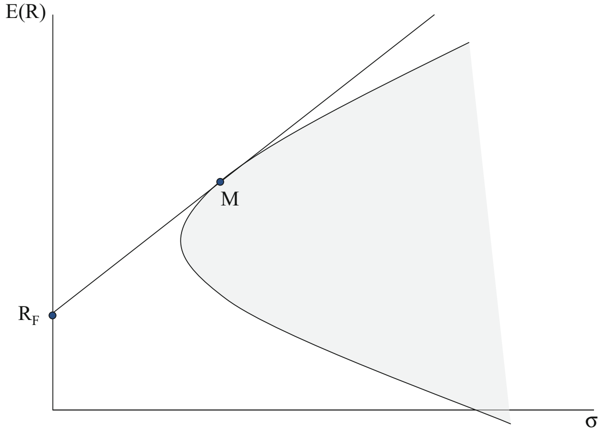
So far we have focused on methods how to properly measure credit risk and approve individual loan transactions. But even if this process is under control and loan underwriting is going well, a prudent bank management must ask the question; "When is enough enough?" Can the bank portfolio grow without limitations, or is there a limit? Moreover, is it optimal to specialize in one client segment, or economic sector, or is it better to split the underwriting activities among more segments and sectors? More specifically, can we optimize the risk/return relationship in the sense of the Markowitz Portfolio Theory (Fig. 4.1)?

The Markowitz Theory is suitable for portfolio equity investments, where the returns may be assumed to be jointly normal, characterized by a covariance matrix. Various portfolios, combinations of the available stocks with different weights then give different risk/expected return profiles. Since the returns are assumed to be normal, their standard deviation is a satisfactory measure of risk that is compared with the expected returns. The point is that there is a diversification effect if the correlation between two assets with similar return is less than 1. Consequently, it is always better to diversify our investment into two or more assets than to put all the money into one asset. And because investors prefer to minimize risk and maximize return, not all portfolios are optimal; the portfolios inside the shaded area in Fig. 4.1 are not optimal, since either the expected return can be improved without increasing the risk, or the risk can be reduced without reducing the return. The top part of the borderline of the shaded area is called the Efficient Frontier, where the optimal portfolio risk/return combinations lie. When stock investments are combined with a risk free investment with zero risk, and return R_F, we obtain the Capital Market Line (the straight line R_FM in Fig. 4.1); with the Efficient Market Portfolio M playing a key role in the Capital Asset Pricing Model (CAPM).

To apply the portfolio theory to a loan portfolio, we first need to define an appropriate risk measure. The difference is that loan portfolio returns are heavily asymmetric, and certainly not normal. Nevertheless, intuitively it should also be better to diversify the available financing resources among more debtors and economic sectors than bet all the money on one debtor, or economic sector. The

J. Witzany, *Credit Risk Management*, DOI 10.1007/978-3-319-49800-3_4

Fig. 4.1 Risk versus return optimization



risk measure applied here, to make the intuition numerically tractable, is the credit unexpected loss or equally, the economic capital.

We have already seen that there is a banking regulation requiring that the capital ratio; i.e., capital divided by the risk weighted assets, must be at least 8%, so, given a capital level, the loan portfolio (with positive risk weights) cannot grow indefinitely. The regulation also puts certain limits on concentrations, i.e. on exposures with respect to single borrowers, or economically connected groups. Yet the question should first be answered independently of the existing regulation. We will explain the notions of expected and unexpected risk applied to a credit portfolio, and discuss the various approaches to their modeling and estimation. We shall see that the Basel regulation has been motivated by intuition and the modeling progress in this area.

4.1 Economic Capital, Expected, and Unexpected Losses

Let us formulate more precisely the definition of the expected and unexpected losses, already indicated in Fig. 2.8. Let us have a portfolio of assets (and liabilities), and let X denote the loss (with a positive sign if there is loss, and a negative sign if there is a profit) on the portfolio in a fixed time horizon T. Today we do not know the value X, which depends on the future "state of the world", and so we model it as a random variable. The expected loss is then obviously defined as $EL = E[X]$, and the unexpected loss as the difference between a quantile of the random variable X and the expected loss. Let $F_X(x) = \Pr[X \leq x]$ be the distribution function of X, then, given a probability level α (e.g. 95 or 99%), define the quantile $q_\alpha^X = \inf\{x | F_X(x) \geq \alpha\}$. The absolute Value at Risk $VaR_\alpha^{abs} = q_\alpha^X$ being the same as the quantile expresses the potential maximum absolute loss that can be realized on the probability level α. The unexpected loss, or relative Value at Risk VaR_α^{rel}, measuring the probability level α potential excess loss over EL, is then precisely expressed as $UL = q_\alpha^X - E[X]$.

The concept of VaR was first developed on investment or trading portfolios. The situation is simplest if the variable X can be assumed normally distributed with mean μ and standard deviation σ, i.e. $N(\mu, \sigma^2)$. Then the unexpected loss on any probability level is simply proportional to σ, and so, the standard deviation is a sufficient measure of risk. Specifically, $q_\alpha^X = \mu + q_\alpha^N \cdot \sigma$ where q_α^N is the quantile of the standardized normal $N(0, 1)$ distribution [for example, in the Excel application given by the function NORMSINV (α)]. Hence, $UL = VaR_\alpha^{rel} = q_\alpha^N \cdot \sigma$. The normality assumption is approximately valid for equity and foreign currency portfolios, where the portfolio standard deviation can be obtained analytically from individual asset return variances, mutual correlations, and the asset weights. The situation becomes more complex when the returns are not normal, or depend nonlinearly on the normal returns of the underlying assets, as in the case of options, or other derivatives. This is, in fact, also the case of debt instruments, bonds or loans that have implicit embedded options, as we shall see. Then the estimations of Value at Risk become more complex, generally leading to a Monte Carlo Simulation approach, where the distribution of losses is generated numerically, calculating the portfolio values in many simulated scenarios.

The Unexpected Loss, or Value at Risk concept can be directly applied to a portfolio of corporate bonds valued with market prices. To analyze the future loss $X = V_0 - V_1$, we need to model the distribution of the future bond portfolio values V_1. If there are historical bond price series (which is rarely the case), then the covariance matrix could be applied for a short time interval where the price changes are not too large. The short term bond price changes are mostly related to movements in risk free interest rates and to changes in the issuers credit quality (spreads) from the perspective of the financial markets. The normal distribution based approach will not be appropriate for a longer time horizon when the bond returns are not symmetrical. In that case, we need to find a more sophisticated approach.

The situation is even more difficult for ordinary banking loan portfolios. First, there is no market value. The accounting value is defined as the outstanding receivable amount, minus provisions if the receivable is impaired. In fact, there is almost no profit potential (besides the interest revenue), and the total loss at the end of a period is just the sum of net provisions created on the portfolio (provisions can be also released, therefore, exceptionally, there could be a profit). This is the loss that enters the Profit/Loss Statement, and that should be the most important for any bank management. Thus, firstly the approaches based on market values can be applied only approximately. Secondly, we usually do not have any historical market price series data that could be directly applied to our assets.

If a bank knows how to estimate the unexpected loss on its portfolio of all the operations (market, credit, and operational) on a probability level, then it should be, first of all, compared to the bank's available capital. For example, if we estimate the unexpected loss to be 20 billion CZK in a 1 year horizon, on the 95% probability level, and the bank's capital is just 15 billion CZK, then there is obviously a problem. The bank will go completely bankrupt within 1 year with a probability of at least 5%, which should not be acceptable to the shareholders, prudent bank management, or the regulators either. The bank should choose a relatively high confidence level α; e.g. 99, 99.5, or even 99.9%, on which the unexpected losses are calculated and compared to the capital. The probability $1 - \alpha$ should correspond to the targeted annual probability of default of the bank itself; i.e., to its own targeted credit rating. The unexpected loss should then be strictly less than, or equal to, the available capital. If it is already equal, and some new investments are proposed, then new capital, corresponding to the incremental unexpected loss will have to be raised. Therefore it is natural to define the economic capital of a new operation as the amount equal to the incremental unexpected loss, and compare the expected profit to it. The ratio should satisfy a condition on the minimum return on capital, and, moreover, the approach can be used to compare consistently a number of proposed projects (loan applications), maximizing the expected return/economic capital ratio. Note, however, that the capital allocation mechanism is incremental—the diversification effect of the new investment depends on the existing portfolio. The allocation of capital in a portfolio becomes a little bit problematic if the order of investments is not given, since it is not clear how to split "the diversification portfolio effect." A proportional approach is, then, probably the most natural.

A leading role in the implementation of economic capital allocation has been played by the Bankers Trust, which in 1995 introduced the concept of RAROC—Risk Adjusted Return on Capital. The return is compared to the RAROC economic capital coefficient, expressed for a market product simply as $2.33 \times \text{Weekly volatility} \times \sqrt{52} \times (1 - \text{Tax rate})$. The term 2.33 corresponds to the 99% standardized normal distribution quantile, $\sqrt{52}$ scales the weekly volatility of returns to a 1-year horizon, and the term $(1-\text{Tax rate})$ says that the bank is interested, primarily, in net after tax profit (or loss). The formula differentiates between less and more risky assets, but it does not take into account the diversification effect. Later, the Bankers Trust expanded the original RAROC concept into a "comprehensive risk management system", based on the concept of Value at Risk, termed RAROC 2020.

Our ability to evaluate the unexpected loss at portfolio level enables us to optimize the risk–return relationship as in the Markowitz portfolio model shown in Fig. 4.1, but with the risk measured, not by the standard deviation, but by the unexpected loss; i.e., economic capital.

4.2 CreditMetrics

The CreditMetrics methodology, published by JP Morgan (1997), has become a standard in the field of portfolio credit risk measurement. The model is based on ratings that are assumed to determine the values of individual debt instruments. It is a Monte Carlo simulation based approach requiring relatively extensive data inputs.

The methodology was originally designed for bonds priced by their market values, but the approach can be modified to a loan portfolio, and loss defined in terms of accounting provisions based on a classification system.

The model can be described by the following key principles:

1. *Today's prices of bonds are determined by their ratings*. And, of course, by the term structure of risk-free interest rates that is fixed throughout the calculation. The rating scale is, for example, S&P's or Moody's. Therefore, for each rating there is a rating specific term structure of interest rates allowing us to consistently value bonds with different maturities. The term structure is obtained from the market prices of bonds with the given rating, and with different maturities in a standard manner. The curves are provided by financial data information companies, such as Bloomberg, and there is, in any case, enough market data to perform the calculations of the term structure. The difference between the rating specific interest rates incorporating certain credit losses and the risk-free interest rates forms the rating specific term structure of credit spreads.
2. *Future prices of bonds (*e.g.*, in a 1-year horizon) are determined by their future ratings*. The point is, that we use today's rating specific term structure of interest rates to obtain the forward rating specific term structure of interest rates. This allows us to determine the forward price of a bond, conditional on its future rating. The market value of a defaulted bond is determined by a recovery rate parameter, depending on its seniority. Note that in this approach we, intentionally, do not take into account interest rate risk. Our goal is, indeed, to analyze only the credit risk. Nevertheless, the model can be extended, simulating also the future risk-free rates, using an appropriate interest rate model, and combining the risk-free rates with the forward credit spreads.
3. *Rating migration probabilities are obtained from historical data*. The rating transition probabilities are regularly published and monitored by all major rating agencies. So, for a single bond portfolio of a given initial rating, it is no problem to simulate the distribution of future market values.
4. *Rating migration correlations are modeled through asset correlations*. To simulate joint migrations of many bonds from different issuers, we have to take into account their correlations. This is the key part of the model that utilizes Merton's credit risk option model, connecting the credit risk with a relationship between the firm's asset value and its total debt.
5. *The asset correlations are estimated by mapping the firms into various economic sector indices*. Moreover, in order to estimate the individual firm correlations, we need the sector indices' correlations which can be obtained, for example, from equity markets data.
6. *Finally, simulate future ratings and market values of all bonds in the portfolio to obtain the portfolio value empirical distribution, expected, and unexpected loss on any given probability level*. In practice we simulate, first of all, the indices' returns with a given correlation structure, as there are a limited number of indices used. Then the idiosyncratic (firm specific) independent factors are

simulated. The calculated firm specific asset returns are translated into new ratings and market values.

Let us now give more technical details on the steps outlined above.

Bond Valuation

Regarding valuation, we assume that for each rating s, there is a term structure of zero-coupon interest rate $r_s(t)$ for every maturity t, allowing us to value an s-rated bond, paying the cash flow $CF(t_i),\ \ i = 1, \ldots, n$. The risk adjusted present value is given by the standard formula

$$P = \sum_{i=1}^{n} \frac{CF(t_i)}{(1 + r_s(t_i))^{t_i}}.$$

The same principle can be used to value the bond at a future time t_0, using the forward rates implied by the current zero coupon rates i.e., by solving the equations

$$(1 + r_s(t))^{t} = (1 + r_s(t_0))^{t_0}(1 + r_s(t_0, t))^{t - t_0}$$

for the forward rate $r_s(t_0, t),\ \ t > t_0$. If we assume that the given bond has a new rating u at time t_0 then the simulated forward value is

$$P_u = \sum_{t_i \geq t_0} \frac{CF(t_i)}{(1 + r_u(t_0, t_i))^{t_i - t_0}}. \tag{4.1}$$

The situation might be complicated by coupon payments between today and time t_0. Here, we assume, for the sake of simplicity, that the first coupon is paid just at t_0; typically in 1 year, and so it is included in (4.1).

Rating Migration

Simulated migration of a bond's rating, with a given initial rating, should respect the historical transition probabilities as shown, for instance, in Table 4.1.

Another key input factor is the recovery rate that is used in the case of default represented by the worst rating. Table 4.2 shows the average recovery rates by seniority classes (the order in which liabilities are satisfied in the case of bankruptcy; senior bonds being the first, and junior subordinated the last). The standard deviations show that the recovery rates in the case of default are not deterministic. In general, the recovery rates should be simulated based on an appropriate probability distribution (see Sect. 3.5), but the average values are often used in a simplified approach. The simplification does not cause a significant error on a large portfolio if the rate of default is not correlated with the recovery rates. However, a number of studies (Altman et al. 2002) indicate that there could be a negative correlation, explained by the economic cycle simultaneously pushing default rates up and recovery rates down. It is shown in Witzany (2009d) that the

Table 4.1 One-year transition probabilities (Source: Crouhy et al. 2000)

	Rating at year-end (%)							
Initial rating	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	90.81	8.33	0.68	0.06	0.12	0	0	0
AA	0.70	90.65	7.79	0.64	0.06	0.14	0.02	0
A	0.09	2.27	91.05	5.52	0.74	0.26	0.01	0.06
BBB	0.02	0.33	5.95	86.93	5.30	1.17	0.12	0.18
BB	0.03	0.14	0.67	7.73	80.53	8.84	1.00	1.06
B	0	0.11	0.24	0.43	6.48	83.46	4.07	5.20
CCC	0.22	0	0.22	1.30	2.38	11.24	64.86	19.79

impact of PD and Recovery Rate correlation may be quite significant, even in the case of a large homogenous portfolio. The original CreditMetrics methodology does not take the PD—Recovery Rate correlation into account. One possible way of incorporating it into the model is proposed in Witzany (2009d).

Based on the two tables above, it is straightforward to simulate single bond portfolio future values. Figure 4.2 shows the probability distribution of a 5-year BBB bond's values after 1 year with a deterministic recovery rate.

To simulate credit migrations of a two, or more, bond portfolio, we have to take correlations into account. It is possible to tabulate historical migrations of pairs of bonds with a given rating to all possible pairs of ratings, for example, the historical transition probability of two bonds with ratings (A, BB) going to (BB, BBB). This approach could be applied to a two bond portfolio, but hardly generalized to a many bond portfolio. More importantly, we have to realize that the correlation between the rating migration of two bonds depends more on shared systematic (macroeconomic) factors, than on specific initial ratings.

The CreditMetrics solution is to parameterize the change of rating of a bond b by a continuous random variable $r(b)$ with the standard normal distribution $N(0,1)$. The variable can be interpreted as a standardized asset return based on Merton's structural default model, but we can also interpret it purely technically as an appropriately scaled credit scoring change. Given an initial rating, we may define, based on rating transition probabilities, a sequence of thresholds for $r(b)$ to trigger possible rating migrations. For example, according to Table 4.1, the transition probability from BB to Default is 1.06%. Since positive values of $r(b)$ should intuitively mean credit quality improvement, and negative values deterioration, we are looking for a threshold Z_{Def}, so that $\Pr\left[r(b) \le Z_{Def}\right] = 1.06\%$. Consequently, $Z_{Def} = \Phi^{-1}(1.06\%) = -2.3$. The threshold for CCC must satisfy the condition $\Pr\left[Z_{Def} < r(b) \le Z_{CCC}\right] = 1\%$, since 1% is the migration probability from BB to CCC, thus: $Z_{CCC} = \Phi^{-1}(1.06\% + 1\%) = -2.04$. Similarly, using the BB rating migration probabilities, we may evaluate Z_{CC} up to Z_{AAA} (Fig. 4.3).

Table 4.2 Recovery rates by seniority (Source: JP Morgan 1997)

Seniority class	Mean (%)	Standard deviation (%)
Senior secured	53.80	26.86
Senior unsecured	51.13	25.45
Senior subordinated	38.52	23.81
Subordinated	32.74	20.18
Junior subordinated	17.09	10.90

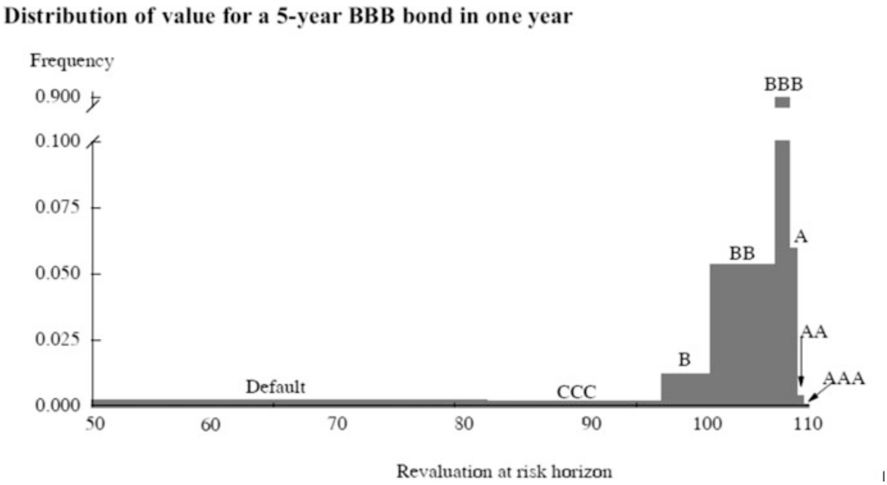


Fig. 4.2 Simulation of a single bond portfolio (Source: JP Morgan 1997)

Merton's Structural Model

The rating migration model described above is, in fact, based on Merton's structural model. The idea of the model is that the default of a firm on its debt D happens if the value of the firm's assets A falls below the debt; i.e. $A < D$. Since the assets change in time stochastically similarly to the firm's equity, we can use a stochastic model for $A(t)$, starting at an initial value $A(0) > D$. To simplify the situation, let us assume that there is just a loan with bullet repayment of an amount D at time T, and we are looking only at the time of maturity T. If $A(T) \geq D$, then the assets are sufficient to pay the debt back in full, and there is no default. The remaining shareholders' value is $A(T) - D$. On the other hand, if $A(T) < D$, then there is a default, the firm goes into bankruptcy or liquidation, the creditors receive just the asset value $A(T)$, and the shareholders value is 0 (Fig. 4.4).

The model has been formulated not only to theoretically define the probability of default, but, in fact, primarily to apply the theory of option valuations to the debt and equity market valuation. The final payoff for creditors at maturity can be expressed as

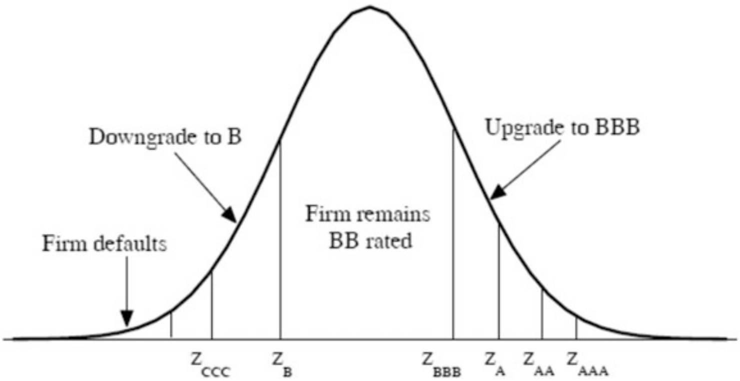


Fig. 4.3 Distribution of asset returns with rating change thresholds (Source: JP Morgan 1997)

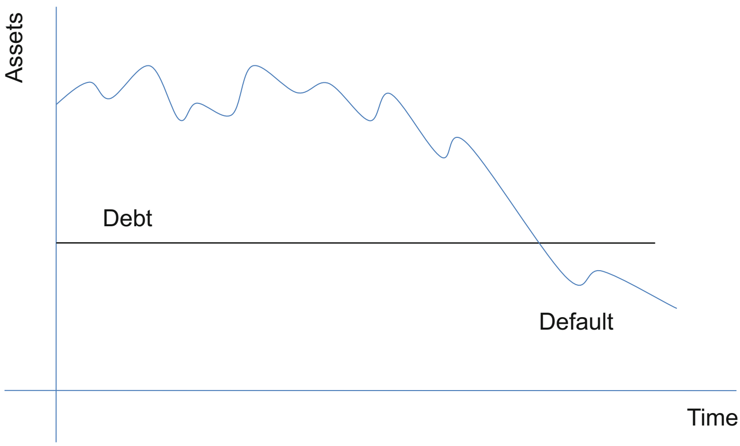


Fig. 4.4 Merton's structural model

$$D(T) = \min(D, A(T)) = D - \max(D - A(T), 0),$$

and so the value of the risky debt can be theoretically valued as the value of the risk free debt, minus the value of the European put option on the firm's assets, with the exercise price D, and maturity T, sold to the shareholders for the credit margin paid over a risk free interest rate. The shareholders payoff at time T can, on the other hand, be expressed as $E(T) = \max(A(T) - D, 0)$ and so the equity value $E(0)$ can be theoretically valued just as the European call option on the assets, with the exercise price D. If the asset value follows, for example, the geometric Brownian motion given by the stochastic differential equation:

$$dA(t) = \mu A(t)dt + \sigma A(t)dW(t),$$

then the call and put options can be evaluated with the Black-Scholes-Merton formula. It is clear that the model has a number of shortcomings. Firstly, the asset market value process is latent; i.e., it is not empirically observable in the majority of cases. This issue can be overcome using stock market data. We can argue that the asset value follows approximately the equity value, and so the volatilities and correlations estimated from stock price data can be used for the assets. In fact, there is a functional relationship between the equity prices and the asset prices given by the model as explained above, and so, using the stochastic calculus, the parameters estimated from stock data can be transformed into the appropriate parameters for the latent asset prices. This principle applied by the KMV model is discussed in more depth in Sect. 4.5. Another issue is that the assets are generally of different liquidity—short-term, long-term, financial and non-financial—and should be discounted in different ways during a distress situation. In addition, default can happen generally at any time before maturity, because of a missed coupon payment or bankruptcy declared due to indebtedness, so this is an American rather than European option. Some of these problems are handled in more elaborated models like the KMV one.

Let us now focus on the rating migration modeling. In the basic structural model the probability of default is determined by the actual asset value $A(t)$, and the asset volatility σ; in fact, by the distance of $\ln A(t)$ from $\ln P$, since the log value follows an ordinary Brownian motion with a drift stochastic differential equation:

$$d(\ln A) = \left(\mu - \sigma^2/2\right)dt + \sigma dW.$$

The larger the distance is, the better the credit rating should be. Thus, starting from an initial rating, the rating migration at time 1 is determined by the standardized $N(0, 1)$ asset return:

$$r = \frac{1}{\sigma}\left(\ln\frac{A(1)}{A(0)} - \left(\mu - \sigma^2/2\right)\right).$$

If the return is positive, then there is a rating improvement, and if the return is negative, then there is a rating deterioration based on the rating migration threshold evaluated from the historical migration probabilities as explained above. Notice that at the end we do not need to estimate the volatility σ, as the rating migrations depend only on the values of the standardized return variable r.

Rating Migration and Asset Correlations

Clearly, it would be incorrect to simulate rating migrations as independent events, since the individual debtor's situations often depend on the same macroeconomic factors, or there could even be a mutual specific economic interdependence between two or more debtors. We have seen that it is practically impossible to model the joint correlation migration of all the initial combinations of two, or more, ratings into new sets of ratings. Alternatively, one could try to replace the alphabetical ratings “Default”, “CCC”,. . ., “AAA” by some ordinals, e.g.: 1,. . .,8, and estimate the correlation between the changes of those variables. This approach is, nevertheless, hardly econometrically justifiable. A natural solution is to use the asset correlations based on Merton’s structural model. Given debtors $i = 1, \ldots, N$, all we need to know is the matrix Σ of correlations $\rho_{ij} = \rho(r_i, r_j)$ between the standardized asset returns, and the rating migration thresholds. The correlations could be estimated from the equity returns data, if all the companies are liquidly traded on stock markets. Since this is usually not the case, CreditMetrics proposes the use of a single-, or multi-, factor model, breaking down debtors’ returns into a combination of systematic factors and independent, idiosyncratic, debtor specific factors; i.e.,

$$r_i = \sum_{j=1}^{k} w_{i,j} r(I_j) + w_{i,k+1}\epsilon_i, \tag{4.2}$$

where $r(I_j)$ is the standardized return of the systematic factor (e.g. sector or country index) I_j, and ϵ_i is the standardized debtor specific factor. Since the systematic factors can be correlated, it is generally insufficient to require $\sum_{j=1}^{k+1} w_{i,j}^2 = 1$, in order to have r_i standardized. Determination of the weights, and thus, of the correlations, is a key step in the model and, unfortunately, also one of the most problematic parts of the CreditMetrics methodology. The Technical Document, in fact, proposes only an expert approach: estimate the weights of the systematic factors (combination of all systematic factors), and of the complementary idiosyncratic factor. In the case of more systematic sector, or country factors, specify the “participation” of the debtor and combine the indices appropriately.

Example Let us consider just one systematic factor I, e.g., a general market index, and let us say that for the first debtor we expertly estimate that 90% of firms’ asset return volatility is explained by the systematic factor. Therefore:

$$r_1 = 0.9r(I) + \sqrt{1 - 0.9^2}\epsilon_1 = 0.9r(I) + 0.44\epsilon_1.$$

Similarly, for the second debtor, we estimate that just 70% of the volatility is explained by the systematic factor, i.e.

$$r_2 = 0.7r(I) + \sqrt{1 - 0.7^2}\epsilon_2 = 0.7r(I) + 0.71\epsilon_1.$$

Since the idiosyncratic factors ϵ_1 and ϵ_2 are mutually independent, and also independent of the systematic factor $r(I)$, the implied correlation is simply

$$\rho(r_1, r_2) = 0.9 \cdot 0.7 \cdot \rho(r(I), r(I)) = 0.63. \square$$

Consequently, the expertly set systematic weights determine the mutual correlations, and significantly influence the final output of the model; i.e., the estimated unexpected loss. If the implied correlations are too far from the "real" correlations, then the output of the model can be completely wrong.

Let us now break down the systematic part into two indices. Assume that our experts estimate that the first debtor has a 60% participation in the automotive industry, and 40% participation in the electronics industry. This is interpreted by CreditMetrics as if the systematic (not standardized) part of the return were expressed by a linear combination of the two industry index returns:

$$R(I) = 0.6 \cdot R(I_{Aut}) + 0.4 \cdot R(I_{El}).$$

To normalize the return $R(I)$, we need to calculate its standard deviation:

$$\sigma(R(I)) = \sqrt{0.6^2\sigma(R(I_{Aut}))^2 + 0.4^2\sigma(R(I_{El}))^2 + 2 \cdot 0.6 \cdot 0.4 \cdot \rho(R(I_{Aut}), R(I_{El})) \cdot \sigma(R(I_{Aut})) \cdot \sigma(R(I_{El}))}$$

Thus, given the volatilities $\sigma(R(I_{Aut})) = 20\%$, $\sigma(R(I_{El})) = 35\%$ and the correlation $\rho(R(I_{Aut}), R(I_{El})) = 30\%$ of the sector indices, we obtain $\sigma(R(I)) = 19.3\%$. Finally, we express the standardized return $r(I) = R(I)/\sigma(R(I))$ as a combination of the returns $r(I_{Aut}) = R(I_{Aut})/\sigma(R(I_{Aut}))$ and $r(I_{El}) = R(I_{El})/\sigma(R(I_{El}))$

$$\begin{aligned} r(I) &= \frac{0.6 \cdot \sigma(R(I_{Aut}))}{\sigma(R(I))} r(I_{Aut}) + \frac{0.4 \cdot \sigma(R(I_{El}))}{\sigma(R(I))} r(I_{El}) \\ &= 0.57 \cdot r(I_{Aut}) + 0.67 \cdot r(I_{El}). \end{aligned}$$

Since the weight of the combined systematic factor is 0.9, we finally get:

$$\begin{aligned} r_1 &= 0.9 \cdot 0.57 \cdot r(I_{Aut}) + 0.9 \cdot 0.67 \cdot r(I_{El}) + 0.44\epsilon_1 \\ &= 0.51 \cdot r(I_{Aut}) + 0.6 \cdot r(I_{El}) + 0.44\epsilon_1. \end{aligned}$$

The second debtor is classified as being 90% in electronics and 10% in the automotive sector. Applying the same calculations on r_2, we obtain:

$$r_2 = 0.04 \cdot r(I_{Aut}) + 0.69 \cdot r(I_{El}) + 0.71\epsilon_2.$$

The implied correlation between r_1 and r_2, then, is

$$\begin{aligned} \rho(r_1, r_2) &= 0.51 \cdot 0.04 + 0.6 \cdot 0.69 + (0.51 \cdot 0.69 + 0.04 \cdot 0.6) \cdot \rho(r(I_{Aut}), r(I_{El})) \\ &= 0.55. \end{aligned}$$

Although the expert classification of a firm into one or more sectors is plausible, we do not see it as a realistic approach to the estimation of the systematic and

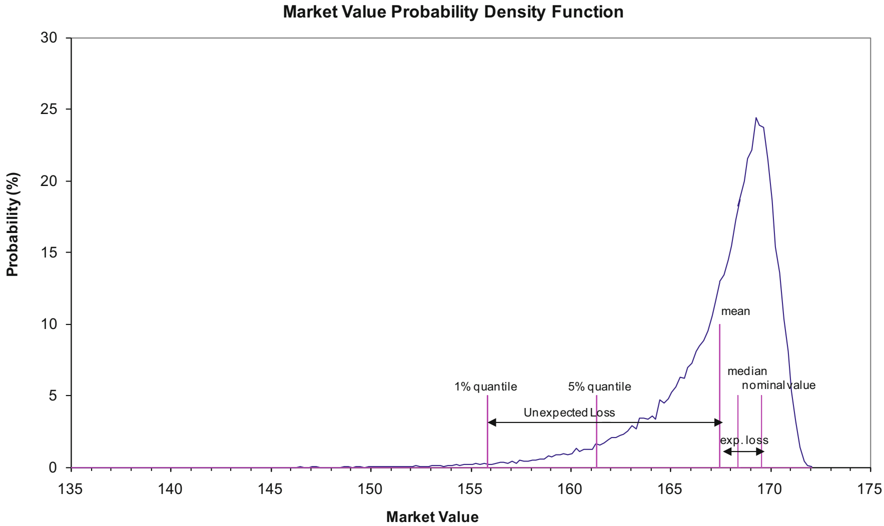


Fig. 4.5 Credit portfolio Monte Carlo simulation

idiosyncratic factor weights determining an overall correlation level. The systematic weight should be based rather on an empirical estimation, for example, using stock returns of a similar company traded on the markets.

Portfolio Simulation

Given an asset return correlation matrix Σ, the scenarios can be generated sampling a vector of the standardized normal variables $\mathbf{u}$, and multiplying it by the Cholesky matrix A; i.e., the lower triangular matrix, so that $A \cdot A^T = \Sigma$. Given the vector of standardized asset returns $\mathbf{r} = A\mathbf{u}$, we determine, based on the thresholds, the rating migrations and the simulated portfolio value $V(\mathbf{r})$ at the end of the period. Repeating the procedure we obtain a large number of sampled values $V_1, \ldots, V_M$, and an empirical distribution of the portfolio market value as in Fig. 4.5. In the case that there is only one systematic factor, or a few of them, it is computationally more efficient to sample first of all the systematic factors, and then the independent idiosyncratic factors generating r_i according to (4.2) for all the exposures.

The empirical distribution can then be used to estimate the mean $\hat{\mu} = \frac{1}{M}\sum_{i=1}^{M} V_i$, the standard deviation $\hat{\sigma} = \sqrt{\frac{1}{M-1}\sum_{i=1}^{M} (V_i - \hat{\mu})^2}$, and, for a given probability α, the α quantile as the $\alpha M-$ lowest value when the sequence $V_1, \ldots, V_M$ is sorted from the

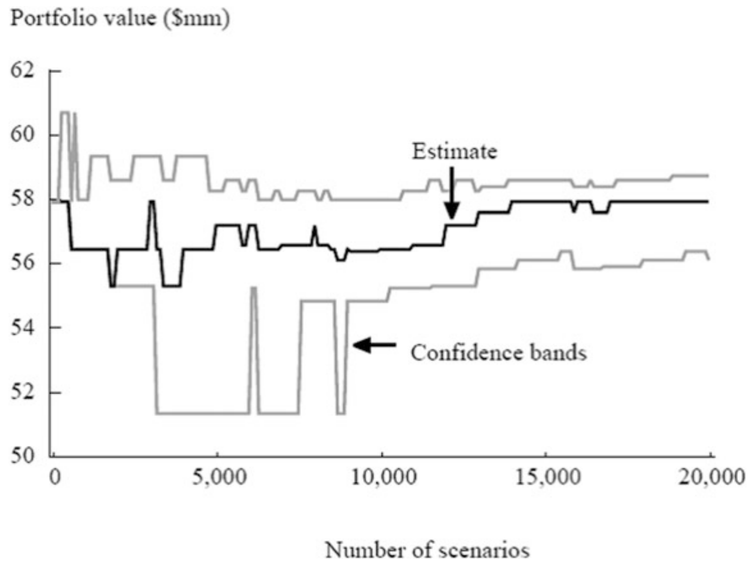


Fig. 4.6 Evolution of the 90% confidence bands for the 0.1 percentile (Source: JP Morgan 1997)

least to the largest value. When doing this, we also have to keep track of the confidence intervals of our estimations. Generally, the order of precision of a plain Monte Carlo simulation based estimation is $\frac{1}{\sqrt{M}}$, and so the naïve minimum number of simulations should be at least 10,000. Regarding the sample quantiles, the confidence intervals can be estimated using a principle similar to the asymptotic normal binomial test (see JP Morgan 1997). For example, if we wish to get the 90% confidence interval of the α quantile set:

$$i_1 = \left\lfloor \alpha M - 1.65\sqrt{M\alpha(1-\alpha)} \right\rfloor, \text{and}$$

$$i_2 = \left\lceil \alpha M + 1.65\sqrt{M\alpha(1-\alpha)} \right\rceil.$$

Then $[V_{i_1}, V_{i_2}]$ is the 90% confidence interval for q_α. For a very small probability level α, such as 0.1%, we need, first of all, enough scenarios, so that $i_1 > 0$, and, to achieve satisfactory precision, the number of scenarios must be quite large. See Fig. 4.6 as an example of the slow convergence of the 0.1 percentile estimate.

Distribution of Recovery Rates

So far, we have implicitly assumed that the market value assigned to a defaulted bond, i.e., its recovery rate, is deterministic. However, individual recovery rates are characterized not only by their mean expected values, but also by their wide

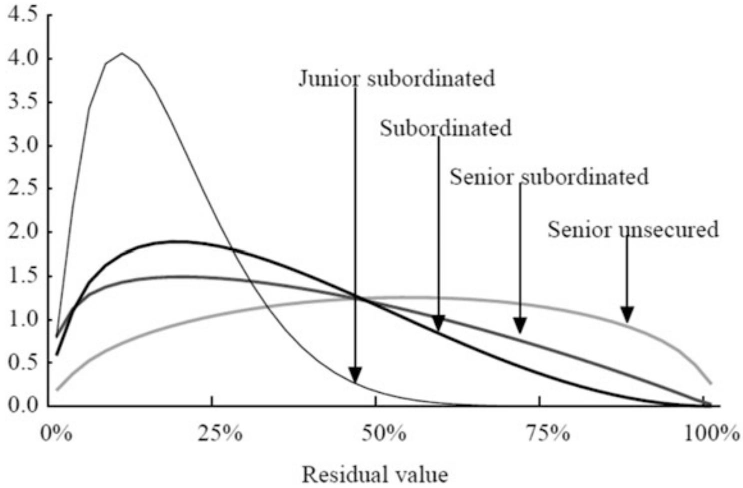


Fig. 4.7 Example beta distributions for seniority classes (Source: JP Morgan 1997)

uncertainty, as indicated by the large standard deviations shown in Table 4.2. One popular possibility, also proposed by the CreditMetrics document, is to model the recovery rate distribution by a beta distribution. Beta distribution is fully characterized by its minimum, maximum, mean, and standard deviations. Since recovery rates (at least in case of bonds) are normally between 0 and 100% of their face value, we can use the beta distributions with minimum 0, maximum 1, mean and standard deviations, according to Table 4.2. Then, for example, the distributions shown in Fig. 4.7 can be used in the Monte Carlo simulation to sample random recovery rates.

Default Rate and Recovery Rate Correlation

However, as it has already been pointed out, there is growing literature (see; e.g., Altman et al. 2002, and Fig. 4.8, or Altman et al. 2004 for an overview) showing that there is empirical evidence of the negative default rate and recovery rate correlation.

One possible approach to extending the CreditMetrics model, incorporating negative PD and RR correlation, is proposed in Witzany (2009d). Let us, for the sake of simplicity, consider the two-systematic-factor model, where defaults are driven by one systematic factor, and by idiosyncratic factors, while the recovery rates are driven by the default systematic factor, by an additional recovery systematic factor, and also by the recovery rate idiosyncratic factors. The idea is that the recovery rates are influenced, partially, by the same macroeconomic factors as default rates, but also by additional factors, for example, related to real estate collateral values. The mix of those two factors allows us to model

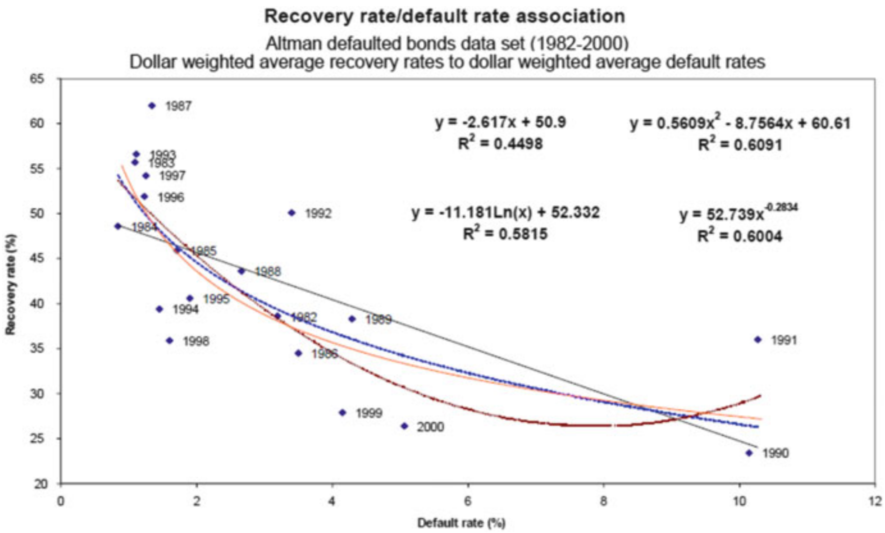


Fig. 4.8 Empirical evidence of negative PD and LGD correlation (Source: Altman et al. 2002)

different levels of PD and RR correlations. Formally, the two factors can be expressed as

$$r_i = \sqrt{\rho_1}X_1 + \sqrt{1-\rho_1}\epsilon_{i,1} \quad \text{and}$$
$$y_i = \sqrt{\rho_2}\left(\omega X_1 + \sqrt{1-\omega^2}X_2\right) + \sqrt{1-\rho_2}\epsilon_{i,2}, \tag{4.3}$$

where the systematic factors X_1, X_2, and the idiosyncratic factors $\epsilon_{i,1}, \epsilon_{i,2}$, are independent standardized normal variables. The mutual PD correlation ρ_1, and the mutual LGD correlation ρ_2, are assumed to be positive, but the correlation ω, between the PD systematic factor and the RR systematic factor, is allowed (and, in fact, expected) to be negative. The first factor r_i, drives the rating migration and default as above, while the second factor y_i, sampled in the case of default, drives the recovery rate. Since y_i has a standardized normal $N(0, 1)$ distribution, we need to know the individual recovery rate distribution. This can be either a beta distribution (Fig. 4.7), or any other parametric or empirically derived distribution. Let Q be the RR cumulative distribution function, then, given the factor y_i, applying the quantile-to-quantile transformation, the corresponding recovery rate is:

$$RR_i = Q^{-1}(\Phi(y_i)). \tag{4.4}$$

To simulate credit portfolio losses, first of all, draw the two systematic factors X_1 and X_2, then the idiosyncratic factors $\epsilon_{i,1}$, for all exposures in the portfolio, and finally $\epsilon_{i,2}$ for the exposures which default. The method can be easily generalized to incorporate more systematic PD factors replacing X_1, and possibly even more systematic factors, instead of X_2.

A reliable estimation of the correlation coefficient ω is certainly the difficult part of the model. Witzany (2009d) uses a maximum likelihood method to estimate $\hat{\omega} = -11.2\%$ on a large historical dataset of unsecured retail loans. The idea is to look at the time series of monthly observed default rates, and recovery rates, on the defaulted loans. Assuming that the portfolio is large, we may express, using the Vasicek formula (see Sect. 4.7), the observed portfolio PD as a function of the latent (time dependent) systematic factor X_1, and the observed portfolio RR, averaging (4.4) over the idiosyncratic factor, as a function of the combined latent factor $\omega X_1 + \sqrt{1-\omega^2}X_2$. The functions, at the same time, depend on the correlations ρ_1 and ρ_2, which can be estimated by the MLE method, together with the parameter ω. The article demonstrates that the impact of the estimated correlation to a 99.9% unexpected credit loss (economic capital), on a large asymptoticz portfolio, is almost a 30% increase of capital, compared to the zero PD–RR correlation model. Similar results in a simpler model were obtained by Altman et al. (2002).

4.3 CreditRisk+

CreditRisk+ is a methodology proposed by Credit Suisse (1997) involving an application of an actuarial science framework to the derivation of the loss distribution of a credit portfolio. A significant advantage of the approach, compared to CreditMetrics, is that all the calculations can be done analytically, without Monte Carlo simulations. The ingenious statistical technique based on probability generating functions is, however, quite complex, and can discourage users, who might tend to view it as a black-box methodology. However, the principles of the approach can be easily illustrated, and, in fact, applied within the framework of the Monte Carlo simulations.

First, let us consider a homogenous portfolio of N receivables of the same size, economic sector, and credit risk with an actual expected probability of default PD_0 in the 1 year horizon. The main difference, compared to CreditMetrics, is that the probability of default is not fixed, but is itself a stochastic random state variable, changing at the end of the modeling horizon to an unknown value PD_1, and the defaults are realized independently conditional on that PD_1 probability of default. The value itself becomes an indicator of the overall macroeconomic situation. The model is called “reduced-form”, since it does not try to capture any of the internal mechanics of defaults. On the other hand, CreditMetrics belongs, rather, to the class of “structural models”, as it is based on Merton’s approach, where the stochastic asset value falling below that of the liabilities explains the event of default. The CreditRisk+, like a simulation in the simplified framework, can be performed in the following two steps:

1. Simulate the future probability of default PD_1 based on the actual predicted value PD_0, and on an appropriate probability distribution (e.g., Gamma), calibrated to the observed or predicted variation of historical default rates.
2. Simulate the number of defaults in the portfolio of N receivables, assuming the probability of default is PD_1 and the default events are independent. The precise distribution would be the binomial one, but for the sake of an analytical solution, the Poisson distribution approximating well the binomial one for large values of N and low PD_1 can be used. The number of defaults multiplied by a fixed expected loss given default (LGD) parameter is, then, the final credit loss given by one simulation run.

Note that the model based only on independent defaults, generated with a fixed ex ante probability PD_0 on a portfolio of N receivables, would be completely unrealistic, as the standard deviation of the simulated portfolio default rate would be $\frac{PD_0}{\sqrt{N}}$; i.e., almost zero if N is large, contrary to the empirical observation of annual default rates fluctuating in an order of the average default rate (see Table 4.3). Thus, step 1 simulating the PD itself is of key importance.

The analytic solution is based on the concept of probability generating functions. Let X be a random variable attaining nonnegative integer values; i.e. $X \in \{0, 1, 2, 3, \ldots\}$, with probabilities $p_n = \Pr[X = n]$. The corresponding probability generating function is formally defined as

$$G_X(z) = \sum_{n=0}^{\infty} p_n z^n.$$

One of the most important tricks is that the sum of the two independent variables $X + Y$ corresponds to the multiplication of the respective probability generating functions:

$$G_X(z) \cdot G_Y(z) = \left(\sum_{n=0}^{\infty} p_n z^n\right) \times \left(\sum_{n=0}^{\infty} q_n z^n\right) = \sum_{n=0}^{\infty} \left(\sum_{i=0}^{n} p_i q_{n-i}\right) z^n = G_{X+Y}(z).$$

If the two probability generating functions have simple analytical forms, then the same is true for $G_{X+Y}(z)$ and the coefficients, i.e., probabilities, can be obtained from the infinite Taylor's expansion. This is, in particular, the case of the Poisson distribution, with a nice generating function in the exponential form:

$$G_X(z) = e^{\mu(z-1)} = \sum_{n=0}^{\infty} e^{-\mu} \frac{\mu^n}{n!} z^n, \text{ i.e., } \Pr[X = n] = e^{-\mu} \frac{\mu^n}{n!}. \tag{4.5}$$

The Poisson distribution mean and variance are both equal to μ. It approximates well to the binomial distribution of a variable $X \in \{0, 1, 2, 3, \ldots, N\}$, with the mean $\mu = N \cdot p$ for a large N, and a small probability p. If X and Y are independent, with

Table 4.3 One-year default rates by rating, 1970–1995 (Source: Crouhy et al. 2000)

Credit rating	One-year default rate average (%)	Standard deviation (%)
Aaa	0.00	0.00
Aa	0.03	0.10
A	0.01	0.00
Baa	0.13	0.30
Ba	1.42	1.30
B	7.62	5.10

the Poisson distribution means μ_1 and μ_2, then, $X + Y$ has the Poisson distribution with the mean $\mu_1 + \mu_2$, since $e^{\mu_1(z-1)}e^{\mu_2(z-1)} = e^{(\mu_1+\mu_2)(z-1)}$. This rule can be applied to generalize the approach for a portfolio of receivables, homogenous in terms of size and sector, but possibly with different levels of risk. In this case, just set $\mu_0 = \sum_{a \in A} PD_a$. Indeed, if the PDs are determined by rating grades, let μ_r be the expected number of defaults in the rating grade pool $r = 1, \ldots, R$, then the probability generating function of the total number of defaults has the form $G(z) = e^{\mu(z-1)}$ of the Poisson distribution with the mean

$$\mu = \sum_{r=1}^{R} \mu_r.$$

The Poisson distribution can, moreover, be analytically combined with the Gamma distribution $\Gamma(\alpha, \beta)$, with parameters α and β, defined by the density function:

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{x}{\beta}} x^{\alpha-1}, \text{ where } \Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx.$$

The shape and scale parameters α and β can be calculated from the mean μ and the standard deviation σ:

$$\alpha = \frac{\mu^2}{\sigma^2}, \quad \beta = \frac{\sigma^2}{\mu}. \tag{4.6}$$

Figure 4.9 shows, for illustration, the shape of the Gamma distribution, with the mean 100, and standard deviations set to 10, 50, and 90.

Let PD_0 be the initial (average) default probability, and σ_{PD} the estimated standard deviation of observed overall annual probabilities of default. Since we focus on the number of defaults in a given portfolio with N receivables, let us set $\mu_0 = N \times PD_0, \sigma_0 = N \times \sigma_{PD}$, calculate α and β according to (4.6), and sample the

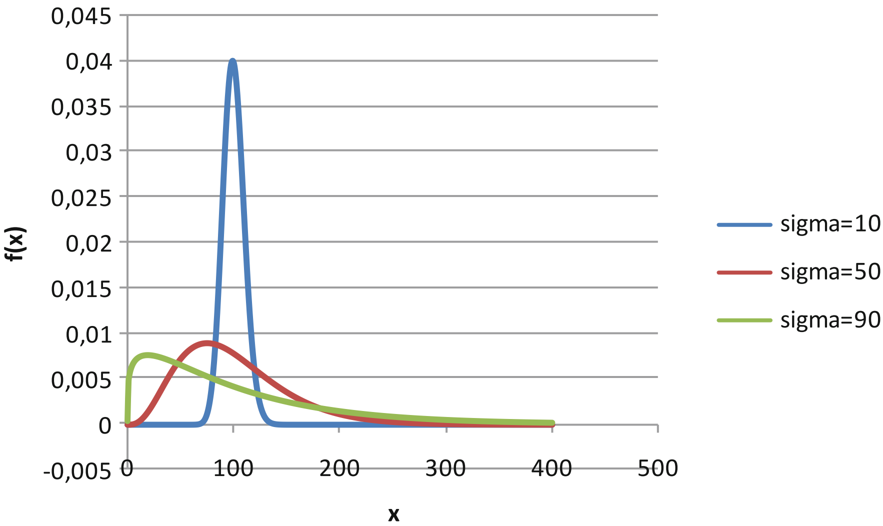


Fig. 4.9 Gamma distribution with mean 100 and various standard deviations

mean number of defaults μ_1 from $\Gamma(\alpha,\beta)$. Finally, generate the number of defaults X from the Poisson distribution with the mean μ_1. Since:

$$\Pr[X=n] = \int_0^\infty \Pr\left[X=n|\mu=x\right]f(x)dx = \int_0^\infty e^{-x}\frac{x^n}{n!}f(x)dx,$$

the probability generating function for X can be expressed as an integral that can, fortunately, be analytically solved:

$$\begin{aligned} G(z) &= \sum_{n=0}^{\infty}\left(\int_0^\infty e^{-x}\frac{x^n}{n!}f(x)dx\right)z^n = \int_0^\infty\left(\sum_{n=0}^{\infty}e^{-x}\frac{x^n}{n!}z^n\right)f(x)dx = \\ &= \int_0^\infty e^{x(z-1)}f(x)dx = \frac{1}{\beta^\alpha\left(1+\beta^{-1}-z\right)^\alpha}. \end{aligned} \tag{4.7}$$

The function on the right hand side can be differentiated at $z=0$, and expressed by the Taylor expansion. The coefficient of z^n is, finally, the desired formula for the probability of n defaults in the context of our elementary model, specifically:

$$\Pr[X=n] = (1-q)^\alpha\binom{n+\alpha-1}{n}q^n, \text{where } q=\frac{\beta}{1+\beta}.$$

This result over $n=0,\ldots,N$ can be identified as the probability density function of the Negative Binomial distribution.

The method above needs to be generalized in two directions: firstly we need to allow different exposure sizes, and secondly different economic sectors. In terms of a Monte Carlo simulation, the CreditRisk+ approach can be formulated as follows:

1. Adjust exposures with recovery rates that are considered to be deterministic. Moreover, split the adjusted exposures into size bands representing multiples of a large unit exposure. Divide expertly the portfolio according to the economic sectors (one exposure could also be split into more sectors). The actual mean number of defaults in a sector/size band portfolio is calculated as the sum of probabilities of default over all the exposures in the portfolio.
2. The sector portfolios are treated as independent.
3. Simulate the future mean number of defaults independently for each sector portfolio, according to an appropriate (Gamma) distribution. For each sector, divide the simulated number of defaults among the size bands, proportionately, according to the expected values.
4. Conditional on the sampled numbers of defaults in each sector portfolio and size band, generate the number of realized defaults treated as independent events (e.g., using the Poisson distribution).
5. Calculate the total loss (as a multiple of the basic exposure unit) over all sectors and size bands for each simulation scenario.

The description shows a major shortcoming of the model, which assumes independence between sectors. Empirically, there is a positive correlation between sector default rates, although lower than within the sector correlations. Moreover, the credit portfolio risk will be significantly reduced if the modeler chooses a fine sector classification compared to just a few sector classifications. This issue can be easily overcome in the Monte Carlo approach, admitting a correlation structure between the sector default rates. The non-zero correlation assumption, however, aborts the nice analytical solution, using the probability generating functions.

Therefore, assuming sector default rates independence, analytically, all we need to solve is the combination of different size bands within a sector portfolio. Once we have sector generating functions, the overall portfolio probability generating function is obtained just as their product. Let us fix a base exposure amount L, and assume that every (expected recovery rate adjusted) exposure in the portfolio is a multiple $m \times L,\ \ m \in \{1, \ldots, M\}$. It means that, in practice, the adjusted exposures must be rounded to integer multiples of L. A credit portfolio loss will be also always an integer multiple of L, and so we seek to find the probability generating function:

$$G(z) = \sum_{n=0}^{\infty} \Pr[\text{portfolio loss} = n \times L] z^n.$$

Let us consider a sub-portfolio of exposures with the size $m_j \times L$, and assume that μ_j is the number of expected defaults in the band $j = 1, \ldots, k$. In an analogy to (4.5), the loss generating function corresponding to the Poisson distribution number of defaults, with the mean μ_j, can be expressed as:

$$G_j(z) = e^{\mu_j(z^{m_j}-1)} = \sum_{n=0}^{\infty} e^{-\mu_j} \frac{\mu_j^n}{n!} z^{m_j n}.$$

Since defaults conditional on the mean number of defaults μ_j, $j = 1, \ldots, k$ are independent, the sector portfolio loss generating function is:

$$G(z) = \prod_{j=1}^{k} G_j(z) = \prod_{j=1}^{k} e^{\mu_j(z^{m_j}-1)} = e^{\sum \mu_j + \sum \mu_j z^{m_j}} = e^{\mu(P(z)-1)},$$

setting $\mu = \sum_{j=1}^{k} \mu_j$ and $P(z) = \sum_{j=1}^{k} \frac{\mu_j}{\mu} z^{m_j}$. Note that changing μ and all μ_j in the same proportion does not change the polynomial $P(z)$; in other words we implicitly assume that the default rate is the same for all exposure bands. Consequently, the procedure of sampling μ from a Gamma distribution with the mean $\mu_0 = N \times PD_0$, standard deviation $\sigma_0 = N \times \sigma_{PD}$, and adjusting μ_j proportionately as $\mu \frac{\mu_{0,j}}{\mu_0}$, corresponds, as in (4.7), to the integral:

$$G(z) = \int_0^{\infty} e^{x(P(z)-1)} f(x) dx = \frac{1}{\beta^{\alpha}\left(1 + \beta^{-1} - P(z)\right)^{\alpha}}. \tag{4.8}$$

If $G_s(z)$ denotes the probability generating function in the form (4.8) for sector $s = 1, \ldots, S$, then the overall probability generating function

$$G_{\text{final}}(z) = \prod_{s=1}^{S} G_s(z)$$

has an analytic form, and its coefficients can be obtained by differentiation at $z = 0$. The symbolic differentiation certainly is not simple, but computationally much more efficient than a full Monte Carlo simulation. The cost paid for this computational efficiency is that there is a limited correlation structure imposed by the model; in particular the assumption of independent sectors. The model can, however, be recommended for a homogenous retail portfolio that could be treated as a single sector, and where a full Credit Metrics debtor rating based simulation would not make too much sense.

Figure 4.10 shows an example of the loss distribution for a single sector portfolio of 10,000 recovery rate adjusted exposures with a total volume of 3 billion CZK and total expected loss 120 million CZK. The probability distribution, i.e. the probabilities $\Pr[\text{Loss} = \text{n} \times L]$ for $n = 0, \ldots, 60\ 000$, has been calculated

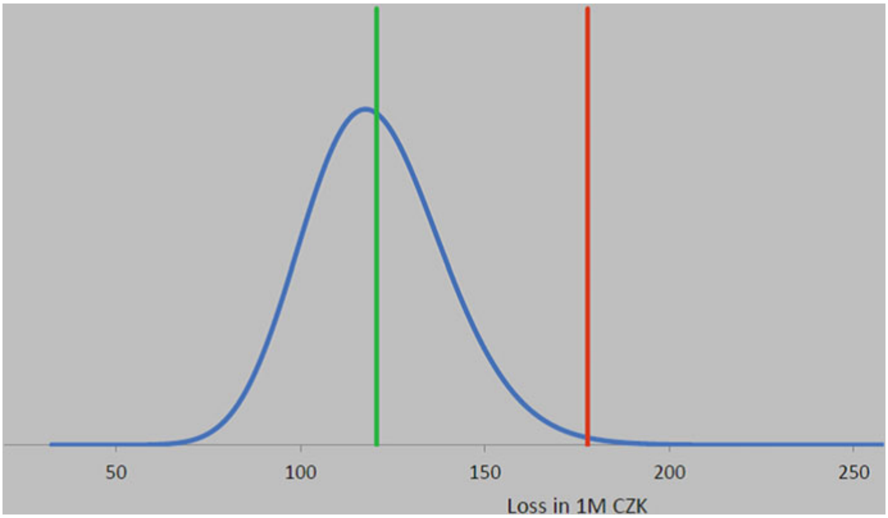


Fig. 4.10 Example of the loss distribution generated by the CreditRisk+ model (the *green line* indicates the expected loss and the red line the 99% quantile of the distribution)

implementing the CreditRisk+ model with $L = 50\,000$ implemented in Excel in a few seconds. The distribution allows us to find quantiles easily on any probability level. For example, the 99% quantile indicated by the red line is 178 million CZK. Therefore, the unexpected loss on the 99% probability level (Credit VaR) equals $178 - 120 = 58$ million CZK. The distribution can be also used to easily calculate other characteristics such as the *expected shortfall,* or *conditional Value at Risk,* defined as the average loss exceeding the quantile on a probability level α. Formally, $CVaR_{\alpha}^{\text{rel}} = E\left[X \middle| X \geq q_{\alpha}^{X}\right] - E[X]$ where X denotes the loss. For example, the shortfall of the sample portfolio on the 99% level is 67 million CZK, which is slightly higher than the Value at Risk.

4.4 CreditPortfolioView

The CreditPortfolioView model, developed by Wilson (1997a, b) and proposed by the McKinsey Consulting Company, is a macroeconomic multi-factor model for joint conditional distributions of default and migration probabilities. While the CreditRisk+ uses a very simple estimation of the initial rating/sector probability of default $PD_{j,0}$ and its standard deviation, the CreditPortfolioView works with a model where macroeconomic variables explain the observed rates of defaults. The main idea is that the next period default rate can, to certain extent, be predicted from the known macroeconomic indicators, and what needs to be modeled is the residual risk; i.e., the difference between the default rate realized in the future and its prediction (see Fig. 4.11).

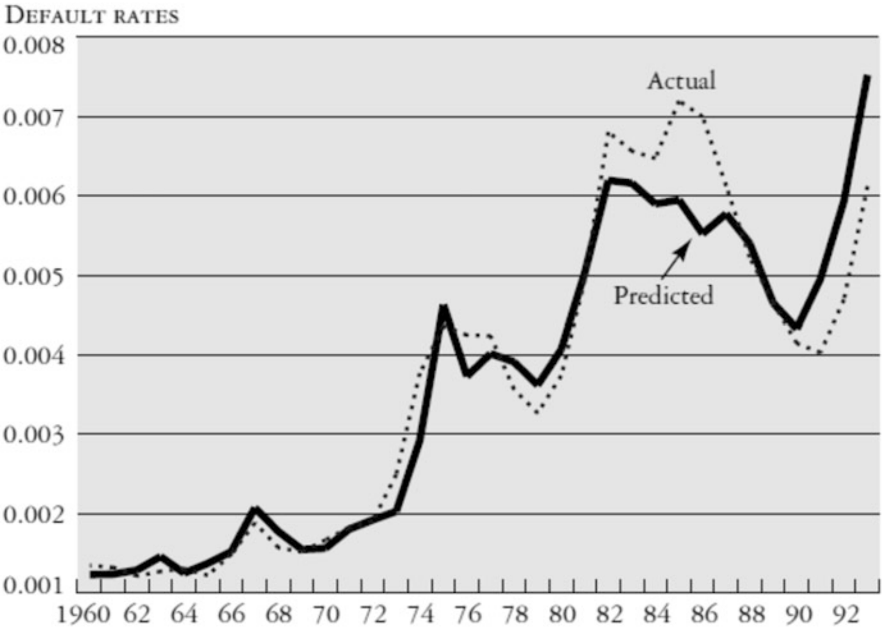


Fig. 4.11 Actual versus predicted default rates in Germany (Source: Wilson 1998)

Specifically, since macroeconomic variables follow a certain auto-regressive process, the model produces a distribution of future default rates, conditional on the actual, and recently observed, macroeconomic variables. The macroeconomic approach also improves the CreditRisk+ methodology, incorporating losses implied by rating deterioration, not just by defaults. On the other hand, the complex approach requires a Monte Carlo simulation.

The macroeconomic model is a relatively standard one, explaining a given sector j, and a speculative grade debtor's probability of default in the period t (e.g., 1 month or a quarter) as $PD_{j,t} = \Lambda(Y_{j,t})$, where Λ is the logit function, and $Y_{j,t}$ is a macroeconomic score capturing the state of the economy by the multi-factor model

$$Y_{j,t} = \boldsymbol{\beta}_j{}'\mathbf{X}_{j,t} + \epsilon_{j,t}; \tag{4.9}$$

where $\mathbf{X}_{j,t}$ are selected basic macroeconomic variables for the sector, and $\boldsymbol{\beta}_j$ are the regression coefficients to be estimated. In the proposed implementation, the macroeconomic variables, indexed by i, moreover, follow a univariate auto-regressive model of the order 2 (AR2):

$$X_{j,t,i} = \gamma_{j,i,0} + \gamma_{j,i,1}X_{j,t-1,i} + \gamma_{j,i,2}X_{j,t-2,i} + e_{j,t,i}. \quad (4.10)$$

The error terms $e_{j,t,i}$ are assumed to be independent and identically distributed with a normal distribution, while $\epsilon_{j,t}$ are assumed to have joint normal distribution characterized by a covariance matrix. Once the model has been calibrated, based on historical macroeconomic variables and observed rates of default, the assumptions can be used to simulate $PD_{j,t}, PD_{j,t+1}, \ldots, PD_{j,t+T}$ based on the information available at time $t-1$.

CreditPortfolioView, in addition, proposes adjustments to the unconditional Markov transition matrix M, based on historical averages covering several business cycles, across many different industries. The matrix implies certain unconditional probability of default PD_0 for speculative grade debtors. On the other hand, $PD_{j,t+s}$ is a PIT estimate, conditional on the state of economy, and of a specific industry, that is expected to be larger than PD_0 in economic recession, and less than PD_0 in economic expansion. The ratio $PD_{j,t+s}/PD_0$ can then be used to adjust the migration probabilities in order to produce the transition matrix:

$$M_{j,t+s} = M(PD_{j,t+s}/PD_0),$$

conditional on the state of the economy and the sector j. One possible simple approach is to set $M(PD_{j,t}/PD_0)$ equal to the recession historical data based matrix if $PD_{j,t+s} > PD_0$ and equal to the expansion matrix otherwise. The matrices for sectors $j = 1, \ldots, J$

$$M_{j,T} = \prod_{s=1}^{T} M(PD_{j,t+s}/PD_0)$$

are then used to generate, independently (conditional on the simulated macroeconomic development), rating migrations of debtors over the time horizon T. The simulated market values are then obtained as in the CreditMetrics model.

Note that the CreditPortfolioView presents a combination of the CreditRisk+ and the CreditMetrics models. In the first step, the probabilities of default are simulated, and in the second step, the independent defaults (rating transitions) are sampled, conditional on the simulated probabilities like in CreditRisk+. But the simulated probabilities are tied to macroeconomic variables, as well as being similarly tied to the asset values driving individual defaults in the CreditMetrics model.

4.5 KMV Portfolio Manager

The KMV portfolio model combines the basic principles of the CreditMetrics approach and the KMV Expected Default Frequency (EDF) methodology. The KMV EDF estimates the probabilities of default, and the market values, of risky debts from stock market data by applying the option pricing theory. The model does not use historical rating transition probabilities, or risk adjusted discount factors. The KMV argues that the actual (risk neutral) credit migration probabilities are much larger than those shown by historical migration frequencies. Indeed, as we have discussed in Sect. 3.3, the external rating agencies' systems are somewhere between PIT and TTC, and so the rating grade probabilities of default are not fixed, and changes of risk are not always reflected in the rating migration as assumed by CreditMetrics. The disadvantage of the KMV approach is that it is appropriate only for a portfolio of loans, or bonds, of publicly listed corporations, not for retail or SME portfolios, where stock market data are not usually available.

Similarly to CreditMetrics, we assume that a debtor's default is driven by the asset value A, following the geometric Brownian motion stochastic differential equation (see also Sect. 4.2):

$$dA = \mu_A A dt + \sigma_A A dW.$$

In a discrete setting, a path of A of the process can be simulated step by step:

$$A(0) = A_0, \quad A(t + \Delta t) = A(t)(1 + \mu_A \Delta t + \sigma_A \epsilon(t)) \tag{4.11}$$

where $\epsilon(t) \sim N(0, \Delta t)$. Since the valuation of derivatives may be done in the risk-neutral world, with investors requiring the risk free return r on any investment, regardless of risk, we may assume that the drift is $\mu_A = r$ and discount at the same time all the cash flows with r. To determine the actual market value of a risky claim paying the notional N at time T, if the assets value ends up above a default threshold $A(T) \geq K$, and paying just the deterministic recovery rate $(1 - LGD)N$ in the case of default, i.e. if $A(T) < K$, all we need to know is the initial asset value A_0 and the volatility σ_A. Therefore, the value of the risky claim is a function of the two parameters:

$$P(0) = f(A_0, \sigma_A).$$

The asset values are not directly observable, but what we may observe is the corporate equity price E_0 and its volatility σ_V.

As explained below, in detail, there is a one-to-one functional relationship between the pairs (A_0, σ_A) and (E_0, σ_E), so consequently, the unknown asset value parameters can be obtained from the observed equity price parameters:

$$A_0 = h_1(E_0, \sigma_E) \quad \text{and} \quad \sigma_A = h_2(E_0, \sigma_E). \tag{4.12}$$

Our goal is, however, to simulate, or determine, the distribution of the market values of claims in the given portfolio in a time horizon $H < T$. In a simulation approach for a single debtor starting from known parameters (A_0, σ_A), we may generate, according to (4.11) (or just sample, using the lognormal property of the geometric Brownian motion; see e.g. Hull 2009), the future asset values $A(H)$, and the corresponding future claim values given by:

$$P(H) = f(A(H), \sigma_A). \tag{4.13}$$

For a portfolio of multiple debtors, we must incorporate a correlation structure of the asset returns similar to CreditMetrics. To recapitulate, there are the following basic steps:

1. Based on equity prices, determine for every debtor in the portfolio the initial asset value and the asset volatility given by (4.12).
2. Estimate correlations between debtor asset returns, as in Sect. 4.2.
3. Simulate many times the future asset values, evaluate the loan values according to (4.13), and obtain the simulated total credit portfolio value.

The KMV also proposes an analytic solution that is applicable under the assumption of a constant pair-wise correlation ρ, and other simplifying assumptions. The portfolio credit loss is then shown to have a normal inverse distribution, for which it is relatively easy to compute the required percentiles. In the case of more complex correlation modeling, the simulation technology is nevertheless needed.

Having reviewed the principles of the KMV credit portfolio modeling approach, let us look in detail at its most important elements.

Estimation of the Asset Value and the Asset Volatility

As explained in Sect. 4.2, if there is only one loan in the amount D, payable at time T, then the equity value at the loan maturity is $E(T) = \max(A(T) - D, 0)$; i.e., the payoff of the European call option with the exercise price D, and maturity T (see e.g. Hull 2009). If the asset value follows (in the risk neutral world) the standard geometric Brownian motion:

$$dA = rAdt + \sigma_A AdW, \tag{4.14}$$

then the call option at time t can be valued by the Black-Scholes formula:

$$E_0 = A_0\Phi(d_1) - De^{-r(T-t)}\Phi(d_2) \tag{4.15}$$

where

$$d_1 = \frac{\ln A_0/D + (r + \sigma_A^2/2)(T-t)}{\sigma_A\sqrt{T-t}} \text{ and } d_2 = d_1 - \sigma_A\sqrt{T-t}. \tag{4.16}$$

Generally, we have the relationship $E = G(A, t)$ fixing the parameters σ_A, r, and T. According to Ito's Lemma, E follows the process

$$dE = \left(\frac{\partial E}{\partial A} rA + \frac{\partial E}{\partial t} + \frac{1}{2} \frac{\partial^2 E}{\partial A^2} \sigma_A^2 A^2 \right) dt + \frac{\partial E}{\partial A} \sigma_A A dW,$$

consequently, the stock volatility can be solved from the equation $\sigma_E E_0 = \frac{\partial E}{\partial A} \sigma_A A_0$. Since it can be shown that $\frac{\partial E}{\partial A} = \Phi(d_1)$ we obtain:

$$\sigma_E = \frac{A_0}{E_0} \Phi(d_1) \sigma_A. \tag{4.17}$$

So indeed, (4.15) and (4.17) give a transformation $E_0 = f_1(A_0, \sigma_A)$, $\sigma_E = f_2(A_0, \sigma_A)$, which can be inverted, as the relationship can be shown to be one-to-one. Empirically, given the values E_0 and σ_E, obtained from historical stock price data, the two equations are solved for A_0 and σ_A by an iterative numerical method.

Distance to Default and the Expected Default Frequency

If the borrower had only one loan payable at time T in the amount D, then the risk neutral probability of default could be expressed as $Q_T = \Pr[A(T) < D] = \Phi(-d_2)$, where d_2 is given by (4.16). However, in practice, the situation is different. A typical firm has many obligations, short-term, medium-term, and long-term loans payable in installments at different times. Default is the event whereby the firm misses a payment on a coupon, and/or reimbursement of the principal. Cross default clauses in debt contracts are such that when the firm misses a single payment, it is declared in default on all its obligations. If bankruptcy is proclaimed, the firm is liquidated, and the proceeds from the sale of the assets are distributed among creditors according to priority rules.

For all these reasons it is difficult to define exactly a general default threshold that would correspond to the real world event of default. Instead, KMV proposes to use an expertly set default point, defined as the firm short-term debt (STD), plus one half of the long-term debt (LTD), classifying all financial obligations into these two categories; i.e., $DPT = STD + LTD/2$. The default point, and the probability of breaching the threshold $\Pr[A(1) < DPT] = \Phi(-d_2)$, calculated in the 1 year horizon d_2, is given by (4.16) with $D = DPT$ (see Fig. 4.12).

As explained above, the probability $\Pr[A(1) < DPT]$ is not identical with the real probability of default, but it is, clearly, a good indicator of credit quality in the context of the structural model. The idea is to use the probability $\Phi(-d_2)$, or for practical reasons, the indicator $DD = d_2$, called the Distance to Default, as a score that still needs to be calibrated to empirical default rates (note that $\Phi(-DD)$ is a decreasing function of DD). Indeed, the KMV uses a database of a large sample of firms to map the DD, for each time horizon, to the calibrated probabilities of default, called the Expected Default Frequencies—*EDF* (see Fig. 4.13).

The figure shows default frequencies in the 1-year horizon, for example, based on the historical observations of the 1-year default rate of firms with $DD = 4$ is

Fig. 4.12 Probability distribution of assets value A (1) at $T = 1$, and the distance to default DD

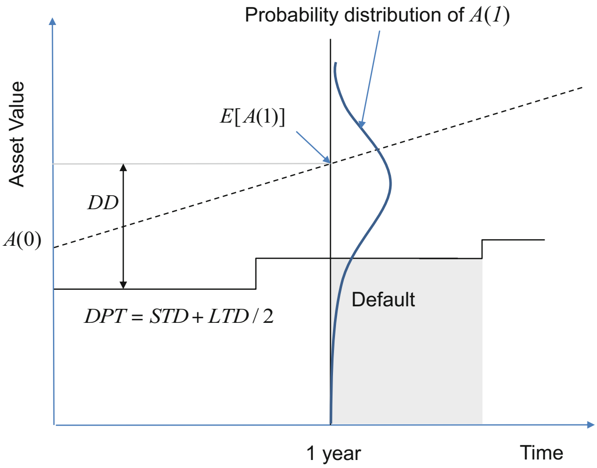
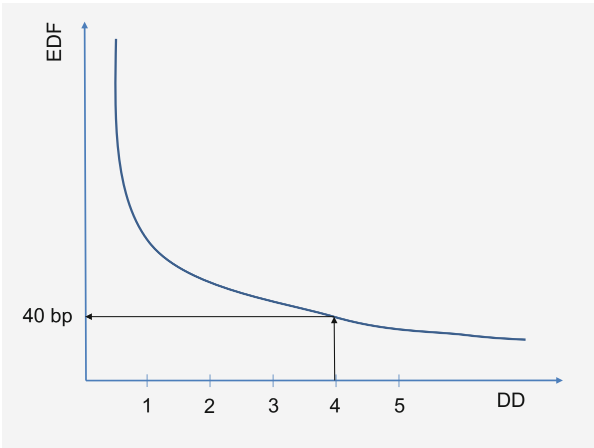


Fig. 4.13 Mapping of the distance to default to the expected default frequencies



0.4%, and so the assigned Expected Default Frequency is $EDF_1 = 0.4\%$. In this way, the expected probability of default, $EDF_T(DD)$, is a function of DD and the time horizon T.

The KMV company has been providing the service of "Credit Monitor", calculating the estimated EDF values since 1993. A number of studies have shown that the EDF is a useful leading indicator of default. Changes in EDF also tend to anticipate, at least 1 year ahead, the typical downgrading dynamics of issuers by rating agencies as illustrated in Fig. 4.14.

Indeed, the EDFs derived from stock market data are much more reactive to changes in firms' financial, and overall economic, situations. The KMV has also constructed the EDF implied (S&P scale) ratings based on non-overlapping ranges of default probabilities that are typical for individual rating classes. For example, the firms with an annual EDF of less than, or equal to, 2 bps are ranked

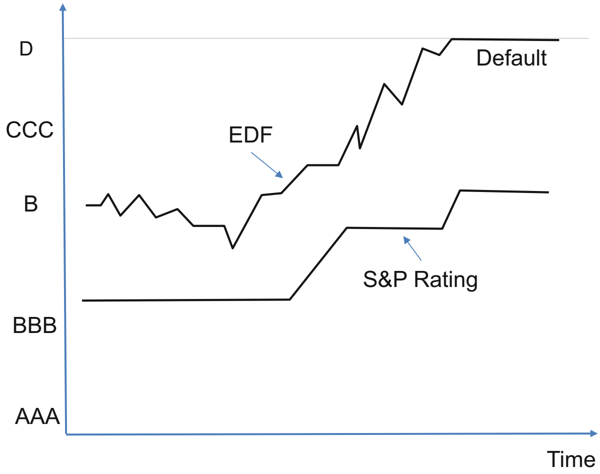


Fig. 4.14 EDF of a firm versus S&P rating

AAA, firms with EDF between 3 and 6 bps are AA, and so on. The time series of EDF ratings can then be used to produce a transition matrix shown in Table 4.4. There is a striking difference when the matrix is compared to the transition probabilities based on actual S&P rating changes in Table 4.5. The probabilities of the KMV rating staying in the same class are much lower than in the case of the S&P probabilities, while the migration probabilities are much larger. On the other hand, the speculative grade probabilities of default are larger in the case of the S&P transition matrix. The phenomenon can be explained by a slower reaction of S&P ratings to actual credit quality changes, and also by the fact that the probabilities of default are not tied to the S&P ratings and partially fluctuate with the cycle. The objection can be raised that the KMV rating migrations are overestimated, being, as they are, driven by the stock market, where the prices may also fluctuate up and down, due to psychological and other non-fundamental reasons.

Risk Neutral Probabilities of Default and Valuation of Contingent Claims
In calibrating the Distance to Default to Expected Default Frequencies, based on historical probabilities of default, we have unfortunately departed from the principle of risk neutral probabilities of default. The EDF's are estimations of real world probabilities, yet we still need estimations of calibrated risk neutral probabilities of default that allow us to discount risky cash flows with the risk free interest rate.

Let us consider a zero coupon bond paying the full nominal A amount at maturity T if there is no default, and just the recovery rate $RR \times A = (1 - LGD) \times A$ if there is a default. If EDF_T is the real world probability of default at time T, then we know how to calculate the mean cash flow at time T, but we do not know precisely how to discount it to obtain its present (market) value, since it is a risky cash flow, and a

Table 4.4 KMV rating transition matrix (Source: Crouhy et al. 2000)

	Rating at year-end (%)							
Initial rating	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	66.26	22.22	7.37	2.45	0.86	0.67	0.14	0.02
AA	21.66	43.04	25.83	6.56	1.99	0.68	0.20	0.04
A	2.76	20.34	44.19	22.94	7.42	1.97	0.28	0.10
BBB	0.30	2.80	22.63	42.54	23.52	6.95	1.00	0.26
BB	0.08	0.24	3.69	22.93	44.41	24.53	3.41	0.71
B	0.01	0.05	0.39	3.48	20.47	53.00	20.58	2.01
CCC	0.00	0.01	0.09	0.26	1.79	17.77	69.94	10.13

Table 4.5 S&P rating transition matrix (Source: Crouhy et al. 2000)

	Rating at year-end (%)							
Initial rating	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	90.81	8.33	0.68	0.06	0.12	0	0	0
AA	0.70	90.65	7.79	0.64	0.06	0.14	0.02	0
A	0.09	2.27	91.05	5.52	0.74	0.26	0.01	0.06
BBB	0.02	0.33	5.95	86.93	5.30	1.17	0.12	0.18
BB	0.03	0.14	0.67	7.73	80.53	8.84	1.00	1.06
B	0	0.11	0.24	0.43	6.48	83.46	4.07	5.20
CCC	0.22	0	0.22	1.30	2.38	11.24	64.86	19.79

certain risk margin should be added to the standard risk free discount rate. On the other hand, if Q_T is the cumulative risk neutral probability of default in the time horizon T, then the mean cash flow in the risk neutral world can be discounted just with the risk free interest rate:

$$\begin{aligned} PV &= e^{-rT} E_Q[\text{cash flow}] = e^{-rT} A((1 - Q_T) + Q_T(1 - LGD)) \\ &= e^{-rT} A((1 - LGD) + (1 - Q_T)LGD) \end{aligned}$$

The same principle can be used for a general cash flow $[CF_1, \ldots, CF_n]$ representing a loan, bond or other instrument. If $Q_1, \ldots, Q_n$ are the cumulative risk neutral probabilities of default, and $r_1, \ldots, r_n$ the risk free interest rates in continuous compounding at the payment time horizons $T_1, \ldots, T_n$, then:

$$PV = E_Q[\text{discounted cash flow}] = \sum_{i=1}^{n} e^{-r_i T_i} E_Q[\text{Cash flow}_i] = \\ = \sum_{i=1}^{n} e^{-r_i T_i} CF_i(1 - LGD) + \sum_{i=1}^{n} e^{-r_i T_i}(1 - Q_i) CF_i LGD. \tag{4.18}$$

Example What is the value of a 3-year bond with a face value of \$100,000, which pays an annual coupon of 5%? Let us assume that the risk-free interest rate is 3% for all maturities, $LGD = 60\%$, and the cumulative risk neutral probabilities are given in Table 4.6.

Note that the present value in (4.18) is on the right hand side of the table, and broken down into the risk free part calculated in the column PV1 and the risk part evaluated in the column PV2. The final present value, 99.552% of the face value calculated in the risk neutral framework, should be identical to the market value evaluated using real world EDFs, and discounted with risk adjusted interest rates.□

What remains is to evaluate all debt instruments in our portfolio corresponding to different borrowers in order to transform the real world EDFs into risk neutral probabilities. Let us fix a borrower, time horizon T, and the estimated Expected Default Frequency EDF_T. The real world asset value process is governed by the stochastic differential equation:

$$dA^* = \mu A^* dt + \sigma A^* dW,$$

with the drift $\mu > r$ reflecting a positive price of risk. There is an exact default threshold K_T, so that $EDF_T = \Pr[A^*(T) < K_T]$. When the probabilities are transformed from the real into the risk neutral world, the present value of contingent cash flows is preserved; in particular the present value contingent on the asset value falling below the threshold K_T. Therefore, the risk neutral probability is $Q_T = \Pr[A(T) < K_T]$, with the same K_T and the asset value process given by the risk neutral process

$$dA = rAdt + \sigma AdW,$$

starting at the same initial value $A(0) = A^*(0) = A_0$. Both probabilities can be expressed using the appropriate Black-Scholes-Merton formulas, and compared:

$$EDF_T = N(-d_2), d_2 = \frac{\ln A_0/K_T + (\mu - \sigma^2/2)T}{\sigma\sqrt{T}},$$

$$Q_T = N(-d_2^*), \, d_2^* = \frac{\ln A_0/K_T + (r - \sigma^2/2)T}{\sigma\sqrt{T}}.$$

Table 4.6 Risk neutral present value calculation example

Time	CF_i	Q_i (%)	e^{-rT_i}	PV_1	PV_2	PV
1	5000.00	3.00	0.9704	1940.89	2824.00	4764.89
2	5000.00	6.50	0.9418	1883.53	2641.65	4525.18
3	105,000.00	9.90	0.9139	38,385.11	51,877.48	90,262.59
Total				42,209.53	57,343.12	99,552.65

Since $d_2 = d_2^* - (\mu - r)\sqrt{T/\sigma}$, we may eliminate the unknown default threshold K_T, and express the risk neutral probability Q_T in terms of the known Expected Default Frequency:

$$Q_T = N\left(N^{-1}(EDF_T) + \frac{\mu - r}{\sigma}\sqrt{T}\right). \tag{4.19}$$

The remaining unknown parameter is the risk adjustment $\mu - r$, which can be estimated according to the Capital Asset Pricing Model (CAPM); i.e., $\mu - r = \beta\pi$, where β is the borrower's asset beta with respect to the market, and $\pi = \mu_M - r$ is the risk premium of the market portfolio. Since the beta parameter can be calculated as $\beta = \rho\frac{\sigma}{\sigma_M}$, where ρ is the correlation of the asset return and the market portfolio returns, the adjustment in (4.19) can be expressed as:

$$\frac{\mu - r}{\sigma}\sqrt{T} = \rho\frac{\pi}{\sigma_M}\sqrt{T} = \rho U\sqrt{T} \tag{4.20}$$

with $U = \frac{\pi}{\sigma_M}$ being Sharpe's ratio. In practice, Sharpe's ratio can be calibrated from the bond market data, and the correlation ρ can be estimated from the stock market data.

Example Calculate the risk neutral probability of default if the Expected Default Frequency in the 2 year horizon is $EDF_2 = 5\%$. Assume that Sharpe's ratio of the S&P 500 index is $U = 0.406$, and the debtor stock versus S&P 500 returns correlation $\rho = 0.7$. Since the same correlation is applicable to the asset returns, we may calculate the risk adjustment according to (4.20) as $\rho U\sqrt{T} = 0.7 \times 0.406 \times \sqrt{2} = 0.402$. Finally, according to (4.19), we get

$$Q_2 = N\left(N^{-1}(0.05) + 0.402\right) = 0.107.$$

Therefore, the risk neutral probability $Q_2 = 10.7\%$ is more than twice the real world probability $EDF_2 = 5\%$.□

4.6 Comparison of the Models

The four presented models, i.e. CreditMetrics, CreditRisk+, CreditPortfolioView, and KMV Portfolio Manager, aim to estimate the same theoretical value defined (in Sect. 4.1) as the unexpected loss $UL = q_\alpha^X - E[X]$ on a portfolio of receivables, given a time horizon T and probability level α. Since the models use different mechanics to simulate future portfolio values and different data sources to calibrate the calculation, we should not be surprised by relatively different outcomes, if the models are applied to the same portfolio.

Recall that the CreditMetrics model requires rating transition probabilities, recovery rates, yield curves for different rating classes, and an estimation of mutual asset correlation matrix. The model is appropriate for portfolios of bonds, corporate loans, or even larger retail receivables. The CreditMetrics methodology is refined in the KMV Portfolio Manager where we need, instead of historical transition probabilities and the rating yield curves, historical stock market data and the capital structure for each debtor in the portfolio. Moreover, the methodology requires a mapping of EDF values to historical default probabilities. Such a database is maintained essentially only by the company KMV Moody's. Both models use a market valuation of receivables at the time T, the values are nevertheless obtained in slightly different ways. The KMV Portfolio Manager does not utilize historical rating transition probabilities, but implicitly generates transition probabilities based on the stock market price volatility. It is generally argued that the KMV implied transition probabilities are larger than the historical probabilities, as stock returns reflect changes in credit quality much more quickly than the external ratings.

While CreditMetrics is a universal model that can be applied to portfolios of different assets, KMV is limited only to debtors with liquid stocks traded on the markets, thus it is very difficult to be applied in countries with less developed equity markets. CreditMetrics could be used even on consumer loan or mortgage portfolios, but simulation of the rating migration on each single retail receivable might be computationally quite demanding. In that case an efficient solution is provided by the CreditRisk+ methodology that needs just individual ex-ante probabilities of default, historical annual default rate volatilities, recovery rates, and possibly a segmentation into economic sectors. The disadvantage of the model is that there is limited flexibility of the correlation structure. This is usually not an issue for the retail portfolios of small homogenous receivables.

Finally, we should mention the CreditPortfolioView Model, which tries to separate expected and unexpected macroeconomic development. The model therefore needs a historical series of appropriate macroeconomic variables as well as historical rates of default for different rating classes. The macroeconomic model then allows the simulation of future probabilities of default (depending on ratings). The defaults, or rather rating changes, are then simulated conditionally on those probabilities. The model is not frequently used in practice due to the difficult macroeconomic modeling, and an ad hoc procedure is used to define the conditional transition matrix and loss distribution.

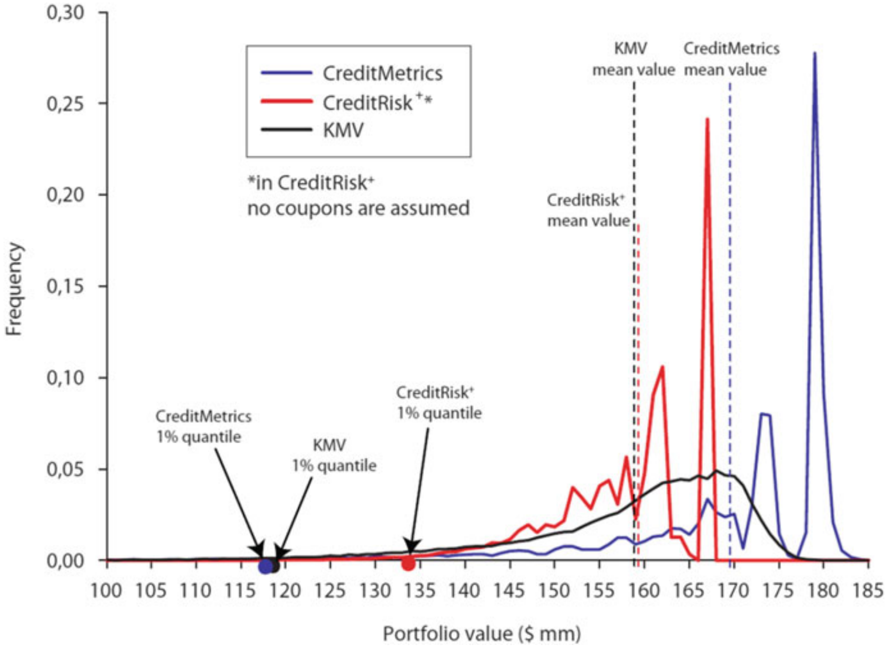


Fig. 4.15 Distribution of future portfolio values simulated by the three models (Source: Kolman 2010)

There is an interesting empirical study by Kolman (2010) comparing the results of the CreditMetrics, CreditRisk+, and KMV Portfolio Manager on a specific portfolio. The portfolio comprises 20 bonds of various issuers from different sectors in the U.S.A. The par value of the individual bond investments is around 10 million USD and total portfolio par value is 164 million USD. The bonds are rated by both Moody's and Fitch, with most of the ratings grades being speculative. Moreover, all the issuers have stocks traded on the markets and capital structure data publicly available. The CreditMetrics model is implemented based on Moody's historical transition probabilities, and correlations derived from the stock market data. CreditRisk+ is based on historical average default rate volatility, and just a single-sector approach is applied. In the case of the KMV model, besides the stock market data, the author uses a publicly available study mapping the EDF values to historical PDs (for details see Kolman 2010). Thus, all the models are applied on the same bond portfolio, use real historical and financial data, and simulate the portfolio loss in the same 1-year horizon. The empirical probability distributions shown in Fig. 4.15 and the relevant quantiles are, however, far from being identical.

The jumpy shape of the CreditRisk+ and CreditMetrics distributions is explained by the discrete character of the methodologies and by the fact that the portfolio is relatively small, having just 20 debtors. CreditRisk+ models just different numbers of defaults, while CreditMetrics is slightly finer, simulating all possible rating

Table 4.7 The unexpected losses estimated by the three models (Source: Kolman 2010)

	Mean value	1% quantile	99% quantile	Unexpected loss
CreditMetrics	169.141	118.069	180.755	45.931
CreditRiks+	158.867	134.000	168.000	24.867
KMV model	158.598	118.771	174.506	39.827

migrations, yet the largest losses are caused by downgrades of the largest bond exposures to the state of defaults. The KMV model, on the other hand, simulates all possible continuous changes of the market values of the individual bonds driven by the underlying asset value processes. Consequently the KMV probability distribution turns out to be much smoother. Table 4.7 shows the means of the portfolio simulated future market values, the 1% quantile, and the corresponding unexpected loss, i.e. 99% Credit Value at Risk.

While CreditMetrics and the KMV model estimated 99% unexpected losses (i.e. the 1% quantile minus the mean portfolio value) are relatively close, the CreditRisk+ unexpected loss is much smaller. This can be explained by the fact that CreditRisk+ uses the Poisson distribution to simulate the number of defaults, which works well for portfolios with much larger numbers of assets than 20. Moreover, CreditRisk+ does not capture rating migrations which may cause additional unexpected losses, or there might be an inconsistency between the historical default rate volatility and the historical transition matrix. Generally, we would rather expect the KMV estimate, compared to CreditMetrics, to be more conservative. The fact that the CreditMetrics 99% unexpected loss came out larger than the KMV result could be explained by the jumpy shape of the CreditMetrics distribution and the smooth character of the KMV distribution.

To conclude, the empirical study of Kolman (2010) shows that it is questionable to apply the CreditRisk+ model to a portfolio with a limited number of bonds or large corporate exposures. On the other hand, it was confirmed that the CreditMetrics and KMV models give similar results, if calibrated with consistent historical data. But even for these two models, the empirical study gave a relative difference of approximately 10% between the estimated values. Thus, the credit risk analysts implementing a credit portfolio model should not take the resulting unexpected loss estimate as the absolute truth. Optimally, more portfolio models should be implemented, and the outcomes in terms of loss distributions should be critically compared.

4.7 Vasicek's Model and the Basel Capital Calculation

The Basel I capital requirement (BCBS 1988) was based on a simple credit risk capital calculation, set as 8% of the assets multiplied by risk weights from the range 0, 20, 50, and 100%. Domestic, OECD governments', and central banks' exposures received 0%, OECD commercial banks 20%, and residential mortgages 50%, while all corporate, business, and consumer loans received 100%. There was some differentiation according to the riskiness of the exposures, but in the most important corporate, business, and retail segments, the 100% weight did not differentiate the risk at all. The guiding principle of the new Basel II regulation was to improve the credit risk calculation formula that should provide a reasonable estimate of the unexpected loss in the context: Credit Value at Risk (see Fig. 2.8).

In the previous section we reviewed several portfolio credit risk models. It could be proposed to allow the banks to use their internal portfolio credit risk model, which would have to be approved by the regulator, as in the case of market risk. Nevertheless, the models use quite different approaches and, in fact, may lead to a wide range of results, as we have seen in Sect. 4.6. The models are very sensitive to asset correlation, rating transition matrices, and other parameters which are difficult to estimate. In this situation, when the credit risk modeling techniques are still being developed, BCBS has decided for a compromise, applying a simple approach based on the Vasicek (1987) model, which has some of the key elements of the more complex models described in the previous sections, but provides an analytic formula that needs just a few key input parameters (PD, LGD, EAD), and may be applied separately to each exposure.

Vasicek's Model

Let us consider a non-defaulted borrower j, and let T_j be the time to default on the borrower's exposure. It is assumed that everyone will default once, and as the time of the future event is unknown at present, the time $T_j < \infty$ is treated as a random variable. If Q_j denotes the cumulative probability distribution of T_j, then, it can be easily verified that the quantile-to-quantile transformed variable $X_j = \Phi^{-1}\left(Q_j\left(T_j\right)\right)$ has the standardized normal distribution. Note that default in the 1-year horizon happens if $T_j \leq 1$, or, equally, if, and only if, $X_j \leq \Phi^{-1}(PD)$, where $PD{=}Q_j(1)$ is the 1-year probability of default. The advantage of the transformation is that we can take the assumption that the variables are multivariate normal with a fixed mutual correlation ρ. The properties of normal variables can be used to obtain a nice analytical result. Effectively, we are applying the one-factor Gaussian copula model. If the mutual correlation is ρ, then each factor X_j can be broken down into one common systematic factor M, and an idiosyncratic; i.e., independent specific factor Z_j, both with the standard normal distribution:

$$X_j = \sqrt{\rho}M + \sqrt{1-\rho}Z_j. \tag{4.21}$$

Let us assume for a moment that the ex-ante probability of default PD is the same for all the exposures $j = 1, \ldots, J$ in a large portfolio. In a Monte Carlo approach, similar to CreditMetrics, we first of all generate on the portfolio level the systematic factor $m \sim N(0,1)$, and then for each exposure $j = 1, \ldots, J$, the idiosyncratic factor $z_j \sim N(0,1)$. Since J is large, the rate of default; i.e., the relative number of cases where $x_j = \sqrt{\rho}m + \sqrt{1-\rho}z_j \le \Phi^{-1}(PD)$ approaches, by the law of large numbers, the conditional default probability:

$$\Pr\left[\sqrt{\rho}M + \sqrt{1-\rho}Z_j \le \Phi^{-1}(PD)\middle|M = m\right] = \Pr\left[T_j \le 1\middle|M = m\right] = PD_1(m);$$

i.e., the rate of default PD_1, conditional on the systematic factor $M = m$. Note that the events of defaults conditional upon $M = m$ are independent, and so the standard deviation of the simulated rate of default is of the order $PD_1/\sqrt{J}$. In Vasicek's model this variance is neglected, assuming that the portfolio is asymptotic; i.e., J infinite. Compare this to credit CreditRisk+, where we simulate, first of all, the overall PD, and then the independent events of defaults conditional upon the probability of default PD.

Finally, the conditional rate of default can be expressed as:

$$\begin{aligned} &PD_1(m) = \Pr\left[T_j \le 1\middle|M = m\right] = \Pr\left[X_j \le \Phi^{-1}(PD)\middle|M = m\right] = \\ &= \Pr\left[\sqrt{\rho}M + \sqrt{1-\rho}Z_j \le \Phi^{-1}(PD)\middle|M = m\right] = \Pr\left[Z_j \le \frac{\Phi^{-1}(PD) - \sqrt{\rho}m}{\sqrt{1-\rho}}\right] = \\ &= \Phi\left(\frac{\Phi^{-1}(PD) - \sqrt{\rho}m}{\sqrt{1-\rho}}\right) \end{aligned} \tag{4.22}$$

Since the conditional portfolio rate of default $PD_1(m)$ monotonically depends just on the one factor, $m \sim N(0,1)$, we can use the quantiles of m to obtain the unexpected default rate on any desired probability level α (e.g., 99%), setting $m = \Phi^{-1}(1-\alpha) = -\Phi^{-1}(\alpha)$. Thus, the unexpected default rate on the probability level α can be expressed as:

$$UDR_\alpha(PD) = \Phi\left(\frac{\Phi^{-1}(PD) + \sqrt{\rho}.\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right). \tag{4.23}$$

The same argument may be used on a portfolio of exposures with different probabilities of default, driven by a rating scale, assuming that for each rating grade the number of exposures is large. Since there is still just one systematic factor, the Vasicek's formula (4.23), applied on exposure level, gives an individual contribution to the overall unexpected default rate. Moreover, if it is multiplied by the Exposure at Default (EAD), and estimated Loss Given Default (LGD), we get an individual loan contribution to the overall portfolio unexpected loss.

Basel II/III Capital Formula, Its Advantages, and Disadvantages

Even though the first Consultative Papers contained proposals attempting to capture portfolio granularity; i.e., penalizing less diversified portfolios, the final version of the capital calculation formula is portfolio invariant based on (4.23), defining the percentage capital requirement (K), and the risk weighted assets independently of the portfolio context as

$$\begin{aligned} &RWA = EAD \times w, \\ &w = K \times 12.5, \\ &K = (UDR_{99.9\%}(PD) - PD) \times LGD \times MA. \end{aligned} \tag{4.24}$$

The banks applying IRB must estimate the PD parameter and, in the advanced approach, even the LGD and EAD parameters, using their internal models satisfying a number of qualitative requirements. Correlation, the key parameter in (4.23), is given by the regulation according to individual segments. For corporate, sovereign, and bank exposures, the correlation is set to be a weighted average between 0.12 and 0.24, depending on the PD:

$$\rho = 0.12\frac{1 - e^{-50PD}}{1 - e^{-50}} + 0.24\frac{e^{-50PD} - e^{-50}}{1 - e^{-50}}.$$

The correlation is slightly reduced for SME exposures, reflecting lower sizes and higher diversification. Similarly, for consumer loans, the correlation is a weighted average between 0.03 and 0.16, while for mortgages it is fixed at $\rho = 0.15$, and finally for revolving loans (e.g. credit cards) it is $\rho = 0.04$. The maturity adjustment is applied only to corporate, sovereign, and bank exposures, in an attempt to differentiate between shorter maturities with a lower risk, and longer maturities with a higher risk, by the following relatively complex econometric formula:

$$MA = \frac{1 + (M - 2.5)b}{1 - 1.5b}, b = (0.1182 - 0.05478 \ln PD)^2.$$

For retail receivables the adjustment is not used, or equivalently, $MA = 1$. The regulators have also decided to apply the high probability level $\alpha = 99.9\%$, which appears to be over-conservative. For example, the analogous insurance industry regulation, Solvency II, works with the probability level of 99.5%. Since the formula (4.23) assumes a perfectly diversified portfolio, the high probability level may be intended as a cushion, making up for an imperfect diversification not reflected in the model. Different levels of ordinary diversification are also clearly incorporated into the correlation coefficients, which are reduced for SME exposures, compared to corporate ones, and for revolving credit compared to consumer loans and mortgages, i.e. for exposures where we normally expect a well-diversified portfolio. In fact, the coefficients had to be calibrated to differentiate the risk and, at the same time, were based on a number of Quantitative Studies (see BCBS 2006b), so that the overall capital level in the system did not change significantly. Moreover, the intention of the regulators was to motivate banks to adopt the IRBF, or IRBA, approach and the higher credit risk management standards, and so the IRB capital requirements should be, under normal conditions, slightly below the Standardized Approach required capital calculated on an identical portfolio.

To summarize, compared to the Basel I RWA calculation, the Basel II formula provides much better risk sensitivity. Yet, it remains still relatively simple, requiring banks to estimate the key parameters without the need for any sophisticated simulation, or analytical, portfolio modeling. We can conclude that certain issues remain unresolved, and some new problems have arisen.

From the perspective of the models presented in Sects. 4.2–4.5, Vasicek's model makes significant simplifying assumptions, neglecting, in particular, the effect of low diversification due to a limited number of exposures, and the possibility of a more complex correlation structure. The regulatory correlations are necessarily calibrated in the spirit of "one size fits all."

Another issue is an underestimation of the unexpected recovery risk. As already discussed in Sect. 3.5, a number of studies have empirically shown (Altman et al. 2004) that recovery rates tend to be high when the economy is doing well, and low when the economy is down. Thus, there is also a negative PD x LGD correlation. It is demonstrated in Witzany (2009d) that the impact on economic capital might be quite significant. The regulation (§468 BCBS 2006a) tries to solve the problem, by requiring that LGD be estimated with a margin of conservatism, reflecting the impact of economic downturns. But the formulation remains vague, allowing different interpretations. It should be made more specific in the future. The precise modeling of unexpected default rates, and no, or limited, modeling of unexpected LGD has another surprising effect, as pointed out in Witzany (2009e). The banks have a certain, though limited, freedom, in defining the event of default. If the definition is too soft, flagging exposures as defaulted relatively early, then many may become 'cured'; i.e., debtors start to pay regularly again, and after some time are flagged as non-defaulted with a 100% observed recovery rate. The soft default definition causes the empirically observed PD to be higher, while the empirical LGD is lower (as many defaulted cases show zero loss). It turns out that this leads to a lower capital requirement, calculated according to (4.24) and (4.23), compared to a standard definition of default. The effect is caused by insufficient modeling of unexpected recovery risk and, on the other hand, by the Vasicek formula, which gives an unexpected default rate estimate that is proportionally lower for higher PD values (see Fig. 3.35).

Pro-cyclicality has become a rising concern, in particular, in the context of the recent crisis. A number of studies (Gordy and Howells 2006; Repullo et al. 2009), have pointed out that the Basel II capital requirement on an average portfolio is relatively low when the economy is doing well, and goes up when the economy is in a recession (see Fig. 4.16). The effect is caused by the general PIT property of internal rating systems, and the implied PDs as already explained in Sect. 3.6. Note that the pro-cyclical effect did not exist in the Basel I framework, with the simple structure of constant risk weights. The higher capital requirements discourage banks from providing more loans in a recession, and the low capital requirement encourages more loans in a period of economic expansion. Although this might be correct from the microeconomic point of view, it has a serious macroeconomic effect, worsening the crises and overheating the economy during expansions. There has been an ongoing discussion on how to "treat the disease without killing the

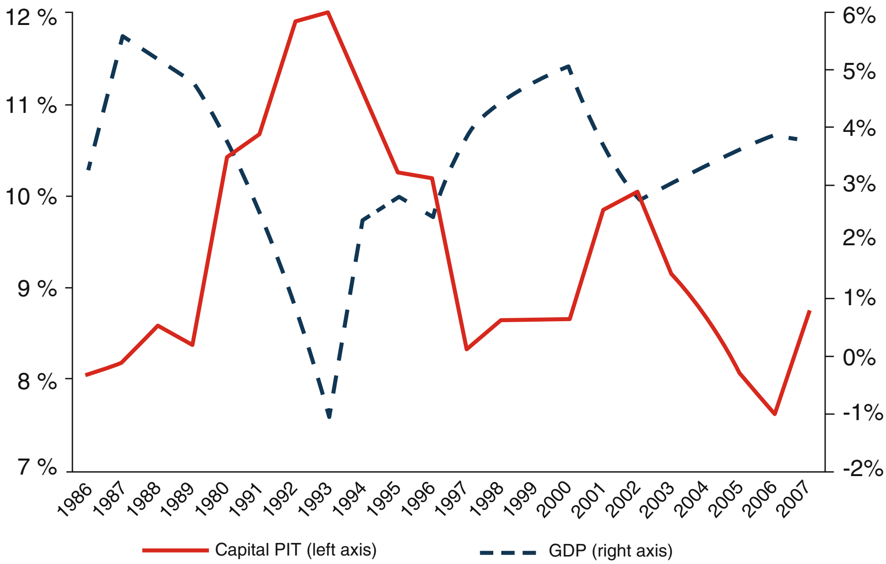


Fig. 4.16 Basel II capital requirement and GDP growth, Spain 1986–2007 (Source: Repullo et al. 2009)

patient," using the phrase of Gordy and Howells (2006) which has led to some conclusions partially incorporated into the Basel III reform (BCBS 2010). As explained in Sect. 2.3, Basel III did not change the capital requirement formula (4.24), but introduced capital conservation and countercyclical buffers in order to mitigate the Basel II pro-cyclical effect. Other Basel III key components are related to capital quality, the risk coverage of market products (see also Sect. 5.6), leverage, and liquidity risk management standards.

Economic Capital and Stress Testing According to Basel II/III

Even though the new regulatory capital calculation tries to approximate the concept of economic capital; i.e., unexpected credit portfolio loss; banks are still required under Pillar II (§725 Principle 1, BCBS 2006a) to "*have a process for assessing their overall capital adequacy in relation to their risk profile and a strategy for maintaining their capital levels*." That is optimally achieved by modeling overall credit, market, and operational risk, using the best available techniques and all available data. The results should be compared to the regulatory capital, as well as to the available capital, and additional capital charges may be required by the regulator if the latter appears too low. Banks are required, explicitly, under Pillar I and Pillar II, to perform scenario based stress testing.

Credit Derivatives and Counterparty Credit Risk 5

Financial derivatives are generally contracts whose financial payoffs depend on the prices of certain underlying assets. The contracts are traded Over the Counter (OTC), or in a standardized form on organized exchanges. The most popular derivative types are forwards, futures, options, and swaps. The underlying assets are, typically, interest rate instruments, stocks, foreign currencies, or commodities. The reasons for entering into a derivative contract might be hedging, speculation, or arbitrage. Compared to on-balance sheet instruments, derivatives allow investors and market participants to hedge their existing positions, or to enter into new exposures with no, or very low, initial investment. This is an advantage in the case of hedging, but at the same time, in the case of a speculation, a danger, since large risks could be taken too easily. Derivatives are sometimes compared to electricity; something that is very useful if properly used, but extremely dangerous if used irresponsibly. In spite of those warnings, the derivatives market has grown tremendously in recent decades, with OTC outstanding notional amounts exceeding 650 trillion USD, as of the end of 2014, and exchange traded derivatives' annual turnover exceeding 1450 trillion USD in 2014.

Similarly, credit derivatives are contracts with payoffs that depend on the creditworthiness of one, or more, counterparties. The creditworthiness is expressed either by binary information indicating default or no default, or by a credit rating, by a market credit spread, or by the market value of credit instruments, for example, corporate bonds. Currently, credit derivatives are traded on OTC markets only, although after the financial crisis, the new regulation (Dodd-Frank Act in the U.S. and EMIR—European Market Infrastructure Regulation) has introduced mandatory centralized clearing for the most important products (e.g., index CDS). As in the case of the classical derivatives, there is a multitude of special credit derivatives instruments, but we will focus only on a few of the most important: Credit Default Swaps (CDS), Total Return Swaps (TRS), Asset Backed Securities (ABS), Mortgage Backed Securities (MBS), and the Collateralized Debt Obligations (CDO). CDS and TRS can be compared to credit protection, or insurance contracts. Investors can buy or sell protection against defaults of one or more reference credit

J. Witzany, *Credit Risk Management*, DOI 10.1007/978-3-319-49800-3_5

entities. ABS, MBS, and CDO, on the other hand, allow financial institutions to pack, slice into different risk categories, and sell a portfolio of loans to other investors. The underlying assets of ABS are retail exposures like auto, credit card, equipment, housing related, student, and other loans, while MBS underlying assets are residential and commercial mortgages. As we explain later ABS and MBS are issued in tranches, and the tranches might be used as underlying assets of CDOs, which makes the analysis of CDOs even more difficult than in the case of ABS or MBS.

While classical derivatives have been actively traded for over 100 years, the trade in credit derivatives started relatively recently, in the late 1990s, with an exponential growth until 2007. Figure 5.1 shows that the trend has been reversed during the recent financial crisis, and yet the global outstanding notional amounts still exceeded 15 trillion USD at the end of 2014. The decline can be attributed partially to lower trading activity, but also to the effect of a centralized settlement allowing CDS positions to be closed out before maturity. The fast growth in credit derivatives trading, and in particular, of asset securitization, could, unfortunately, be blamed, to a certain extent, for the crisis.

The financial crisis has, in particular, had a significant impact on the development of global CDO issuance, as shown in Fig. 5.2. The sharp increase in CDO issuance went hand in hand with the strong growth of the subprime mortgage market in the U.S. until 2006–2007, and contributed to the creation of the real estate bubble. Other issues were the difficult valuation models, which were understood only by a few market participants, as well as too large dependence on rating agencies. The catastrophic implications of the financial crisis on some major market players almost caused a fear of CDOs, and a sharp decline of issuance to relatively negligible volumes (4.3 billion USD for 2009). A significant reduction of volumes can also be seen in the issuance of ABS and MBS (Fig. 5.3). Nevertheless, the most recent numbers indicate a partial recovery of the CDO as well as the ABS market.

The mechanics of CDS, TRS, ABS/MBS and CDO will be described in more detail in Sect. 5.1. The section also gives an introduction to elementary valuation principles. More advanced modeling and valuation techniques are then outlined in Sects. 5.2–5.4. and Sect. 5.5 discusses credit derivatives regulatiry issues.

Counterparty credit risk is the risk of loss on a derivative position due to default of the counterparty. The modeling of the risk incorporates both the task of the derivative pricing and the credit risk modeling. The topic has become recently, especially during and after the financial crisis, very important and so we focus on it in the last Sect. 5.6.

5.1 Credit Derivatives Market

Credit Default Swaps

The most popular credit derivative is a *credit default swap* (CDS). A single name CDS contract provides insurance (credit protection) against losses in the case of the

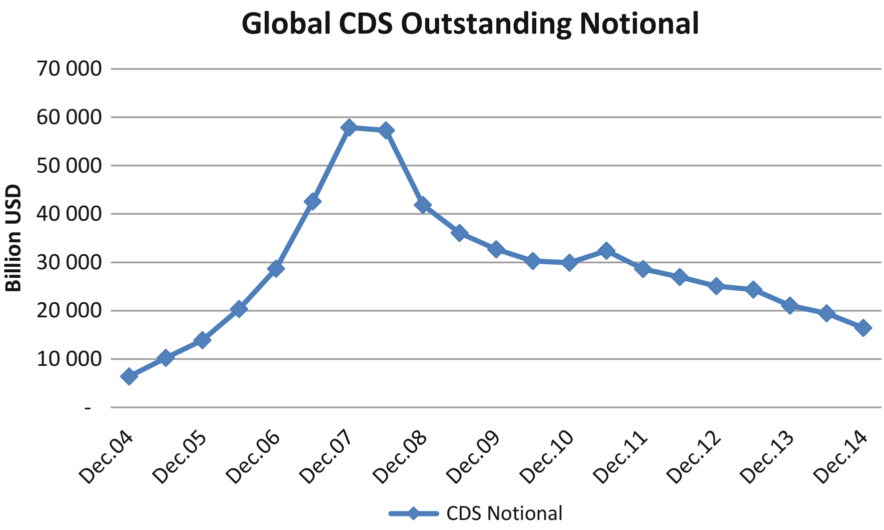


Fig. 5.1 Development of the global CDS market (Source: BIS Derivatives Statistics)

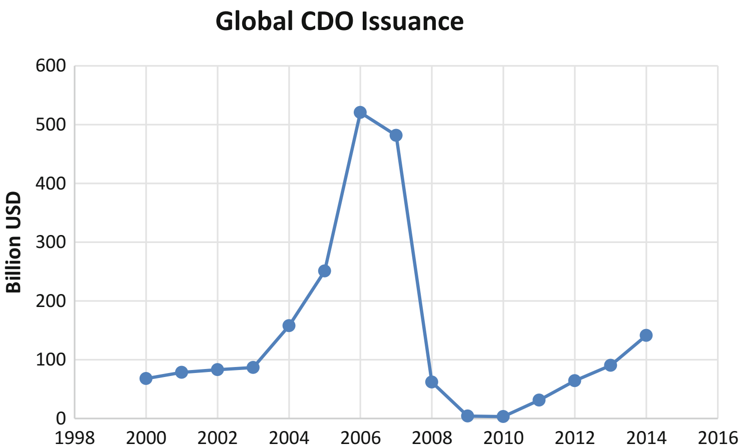


Fig. 5.2 Development of global CDO issuance (Source: SIFMA)

default of a specified subject, known as the *reference entity*. The buyer of the insurance makes periodic payments to the insurance seller and receives compensation in the case of default, called the *credit event*. In case of financial settlement, the compensation may be calculated as the loss on a notional amount, based on quotations collected by an independent agent. Another possibility is a physical settlement, which could be stipulated by the contract. This is an obligation of the insurance seller to buy the underlying bonds from the insurance buyer for their face value.

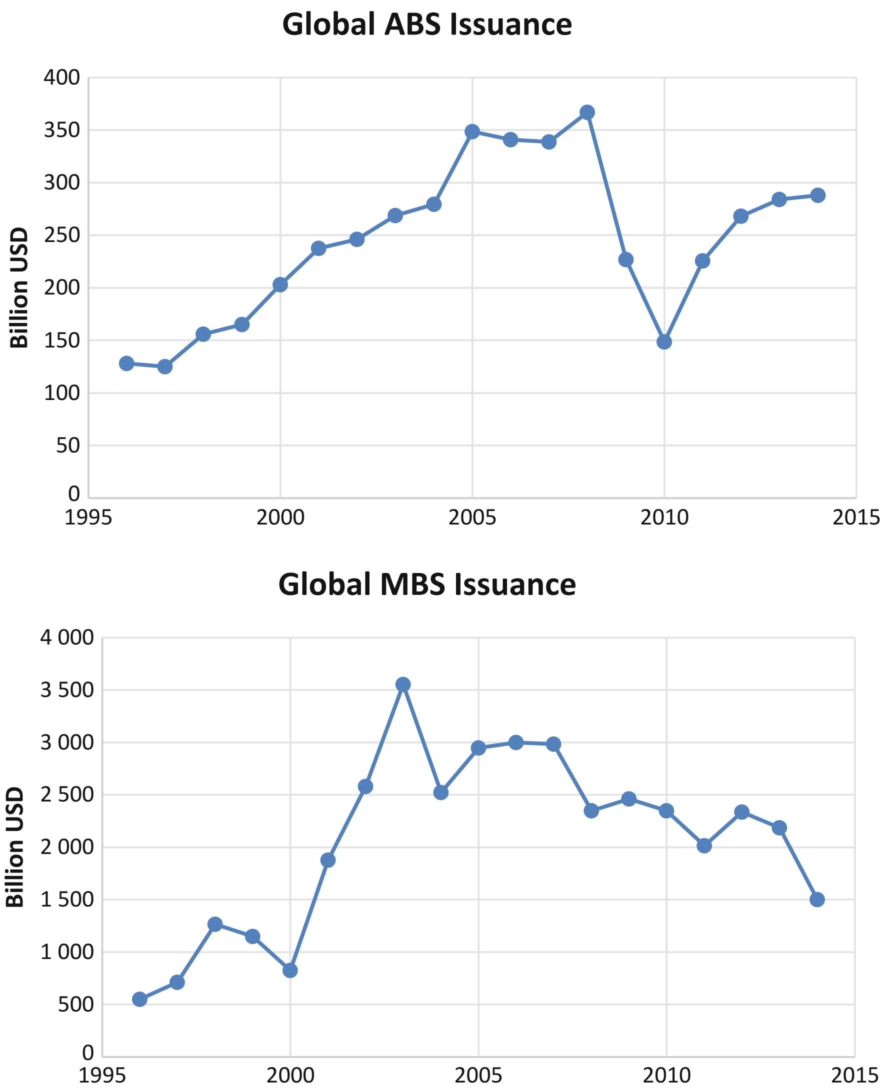


Fig. 5.3 Global (US + Europe) ABS and MBS issuance (Source: SIFMA)

Example Figure 5.4 shows an example of a CDS contract, where the credit protection buyer pays, annually, 200 basis points (i.e. 2%) on the notional amount of 10 million EUR (corresponding to 1000 bonds with 10,000 EUR face value), against the possibility of default during the next 5 years. Those were approximately the quotes on the Greek government debt CDS at the end of November 2009. Let us assume that the CDS starts on December 1, the credit spread is, as usual, payable in arrears on December 1 of 2010, 2011, 2012, 2013, and 2014; the reference entity is a country G, and the settlement, in the case of default, is in cash.

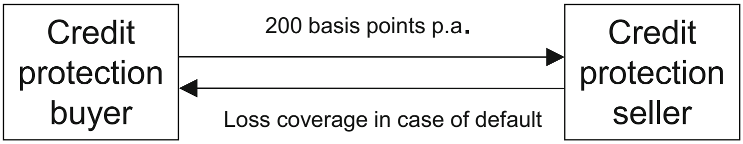


Fig. 5.4 Credit default swap example

If there was no default, then the credit protection seller receives $5 \times 0.2 = +1$ million EUR, and there is no payment to the credit protection buyer. Let us now assume that on June 1, 2013, the reference entity G defaults. The calculation agent, based on quotations usually 1 month after default, determines the loss at 45%. The credit protection seller then must pay 4.5 million EUR to the credit protection buyer to cover the loss. The credit protection buyer must still pay the accrued credit protection spread for the period December 1, 2012–June 1, 2013; i.e., 100,000 EUR, but the remaining credit spread payments are terminated due to the event of default. In this case, the net result is negative, i.e. $0.7 - 4.5 = -3.8$ million EUR from the perspective of the credit protection seller. □

The CDS contracts are not as innovative as is sometimes claimed. In fact, a CDS contract is essentially the same as a banking guarantee. The single-name and multi-name CDS also share a number of principles with reinsurance contracts. The main difference is that CDS contracts are, to a large extent, standardized through an ISDA (International Swap and Derivatives Association) framework documentation, allowing for fast OTC trading. Moreover, CDS contracts are not subject to insurance regulation.

Credit Spreads and Risk Neutral Probabilities

Risk neutral probabilities have already been introduced in Sect. 4.5 as a key theoretical framework for the valuation of credit event contingent claims. The relationship between risk neutral PDs, historical PDs, and the market values of bonds or CDS is outlined in Fig. 5.5.

We have seen in Sect. 4.5 that the risk neutral PDs could be derived directly from the historical PDs. However, it is more usual to derive risk neutral PDs from market bond prices if those are available. Corporate bond prices reflect not only expected losses due to a real probability of default, but, moreover, the uncertainty of the estimates, and the price of risk charged by the investors for the possibility of losses in the case of default. Remember that the risk-neutral valuation principle, based on applied stochastic calculus, says that given one, or more, sources of uncertainty, we can adjust the probabilities of all the scenarios so that the value of any derivative security today can be calculated as the expected future value, discounted by the risk free interest rate. Given the market value of a risky bond, and the risk-free interest rate, the risk neutral PDs can be relatively easily calculated. In turn, risk neutral PDs can be used to value different credit derivatives and to determine the equilibrium of CDS spreads.

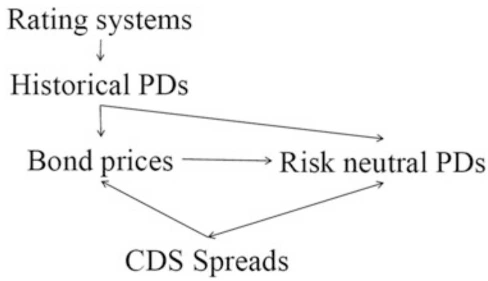


Fig. 5.5 Relationship between different sources of credit risk valuation

CDS spreads can be, however, directly obtained from bond yields using asset swaps. We are going to give an example of the relationship, before illustrating, in detail, the risk neutral PDs calculation.

Example Let us firstly assume that the market value of a 5 year corporate bond, with a 5% annual coupon, is 100%. If government bonds can be currently issued with a 4% coupon, representing the 5 year risk free interest rate, then the market 5 year CDS spread (with physical, or cash, settlement, not binary) on the same corporation should be close to $s = 1\%$. The argument is that the investment in the corporate bond, combined with the CDS protection paying s on the same notional amount, is equivalent to becoming a risk free investment. A small complication is that in the case of default, for instance after 3 years, the loss is covered by the protection seller, i.e. effectively 100% of the notional amount is paid back, and the coupon payments are terminated. Moreover, in practice, the bond market values deviate from the 100% par value. The second issue is resolved with an asset swap. Let us say that counterparty A buys the bond value for 94%. A corresponding asset swap with counterparty B is defined as follows: first the counterparty A pays B the 6% difference between the par and market value of the bond. A, moreover, pays the coupon 5% annually to B, and B pays Libor + s of the bond principal to A until maturity, independently (!) of whether the bond issuer defaults or not. From the perspective of A, the cash flow is transformed to a 100% principal investment into a corporate bond, with the same issuer paying Libor + s. Again, we can argue that the spread s must be (approximately) equal to the market CDS spread. □
Generally, risk neutral probabilities can be calculated from the valuation formula (4.18), which may be written in the simplified form:

$$P = \sum_{i=1}^{n} e^{-r_i T_i} CF_i \times (1 - LGD) + \sum_{i=1}^{n} e^{-r_i T_i} (1 - Q_i) \times CF_i \cdot LGD \qquad (5.1)$$

where Q_i is the cumulative risk neutral probability of default with maturity T_i, r_i the risk free interest rate with maturity T_i, CF_i the payment due at T_i, and LGD is the fixed loss given default ratio. Starting from the market values of fixed coupon

Table 5.1 Calculation of risk neutral probabilities from bond prices

Bond value	Coupon	Maturity	r_i (%)	Q (%)
101.00	3.50	1	2.00	1.11
102.50	5.00	2	3.00	3.20
102.00	5.00	3	3.50	5.45

bonds, maturing in 1, then 2, or more years, we can bootstrap the risk neutral probabilities for one, two, or longer maturities.

Example Table 5.1 shows the input market prices of bonds with different coupons, and maturities of 1, 2, and 3 years. The fourth column gives the risk free annual interest rates for the respective maturities. The risk-neutral cumulative probabilities of default in the last column are calculated as follows. The loss given default is assumed to be a fixed $LGD = 0.4$. The 1-year probability $Q_1 = 0.0111$ is calculated solving the Eq. (5.1) for the 1-year bond:

$$101 = e^{-0.02} \times 103.5 \times 0.6 + e^{-0.02} \cdot (1 - Q_1) \times 103.5 \times 0.4.$$

The 2 year probability Q_2 is obtained from the Eq. (5.1) for the 2 year bond, with Q_1 already known:

$$\begin{aligned}102.5 = e^{-0.02} \times 5 \times 0.6 + e^{-0.03} 105 \times 0.6 + e^{-0.02}(1 - 0.0111) \times 5 \times 0.4 \\ + e^{-0.03}(1 - Q_2) \times 105 \times 0.4.\end{aligned}$$

Similarly, we continue to get the 3 year probability Q_3 from the Eq. (4.18) for the 3 year bond based on the already calculated values for Q_1 and Q_2. □

The procedure illustrated above gives us a number of cumulative risk neutral probabilities of default for a discrete set of maturities. In order to value a general contingent cash flow, we need to interpolate, or extrapolate, the cumulative probabilities of default for all possible maturities. This can be consistently done using the concept of survival analysis outlined in Sect. 3.3. Recall the hazard rate, or rather *intensity of default*, $\lambda(t)$, that is defined as the annualized probability of default of the given entity over the time interval $[t, t + dt)$, provided there was no default until t. Thus, if $Q(t)$ is the cumulative probability of default, and $S(t) = 1 - Q(t)$ the corresponding cumulative probability of survival, then:

$$\lambda(t) = \frac{dQ(t)}{1 - Q(t)} \frac{1}{dt}, \text{ i.e. } \frac{dS(t)}{dt} = -\lambda(t) S(t),$$

and so

$$Q(t) = 1 - e^{-\int_0^t \lambda(s)ds}, \tag{5.2}$$

or

$$Q(t) = 1 - e^{-\overline{\lambda}(t)t} \tag{5.3}$$

where $\overline{\lambda}(t) = \frac{1}{t}\int_0^t \lambda(s)ds$ is the average default intensity between time 0 and time t.

It makes sense to assume that the default intensity is constant, independent of time (exponential model), or that the intensities follow a certain parametric, or nonparametric, pattern. So, given a set of calculated cumulative probabilities of default of an entity (or rating grade), e.g., $Q(1)$, $Q(2)$, $Q(3)$, ... we may calculate the corresponding average default intensities $\overline{\lambda}(1)$, $\overline{\lambda}(2)$, $\overline{\lambda}(3)$, ... from (5.3), and interpolate or extrapolate $\overline{\lambda}(t)$ between those values for a general t. Finally, $Q(t)$ is obtained again by (5.3). Alternatively, we may assume that the intensities are, piecewise, constant, e.g., that there is constant intensity of default λ_1 over the first year that can be calculated from (5.2) given $Q(1)$, λ_2 for the second year obtained from $Q(2)$ and λ_1, etc. The Eq. (5.3) again allows us to calculate $Q(t)$ for arbitrary maturity.

Example The resulting average intensities, and annual constant intensities of default based on cumulative risk-neutral probabilities of default, obtained in Table 5.1, are shown in Table 5.2. The solution of (5.3) for $\overline{\lambda}(t)$ is straightforward:

$$\overline{\lambda}(t) = \frac{-1}{t}\ln(1 - Q(t)).$$

Regarding constant intensities, note that if $\lambda_1 = \overline{\lambda}(1)$, then $\lambda_2 = -\ln(1 - Q(2)) - \lambda_1$, and so on.

To estimate the cumulative probability of default for $t = 2.5$, we may linearly interpolate the average intensity

$$\overline{\lambda}(2.5) = \left(\overline{\lambda}(2) + \overline{\lambda}(3)\right)/2 = 1.75\%,$$

and obtain $Q(2.5) = 4.27\%$. Alternatively, we could use (5.2) and the constant intensities of default, $Q(2.5) = 1 - e^{-\lambda_1 - \lambda_2 - 0.5\lambda_3} = 4.33\%$. The results are close, but not identical, since the first approach implicitly assumes a smooth (not piecewise constant) shape of the intensities of default. □

The average, or constant, annual intensities of default also allow us to analyze better the term structure of the probabilities of default. The average intensities could be put into an analogy with annualized interest rates in continuous compounding,

Table 5.2 Calculation of intensities from the cumulative probabilities of default

Maturity	Q (%)	$\overline{\lambda}$ (%)	Const. λ (%)
1	1.11	1.12	1.12
2	3.20	1.62	2.13
3	5.45	1.87	2.36

while the annual intensities are parallel to the forward annual interest rates implied by the term structure of the interest rates. For example, Table 5.2 shows that the risk neutral intensities of default expected by the market are increasing, which is a normal situation for higher rating grades.

The average risk neutral default intensity $\overline{\lambda}$ can be also estimated in a simplified way from the spread s of a bond over a yield of a similar risk free bond. Since the spread must cover the bond's average annual losses over its life, we may use the approximate equation $s=\overline{\lambda}\cdot LGD/\left(1-\overline{\lambda}\right)$ or $\overline{\lambda}=s/(LGD+s)$. Note that the approach neglects the effect of discounting. For example, if the spread is 200 basis points (i.e. 2%), and $LGD=0.4$, then $\overline{\lambda}=0.02/0.42=7.76\%$.

Hull (2009) compares historical default intensities and risk neutral intensities obtained from bonds of various rating classes. Table 5.3 shows that there is a dramatic difference between the historical intensities and risk-neutral intensities obtained from bond prices, in particular for the best ratings, where the risk-neutral intensities are 10, or more, times the historical frequencies. The phenomenon is explained by the existence of systematic non-diversifiable risk, which does have a price. If the economic capital, in the sense explained in Sect. 4.1, allocated to a bond were, for example 8%, and the required annual return on economic capital were 15%, then investors would have to add $8\%\times 15\%=1.2\%$ to the expected annualized losses and to the risk free yield. Moreover, corporate bonds are relatively illiquid, and so investors also require a premium over the yield on liquid government bonds. It could be also argued that the unexpected loss priced by the market is not only the systematic one caused by the positive correlation, but is also caused, partly, by the low diversification risk, due to the fact that full diversification is difficult, or impossible, to achieve on corporate bond markets.

Valuation of Single Name CDS

Valuation of single name CDS is relatively straightforward. One applies the formula (5.1), and a given term structure of risk neutral default probabilities $Q(t)$ for the CDS reference credit entity. Nevertheless, as for other OTC derivatives, we need to distinguish the valuation of an outstanding CDS position, and determination of the market CDS spread, which makes the market value of a new CDS contract equal to zero.

Table 5.3 Seven-year default intensities in % p.a. (Source: Hull 2009)

Rating	Historical default intensity	Default intensity from bonds	Ratio	Difference
Aaa	0.04	0.60	16.7	0.56
Aa	0.05	0.74	14.6	0.68
A	0.11	1.16	10.5	1.04
Baa	0.43	2.13	5.0	1.71
Ba	2.16	4.67	2.2	2.54
Caa and lower	13.07	18.16	1.4	5.5

Example Let us consider an outstanding CDS position, with 3 remaining years to maturity, where we pay 120 basis points on \$100 million notional, and the reference entity has the risk neutral default probabilities shown in Table 5.1. We make the simplifying assumption that defaults always happen halfway through a year. Therefore, the spread \$1.2 million is paid at the end of years 1, 2, and 3, provided the reference entity survives, and the payoff, in the amount of \$40 million (assuming $LGD = 0.4$), is paid at time 0.5, 1.5, or 2.5, if the reference entity defaults by the end of the year 1, 2, or 3. In the case of default, we also need to deduct the accrued semi-annual spread payment of \$0.6 million. For the payoff, we need to use the unconditional default probabilities for the respective years; i.e., $Q(1)$ for the first year, $Q(2) - Q(1)$ for the second, and $Q(3) - Q(2)$ for the third; since the swap is terminated by the default event. The cash flows, with the annual default and survival probabilities, and with the risk free rates used for discounting, are shown in Table 5.4. The resulting mark-to-market value of −\$1.21 million shows that the credit spread is too high under current market conditions; i.e. under market implied risk neutral probabilities of default. This can also be verified in a simplified way, comparing the annualized expected loss estimate

$$\bar{\lambda}(3) \times LGD = 1.87\% \times 0.4 = 0.75\%,$$

and the spread payment of 1.2%.

A slightly different task is to determine the spread s, which makes the market value of a 3 year CDS, entered into under current market conditions, equal to zero. Then, we must replace the fixed payment of \$1.2 million in Table 5.4, by the expression \$ $100s$ million, the payoff by \$$(40 - 50s)$ million, and express the expected present value as a function of the unknown variable s:

$$PV = 2{,}098 - 275.8 \times s$$

Finally, the equilibrium spread is the solution of the equation $PV = 0$; i.e., $s = 2.098/275.8 = 0.76\%$. It can be verified, by recalculating Table 5.4 with the 76 basis point CDS payment, that the resulting market value is then indeed almost zero. □

Table 5.4 Valuation of an outstanding CDS

Time	Q_i (%)	r_i (%)	CF_i	Probability (%)	Expected PV
0.5		2.00	39.40	1.11	0.43
1	1.11	2.00	−1.20	98.89	−1.16
1.5		2.50	39.40	2.09	0.79
2	3.20	3.00	−1.20	96.80	−1.09
2.5		2.17	39.40	2.26	0.84
3	5.45	3.50	−1.20	94.55	−1.02
Total					−1.21

The CDS valuation, shown above, starts with bond market values that are usually strongly determined by the external, or internal, ratings of the reference entity performed by rating agencies and other market players. However, once the CDS market becomes liquid, new credit information might be incorporated into market CDS prices much faster than to the updated ratings or bond prices. The CDS prices, then, in a sense, emancipate and become a primary source of the actual market view of the credit quality of a subject, which can be used to value other credit instruments like bonds or loans (see Fig. 5.5). For example, Fig. 5.6 shows the development of Ireland's, Greece's, Spain's, and Portugal's sovereign CDS spreads during the period 1/2009–12/2010, and it is the best characterization of recent dramatic developments in the Greek and other European government credit problems, as viewed by the financial markets.

Total Return Swaps

Total return swap (TRS) is a credit derivative contract that directly allows us to transform a risky investment into an almost risk-free cash investment paying a floating reference rate. It is an agreement between *a total return payer* and a *total return receiver* to exchange the total return on a risky bond (or other assets) for Libor (or another reference rate), plus certain spread. If there is a default on the bond, then the swap is usually terminated. By the total return, we understand not only coupons and interest, but also the change in market value of the asset over the life of the swap.

As an example, consider an investment of $95 million into 5-year bonds with 4% coupon purchased at 95% of their face value. The investor may enter into a corresponding 5-year total return swap (Fig. 5.7), exchanging the $2 million semi-annual coupons for Libor + 0.1% on the $100 million bonds' notional. At maturity, if there is no default, the investor, as the total return payer, will also pay the $5 million difference between the repaid bonds' face value ($100 million) and the initial $95 million market value. If there is a default, then the swap is terminated and the total return payer receives the difference between $95 million and the investment value at default, plus the accrued Libor + 0.1% interest. The result is

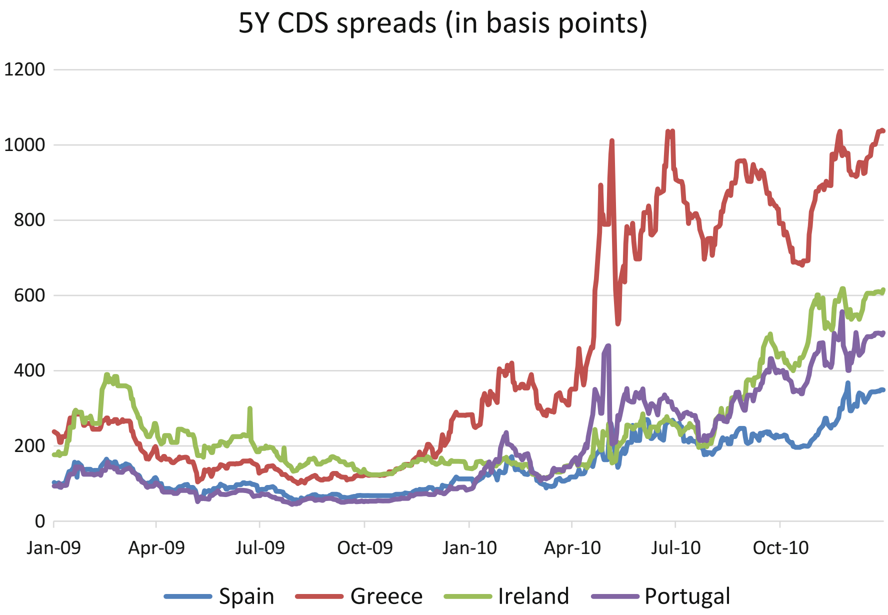


Fig. 5.6 CDS spreads of selected EU countries 10/2009–5/2010 (Source: Standard & Poor's)

similar to a combination of an asset swap and a CDS protection. However, the total return swap can also be entered into for a period shorter than the bond's maturity. The change in market value, then, reflects not only the change in the issuers' credit quality, but also the possible risk free yield curve movement.

Note that the 10 basis points spread over LIBOR is not an analogy of a CDS spread on the issuer's risk. The total return payer retains the bonds, but his return is risk-free, provided the total return receiver does not default in the case when the bond issuer defaults, and there is a loss to be covered by the swap. Therefore, the 10 point spread reflects, rather, the total return swap receiver's (in fact, the credit protection provider's) probability of default, and it is also related to the correlation with the issuer's credit risk.

The example shown describes the possible mechanics of a total return swap. There are, however, many alternatives; for instance, the physical settlement at maturity of the swap, or at the default of the underlying bond, periodic payments based on the underlying asset value gain/loss, etc.

CDS Indices

A simple CDS index can be defined in a similar way to equity indices, as an average quoted CDS spread taken over a defined portfolio of reference entities. However, the standard CDS indices, like CDX NA IG; covering 125 investment grade companies, and iTraxx Europe, covering 125 investment companies in Europe,

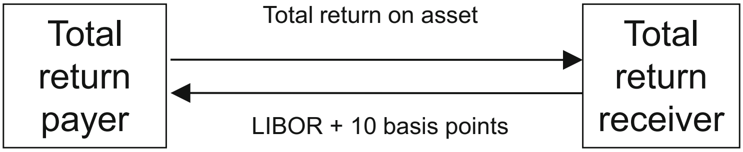


Fig. 5.7 Total return swap

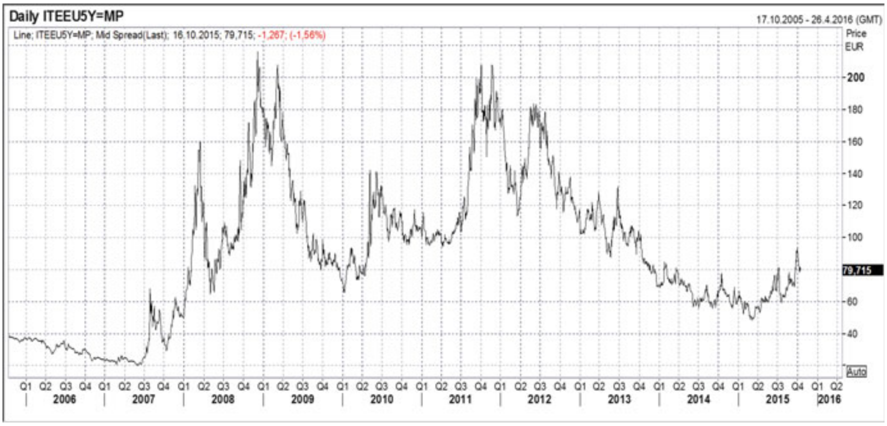


Fig. 5.8 Development of the iTraxx Europe index 2006–2015 (Source: Thomson Reuters)

are more than just statistical figures. The indices are traded as multi-name credit default swaps where the protection seller pays for the credit losses of the counterparties in the underlying portfolio. Roughly speaking, the contract can be broken down into a collection of 125 credit default swaps on the individual names with the same notional. Consequently, the index quotation should be (almost) equal to the average CDS spread over the portfolio in an arbitrage free market.

The CDX NA IG, and iTraxx Europe OTC contracts, are quite standardized and can be compared, in a sense, to the exchange traded contracts, such as futures or options. In fact, the NYSE Euronext made iTraxx CDS contracts available for processing and clearing on its Bclear platform at the end of 2008. There is an active market for maturities of 3, 5,7, and 10 years for contracts ending usually December 20, or June 20, and with a typical notional of $100 million. Moreover, for each index and maturity there is a "coupon" specified, and the standardized contracts are traded based on the CDS market value settlement. For example, for the Series 24 5-year iTraxx Europe, maturing on June 20, 2020, the coupon has been set at 100 basis points (Source: Markit Financial Information Services).

Let us assume that the actual market iTraxx Europe quote is 79 basis points as indicated by Fig. 5.8. It means that the market value of the CDS contract, with its fixed 100 basis points coupon, is positive from the perspective of the credit protection seller, and so the seller must pay an initial compensation to the credit

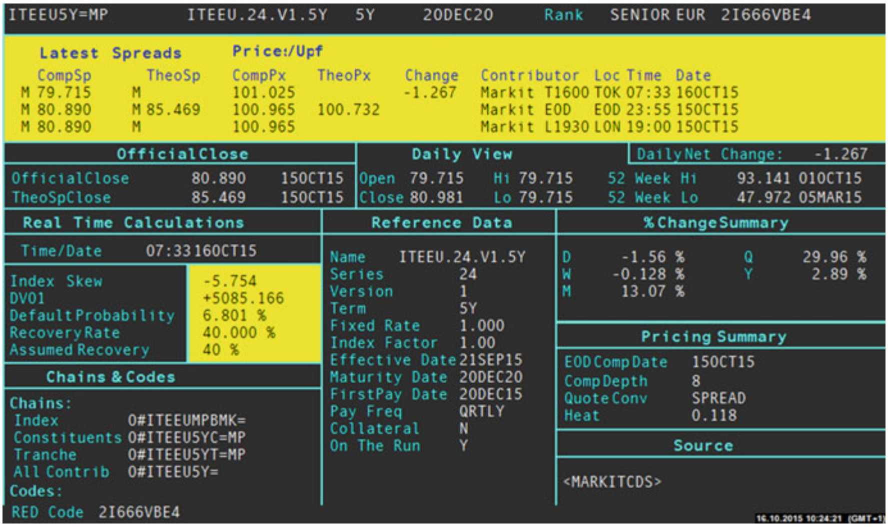


Fig. 5.9 Series 24 5-year iTraxx Europe quotation (Source: Thomson Reuters)

protection buyer. This mechanism allows counterparties to enter into opposite positions in order to close-out the contracts with a net gain/loss given by the initial settlement payments. The compensation is the market value of the CDS position calculated by a standardized methodology based on the risk neutral probabilities of default, and discounted by risk free interest rates, as explained above. The standard recovery rate used for the valuation is currently set at 40%. The market value is usually expressed as the duration, multiplied by the difference between the quote and the coupon; i.e., in our case, 21 basis points. If the duration was 4.9, then the market value would be $4.9 \times 0.21\% = 1.029\%$ of the CDS notional amount payable at the initial settlement date by the protection buyer (compare with the quotations in Fig. 5.9). The credit protection buyer then pays, in arrears, the fixed 100 basis point multiplied by the notional multiplied by $n/125$, where n is the number of companies that have not defaulted yet.

Multi-name CDS

The CDS indices characterize not only the average level of credit risk on a portfolio of counterparties, but are traded, in fact, as multi-name CDS contracts. Generally, basket or multi-name CDS are contracts based on a number of reference entities. The index contracts described above belong to the category of *add-up* CDS, where the protection seller provides a payoff when any of the reference entities default. The add-up CDS is terminated only if all the entities default, otherwise, it continues with a reduced notional. As we have argued, add-up CDS can be broken down into a set of corresponding single name CDS. There is, however, one small difference in their valuation: the single name swaps would have different market spreads, while the add-up CDS quote represents a flat spread, applicable to all the entities. Since entities with higher spreads have lower survival probabilities, the add-up basket CDS spread is usually slightly below the average of the corresponding single name spreads.

The valuation becomes more complex for the *first-to-default, second-to-default,* and generally, the *nth-to-default* basket CDS that provides a payoff only when the n-th default occurs. The payoff paid by the credit protection seller is calculated as the loss on the n-th default, and the swap is then terminated; i.e., there are no further payments by either party.

Example Let us consider a 5-year second-to-default swap on the basket of ten counterparties (bonds) with the same \$1 million notional, and with the spread of 200 basis points. The protection buyer pays the 2% p.a. spread, as long as there is no or only one, default. If no, or just one, entity defaults during the 5 year period, then there is no payoff from the protection seller, and the total payment by the credit protection buyer is $5 \times 2\% \times \$1$ million = \$0.1 million. If there is a second default, then the loss on the defaulted counterparty bond is settled, physically or in cash, the spread accrued until the time of the second default is paid, and the swap is terminated. Therefore, if $LGD = 0.6$ then the payoff is \$0.6 million.

Before we enter into such a swap, we should be able to decide whether the offered spread is too high or too low. It turns out that the answer depends on a new variable, on the correlation between the defaults of the debtors in the portfolio. Let us assume, in the context of the example, that the 5-year risk neutral cumulative default probability is 10% for each of the counterparties. Let us consider, for the sake of illustration, the following two possibilities:

1. Defaults of all the counterparties are fully dependent (100% correlation). In this case, there is either no default (no payoff), or ten defaults (\$0.6 million payoff). The probability of the joint default is 10%, and so the probability weighted, i.e. expected, payoff would be \$0.06 million without discounting. The expected payoff is lower than the total spread payment of \$0.1 million (without the effect of discounting and survival probabilities), and so, the 200 basis points spread appears to be too high from the perspective of the protection buyer.
2. The defaults are independent (0% correlation). Then we can easily calculate that the probability of no default is $q_0 = 0.9^{10} = 0.3487$, the probability of one default $q_1 = 10 \cdot 0.1 \cdot 0.9^9 = 0.387$, and the probability of two, or more, defaults is $q_{\geq 2} = 1 - q_0 - q_1 = 0.264$. Therefore, the expected payoff, without the effect discounting, is $q_{\geq 2} \times \$0.6\text{million} = \$0.158\text{million}$. Thus, under the zero correlation assumption, the 200 basis points spread corresponding to a maximum \$0.1 million total spread payment becomes quite attractive. □

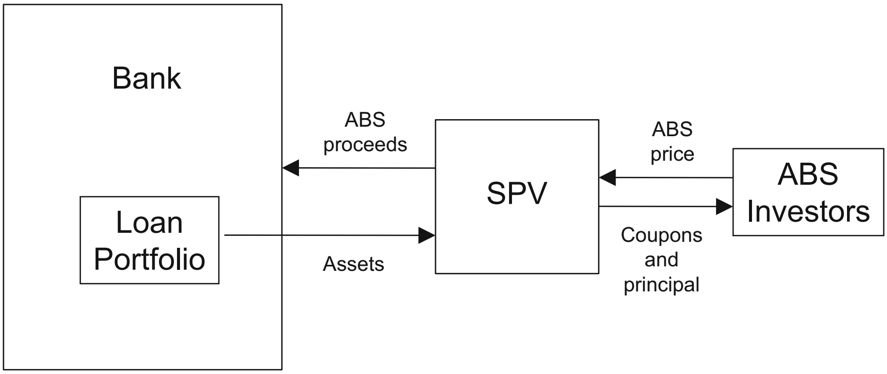


Fig. 5.10 Pass-through ABS scheme

It turns out that valuation of a basket CDS is very sensitive to the default correlation parameter, as well as the probabilities of default. Section 5.2 will discuss advanced default correlation modeling and the valuation techniques.

Asset-Backed Securities and Collateralized Debt Obligations

An asset-backed security (ABS) is, in general, a security created from a portfolio of loans, bonds, or other financial assets. The assets are sold to a special entity called the Special Purpose Vehicle (SPV), which issues and sells to investors the ABS which is financing the purchase of the asset portfolio into the SPV. The ABS investors receive coupons and principal repayments that depend solely on the cash flows from the asset portfolio, not on the credit quality of the creator. For example, a bank providing consumer loans may need to free up its capital to sell more loans. Rather than keeping the existing consumer loan receivables on its balance sheet, the bank may decide to set up an SPV, and sell the loan portfolio to the SPV. The bank is then insulated from the credit risk of the loan portfolio, which is borne solely by the ABS investors and, moreover, it earns a fee for originating and servicing the loans.

Figure 5.10 shows a scheme of a "pass-through" ABS where the loan portfolio cash flows are directly passed, after subtracting the fees, to the ABS investors, and there is no prioritization. The coupons and principal of the ABS bonds should be paid in full if the assumptions regarding defaults and losses are met. The investors may even receive a premium, but if the loan portfolio losses exceed the assumptions, then, the investors suffer a loss.

A mortgage-backed security (MBS) is a type of asset backed security representing a claim on the cash flows from mortgage loans, most commonly on residential property. The securitization process is often facilitated by government agencies, or government-sponsored enterprises, which may offer features to mitigate the risk of default associated with these mortgages. In the United States, most

MBS's are issued by the government sponsored Federal National Mortgage Association (Fannie Mae), and the Federal Home Loan Mortgage Corporation (Freddie Mac), or by the Government National Mortgage Association (Ginnie Mae). Ginnie Mae is fully backed by the government guarantees, so that investors receive timely payments. Fannie Mae and Freddie Mac also provide certain guarantees and, while not backed fully by the government, have special authority to borrow from the U.S. Treasury. Investors in MBS still bear the risk of higher losses on the mortgage portfolio, as well as an interest rate risk of early mortgage prepayments. Huge financial losses suffered by government sponsored enterprises, caused by the increased numbers of subprime mortgage defaults, marked the beginning of the global financial crisis in late 2007. In 2008 Fannie Mae and Ginnie Mae were placed into the conservatorship (forced administration) of the Federal Housing Finance Agency. The U.S. Government effectively took over both agencies and provided them with a financial support exceeding 100 billion USD. It should be noted that the agencies repaid the subsidy in dividends paid to the Treasury in the years 2013–2014.

The credit rating of an ABS is often enhanced by issuing one, or more, subordinated tranches which cover credit losses with priority over the most senior tranche, as shown in Fig. 5.11. The portfolio cash flows are used to repay coupons and principal, using a set of rules known as the *waterfall*: the cash flows are, first of all, used to pay the Senior Tranche investors' promised return of 5% in the given example, then, as far as possible, the Mezzanine Tranche investors' return of 10%, and finally the Equity Tranche investors' return, up to 25%. From a different perspective, the loan portfolio losses are first of all, allocated to the Equity Tranche,

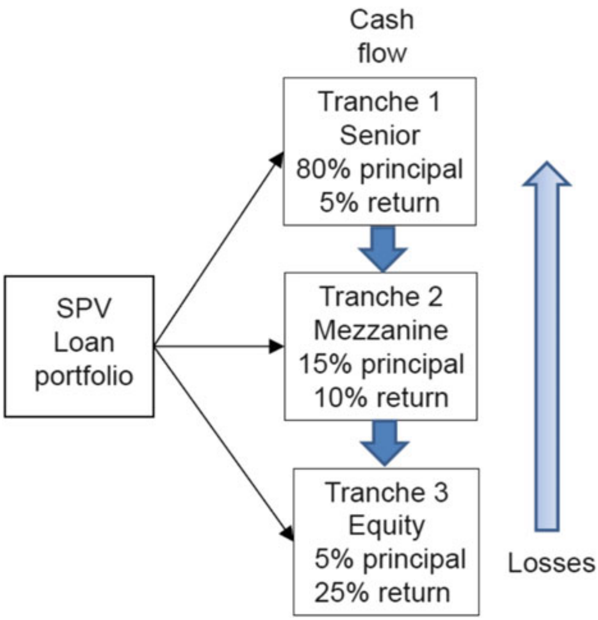


Fig. 5.11 Possible structure of an ABS with three tranches

then, if exceeding 5%, to the Mezzanine Tranche, and finally, the losses over 20% of the portfolio value hit the Senior Tranche. If the portfolio expected loss was somewhere around 2%, then, unexpected losses over 20% should be highly improbable, and so the Senior Tranche may receive a very high rating. Thus, the Senior Tranche is collateralized by the subordinated tranches. The equity tranche on the other hand becomes quite risky, and usually is not rated.

The sophisticated construction is motivated by the demand for investments into higher grade bonds by many institutional investors, such as pension funds, or insurance companies, which have investment limits linked to the ratings. The conservative investors may go into the senior ABS, and less risk averse investors may invest in the mezzanine, or equity, ABS. The equity tranche is also sometimes retained by the ABS creator.

Collateralized Debt Obligations

A type of asset-backed security that has become quite popular, and that, at the same time, has been blamed for the recent financial crisis, is the collateralized debt obligation. In this case, the assets that are securitized are bonds or corporate loans. There is always the element of tranching; i.e., there is a number of tranches from the senior one to equity. The number of tranches is often larger than three. According to this definition, it may be hard to understand what the connection between CDOs and the subprime mortgage crisis is. Let us look at the development of the issuance of those securities by collateral type since 2000, in Table 5.5, provided by SIFMA (Securities Industry and Financial Markets Association). In 2004–2007, the majority of issued CDOs had collateral classified as structured finance, meaning MBSs, ABSs, or their tranches. The tranches used for securitization were typically mezzanine MBS/ABS tranches, which were difficult to sell as such (Hull 2009). Thus, in this way, subprime mortgages have been packed into MBSs and repacked into CDOs. In fact, there is a double securitization, and we should speak about ABS CDOs. Similarly, there could also be CDO-squared CDOs, where the securitized assets are again CDOs. The valuation of those ABS CDOs or CDOs-squared becomes very difficult, not only because of the mathematical model, but also in terms of understanding and decoding all of the legal documentation related to the issues. The market has become dependent on rating agencies that were often too optimistic, partly due to a conflict of interest (being paid by the CDOs issuers), and partly due to the various weaknesses of the models, or data used. High uncertainty in the pricing of those securities, and a panic on the market during the crises, led to large losses, even on the senior tranches. Table 5.5 shows that Structured Finance CDOs were almost totally abandoned in 2009. However, it seems that the CDO market has recovered since then and we have to hope that the market participants have taken a lesson from the crisis.

An alternative way of CDO classification is into *cash flow, hybrid, synthetic*, and *market value*. The structure we have described above is a cash CDO. The synthetic funded CDO is based on a short position in one, or more, CDSs, and a

Table 5.5 Global issuance of CDOs by collateral type in USD millions (Source: SIFMA)

Year	High yield bonds	High yield loans	Investment grade bonds	Mixed collateral	Other	Other swaps	Structured finance	Total
2000	11,321	22,715	29,892	2090	932		1038	67,988
2001	13,434	27,368	31,959	2194	2705		794	78,454
2002	2401	30,388	21,453	1915	9418		17,499	83,074
2003	10,091	22,584	11,770	22	6947	110	35,106	86,630
2004	8019	32,192	11,606	1095	14,873	6775	83,262	157,821
2005	1413	69,441	3878	893	15,811	2257	157,572	251,265
2006	941	171,906	24,865	20	14,447	762	307,705	520,645
2007	2151	138,827	78,571		1722	1147	259,184	481,601
2008		27,489	15,955				18,442	61,887
2009		2033	1972				331	4336
2010		1807	4806		321		1731	8666
2011		20,002	1028		8126		1975	31,131
2012		44,062	62				20,246	64,371
2013		26,362					63,911	90,273
2014		70,018	430				70,846	141,294

corresponding amount of risk free investments. For example, instead of creating a portfolio of corporate loans, the originator of the CDO creates a portfolio consisting of short positions (sold protection) in credit default swaps. The bank may retain the relationship with the corporate clients, and at the same time, transfer the credit risk of the existing loans to the SPV and the CDO investors. The volume of funded synthetic CDOs issued was over $48 billion in 2007, but went down to $0.25 in 2009. Hybrid CDOs combine the funding structures of cash and synthetic CDOs. Market value CDOs allow trading in, and market valuation of, the underlying assets.

Another classification of CDO securities is into the arbitrage and balance sheet types. Balance sheet CDOs aim to remove assets, or the risk of assets, from the balance sheet of the originator, while arbitrage CDOs attempt to capture the mismatch between the yields of assets (CDO collateral), and the financing costs of the generally higher rated liabilities (CDO tranches).

Unfunded Synthetic CDOs

The numbers in Table 5.5 do not include what is called unfunded synthetic CDOs, because those are not, in fact, securities, but OTC contracts similar to the basket CDS. An unfunded synthetic CDO is structured so that default losses on the underlying CDS portfolio are allocated to tranches. For example, in an analogy to Fig. 5.11, the equity tranche would be responsible for the first 5% of the losses, the mezzanine tranche for the next 15% of the losses, and the senior tranche for the remaining losses. The principal amounts: 5, 15, and 80% of the total would be reduced by the losses paid to protection buyers. If a notional is wiped out, then the CDO tranche is terminated. The construction is not as new, or exotic, as it might seem. It is, in fact, quite similar to the non-proportional reinsurance contracts that have been used on the reinsurance market since long before the CDS and CDO products were invented.

Besides the standardized indices like CDX or iTraxx, which are traded as CDS contracts, there are even standard single tranches that are traded as synthetic CDOs. Trading in those contracts is known as *single tranche trading*. In the case of CDX, there are six tranches defined by the loss thresholds: 3, 7, 10, 15, 30, 100%, and in the case of iTraxx Europe, the thresholds are: 3, 6, 9, 12, 22, 100%. The contracts usually pay a fixed spread, with an upfront fee that is quoted. For example, the 10Y iTraxx Europe 9–12 contract maturing in June 2021 has been quoted at 109 bp, as of 15th October, 2015, meaning that the protection buyer pays 1.09% of the agreed notional upfront, and then standard 1% p.a. quarterly in arrears. If the cumulative loss on the iTraxx portfolio since the start date of the contract is less than 9%, then there is no payment by the protection seller. If the losses are, for example, 10%, then, 1/3 of the notional amount is paid by the protection seller, and the notional is reduced to 2/3. The protection buyer continues to pay the 1% spread on the reduced notional. If the iTraxx losses exceed 12%, then the full notional will be paid to the protection buyer, and there will be no more spread payment.

Other Credit Derivatives
Similar, to classical derivatives, once the market with "simple" credit derivatives was established, it was natural for derivatives dealers to trade more complex contracts like forwards and options on CDS.

A forward credit default swap is an obligation to buy or sell a specified CDS (reference entity, maturity, notional, settlement mechanism) at a future time T, with a fixed spread s. Unlike forward interest rate swaps, the contract may cease to exist if the reference entity defaults before time T. Similarly, a CDS call/put option would give the option holder an option to buy/sell a specified CDS protection at price s at time T. Again, the option will cease to exist if the reference entity defaults before time T. For example, there are forwards and options on iTraxx, or CDX. Similar to the CDS market, there are forwards and options on CDO or ABS tranches.

5.2 Valuation of Multi-name Credit Derivatives

We have seen that the mathematics of the risk-free discount interest rates, and risk-neutral probabilities of default, is sufficient to value classical bonds, or single name CDS. However, the valuation of multi-name credit derivatives, such as basket CDS or CDOs, brings a new level of difficulty. In principle, we need to model the probability distribution of returns and losses split into different CDO tranches, or basket CDS layers. Essentially, we may apply the credit portfolio modeling techniques outlined in Chap. 4, but in this case, the requirements on modeling precision, having a real financial impact, are much higher than in the case of the pure risk management, where it might be sufficient to obtain just an approximate estimation of unexpected losses. We have seen that there are different models that might give different results, and in all cases there are the key concepts of default, asset return, or correlation, which enter the model. The estimations of the credit portfolio unexpected loss quantiles turn out to be extremely sensitive to the correlation parameters, but these numbers themselves are usually difficult to estimate, due to limited and ever changing data. The model and correlation risk could be blamed for uncertainty in the valuation of the multi-name instruments. Underestimation of the model and parameter risks could also be blamed for the recent financial crisis.

In this section we will look, first of all, at the connection between the valuation of multi-name credit derivatives and the unexpected risk modeling. Then we will outline a general Monte Carlo simulation approach and an analytical approach, in the context of the Gaussian copula correlation modeling. In the next section, we will look at several alternative advanced dependence modeling approaches.

Unexpected Loss Modeling and Multi-name Credit Derivatives Valuation
To demonstrate the close relationship between the two goals, unexpected loss modeling and multi-name credit derivatives valuation, let us consider a simplified, unfunded, synthetic CDO structure, where the payoff, depending on the underlying portfolio losses, is realized only at maturity T. Let us assume that the portfolio value payable at time T is 1 unit, and the tranche thresholds are $0 = x_0 < x_1 < \cdots < x_n = 1$. To calculate the market spreads assumed to be paid in the form of upfront fees we need to introduce the risk-neutral random variable $X \in [0, 1]$ representing losses on an underlying portfolio of certain single name CDS contracts; i.e., on corresponding loans or bonds, at maturity T. Note that the risk-neutral distribution of losses differs from the real world distribution of losses, just as risk-neutral probabilities of default differ from the real probabilities. Given that, we can easily define the payoff of tranche i as

$$\text{Payoff}_i = \begin{cases} 0 \text{ if } X \le x_{i-1} \\ X - x_{i-1} \text{ if } x_{i-1} < X \le x_i \\ x_i - x_{i-1} \text{ if } x_i < X. \end{cases}$$

If r is the risk-free constant interest rate, and $F(x)$ the cumulative distribution function of the random variable X, then the Tranche i market value can be expressed as:

$$\begin{aligned} f_i &= e^{-\gamma T} E[\text{Payoff}_i] = \\ &= e^{-\gamma T}\big((F(x_i) - F(x_{i-1}))E\big[X - x_{i-1} \big| x_{i-1} < X \le x_i\big] + (1 - F(x_i))(x_i - x_{i-1})\big). \end{aligned} \tag{5.4}$$

If $F(x)$ is approximately linear between x_{i-1} and x_i (i.e., X is approximately uniformly distributed in the interval $[X_{i-1}, X_i]$), then,

$$E\big[X - x_{i-1} \big| x_{i-1} < X \le x_i\big] \cong \frac{x_i + x_{i-1}}{2} - x_{i-1} = \frac{x_i - x_{i-1}}{2},$$

and finally, using the linearity of $F(x)$ on the interval $[x_{i-1}, x_i]$, we get that the value of tranche i as a percentage s_i of its notional value $x_{i-1} - x_i$ is approximately equal to the risk-neutral probability that the loss exceeds the midpoint of $[x_{i-1}, x_i]$, discounted by the risk free interest rate r from T to time zero,

$$s_i = \frac{f_i}{x_i - x_{i-1}} \cong e^{-rT}\left(1 - F\left(\frac{x_{i-1} + x_i}{2}\right)\right) = e^{-rT}\Pr\left[X > \frac{x_{i-1} + x_i}{2}\right]. \tag{5.5}$$

On the other hand, if we are given market prices expressed as up-front percentage spreads s_i, then we can estimate the values

$$F\left(\frac{x_{i-1}+x_i}{2}\right) \cong 1 - e^{rT} s_i, \tag{5.6}$$

which allows us to construct an approximation of the cumulative distribution function $F(x)$, assuming linearity on the intervals $[x_{i-1}, x_i]$, and the facts that $F(0) = 0$ and $F(1) = 1$. Similarly, if we consider a funded CDO with a 1 unit underlying portfolio, and a tranche payoff defined as the notional minus the tranche loss at time T, then the market value of tranche i as a percentage of the tranche notional should be approximately equal to the risk-neutral probability of loss being less than the midpoint of $[x_{i-1}, x_i]$, discounted by r, specifically

$$p_i \cong \left(1 - e^{-rT}\right) + e^{-rT} F\left(\frac{x_{i-1}+x_i}{2}\right). \tag{5.7}$$

Consequently, the valuation of a simplified CDO, or basket of CDS tranches, goes hand-in-hand with the underlying portfolio loss modeling. Full knowledge of the risk neutral loss distribution allows us to value precisely the tranches according to (5.4), and the tranche market value gives us an approximation of the distribution function according to (5.6) or (5.7), which improves with an increasing number of tranches.

In practice, the situation is more complex since the cash flows, payoff and spread payments are realized between the start date and maturity, not only at maturity. It means that we need to model not only the distribution of losses at time T, but also their distribution over time. Thus, we need to model the times of defaults and survival probabilities. Yet, looking at cumulative losses and total up-front fees, the approximate relationships (5.5) and (5.7) between the market prices and the risk neutral cumulative probabilities remain valid. Note that in case of CDOs, the situation is also more complex due to the coupons paid by the tranches. By the notional, in the context of the simplified model described in the case of a general CDO, we mean the time T forward value of all the promised payments, and the loss is compared with respect to this value.

General Monte Carlo Simulation Approach

As argued above, the valuation of basket CDS, or CDOs, is in principle equivalent to portfolio loss modeling, for which, in Chap. 4, we have described various Monte Carlo simulation approaches. Since, in practice, we need to model not only the total loss at the end of a horizon, but also the cash flows over time; we need to improve the model, which can be formulated in a general Monte Carlo simulation framework.

Specification:

1. Fix the term structure of the risk-free interest rates. In the basic approach, interest rates can be considered deterministic. In an advanced approach, interest rates should be stochastic, and possibly correlated with intensities of default.
2. For each receivable in the portfolio, determine the cash flows (payment calendar) of coupon and principal repayments. In the case of float coupons, for instance, Libor + 1% p.a., and deterministic interest rates, replace the unknown float rate with the corresponding forward risk free interest rate.
3. For each receivable i, estimate the term structure of risk neutral probabilities of default; i.e., the distribution of the time to default variable τ_i. In Sect. 5.4 we will study possibilities of how to approach the stochastic intensity of default modeling.
4. Use either a flat recovery rate, or recovery rate depending on claim seniority, possibly on individual debtors. An advanced option is a random recovery rate with an appropriate distribution.
5. Specify correlations (dependence) among the times to default. Note that a complete specification would have to involve the joint distribution of the vector of n random variables $\langle \tau_i; i = 1, \ldots, n \rangle$. The standard solution is the Gaussian copula model, where we specify all the correlations $\rho_{i,j}$ of the standardized default times, i.e. of the variables $X_j = \Phi^{-1}\left(Q_j\left(\tau_j\right)\right)$, assuming a joint normal distribution as described in Sect. 4.7. We have to keep in mind that this is just one of many possible correlation models. The copula specification (Gaussian or other) allows simulating times to default accordingly. Moreover, if our ambition is to use stochastic recovery rates, with a dependence on default rates, the correlation structure needs to be specified as well.

Simulation:

6. Simulate the times to default $\langle \tau_i; i = 1, \ldots, n \rangle$, and the recovery rates RR_i if there is a default in the given time horizon; i.e., if $\tau_i \leq T$. Conditional on this scenario, obtain an exact realized cash flow generated by the underlying loan portfolio until the maturity T.
7. Based on the specified CDO waterfall, or basket CDS definition, determine the payoff cash flows conditional on the times of the default scenario given by step 6. Discount the payoff cash flows with the risk-free interest rates, and add the result to a list of simulated values.

Final result evaluation:

8. Since our goal is to get the discounted expected (i.e., average) payoff, we need to run steps 6 and 7 sufficiently many times and then calculate the mean value, approximating the market value of the given instrument (or a number of tranches that can be valued simultaneously). The confidence interval, and

number of simulation runs required may be estimated using the techniques discussed in Sect. 4.2.

Besides the problem of the risk-neutral probabilities of default, recovery rates, and correlation structure specification, it should be pointed out that the apparently simple step 7; i.e., the implementation of a CDO or CDS waterfall, may be quite complex. This is the case, in particular, for the "classical" CDOs with an underlying portfolio of loans and bonds including structured ones. The interest and principal repayments are used to cover management fees, tranche coupons, and principals, based on various complex prioritization schemes (see, e.g., Duffie and Singleton 2003).

Typically, a tranche is repaid as a sinking-fund bond, with a promised coupon c (paid n times in a year), where the principal $F(k), k = 0, \ldots, K$ may be gradually repaid, or contractually reduced over given K periods; and on the other hand, in the case of insufficient coupon payments $Y(k)$, there is an accrued interest variable:

$$U(0) = 0,$$
$$U(k) = \left(1 + \frac{c}{n}\right)U(k-1) + \frac{c}{n}F(k-1) - Y(k).$$

The reduction of the principal may be due to a prepayment $D(k)$, or unpaid contractual reduction $J(k)$, typically due to losses on the underlying portfolio; i.e., $F(k) = F(k-1) - D(k) - J(k)$, with $F(0)$ equal to the initial tranche face value. The scheme also usually involves a reserve account $R(k)$, where any excess cash is deposited and accrues risk free interest. The reserve account is drawn if there is a cash shortfall, and it is fully paid out at maturity. At maturity the reduced principal plus the unpaid accrued interest are to be paid to the extent provided for in the CDO contract. Thus, the accrued unpaid interest effectively enhances the principal. It should be noted that a shortfall on the final payment does not constitute the CDO default, as long as the contractual prioritization rules were met. The equity tranche usually does not have any defined coupon, and receives only the remaining cash at maturity.

There are two basic types of prioritization schemes: *uniform* and *fast*. Under the uniform prioritization, there are no early prepayments of principal. The collected interest is used to pay for coupons from the most senior tranche down to the junior tranches, modifying the unpaid accrued interest, or the reserve account as defined above. On the other hand, the total losses, less collected and undistributed interest income, define the unpaid principal reductions, starting from the equity tranche, and going up to more senior tranches. At maturity, the final collected principal, and the remaining reserve are paid in priority order, covering both coupons and principal, with the equity tranche collecting any remaining amount (the final amount allocated to the equity tranche can exceed the remaining principal).

Under the fast prioritization, the senior tranche is allocated interest and principal repayments as quickly as possible, until the remaining principal is reduced to zero, or until maturity. Meanwhile, the mezzanine tranche accrues unpaid interest which starts to be paid back once the senior tranche has been repaid, and so on. For this scheme, there are no contractual reductions of principal (in fact, reduction of the principal would not make any difference in the fast cash flow allocation scheme).

In practice, the prioritization schemes are usually somewhere between the two simple cases, applying the so called over-collateralization and interest rate coverage tests, which may trigger early principal repayments. For example, the over-collateralization test for the senior tranche may require that the actual underlying portfolio principal, divided by the senior tranche principal, is larger than 120%. The interest rate coverage test may require that the collected interest, divided by the senior tranche coupon, is larger than 150%, etc. If the condition is satisfied, then the collected interest is used as in the uniform prioritization scheme. On the other hand, if the condition is not satisfied, then the fast prioritization, i.e. repayment of the principal of the senior tranche, is applied until the tests are satisfied.

Analytical Valuation of Basket CDS

The single-factor Guassian copula model has been already explained in the context of Vasicek's Model (Sect. 4.7). Under the model, the standardized time to default variable $X_j = \Phi^{-1}(Q_j(\tau_j))$ is broken down into a systematic factor M, and an idiosyncratic; i.e., debtor specific factor Z_j so that the factors are independent, standardized normal, and

$$X_j = \sqrt{\rho_j}M + \sqrt{1-\rho_j}Z_j. \tag{5.8}$$

Thus, the mutual correlation of the standardized times to default i and j is $\sqrt{\rho_i\rho_j}$. Analogously to (4.22) we get that the cumulative probability of default in the time horizon t and conditional on $M = m$ is

$$Q_j(t|m) = \Pr[\tau_j \le t|M = m] = \Phi\left(\frac{\Phi^{-1}(Q_j(t)) - \sqrt{\rho_j}m}{\sqrt{1-\rho_j}}\right). \tag{5.9}$$

Moreover, in the standard market model we assume that the portfolio is homogenous, i.e. that all the correlations are the same $\rho = \rho_j$ and the time-to-default distribution Q_j is the same Q for all debtors in the portfolio (Hull 2009). The key idea is that, conditional on a fixed systematic variable value $M = m$, the time to default variables become independent, and the market value might be expressed analytically. The result then needs to be averaged (i.e., integrated out) over all possible values of M with the standard normal density. The calculation also implicitly assumes that we know the unconditional cumulative default probability function in the form $Q(t) = 1 - e^{-\Lambda(t)}$ where $\Lambda(t)$ is the cumulative hazard (see

Sect. 3.3). It is usually assumed that the default intensity λ is constant, i.e. $Q(t) = 1 - e^{-\lambda t}$, but λ can be also a deterministic function of time (see Sect. 5.1).

Let us consider a kth-to-default CDS on a portfolio of n debtors, each with the same one unit notional, and with the same (risk neutral) intensity of default. Conditional on the systematic factor value $M = m$, the probability of exactly l defaults by the time t is given by the binomial distribution formula:

$$P(l, t|m) = \binom{n}{l} Q(t|m)^l (1 - Q(t|m))^{n-l}.$$

The probability of l defaults could be also expressed in the case where the conditional probabilities of default depend on the individual borrowers, but of course, with a significantly more complex expression.[1] If T is the CDS maturity, then the probability of a positive payoff can be easily expressed as:

$$P(\geq k, T|m) = \sum_{l=k}^{n} P(l, T|m).$$

To obtain the full valuation result we need to model the defaults and the CDS payments in the periods $t_0 = 0, t_1, \ldots, t_N = T$. Suppose that the spread as a percentage of each of the underlying receivables, with a unit exposure, is s, and let $D(t)$ be the maturity t discount factor; i.e., the present value of one currency unit paid at t. In the standard market model, the defaults are assumed to happen in the midpoints of the periods $[t_{j-1}, t_j]$. The loss given default is assumed in the homogenous model to be a constant $L = 1 - R$. Then, the market value of the CDS, conditional on the systematic factor value $M = m$, from the perspective of a protection seller can be expressed as the discounted value of expected spread payments minus the discounted expected value of the payoff in the case of default:

$$\begin{aligned} MV(m) = & \sum_{j=1}^{N} s(t_{j-1} - t_j) D(t_j) P(< k, t_j|m) + \\ & + \sum_{j=1}^{N} 0.5 s(t_{j-1} - t_j) D(0.5t_{j-1} + 0.5t_{j-1}) (P(\geq k, t_j|m) - P(\geq k, t_{j-1}|m)) - \\ & - \sum_{j=1}^{N} L \times D(0.5t_{j-1} + 0.5t_{j-1}) (P(\geq k, t_j|m) - P(\geq k, t_{j-1}|m)) \end{aligned} \tag{5.10}$$

[1] In this case, the conditional default probability n can be calculated recursively with respect to, the number of borrowers in the portfolio, essentially adding them one by one, see Brigo et al. (2010).

Finally, the unconditional market value is obtained integrating (5.10) over m with the standard normal distribution:

$$MV = \int_{-\infty}^{+\infty} MV(m)\varphi(m)dm. \tag{5.11}$$

While the expression (5.10) is analytical, the integral (5.11) needs to be evaluated numerically. Note that based on (5.10) and (5.11), the market value can written in the form: $MV = s(A+B) - C$, where A, B are the integrals of the first and the second sum in (5.10), taking out the variable s, and C is the integral of the third sum. When CDS is entered into under market conditions the value must be equal to zero, and so we can easily solve it the spread $s = \frac{C}{A+B}$.

Valuation of Synthetic CDOs

Let us consider an unfunded synthetic CDO on a portfolio of n receivables each with the same exposure $1/n$, and a tranche with the *attachment* and *detachment points* $0 \le A \le B \le 1$. If the cumulative loss $L(t)$ on the portfolio exceeds the attachment point A then the protection payer pays the difference $L(t) - A$, but not more than $B - A$. Since defaults take place gradually, it is useful to introduce the reduced tranche $[A, B]$ notional cumulative loss function

$$L^{A,B}(t) = \frac{\min\big(\max(L(t), A), B\big) - A}{B - A}. \tag{5.12}$$

The default leg present value, given the loss function, then is

$$DL^{A,B}(0) = \int_0^T D(t)dL^{A,B}(t) \cong \sum_{j=1}^{N} D\left(0.5t_{j-1} + 0.5t_j\right)\left(L^{A,B}\left(t_j\right) - L^{A,B}\left(t_{j-1}\right)\right) \tag{5.13}$$

where we approximately assume as above that the premium payment periods are $t_0 = 0, t_1, \ldots, t_N = T$ and defaults take place only in the middle of the periods.

Regarding the premium leg, there is a slight conventional difference used by the standard synthetic CDOs like DJ-iTraxx or CDX compared to the kth-to-default CDS described above. The premium is, in this case, paid on the tranche notional amount less non-recovered losses (not less defaulted notional amounts) at the end of each premium period; i.e., again given the reduced tranche loss function, we define the "risky annuity" as

$$DV^{A,B}(0) = \sum_{j=1}^{N} \left(t_j - t_{j-1}\right) D\left(t_j\right) \left(1 - L^{A,B}\left(t_j\right)\right). \tag{5.14}$$

If U and s denote the up-front payment and the spread premium, then the premium leg present value is $U + sDV^{A,B}(0)$, conditional on the loss function. Therefore, to find the initial fair market spread we just need to solve the equation

$$E_0\left[DL^{A,B}(0)\right] = U + sE_0\left[DV^{A,B}(0)\right], \text{ and so} \tag{5.15}$$

$$s = \frac{E_0\left[DL^{A,B}(0)\right] - U}{E_0\left[DV^{A,B}(0)\right]}. \tag{5.16}$$

Here, the risk-neutral expected value is taken from the time 0 perspective.

In the standard market model we apply the single-factor Gaussian copula and the homogeneity assumptions as above, i.e.

$$\Pr[L(t) = k(1-R)] = \int_{-\infty}^{+\infty} P\left(k, t \middle| m\right) \varphi(m) dm.$$

This semi-analytical formula allows us to calculate $E_0[L^{A,B}(t)]$, and, therefore, the expected values as in Eq. (5.15).

Large Homogenous Portfolio Model

If we assume that the portfolio is homogenous and that the number of its elements is large (tends to infinity) then the formula for the default and premium leg formulas can be further simplified, and the premium can, in fact, be expressed analytically. In this case, by the law of large numbers, the conditional default rate approaches (can be approximated by) the conditional probability of default $Q\left(t \middle| m\right)$ given by (5.9). Therefore, the conditional expected loss on the large homogenous portfolio (LHP) can be expressed as $\bar{L}\left(t \middle| m\right) = LQ\left(t \middle| m\right)$ and the $[0, B]$ tranche conditional expected loss simply as

$$\bar{L}^{0,B}\left(t \middle| m\right) = \min\left(\bar{L}\left(t \middle| m\right), B\right) / B.$$

Finally, the unconditional loss can be obtained integrating out the systematic factor:

$$\bar{L}^{0,B}(t) = E_0\left[\bar{L}^{0,B}\left(t \middle| m\right)\right] = \int \frac{1}{B} \min\left(\bar{L}\left(t \middle| m\right), B\right) \varphi(m) dm. \tag{5.17}$$

It turns out (Brigo et al. 2010) that this integral can be solved analytically:

$$\bar{L}^{0,B}(t) = \Phi(A_1) + \frac{L}{B}\Phi_2\left(A_1, \Phi^{-1}(Q(t)); -\sqrt{\rho}\right),$$
$$A_1 = \frac{\Phi^{-1}(Q(t)) - \sqrt{1-\rho}\Phi^{-1}(B/L)}{\sqrt{\rho}}. \tag{5.18}$$

For a generic tranche $[A, B]$ we can use the elementary relationship

$$\bar{L}^{A,B}(t) = \frac{1}{A-B}\left[B\bar{L}^{0,B}(t) - A\bar{L}^{0,A}(t)\right]. \tag{5.19}$$

Therefore, according to (5.13) and (5.14) there are analytical formulas for the default leg and for the premium leg, and so for the equilibrium tranche spread

$$s = \frac{\overline{DL}^{A,B}(0) - U}{\overline{DV}^{A,B}(0)}, \tag{5.20}$$

where

$$\overline{DL}^{A,B}(0) = \sum_{j=1}^{N} D\left(0.5t_{j-1} + 0.5t_j\right)\left(\overline{L}^{A,B}\left(t_j\right) - \overline{L}^{A,B}\left(t_{j-1}\right)\right)$$

and

$$\overline{DV}^{A,B}(0) = \sum_{j=1}^{N} D\left(t_j\right)\left(t_j - t_{j-1}\right)\left(1 - \overline{L}^{A,B}\left(t_j\right)\right).$$

Compound and Base Correlations

Given the market quotation of a CDS index and its tranches (see, e.g., the iTraxx quotes in Fig. 5.12) we arrive at the concept of *implied correlations* similar to the implied option volatilities. To be specific, let us consider the finite homogenous pool model and assume that we are given the homogenous default probabilities estimated from the underlying index quotations for a set of maturities. Note that the market spread implied by the formula (5.20) is a function $s = s(\rho)$ of the correlation parameter used in (5.9) and (5.18), and so by inverting this relationship given a market spread we can obtain the implied correlation called the *compound correlation*.

However, it turns out that the concept of compound correlations is quite problematic and the market uses rather so called *base correlation* quotes (see Fig. 5.12).

To explain the difference, note that the tranche $[A, B]$ expected loss function (5.19) can be expressed as a function of two correlations,

$$\bar{L}^{A,B}(t, \rho_A, \rho_B) = \frac{1}{A-B}\left[B\bar{L}^{0,B}(t, \rho_B) - A\bar{L}^{0,A}(t, \rho_A)\right]$$

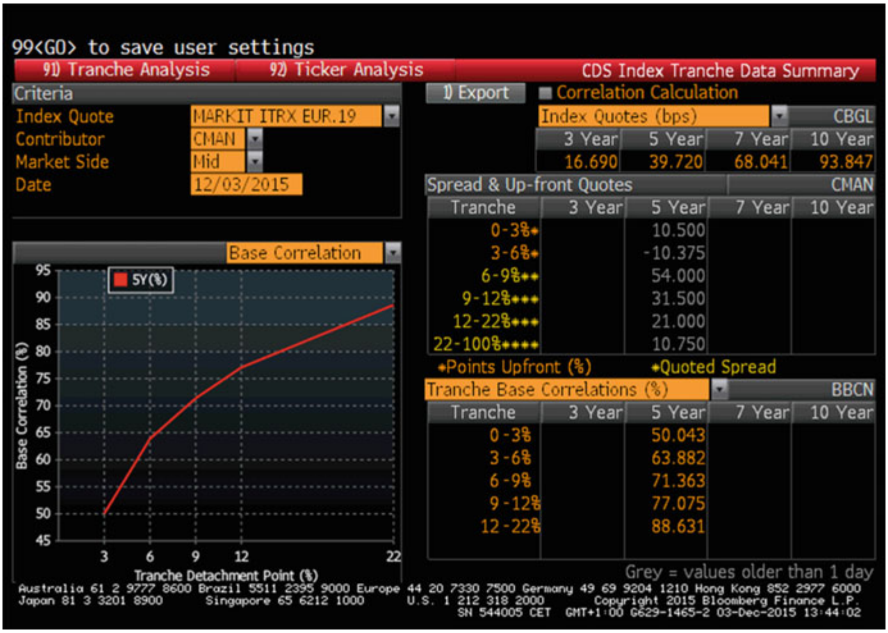


Fig. 5.12 DJ-iTraxx CDS tranche quotations (3. 12. 2015, source: Bloomberg)

where ρ_A is the correlation used to value the tranche $[0, A]$ according to (5.18). Therefore, according to (5.13) and (5.14), the tranche $[A, B]$ default leg and the risky annuity can be written as function of two correlations ρ_A and ρ_B, and so the spread $s^{A,B} = s(\rho_A, \rho_B)$ is a function of the two correlations. Now, we can exactly explain the difference between the two correlation concepts: the compound correlation $\rho_{A,B}$ solves the equation $s^{A,B} = s(\rho_{A,B}, \rho_{A,B})$ and depends on both the attachment and detachment points, while the base correlation ρ_A is the one that correctly values the "base" tranche $[0, A]$ and depends only on the detachment point. Given market quotations, the base correlations can be calculated recursively: for the first tranche $[0, A]$ the correlation ρ_A is implied directly by the quote. Next, given ρ_A find ρ_B solving the equation $s^{A,B} = s(\rho_A, \rho_B)$ where $s^{A,B}$ is the market quote.

One advantage of the base correlations is that they can be more-or-less consistently interpolated (as in Fig. 5.12) and then used to calculate the implied $s^{A,B}$ spread for any nonstandard attachment and detachment points A and B. It is not clear how to value, in a simple way, non-standard tranches given compound correlations. A more serious flaw of the compound correlations is that the market spread is not always a monotonous function of the compound correlations (see, for example, Fig. 5.13), and so the equation $s^{A,B} = s(\rho_{A,B}, \rho_{A,B})$ does not have to have a unique solution! According to an empirical study of Torresetti et al. (2006) this is in fact a very frequent situation. The study shows that this invertibility problem does not arise with the base correlation.

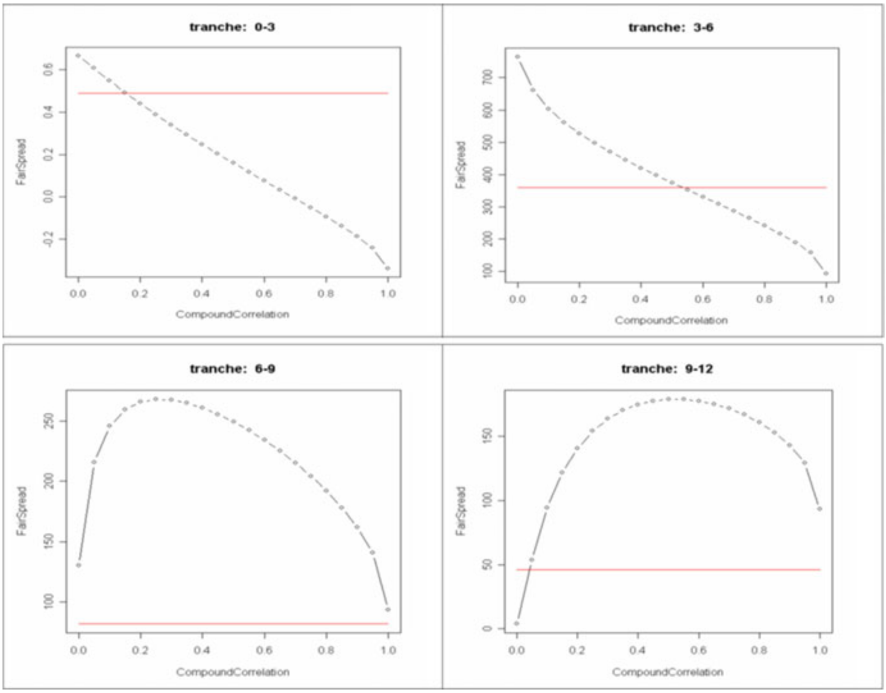


Fig. 5.13 Tranche spread as a function of compound correlation (Source: Torresetti et al. 2006)

However, even the base correlation is not without flaws. Torresetti et al. (2006) point out that for certain tranches and base correlations, the expected tranche loss becomes negative. Note that there is an inconsistency of both concepts, with the single-factor Gaussian model assuming that there is only one correlation. Hence, already using two correlations ρ_A and ρ_B to calculate the expected tranche loss and fair tranche spread is inconsistent with the theoretical model, and one should not be surprised sometimes to obtain strange results.

5.3 Advanced Dependence Modeling

Although the single-factor Gaussian model provides computationally a relatively efficient way to value CDO tranches, we have seen above that the market reality (correlation skew) is not consistent with the assumption of a constant Gaussian time of default correlation. In this section, we will discuss several advanced dependence approaches that aim to fix this problem. The multitude of available methods, each with pros and cons, and with more or less different results, demonstrates the model risk of the CDO tranche valuation. The fact that the Gaussian Copula has been, until recently, the only market standard model, also explains the fragility undergone by the markets during the recent financial crisis.

Double-t and the Logistic Copula

Hull and White (2004) note that the two factors in the decomposition (5.8) of the standardized time to default of an individual debtor into a systematic and idiosyncratic component can have generally any zero-mean and unit-variance distributions leading to various correlation structures. In particular, besides the Gaussian copula, they propose to use the Student t-distributions, possibly with different degrees of freedom. The goal is to find the parameters (correlation and the degree of freedom) that flatten the correlation skew, i.e. a single correlation can be applied to value all CDO tranches at the same time. The argument is that the fatter tails of the Student t-distributions should better fit the market reality.

Specifically, let M and Z_j have the Student t-distributions with d_1 and d_2 degrees of freedom, respectively. Then, unfortunately, the convoluted variable

$$X_j = \sqrt{\rho}M + \sqrt{1-\rho}Z_j \tag{5.21}$$

does not have the Student t-distribution. However, the cumulative distribution function F_{X_j}, depending on the degrees of freedom and on the correlation parameters, can be obtained numerically by a simulation, or by integration (Vrins 2009). Another slightly more computationally efficient way is to employ the technique of Fourier transformations. Since the characteristic function of a scaled Student t-distribution is known, we can use the Gil-Pelaez transformation to retrieve the distribution F_{X_j} from the product of the two characteristic functions corresponding to the convoluted random variable (for more details see, e.g., Kolman 2014). Brigo et al. (2010) propose to use the moment-generating functions and an inverse Laplace transform as an alternative to the Fourier transformation.

The distribution function F_{X_j} is needed not only to calculate the conditional probability of default given the correlation parameter:

$$Q_j(t|m) = \Pr[\tau_j \le t|M = m] = F_{d_2}\left(\frac{F_{X_j}{}^{-1}(Q_j(t)) - \sqrt{\rho}m}{\sqrt{1-\rho}}\right),$$

where F_{d_2} is the Student t-distribution with d_2 degrees of freedom, but also to fit the model, i.e. to estimate the correlation parameter and the degrees of freedom. Since the distribution function must be numerically estimated for each combination of parameters, the fitting procedure becomes computationally quite demanding.

Witzany (2013b) investigates the logistic distribution copula where both variables in (5.21) have the logistic distribution $\Lambda(x) = 1/(1+e^{-x})$. Although a mix of two logistic distributions is not exactly logistic the study proposes to use the logistic distribution as an approximation. It should be noted that the logistic distribution is used not only to calculate the conditional probabilities of default, but also in the fitting phase, presumably offsetting the effect of the approximation. The important advantage of this approach is that both the logistic distribution function and its inverse are analytical, and so the calculations are very efficient. The model is empirically applied to estimate the 99.9% regulatory capital quantile on different loan product portfolios. It turns out that the logistic copula model more than doubles the required capital compared to the Gaussian standard model. Other possibilities tested in the literature are the Normal Inverse Gaussian (NIG) distribution based copula (Kolman 2013) or a copula corresponding to a mix of Gaussian and Generalized Hyperbolic distributions (Gapko and Smid 2012).

Copula Based Correlation Modeling

So far, we have focused on single-factor homogenous models with one single correlation. However, as noted in the full Monte Carlo simulation description in Sect. 5.2, even in the case of the Gaussian model one should specify $n(n+1)/2$ correlations where n is the number of receivables. This large number of correlations is collapsed to one correlation in the case of the homogenous model and to n correlations in case of the non-homogenous single-factor model. The model could be improved introducing more systematic factors, e.g. corresponding to sectors as in Sect. 4.2. Nevertheless, in full generality, we have to work with (at least) $n(n+1)/2$ correlations and, moreover, with different correlation structures that can be described by the concept of copulas.

To be more specific, in Step 5 of the Monte Carlo Simulation approach, the key task is to specify the correlation structure of the time to default variables $\langle \tau_i; i = 1, \ldots, n \rangle$. Even if we use the normalized variables $X_j = \Phi^{-1}\left(Q_j\left(\tau_j\right)\right)$, the multivariate normal distribution with a general correlations matrix $\sum$ is not, by far, the only possibility. Consider, for example, the one-systematic factor model (5.8), breaking X_j down into a systematic factor and an idiosyncratic factor with a fixed correlation coefficient ρ. There is empirical evidence that correlations might be low under normal circumstances, and become high in a financial crisis, when many things go wrong and the systematic factor M is very low. This could be captured by the relationship

$$X_j = \sqrt{\rho(M)}M + \sqrt{1 - \rho(M)}Z_j \tag{5.22}$$

with the correlation $\rho(M)$ depending on the systematic factor.

Example Let us assume that the "true" correlation structure is given by (5.22), with $\rho(M) = 0$, if $M > \Phi^{-1}(0.01)$ and $\rho(M) = 1$, if $M \leq \Phi^{-1}(0.01)$. The ordinary correlation between X_i and X_j for $i \neq j$ can be estimated empirically, or calculated analytically, to be $\rho \cong 6.9\%$. If this correlation is used as an input to the Gaussian Copula single-factor model, on a portfolio of loans with uniform $PD = 1\%$, then, the probability of loss on the senior tranche would be virtually 0. But, according to the "true" model, defaults happen if $M \leq \Phi^{-1}(0.01)$, and in this case all the debtors would go into default due to the 100% correlation. Therefore, the "true" probability of loss on the senior tranche would be 1%.

Note that the mistake in the example above happened, not because we made a wrong estimate of the correlation, but because we chose a wrong correlation model, applying the Gaussian Copula model in the situation of a strongly nonlinear relationship between X_i and X_j for $i \neq j$. □

It turns out that in order to describe a general correlation structure of a set of random variables, it is useful to separate the marginal distributions of the variables and the dependence structure captured by a multi-variate function called the copula function. The application of copula to credit risk modeling (valuation of multi-name credit derivatives) was first proposed by Li (1999). For a more complete treatment of copulas, the reader is referred also to Nelsen (1999), or Cherubini et al. (2004).

Let us consider, for the sake of simplicity, two random variables X_1 and X_2 with the joint distribution function $F(x_1, x_2)$. Let us assume that the marginal distribution functions $F_1(x_1) = F(x_1, +\infty)$ and $F_2(x_2) = F(+\infty, x_2)$ are both continuous and strictly increasing. Since the two functions $F_i : [-\infty, +\infty] \to [0,1]$ are one-to-one and invertible we can define the dependence function $C : [0,1] \times [0,1] \to [0,1]$ as follows: for $u, v \in [0,1]$ simply set

$$C(u,v)=F\left(F_1{}^{-1}(u), F_2{}^{-1}(v)\right) = \Pr[F_1(X_1) \leq u, F_2(X_2) \leq v].$$

Therefore, the function C can be also characterized as the bivariate joined cumulative distribution of the two uniform random variables $F_1(X_1)$ and $F_2(X_2)$. It immediately follows that it has the following elementary properties:

(a) $C(0,v) = C(0,v) = 0$,
(b) $C(u,1) = u$ and $C(1,v) = v$ for every $u, v \in [0,1]$, and moreover
(c) C is 2-*increasing*, i.e. for every $0 \leq u_1 \leq u_2 \leq 1, 0 \leq v_1 \leq v_2 \leq 1$

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

The last inequality follows realizing that the left hand side equals the probability

$$\Pr[F_1(X_1) \in [u_1, u_2], F_2(X_2) \in [v_1, v_2]].$$

Note that it follows from the property (c) that $C(u,v)$ is non-decreasing in both variables u and v.[2] Sklar (1959) defines *bivariate copulas* as functions satisfying the properties (a–c) above. Sklar's famous Theorem says that for any bivariate copula C and continuous marginal distribution functions F_1 and F_2, the function $C(F_1(x_1), F_2(x_2))$ is a joint distribution function with margins F_1 and F_2; i.e. there are random

[2] Let $v_1 = 0$ and $v_2 = v$, then according to the two-increasingness property $C(u_2, v) - C(u_1, v) \geq 0$ whenever $u_2 \geq u_1$.

variables X_1 and X_2 with those marginal distributions and the copula C. The definitions and Sklar's theorem can also be generalized for non-continuous marginal distributions with only slight technical complications. It follows from the definition that the copula functions are invariant to the increasing transformations of the random variables.[3] The invariance property holds for concordance measures like Kendall's tau, Spearman's rho, or Gini's coefficient, but not for the classical linear correlation measure.

It is useful to define the following three important copulas:

– the *maximum copula* $C^+(u,v) = \min\{u,v\}$,
– the *minimum copula* $C^-(u,v) = \max\{u+v-1,0\}$,
– and the *product copula* $C^\perp(u,v) = uv$.

The maximum copula corresponds to perfect dependence between two random variables. For example, if $U = V$ are uniform and perfectly dependent then indeed

$$\Pr[U \le u, V \le v] = \Pr[U \le \min\{u,v\}] = C^+(u,v).$$

On the other hand, if U and V are perfectly negatively dependent, i.e. $U = 1 - V$, then

$$\Pr[1 - V \le u, V \le v] = \Pr[1 - u \le V \le v] = C^-(u,v).$$

The product copula simply corresponds to independent random variables where

$$\Pr[U \le u, V \le v] = uv = C^\perp(u,v).$$

Hence, the copula function is also called the *independent copula*.

It can be easily hown[4] that for any bivariate copula C the following (Fréchet–Hoeffding) inequality holds:

$$C^-(u,v) \le C(u,v) \le C^+(u,v).$$

Therefore the two copulas C^- and C^+ are called the *Fréchet upper* and *lower bounds*.

Sklar's Theorem allows us to define and investigate many interesting dependence structures. Firstly, we define a copula and then we "plug in" any marginal distributions to obtain various joint distributions. For example, it is easy show that any convex linear combination $pC_1 + (1-p)C_2$ of two copulas C_1 and C_2 (with $p \in [0,1]$) must again be a copula. In particular, in this way, we can mix the lower

[3]If C is the copula given by two random variables X_1, X_2 and if α_1, α_2 are two increasing continuous functions, then C is also the copula given by $\alpha_1(X_1), \alpha_2(X_2)$.

[4]The left-hand side of the inequality follows from the copula properties (b) and (c) setting $u_2 = v_2 = 1$, $u_1 = u$, and $v_1 = v$. The right-hand side of the inequality follows from the fact that C is increasing in both variables and from (b).

Fréchet bound, the upper Fréchet bound and the independent copula to obtain the *Fréchet family of copulas*. In particular, we can consider the *Fréchet mixture copulas* in the form $C = pC^{+} + (1-p)C^{\perp}$ as a special case combining the upper Fréchet bound (the perfect copula) and the independent copula. It can be shown that the copula has both upper and lower *tail dependency* measured by p in the sense that the conditional probability $\Pr\left[U \le u \middle| V \le u\right]$ approaches p when u is small, and similarly $\Pr\left[U \ge u \middle| V \ge u\right]$ approaches p when u is close to 1. This copula is very simple and the (tail) dependency parameter p appears to be more relevant in financial applications then the standard linear correlation coefficient ρ.

Another example is the Marshall-Olkin copulas:

$$C(u_1, u_2) = \min\left(u_1^{1-\alpha_1} u_2, u_1 u_2^{1-\alpha_2}\right)$$

where $\alpha_1, \alpha_2 \in [0, 1]$. It can be shown (Nelsen 1999) that the standard concordance measures like Kendall's tau, Spearman's rho, or the upper tail dependency can be expressed in a simple analytical form given the two parameters:

$$\tau = \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2 - \alpha_1 \alpha_2}, \rho_S = \frac{3\alpha_1 \alpha_2}{2\alpha_1 + 2\alpha_2 - \alpha_1 \alpha_2}, \text{ and } \lambda_U = \min(\alpha_1, \alpha_2).$$

There are a number of other well-known parametric copula families like the Gaussian, Student t, the Fréchet, or the Archimedean copulas. Let us look at their detailed definitions in the general multivariate set-up.

In the multivariate case, the copula functions correspond to joint distribution functions of n uniform variables $U_1, \ldots, U_n$, i.e.

$$C(u_1, \ldots, U_n) = \Pr[U_1 \le u_1, \ldots, U_n \le u_n].$$

Now, we must have $C(0, u_2, \ldots, u_n) = 0$, $C(u_1, 1, \ldots, 1) = u_1$, etc. The copula property (c) generalizes to the requirement that C is n-increasing, that is the probability mass on any sub-rectangle of $[0, 1]^n$ must be nonnegative.

In the multivariate case, the product copula $C^{\perp}(u_1, \ldots, u_n) = u_1 \cdots u_n$ again corresponds to n independent random variables; while the Fréchet upper bound $C^{+}(u_1, \ldots u_n) = \min(u_1, \ldots, u_n)$ corresponds to the fully dependent variables. It can be shown that $\max(u_1 + \cdots + u_n - 1, 0)$ is a lower bound, i.e. $\max(u_1 + \cdots + u_n - 1, 0) \le C(u_1, \ldots u_n)$ for any copula C, but unfortunately, for any $n > 2$, the function is never a copula.[5] It means that we cannot generalize the bivariate Fréchet copula family to the multivariate case, but we can still define the multivariate mixture copulas $pC^{+} + (1-p)C^{\perp}$.

[5]Nevertheless, for any $\mathbf{u} \in [0, 1]^n$ there is a copula C so that $\max(u_1 + \cdots + u_n - 1, 0) \le C(\mathbf{u})$. Therefore it is the best lower bound (Nelsen 1999).

The *Gaussian copula* function, which we have used for default correlation modeling so far, is given by a multivariate normal distribution function $\Phi_n(x_1, \ldots, x_n; \Sigma)$, with a correlation matrix Σ:

$$C_G(u_1, \ldots, u_n) = \Phi_n\left(\Phi^{-1}(u_1), \ \ldots, \Phi^{-1}(u_1); \Sigma\right).$$

It is important to note that the copula is defined from the multivariate Gaussian distribution but can be defined by non-Gaussian margins, e.g. lognormal or exponential that we used in time to default modeling. The copula function of a vector of random variables $\langle X_1, \ldots, X_n \rangle$ with general marginal distributions $F_1, \ldots, F_n$ (continuous increasing) is Gaussian, and the quantile-to-quantile (Q-Q) transformation $\left\langle \Phi^{-1}(F_1(X_1)), \ \ldots, \Phi^{-1}(F_n(X_n)) \right\rangle$ is multivariate Gaussian. Specifically, we have used the Gaussian single-factor copula model where the Gaussian structure is simpler: there is a single Gaussian common factor, and the mutual correlations are determined by a correlation with respect to the common factor.

Another useful class of copula functions are the *Student's t copulas*, with v degrees of freedom, similarly to the Gaussian copula given by the relationship:

$$C_S(u_1, \ldots, u_n) = t_{n,v}\left(t_v^{-1}(u_1), \ \ldots t_v^{-1}(u_1); \Sigma\right),$$

where t_v is the univariate Student's t distribution, and $t_{n,v}$ the multivariate Student's t distribution with a correlation matrix Σ. The multivariate Student's t distribution vector of random variables might be obtained from a vector of normal variables $\langle X_1, \ldots X_n \rangle$, with the distribution $\Phi_n(x_1, \ldots, x_n; \Sigma)$ divided by $\sqrt{Y/v}$,where Y is an independent χ_v^2 variable.

A popular universal class of copulas are *the Archimedean copulas* defined by the analytical form

$$C_A(u_1, \ldots, u_n) = \varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_n)),$$

where φ is a strictly decreasing and convex function from $[0, 1]$ onto $[0, +\infty]$. The function φ is called *the generator* of the Archimedean copula. The best known Archimedean copulas are:

1. *Gumbel's Copula*: $\varphi(u) = (-1\text{n}(u))^\alpha, \alpha > 1$,
2. *Clayton's Copula*: $\varphi(u) = u^{-\alpha} - 1, \alpha > 0$, or
3. *Frank's Copula*: $\varphi(u) = 1\text{n}\left(\frac{e^{-\alpha u}-1}{e^{-\alpha}-1}\right), \alpha > 0$.

In order to use a parametric copula for credit risk modeling one needs to estimate its parameters and then simulate (or optimally analytically describe) the distribution of future cash flows of the multi-name credit derivative we want to value.

The classical inference of copula parameters is based on the maximum likelihood estimation (MLE) method. Let us assume that we have a sample of (presumably independent) observations $\{\mathbf{X}_t; t = 1, \ \ldots, T\}$, e.g. financial asset returns, from a multivariate distribution characterized by a copula C and marginal distributions $F_1, \ldots, F_n$ from a parametric family with unknown parameters $\mathbf{\theta}$. In order to estimate $\mathbf{\theta}$ using the exact MLE method we just need to express the multivariate density function:

$$f(\mathbf{x}) = c(F_1(x_1), \ldots, F_n(x_n)) \cdot \prod_{i=1}^{n} f_i(x_i),$$

where the copula density is $c = \frac{\partial^n C}{\partial u_1 \ldots \partial u_n}$ and f_i 's are the marginal densities. The log-likelihood function then is

$$l(\theta) = \sum_t \ln c(F_1(x_{t1}), \ldots, F_n(x_{tn})) + \sum_t \sum_{i=1}^{n} \ln f_i(x_{ti}).$$

The exact MLE estimation $\mathbf{\theta} = \arg\max l(\mathbf{\theta})$ then has the standard asymptotic properties (Cherubini et al. 2004). In particular, the covariance matrix of $\mathbf{\theta}$ (Fisher's information matrix) can be estimated by the inverse of the negative Hessian matrix of the likelihood function.

The MLE procedure might be computationally very intensive, especially if there are many copula and marginal distributions' parameters to estimate. The estimation procedure can be simplified by the so called *inference for the margins* method (IFM) where, in the first step, we estimate the margin parameters maximizing only the likelihood related to the margins and then, in the second step, we estimate the copula parameters maximizing the copula likelihood function. The IFM estimator is not necessarily the same as the MLE estimator; it is only an approximation and could serve as an initial value for the exact MLE procedure. Another possible method called canonical maximum likelihood (CML) transforms the marginal data into uniform variants based on their empirical distributions. Finally, the copula itself can be estimated non-parametrically, typically using a kernel in order to obtain a smoothed empirical copula.

In order to simulate a multivariate distribution, e.g. times of default $\langle \tau_1, \ldots, \tau_n \rangle$, specified by the marginal distributions $Q_1(\tau_1), \ldots, Q_n(\tau_n)$, and a copula $C(u_1, \ldots, u_n)$, all we need to do is to simulate $\langle u_1, \ldots, u_n \rangle$ from C and apply the transformations $\tau_j = Q_j^{-1}(u_j)$. This is relatively simple for the Gaussian or Student t copula. Given a correlation matrix Σ to simulate from Gaussian copula distribution, one needs just to sample $\mathbf{x} = \langle x_1, \ldots, x_n \rangle$ from the corresponding multivariate Gaussian distribution $N(\mathbf{x}; 0, \Sigma)$ (e.g., based on the Choleski decomposition of Σ) and then use the Q-Q transformations $\tau_j = Q_j^{-1}(\Phi(x_j))$. One can proceed similarly in case of the Student t distribution.

For a general copula simulation we need to find the conditional distributions:

$$C_k\left(u_k\middle|u_1,\ \ldots,u_{k-1}\right)=\Pr\left[U_k\le u_k\middle|U_1=u_1,\ \ldots,U_{k-1}=u_{k-1}\right].$$

The conditional distributions can be obtained by partial differentiation of the copula function. For example, for a differentiable bivariate copula $C(u,v)$ it follows that

$$c_u(v)=\Pr\left[V\le v\middle|U=u\right]=\lim_{\Delta u\to 0}\frac{C(u+\Delta u,v)-C(u,v)}{\Delta u}=\frac{\partial C}{\partial u}.$$

Hence, to sample a pair (u,v) from C we can just sample independent variable $u,w\in[0,1]$ from the uniform distribution and set $v=c_u^{-1}(w)$ to get the second draw.

For a multi-variate copula C let $C_k(u_1,\ldots,u_k)=C(u_1,\ldots,u_k,1,\ldots,1)$ where $k=2,\ \ldots,n$. Then

$$C_k\left(u_k\middle|u_1,\ \ldots,u_{k-1}\right)=\Pr\left[U_k\le u_k\middle|U_1=u_1,\ \ldots,U_{k-1}=u_{k-1}\right]= \\ =\frac{\partial^{k-1}C_k(u_1,\ldots,u_k)}{\partial u_1\ldots\partial u_{k-1}}\Big/\frac{\partial^{k-1}C_{k-1}(u_1,\ldots,u_{k-1})}{\partial u_1\ldots\partial u_{k-1}}.$$

The Monte Carlo simulation then proceeds as follows:

- Simulate a random value u_1 from $U(0,1)$,
- Simulate a random value u_2 from $C_2\left(\cdot\middle|u_1\right)$,
- ...
- Simulate a random value u_n from $C_n\left(\cdot\middle|u_1,\ \ldots,u_{n-1}\right)$.

As in the bivariate case, a value u_k is simulated from $C_k\left(u_k\middle|u_1,\ \ldots,u_{k-1}\right)$ by sampling $w\in[0,1]$ from the uniform distribution and setting $u_k=C_k^{-1}\left(w\middle|u_1,\ \ldots,u_{k-1}\right)$. Therefore, the procedure is relatively efficient if the conditional distribution can be inverted. It is easy to see that the distributions C_k are analytical for Archimedean copulas (with φ and φ^{-1} analytical) and can be inverted for the particular copulas listed above (Gumbel, Clayton, and Frank). Without analytical invertibility, the equation $w=C_k\left(u_k\middle|u_1,\ \ldots,u_{k-1}\right)$ must be solved numerically and the procedure obviously becomes computationally more intensive.

Implied Copula

A popular approach to CDO tranche valuation has been proposed by Hull and White (2006). Although the method is called "*implied copula*", it does not, in fact, use the concept of copulas directly. The idea is to focus on the probability distribution of the portfolio losses, or equivalent, given a constant LGD, on the distribution of default rates. The non-parametric loss distribution is calibrated from available tranche quotations. Then, it can be used to value any other non-standard tranche. The "implied copula" is, therefore, in a way implicitly hidden in the loss probability distribution. It is not explicitly specified and calibrated.

Specifically, let us consider a homogenous portfolio with constant intensity of default conditional on a systematic factor M taking finitely many values $m_1, \ldots, m_h$. Hull and White (2006) propose to specify a reasonable list of possible default intensity values $\lambda_1, \ldots, \lambda_h$ and calibrate their probabilities $p_1, \ldots, p_h$. We can assume a constant recovery rate R, or, in a more general approach, conditional recovery rates $R_1, \ldots, R_h$. The recovery rates can be linked to the default intensities via a functional form, e.g. exponential, as proposed in Cantor et al. (2005). Note that the systematic factor numerical values have no meaning; we just have scenarios indexed $1, \ldots, h$ with the specified default intensities (and recovery rates). For any single name i, the conditional probability of default in a time horizon t is

$$Q(t|j) = \Pr[\tau_i \le t | M = m_j] = 1 - e^{-\lambda_j T},$$

and the unconditional probability of default:

$$Q(t) = \Pr[\tau_i \le t] = \sum_{j=1}^{h} p_j \left(1 - e^{-\lambda_j T}\right).$$

In the case of the LHP (Large Homogenous Portfolio) model, all randomness is eliminated and both the default rate and the loss rate are determined by the scenario. Therefore, given the scenario probability distribution, we can express the conditional loss $L(t|j) = (1 - R_j)(1 - e^{-\lambda_j t})$, and hence the tranche $[A, B]$ conditional loss according to (5.12), the conditional default leg $DL^{A,B}(0|j)$ value according to (5.13), and the risky annuity $DV^{A,B}(0|j)$ value according to (5.14). Given the market quotes $U^{A,B}$ of the up-front payment and the spread $S^{A,B}$ we can calculate the conditional premium leg value

$$PL^{A,B}(0|j) = U^{A,B} + S^{A,B} DV^{A,B}(0|j),$$

and the conditional premium receiver tranche market value

$$MV^{A,B}(0|j) = PL^{A,B}(0|j) - DL^{A,B}(0|j).$$

Finally, all the conditional quantities can be easily made unconditional; in particular, the unconditional receiver tranche market value is:

$$MV^{A,B}(0) = \sum_{j=1}^{h} p_j MV^{A,B}(0|j).$$

The calibration goal is to find $p_1, \ldots, p_h$ subject to the standard conditions $\sum p_j = 1, p_j \geq 0$ in order to make $MV^{A,B}(0)$ equal to zero, or at least as close to zero as possible, taking into account all quoted tranches $[A_k, B_k]$. One may also include the index quote condition (corresponding to the tranche [0,100%]). For example, for iTraxx Europe there are, besides the index quote, only six tranches with the attachment/detachment points 0, 3, 6, 9, 12, 22, 100%, hence we should be able to make the market values equal to zero if there are eight or more scenarios. However, Brigo et al. (2010) note that one may need up to 30 scenarios, i.e. $h = 30$, or even 125 scenarios listing all possible default rates (less than 100%) on the index portfolio, in order to fit the quotes with sufficient precision. The number of scenarios has another unpleasant effect in that the optimization has multiple solutions depending on the initial guess. It is shown that this issue can be overcome by imposing an additional smoothing condition on the probability distribution.

Expected Tranche Loss

Even though the implied copula model is a powerful empirical approach, it is still a static model where the loss distribution is obtained only from single maturity quotes. Indeed, it turns out that for the implied copula model, and likewise for any other copula model, the loss distributions might differ for different maturities. The Expected Tranche Loss approach (Walker 2006) is a relatively straightforward, model-free way to extract the key information needed to value CDO tranches based on quotes across maturities and attachment/detachment points. It is based on the simple observation that the key inputs in the equilibrium spread formula

$$S_0^{A,B} = \frac{E_0\left[DL^{A,B}(0)\right] - U_0^{A,B}}{E_0\left[DV^{A,B}(0)\right]} = \\ = \frac{\sum_{j=1}^{n} D\left(0.5t_{j-1} + 0.5t_j\right)\left(E\left[L^{A,B}\left(t_j\right)\right] - E\left[L^{A,B}\left(t_{j-1}\right)\right]\right) - U_0^{A,B}}{\sum_{j=1}^{N}\left(t_j - t_{j-1}\right)D\left(t_j\right)\left(1 - E\left[L^{A,B}\left(t_j\right)\right]\right)} \tag{5.23}$$

are the expected tranche losses $E[L^{A,B}(t_j)]$ where t_j are typically the quarterly premium payment times and $[A, B]$ are the standard attachment/detachment points. Now, in a dynamically consistent model, the expected tranche losses should be the same independently of the CDO tranche maturity. Therefore, the idea is to collect all the available upfront fees and spread quotes $U_{0,T}^{A,B}, S_{0,T}^{A,B}$ for different maturities T and, in a way, bootstrap the expected tranche loss $E[L^{A,B}(t_j)]$ values from the quotes. Then, using appropriate interpolations, the formula (5.23) can be used to value consistently any other non-standard tranche with an arbitrary maturity. One possible approach is to look directly for $E[L^{A,B}(t_j)]$. However, as with the compound and base correlation, we then have an issue with interpolation for non-standard tranches, and it is more appropriate to estimate the expected equity losses $g(t,B) = E[L^{0,B}(t)]$ for the "base" tranches $[0,B]$. Given these quantities we know that

$$E\left[L^{A,B}(t)\right] = \frac{g(t,B) - g(t,A)}{B - A},$$

and so (5.23) can be rewritten as a set of equations for the available spread and up-front fees with unknown values $g(t_j, B_k)$, for standard payment times t_j and detachment points B_k (e.g. 0, 3, 6, 12, 22, 100%). Since the equations do not have to have an exact solution, we have to run an optimization problem minimizing a mispricing function. For a quoted maturity T and tranche $[A,B]$ the standard mispricing formula is

$$MS_{T,A,B} = \frac{S_T^{A,B} - S_T^{A,B,\mathrm{mid}}}{\left(S_T^{A,B,\mathrm{ask}} - S_T^{A,B,\mathrm{bid}}\right)/2}$$

where $S_T^{A,B}$ is the model implied spread and $S_T^{A,B,\mathrm{bid}}$, $S_T^{A,B,\mathrm{ask}}$ are the quotes. Overall, we minimize the sum of squared mispricings

$$\sum_{\text{quoted tranche } (T,A,B)} MS^2_{T,A,B}. \tag{5.24}$$

In addition, we have to formulate some necessary constraints on the expected equity tranche loss values $g(t_j, B_k)$:

$$0 \le g(t_j, B_k) \le 1, g(t_j, 0) = 0,$$
$$g(t_j, B_k) \ge g(t_{j-1}, B_k),$$
$$g(t_j, B_{k-1}) \le g(t_j, B_k),$$
$$g(t_j, B_{k+1}) \le g(t_j, B_k) + (B_{k+1} - B_k)\frac{g(t_j, B_k) - g(t_j, B_{k-1})}{B_k - B_{k-1}}.$$

The last inequality equivalent to the ETL inequality between a more and a less senior tranche:

$$E\left[L^{B_k,B_{k+1}}(t_j)\right] \le E\left[L^{B_{k-1},B_k}(t_j)\right].$$

Brigo et al. (2010) tested the model on DJ-iTraxx pre-crisis data (2003–2006) and concluded that the pricing error never exceeded by more than 20% the bid-ask range with spline interpolations, i.e. the model was able to fit the market data very well.

To conclude, instead of going into arbitrary structural assumptions, like the Gaussian or other copulas, the ETL method focuses on the most direct market quantities that are embedded in the quotations. Compared to the copula models, the ETL approach provides acceptable market fit and at the same time does not lead to inconsistencies in the time dimension.

Generalized Poisson Loss Model

Another idea proposed by Brigo et al. (2006a, b) is to model the dynamics of the portfolio loss as a stochastic process $Z(t)$ characterized as a mix of independent Poisson processes $N_j(t)$ with different amplitudes α_j. The counting Poisson process starts at $N_j(0) = 0$ and from time to time jumps up by one, i.e. $dN_j = 0$ or 1 where the probability of a nonzero jump is $\lambda_j(t)dt$ and $\lambda_j(t)$ is the Poisson intensity. If we had a portfolio of independent exposure with identical default intensity λ, then one Poisson process with this intensity would count random defaults on the portfolio over time. However, since these dependencies and multiple defaults might take place over a short time period, it is proposed to add another independent process multiplied, for example, by the amplitude 2, i.e. modeling the occurrence of double defaults. Then, to improve the model, we can add a third process multiplied by a larger amplitude, and so on. Therefore, the Generalized-Poisson process we want to apply will have the following form:

$$Z(t) = \sum_{j=1}^{n} \alpha_j N_j(t),$$

where the Poisson processes are independent, $\alpha_1 < \alpha_2 < \ldots$, and the intensities $\lambda_j(t)$ are deterministic functions of time. One possibility is to use this process as the default-counting process and multiply it by a fixed LGD. However, Brigo et al. (2006a, b) propose to model the loss directly, defining it as $L_t = \min(Z_t, M')/M'$ where $M' \geq M$ is larger than or equal to the number of names in the portfolio. In this way, we make sure that the loss does not exceed 100% and at the same time allow that the minimum jump size can be less than $1/M$ (e.g., corresponding to LGD/M).

The advantage of the Generalized-Poisson process is that it can be handled relatively well analytically; the marginal distributions of $Z(t)$, i.e. the loss distributions conditional on t, can be obtained from the characteristic function

$$\varphi_{Z(t)}(u) = E_0[\exp(iuZ(t))] = E_0\left[\exp\left(iu\sum_{j=1}^{n}\alpha_j N_j(t)\right)\right] =$$
$$= \prod_{j=1}^{N} E_0\left[\exp\left(iu\alpha_j N_j(t)\right)\right] = \prod_{j=1}^{N} \varphi_{N_j(t)}\left(u\alpha_j\right).$$

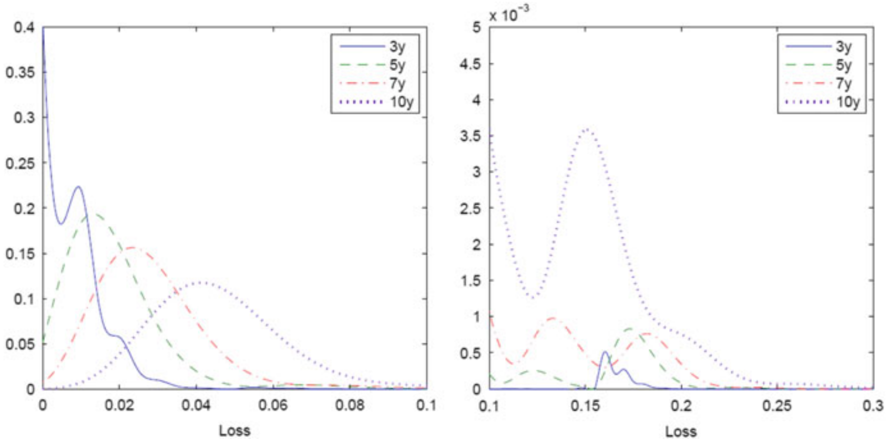


Fig. 5.14 Loss distribution evolution of the GPL model with minimum loss jump size of 50 bp on all the quoted maturities up to 10 years, drawn as a continuous line (source: Brigo et al. 2006b)

Now, the point is that for each Poison process the characteristic function is known leading to

$$\varphi_{Z(t)}(u) = \exp\Big(\sum_{j=1}^{N} \Lambda_j(t)(e^{iu\alpha_j} - 1)\Big),$$

where $\Lambda_j(t) = \int_0^t \lambda_j(s)ds$ is the cumulated intensity. Finally, the probability distribution of $Z(t)$ can be computed via the inverse Fourier transformation of the characteristic function. Note that the technique is analogous to the one employed by the CreditRisk+ methodology (Sect. 4.3).

Once we are able to calculate the loss distribution and the expected tranche losses, the calibration procedure can be formulated, as usual based on available market quotes and minimizing an objective error function. First we need to set the integer $M' \geq M$. Brigo et al. (2006a, b) recommend $M' = 200$ in the case that $M = 125$ (note that $125/200 = 62.5\%$ corresponds to a "normal" LGD value), although the value of M' can be arbitrarily large. The next step is to choose α_1 (typically equal to 1) and a non-decreasing function $\Lambda_1(t)$ piecewise constant in the tranche maturities. Then choose α_2 and $\Lambda_2(t)$, and so on, until the default intensities are negligible or the fit is sufficiently good. The proposed fit function (5.24) is the same as for the ETL model.

Brigo et al. (2006a, b) report very good empirical results based on 2005 market data. Figure 5.14 shows an interesting example of the multimodal loss distributions for various maturities with modes corresponding to different amplitudes of the Generalized-Poisson distribution.

5.4 Dynamic Intensity of Default Modeling

The credit derivatives valuation models discussed so far can be characterized as static models. The probabilities of defaults are estimated for different time horizons from today's perspective, and even future intensities of default are modeled as the fixed forward intensities of default, implied by the current default probabilities. However, looking, for example, at the development of short term market CDS spreads, the intensities of default are not deterministic, but develop apparently stochastically over time. Therefore, default intensities should be modeled as random over time, i.e. stochastic as well. The stochastic approach might improve the valuation of various complex credit derivatives, and it is necessary to value options on CDS and similar products.

The challenging problem of dynamic credit risk modeling can be compared to interest rate modeling. Today, we may derive the static term structure of the (risk free) interest rates from bonds, interest rate swaps, or other instruments which can be valued, like most interest rate products, just by discounting the fixed cash flows. But to value interest rate options and other derivatives, we need to introduce a stochastic model of interest rates. This is more difficult than modeling a stock price or an exchange rate because we need to capture the development of the full term structure of interest, for example, of the curve of zero coupon interest rates with all possible maturities $\langle R(t,T); T \geq t \rangle$, from the perspective of time t. We know today's term structure $\langle R(0,T); T \geq 0 \rangle$, but for $t > 0$ all the variables $R(t,T)$ are stochastic. The advanced models like Heath, Jarrow, and Morton (HJM), or the LIBOR-market model indeed describe the dynamics of the full term structure (see e.g. Hull 2009). We will firstly look in more detail at the popular *Vasicek's model* which belongs to the class of the *short-rate models*, primarily modeling only the instantaneous interest rate $r(t)$, and implying all the other interest rates and variables from the model. The approach is similar to the intensity of default stochastic *reduced form models* outlined below. Moreover, the interest rate and probability of default modeling should go hand in hand in the more advanced models, since there might be a certain degree of interdependence between intensity of default and the interest rate level. We will also look at alternative *structural models* similar to the KMV EDF approach, where default is triggered by the stochastic asset value breaking a default threshold. The probability of default is then determined by the distance of the asset value from the threshold and the process dynamics.

Vasicek's Interest Rate Model

Vasicek's model (Vasicek 1977) can be described by the stochastic differential equation

$$dr = a(b - r)dt + \sigma dz, \tag{5.25}$$

where a, b, and σ are constants, $dr = dr(t)$ is the change of the instantaneous interest rate $r(t)$ over an infinitesimal time interval of length dt, and dz is the random change of the Wiener process; i.e., $dz \sim N(0, dt)$. Without going into the fundamentals of the stochastic processes, we may interpret the Eq. (5.25) as a Monte Carlo simulation algorithm described as follows:

1. Set $r(0) = r_0$, fix a time horizon T, and an elementary time step $\Delta t = T/N$.
2. Given $r(t)$, sample Δz from $N(0, \Delta t)$ and set $r(t + \Delta t) = r(t) + a(b - r(t)) + \sigma\Delta z$.
3. Repeat the step 2 until $T = N\Delta t$, and interpolate linearly $r(t)$ between the calculated values $r((j-1)\Delta t)$ and $r(j\Delta t)$ for $j = 1, \ldots, N$. The resulting function $r : [0, T] \to R$ is one short-rate simulated path.
4. Repeat the steps 2 and 3 to obtain a larger number of simulated paths.

Note that all we know today is the instantaneous interest rate r_0. All the future rates are unknown, i.e. stochastic, but with a distribution governed by the Eq. (5.25). The logic of the equation is that there is a random change of the rate given by σdz, but unlike to stocks, there is no positive (or negative) drift, rather, a tendency to revert to the mean level given by the parameter b, with the speed given by $a > 0$. Figure 5.15 shows a sample path simulated according to (5.25) with the indicated parameters. The arrows indicate the tendency of the stochastic interest rate r to revert to the long term mean $b = 4\%$

In order to value other interest rate instruments, the model needs to be set up in the world that is risk-neutral with respect to the instantaneous interest rate $r(t)$; i.e., investors investing into any security require just the return $r(t)dt$ over the period $[t, t + dt]$ (and not more), regardless of the asset's level of risk. If f denotes the value of the security, then the risk-neutral principle can be formally expressed by the equation:

$$\frac{df}{f} = rdt + \sigma_f dz.$$

Technically we need to change the real-world probability measure to a new measure that is also called *forward risk-neutral*, with respect to a *numeraire* (Hull 2009). Here the numeraire is the *money-market account* with the initial value $g(0) = 1$, and accruing the instantaneous interest rate; i.e. $dg = rgdt$. Therefore, the value of the money-market account at time t can be expressed as:

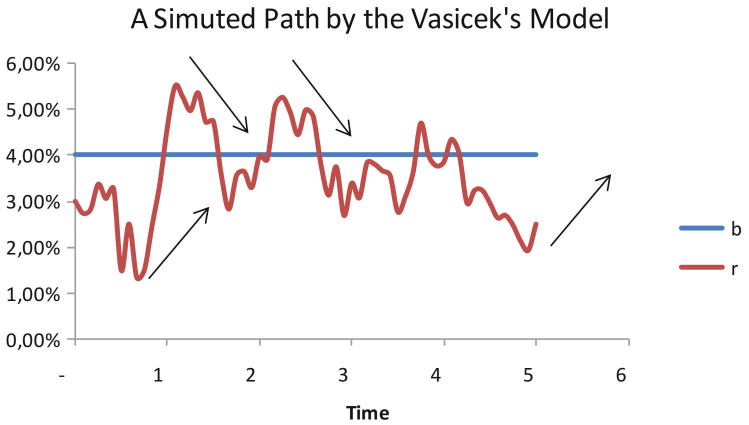


Fig. 5.15 A path simulated by Vasicek's model with the parameters $r_0 = 3\%$, $a = 0.1$, $b = 4\%$, and $\sigma = 2\%$

$$g(t) = \exp\left(\int_0^t r(\tau)d\tau\right). \tag{5.26}$$

Note that $g(t)$ is random, since the integral in (5.26) is taken over a random path of r. We say that a measure is forward risk neutral with respect to g, if, for any security derivative with the same source of uncertainty (i.e., depending on the same underlying Wiener process or processes) f/g is a martingale; i.e.,

$$\frac{f(t)}{g(t)} = \hat{E}\left[\frac{f(T)}{g(T)}|t\right], t < T,$$

where $\hat{E}\left[\cdot|t\right]$ denotes the conditional expected value with respect to the measure, and from the perspective (given information) of time t. In particular, if the payoff of a derivative f_T at time T is known, then its time zero market value can be expressed as:

$$f_0 = \hat{E}\left[\exp\left(-\int_0^T r(\tau)d\tau\right)f_T\right]. \tag{5.27}$$

So, if $P(t,T)$ denotes the market value of the zero coupon bond paying 1 unit at time T, then

$$P(t,T) = \hat{E}\left[\exp\left(-\int_0^T r(\tau)d\tau\right)\right]. \tag{5.28}$$

Once we get the values of $P(t,T), T > t$, we also have the full term-structure of interest rates $R(t,T) = \frac{-1}{T-t}\ln P(\mathrm{t},\mathrm{T})$ in the continuous compounding for all maturities. Note that the expected value in (5.28) depends only on the initial value $r = r(t)$, on times t, T, and on the fixed parameters of the stochastic process a, b, σ. The values in (5.28) and, in general, in (5.27), can be estimated numerically by a Monte Carlo simulation. However, one of the reasons for the popularity of Vasicek's model is that the solution for $P(t,T) = f(r,t)$ can be found in an *affine form*:

$$f(r,t) = e^{-(\alpha(t)+\beta(t)r)(T-t)}, \tag{5.29}$$

or equivalently, $R(r,t,T) = \alpha(t) + \beta(t)r$. The key tool for getting the solution and, in general, to obtain various results in stochastic calculus, is *Ito's Lemma,* which says that if $x(t)$ is a stochastic process satisfying the equation $dx = a(x,t)dt + b(x,t)dz$ and if $f(x,t)$ is a sufficiently differentiable function of two independent variables, then $f = f(x(t),t)$ is a new stochastic process satisfying the equation:

$$df = \left(\frac{\partial f}{\partial x}a + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}b^2\right)dt + \frac{\partial f}{\partial x}bdz.$$

In the forward risk neutral probability measure, for $f(r,t) = P(r,t,T)$, the drift coefficient must be equal to rf and thus, in the case of Vasicek's model, we obtain the classical Black-Scholes-Merton partial differential equation:

$$\frac{\partial f}{\partial r}a(b-r) + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial r^2}\sigma^2 = fr.$$

The partial differential equation can be solved analytically for an f in the form (5.29), and with the boundary condition $f(r,T) = 1$. The final solution is usually written as follows:

$$P(t,T) = A(t,T)e^{-B(t,T)r(t)},$$
$$B(t,T) = \frac{1-e^{-a(T-t)}}{a}, \text{ and}$$
$$\mathrm{A}(\mathrm{t},\mathrm{T}) = \exp\left(\frac{(B(t,T)-T+t)(a^2b-\sigma^2/2)}{a^2} - \frac{\sigma^2 B(t,T)^2}{4a}\right).$$

The interest rate $r(t)$ can also be characterized as a normal variable, with an analytically expressed mean and variance. This indicates a disadvantage of the model, since $r(t)$ may be negative with a positive probability. On the other hand, this significantly simplifies the simulation of future interest rates $r(t)$, and thus, of $P(t,T)$, $R(t,T)$ and so on. Moreover, continuing in the analysis outlined above, it is possible to obtain an analytical solution for options on zero coupon bonds, fixed coupon bonds, and other "plain vanilla" interest rate derivatives.

Besides the possibility of negative interest rates, another shortcoming of the model is its limited flexibility for capturing the initial term structure of interest rates. There are a number of other short rate models like the Cox, Ingersoll, and Ross model, the Ho-Lee model, or the Hull-White; one or more factor models which improve on Vasicek's model, but typically at the cost of lower analytical tractability.

Structural Stochastic Models

Before discussing the intensity of default models sharing certain similarities with the short rate models, let us look at the class of structural models already introduced in Sect. 4.5.

In the Black-Scholes-Merton approach, $A(t)$ is the asset value at time t following the geometric Brownian motion stochastic differential equation:

$$dA = (\mu - \gamma)Adt + \sigma Adz, \tag{5.30}$$

where μ is the mean rate of return on the assets, γ the dividend payout ratio, and default at time T is defined by the condition $A(T) < D$, where D is a default threshold. Then, the probability of default at time T is given by the distance to default as $\Phi(-DD)$ where $DD = d_2$; see (4.16) and Fig. 5.16.

Unfortunately, since we need to capture the random time to default, this model is not sufficient. It could be argued that default happens once the default threshold is passed. In fact, one of the problems of the Black-Scholes-Merton model is that it may happen that $A(s) < D$ for some $s < T$, yet $A(T) \geq D$. So the debtor who, in fact, should have defaulted during the period $(0, T)$, does not default at time T, according to the model. This leads to the concept of *first-passage models*, where default happens in the time interval $[t, T]$, anytime there is an $s \in [t, T]$, so that $A(s) < D$. Given a path for A, the time of default can be defined as $\tau = \inf\{s; A(s) < D\}$. The probability of default, $p(t, T)$ in $[t, T]$, can still be expressed analytically, applying the technique for the valuation of barrier options (Duffie and Singleton 2003)

$$p(t, T) = H(X_t, T - t),$$

where

$$X_t = \frac{\ln A(t) - \ln D}{\sigma},$$

$H(x, s) = \Phi\left(\frac{x+ms}{\sqrt{s}}\right) - e^{-2mx}\Phi\left(\frac{-x+ms}{\sqrt{s}}\right)$, and $m=\frac{\mu-\gamma-\sigma^2/2}{\sigma}$. The most serious flaw of the model is that it implies unrealistically low forward intensities of default in a short horizon if the initial asset value A_t is not close to the debt threshold D (Duffie and Singleton 2003). Figure 5.16 shows the default swap spreads implied by the

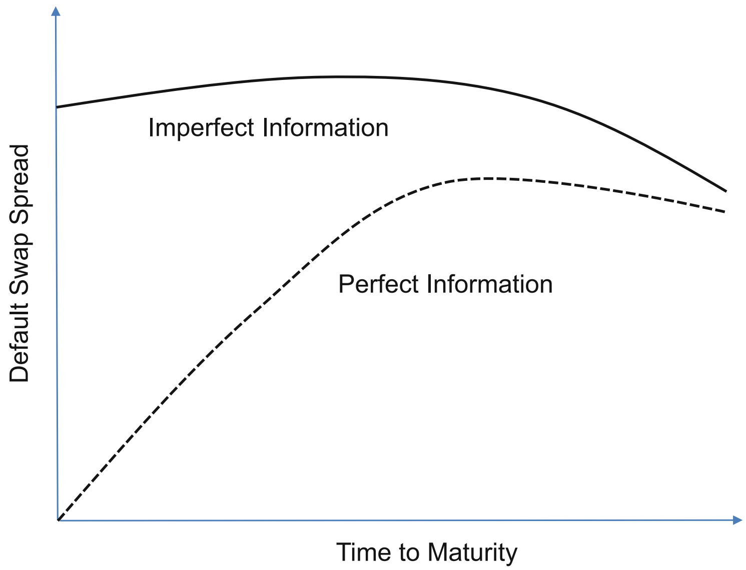


Fig. 5.16 Default swap spreads implied by the first-passage model with perfect and imperfect information

first passage model. The curve starting at the origin, and labeled "Perfect Information", illustrates that the model swap spread is initially very low as the default intensities are close to zero for the first few months, and then it increases rapidly and slightly declines for longer maturities.

The empirical evidence says that the default spreads are more-or-less flat, may increase for good ratings, or decrease for bad ratings, but never start at zero as implied by the first-passage model. This inconsistency can be partially solved by a number of methods. For example, the classical first-passage model assumes perfect knowledge of the assets value and of the default threshold which is not completely realistic. If we allow for imperfect information, i.e. assuming that the initial asset value $A(0)$ (and/or the default threshold D) is drawn from a random distribution, then we may get a more realistic picture (see the curve labeled "Imperfect Information" in Fig. 5.16).

Another way to make the first-passage model more realistic is to add the possibility of jumps into the stochastic model for $A(t)$; i.e.,

$$dA = (\mu - \gamma - k)Adt + \sigma Adz + AdJ,$$

where J is a compound Poisson process, and k is the mean jump size multiplied by the arrival intensity of jumps. The assets of a firm may, indeed, jump up and down for many different reasons, and that may cause a sudden default, even if the initial distance to default was relatively large. The model generally becomes less tractable, but if the jump sizes are assumed to be lognormal, then there is an analytical solution provided by Zhou (2001).

Reduced Form Models

Reduced form models, treating just the default intensity as a stochastic process, present a natural alternative to the structural models. A clear advantage is that probabilities of default are empirically observable (e.g., through CDS quotations), while asset value processes are usually latent. It should be easier to set up, calibrate, and back-test a given stochastic intensity model, while the structural models are usually quite inflexible, or too complex, as discussed above. On the other hand, in the structural approach, the term structure of default probabilities, as well as the arrival of default, is defined by the asset value stochastic model; the value of $A(t)$ and the parameters of the model determine the probabilities of default $p(t,T), T > t$, and at the same time, any particular path of $A(s), s \geq t$ determines the time of default. This is not the case of stochastic intensity models. For a particular path of $\lambda(s), s \geq t$, we just know that the probability of default at any time interval of the form $[s, s + ds]$, conditional on survival until s, is $\lambda(s)ds$, but the specific time of default is not determined. To incorporate the time of default, the concept of a *doubly stochastic process* is introduced (Duffie and Singleton 2003). Hence, there are two layers of uncertainty:

1. the stochastic intensity process, and
2. the Poisson arrival of default process, conditional on the default intensity process.

For example, let us assume that the stochastic intensity is governed in an analogy to Vasicek's short rate model by the mean reversion stochastic differential equation:

$$d\lambda = a(b - \lambda)dt + \sigma dz. \tag{5.31}$$

If our goal is just to simulate random times of default on a time interval $[0, T]$, then we may proceed as follows:

1. Fix a small $\Delta t = T/N$ and set $\lambda(0) = \lambda_0$.
2. For a given $s = j\Delta t, j = 0, \ldots, N-1$, provided there was no default until s, sample Poisson default on $[s, s + \Delta t]$, with the probability $\lambda(s)\Delta t$. If there is a sampled default, set $\tau = s + \Delta t/2$. If there is no default, sample $\Delta z \sim N(0, \Delta t)$, and set $\lambda(s + \Delta t) = \lambda(s) + a(b - \lambda(s)) + \sigma\Delta z$.
3. Repeat Step 2 until a default, or until time T. If no default was sampled until T, set $\tau = +\infty$.

The naïve Monte Carlo procedure, applicable to an arbitrary stochastic default intensity model, can be slightly improved using the concept of a compensator:

Compensator simulation: Simulate the cumulative intensity of default:

$$\Lambda(t) = \int_0^t \lambda(s)ds \ \text{ for } \ t \in [0, T],$$

and independently sample a value u from the standard (unit mean) exponential distribution. Let τ be chosen so that $\Lambda(\tau) = u$, i.e. $\tau = \Lambda^{-1}(u)$, if $u \le \Lambda(T)$, and $\tau = +\infty$ otherwise.

To explain, a random variable U has the standard exponential distribution, if its cumulative distribution function is $\Pr[U \le t] = 1 - e^{-t}, t \ge 0$; i.e., its survival function $\Pr[U > t] = e^{-t}$. Thus, U would be the time to default variable if the intensity of default was 1; therefore, its mean value equals 1. The compensator function $\Lambda(t)$ transforms the real time to the unit intensity time to default time scaling by the given intensities of default $\lambda(s), s \in [0, T]$. Formally,

$$\Pr[\tau > t] = \Pr\left[\Lambda^{-1}(U) > t\right] = \Pr[U > \Lambda(t)] = e^{-\Lambda(t)}.$$

Recall that according to (5.2), the survival probability conditional on a specific default intensity path is:

$$S(t|\langle\lambda(s); 0 \le s \le t\rangle) = \exp\left(-\int_0^t \lambda(s)ds\right) = e^{-\Lambda(t)}, \tag{5.32}$$

consequently, the two step compensator simulation indeed generates the desired distribution of times to default.

Conditioning (5.32) on the information available at time 0, we get that:

$$S(0,t) = \Pr\left[\tau > t \middle| 0\right] = \hat{E}\left[\exp\left(-\int_0^t \lambda(s)ds\right)\right]. \tag{5.33}$$

Note that the survival probability $S(t, T) = \Pr\left[\tau > T \middle| t\right]$ for the time horizon T, from the perspective of time t, has exactly the form of (5.28), as in Vasicek's model, and thus it can be solved analytically in an affine form (5.29).

If the default intensity model can be treated analytically, as in the case of the affine specification above, then we can, in fact, proceed much more efficiently.

Inverse survival function simulation: Generally, if the survival probability $S(t) = S(0, t)$ as a decreasing continuous function $S : [0, +\infty] \to [0, 1]$ can be analytically calculated, then the time to default is easily simulated sampling the variable u from the uniform distribution $U(0, 1)$, and calculating $\tau = S^{-1}(u)$.

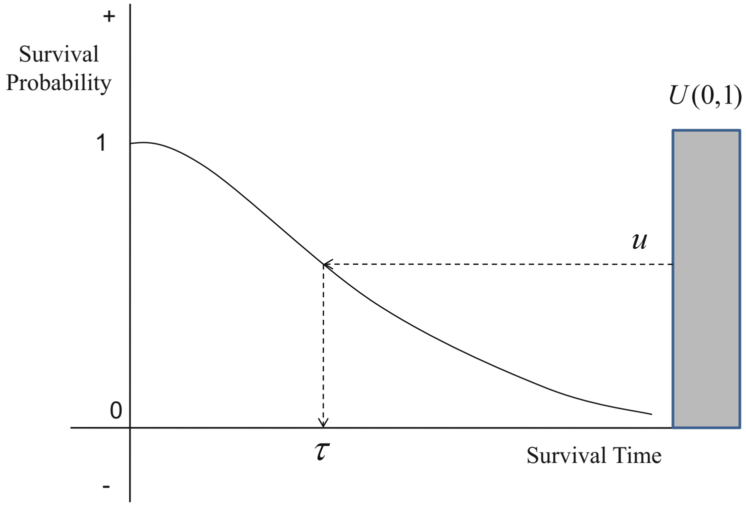


Fig. 5.17 Simulating time to default with the inverse survival function

Indeed, $\Pr[\tau > t] = \Pr\left[S^{-1}(u) > t\right] = \Pr[u < S(t)] = S(t)$; see Fig. 5.17 for an illustration.

As in the case of interest rate modeling, there are many advanced default intensity models that have certain advantages compared to Vasicek's model, but become less analytically tractable. Among the most popular are the mean reverting model with jumps, the Cox-Ingersoll-Ross (CIR) model, and the Heath, Jarrow, and Morton (HJM) models.

The mean-reverting process with jumps can be written as:

$$d\lambda = \kappa(\gamma - \lambda)dt + dJ. \tag{5.34}$$

So, there are independently distributed jumps at Poisson arrival times, with certain intensity c, and between the jumps the process is exponentially reverting to the mean γ at the rate κ. If the jumps are, for example, exponentially distributed, then the process belongs to the class of basic affine processes; i.e., the survival probability from t to T can be expressed in the form:

$$S(t, T) = e^{-(\alpha(t) + \beta(t)\lambda(t))(T - t)}, \tag{5.35}$$

where $\alpha(t)$ and $\beta(t)$ are determininstic functions depending only on the process parameters, and on the argument t (see Duffie and Singleton 2003). The logic of the model is that the worsening of credit quality usually happens in jumps, while improvement is gradual; see Fig. 5.18 for an illustration of a simulated path.

Another popular model used when modeling interest rates and default intensities, is the CIR—Cox, Ingersoll, and Ross (1985) model, for which:

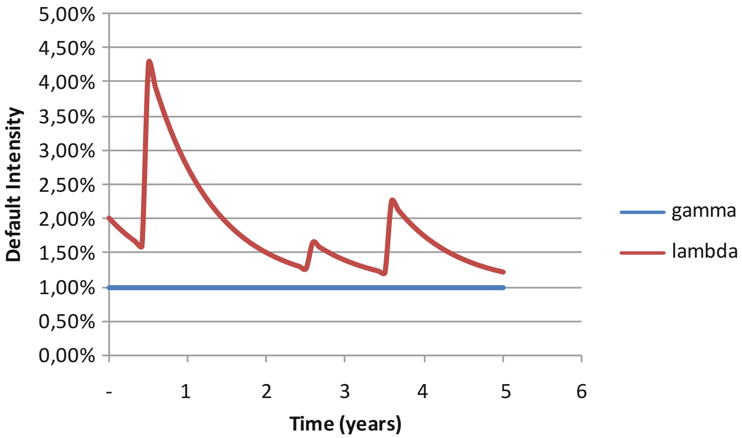


Fig. 5.18 A simulated path of the mean-reverting process with jumps ($\kappa = 0.1, \gamma = 1\%, c = 75\%$, and jump mean 2%)

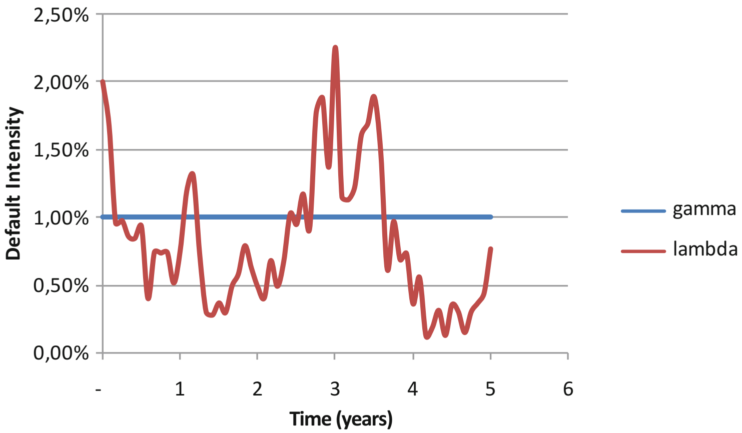


Fig. 5.19 A simulated path of the CIR process ($\kappa = 0.1$, $\gamma = 1\%$, and $\sigma = 10\%$)

$$d\lambda = \kappa(\gamma - \lambda)dt + \sigma\sqrt{\lambda}dz.$$

The model is similar to Vasicek's model (5.31), but the term σdz is replaced with $\sigma\sqrt{\lambda}dz$. It can be argued that this random change better fits the empirical reality, but most importantly, the intensity λ (or interest rate) always stays nonnegative in the CIR model, while Vasicek's model admits negative values. The model also has an affine form solution (5.35), although more complex than in the case of Vasicek's model. The simulated paths are not as jumpy as in the case of the mean reverting process with jumps—compare Figs. 5.18 and 5.19.

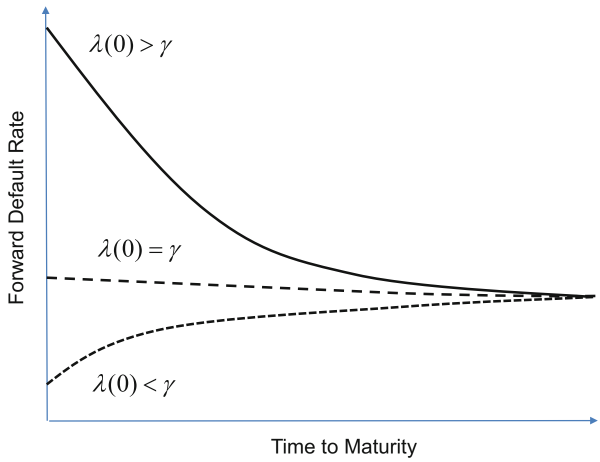


Fig. 5.20 Typical forward default rates imlied by the CIR and the mean reverting jump model with varying initial intensities

Even though the paths of the two processes have different shapes, the forward default rate term structures are very similar. Note that the forward default rates are implied by the expected survival probabilities (5.33), and by the relationship

$$\lambda(0,t) = \frac{-1}{S(0,t)} \frac{dS(0,t)}{dt}. \tag{5.36}$$

Figure 5.20 shows a typical term structure of the forward default rates implied by the CIR model depending on the relationship between the initial intensity $\lambda(0)$ and the long term mean reverting intensity level γ. The forward default rate of course reverts to the base level γ, but it is slightly decresing also if $\lambda(0) = \gamma$ due the convexity effect (Jensen's inequality) as it is given by (5.36) and (5.33).

Duffie and Singleton (2003) generate forward intensities in the CIR model with varying initial intensities, and then calibrate the parameters of the mean-reverting model to match the moments of λ, implied by the CIR model. It turns out that that the resulting forward intensities implied by the mean reverting model are almost identical and have the same term structure pattern as shown in Fig. 5.20.

The model can be further combined, for example, considering a combination of the CIR model and the jump model:

$$d\lambda = \kappa(\gamma - \lambda)dt + \sigma\sqrt{\lambda}dz + dJ.$$

The more complicated CIR model with jumps does not necessarily bring a significant improvement, as indicated by the comparison of the pure jump and the pure CIR diffusion models above. Nevertheless, it still belongs to the class of affine intensity models.

The disadvantage of the intensity models, similar to the short rate models, is the limited flexibility in fitting the initial term structure of probabilities of default and their volatilities. This is solved by the Heath, Jarrow, and Morton (1992)—HJM model, where the term structure is captured by the curve $\langle \lambda(t,T); t \leq T \rangle$ of forward intensities of default from the perspective of time t, which evolves according to the following set of stochastic differential equations:

$$d\lambda(t,T) = \mu(t,T)dt + \sigma(t,T)dz_t, t < T.$$

The initial values $\lambda(0,T)$, are given by the current forward (risk neutral) intensities of default. The volatilities $\sigma(t,T)$ can be estimated from historical, or market, data, but by the same arguments used by the interest rate HJM model, the drift μ must be calculated from the volatilities in the form:

$$\mu(t,T) = \sigma(t,T) \times \int_t^T \sigma(t,s)ds.$$

Reduced Form Pricing

Joint short rate $r(t)$ and default intensity $\lambda(t)$ modeling under the risk neutral measure enable us, as shown by Lando (1988), to price a zero-coupon defaultable bond as

$$P(t,T) = \hat{E}\left[\exp\left(-\int_t^T (r(u) + \lambda(u))du\right) \Big| t\right] \tag{5.37}$$

provided that the default has not already occurred by time t. Indeed, conditional on the path of r and λ, the risk neutral survival probability between t and T is $\exp\left(-\int_t^T \lambda(u)du\right)$, the discount factor is $\exp\left(-\int_t^T r(u)du\right)$, and so the risk neutral conditional discounted expected cash flow is $\exp\left(-\int_t^T (r(u) + \lambda(u))du\right)$. Taking the average over all possible paths we obtain (5.37).

Interest Rate and Default Intensity Correlations

The Eq. (5.37) naturally admits a correlation between the two processes. According to Lando (2005), dependence may be introduced through a correlated Wiener process driving the stochastic differential equations for interest rates and default intensities. For example, if the short rate r and the intensity λ both follow Vasicek's process and the estimated correlation between the short time increments dr and $d\lambda$ is to be ρ, then the stochastic equations can be written as

$$\begin{aligned} dr &= a(b-r)dt + \sigma_r dz_1, \\ d\lambda &= \kappa(\gamma-\lambda)dt + \sigma_\lambda\left(\rho dz_1 + \sqrt{1-\rho^2}dz_2\right). \end{aligned}$$

As Lando (2005) notes, Vasicek's model is easy to work with and gives a closed form solution. Its disadvantage is the possibility of negative interest rate or default intensities. The same approach unfortunately cannot be easily applied to the CIR process. It turns out that the processes are not jointly affine and the model is difficult to work with.

The correlation structure between interest rates and default intensities of one or more obligors is, however, naturally incorporated into affine intensity models (Duffie and Singleton 2003). The idea is that there is an underlying stochastic multidimensional factor $\mathbf{X}=(X_1,\ldots,X_n)$ of independent affine (e.g. Vasicek's or CIR) processes explaining short interest rates and default intensities through affine relationships

$$\begin{aligned} r &= a_r + \mathbf{b}_r \cdot \mathbf{X}, \\ \lambda &= a_\lambda + \mathbf{b}_\lambda \cdot \mathbf{X}. \end{aligned} \tag{5.38}$$

The variables X_i may include industry or economic business cycle indicators, interest rate and yield spread factors, or other drivers. This model allows the correlation between the short rate and default intensity through their joint dependence on $\mathbf{X}$. Moreover, the model is easily extended to more debtors with default intensity correlations specified by a multifactor model. The survival probability and the presented value of a defaultable zero-coupon bond can be shown to have, again, a tractable exponential affine solution:

$$\begin{aligned} S(t,T) &= e^{\alpha_S(t,T)+\boldsymbol{\beta}_S(t,T)\cdot\mathbf{X}(t)} \text{ and} \\ P(t,T) &= e^{\alpha_P(t,T)+\boldsymbol{\beta}_P(t,T)\cdot\mathbf{X}(t)}. \end{aligned} \tag{5.39}$$

CDO and Basket CDS Valuation

The affine multi-dimensional model (5.38) gives us a new way of simulating times to default in order to value a CDO, or basket CDS, in a Monte Carlo simulation, as described in Sect. 5.2:

1. Generate the affine processes $\mathbf{X}=(X_1,\ldots,X_n)$ over a given interval $[0,T]$.

2. Based on (5.38), calculate the implied paths of the short rate r, and of the intensities $\lambda_1, \ldots, \lambda_k$ where k is the number of debtors in our portfolio.
3. Sample independent times of default $\tau_1, \ldots, \tau_k$ conditional on the calculated intensities, using the compensator approach.
4. Set up the CDO, or basket CDS, cash flows conditional on the times of default $\tau_1, \ldots, \tau_k$, and calculate their present value with the discount factors calculated according to (5.39); i.e., again conditional on the simulated path for $\mathbf{X}$.
5. Running the Steps 1–4 sufficiently, calculate the mean and other moments of the simulated present values many times.

Note that the defaults themselves are independent, conditionally on the simulated intensities. The dependence is incorporated into the process for the intensities of default, similar to how it is in the CreditRisk+ model, while in the structural CreditMetrics model, the correlated factors drive the arrivals of default directly. The difficult part in implementing the model is, as usual, the calibration of the processes for $\mathbf{X}$, and the estimation of the coefficients in (5.38). For a more detailed treatment of the implementation issues, the reader is referred to Lando (2005), Duffie and Singleton (2003); or Fong (2006).

5.5 Basel on Credit Derivatives and Securitization

Credit derivatives and securitized products allow the transfer of credit risk between different market players, and must, of necessity, be dealt with by banking regulation. Basel II (BCBS 2006a) treats credit derivatives in sections related to credit risk mitigation (i.e., reduction of exposures due to collateral or credit protection, see §109–210 of BCBS 2006a), and in parts related to off-balance sheet exposures (see e.g. §82–89 of BCBS 2006a). The capital requirements of securitized transactions, including synthetic deals, are described in detail in the special Part IV of the first Pillar, called “Securitization Framework” (see Fig. 2.9).

The Basel III reform (BCBS 2010) does not change this treatment directly. However, there is an indirect impact of the new capital requirement related to counterparty credit risk, namely the CVA capital charge discussed in the following section that can be quite significant in the case of credit derivatives. Another indirect limitation is created by the newly introduced leverage ratio that includes all off-balance sheet exposures. Basel III also substantially increases capital requirements for re-securitized exposures.

The essential approach of Basel II on credit derivatives is the substitution approach; that is, the risk weight of the protection seller substitutes the risk weight of the underlying asset. On the other hand, the protection seller assumes the credit exposure of the reference entity. The regulation sets down a number of strict qualitative criteria for the credit event definition, asset and maturity mismatches, and gives a list of eligible protection providers, including sovereign entities, private sector entities (PSE), banks, and securities firms with a lower risk weight than the counterparty. The credit protection must be irrevocable and unconditional in order

Table 5.6 Risk weights of securitized and re-securitized assets by long-term rating category in the Standardized Approach

External credit assessment	AAA to AA− (%)	A+ to A− (%)	BBB+ to BBB− (%)	BB+ to BB− (%)	B+ and below or unrated
Risk weight (securitized exposures)	20	50	100	350	Deduction
Risk weight (re-securitized exposures)	40	100	225	650	Deduction

to qualify for the substitution approach. Interestingly, the regulation also allows one to apply first and second to default swaps. The first to default swap may substitute, in terms of the risk weight, only the exposure in the protected portfolio with the least risk weight, and with an amount not larger than the CDS notional. The second to default can be used only in connection with the first to default swap on the same portfolio. Regarding the capital requirement of the protection seller, the bank has to use either an external credit assessment of the product and apply a corresponding regulatory weight, or, if there is no such rating, it has to calculate the sum of all risk weights in the portfolio up to 1250%, and apply it to the protection notional. In case of the second to default swap, the best risk weight in the portfolio can be excluded (§207–210, BCBS 2006a).

According to Part IV of the Basel II, first Pillar (§538–643): "Banks must apply the securitization framework for determining regulatory capital requirements on exposures arising from traditional and synthetic securitizations or similar structures that contain features common to both. Since securitizations may be structured in many different ways, the capital treatment of a securitization exposure must be determined on the basis of its economic substance rather than its legal form." So, although the document describes a number of securitization techniques, including ABS, ABCP, credit enhancements, and others, it tries to foresee the dynamic development of new types of securitized products which might not be formally covered by the text, yet should be treated on the basis of economic substance in an analogy with the given principles. Due to the complexity of the securitized products, the regulation emphasizes external ratings. The risk weights in Table 5.6 illustrate the attractiveness of investments in securitized senior tranches, in terms of capital requirements, even though the underlying portfolio might receive a much higher risk weight. This, maybe, also explains the excessive pressure put on the rating agencies by the new regulation (or temptation for them) to assign higher grades. This effect has been to a certain extent mitigated by the Enhancements to the Basel II Framework (BCBS 2009a), belonging to a set of BCBS documents sometimes called "Basel 2.5", which significantly increases the requirement for re-securitized exposures, i.e. exposures like CDO based on pools of assets containing securitized exposures. In order to limit dependence on external ratings, BCBS (2009a) adds a new general rule, according to which a bank using the securitization framework must, on an ongoing basis, have a comprehensive understanding of the risk characteristics of its individual securitization exposures, whether on-balance sheet or off-balance sheet, as well as the risk characteristics of the pools underlying its securitization exposures.

Nevertheless, even in the case of securitized exposures, banks are also allowed to use an Internal Rating Based (IRB) approach. The IRB approach is applicable only if the bank has been approved for applying the IRB approach to all classes of the securitized underlying assets. The banks can then apply the Rating Based Approach (RBA), using an internal, or external or inferred credit rating. The RBA risk weights refine Table 5.6, depending on the rating type and the granularity of the underlying portfolio. If external or inferred ratings are not available, then the banks may apply the Supervisory Formula (SF), or the external rating agency methodology based Internal Assessment Approach (IAA).

Under the relatively complex Supervisory Formula Approach, the capital charge for a securitization tranche depends on seven bank-supplied inputs: the amount of the underlying exposures (UE); the securitization exposure's proportion of the tranche that contains the securitization exposure (TP); the IRB capital charge which would be applicable if the underlying exposures were not securitized (KIRB); the tranche's credit enhancement level (L) and thickness (T); the pool's effective number of exposures (N); and the pool's exposure-weighted average loss-given-default (EWALGD). As in the case of ordinary exposures, the formula tries to approximate the unexpected loss of the products. Compared to the stochastic models described in the previous sections, the formula is based on a number of simplifying assumptions and, not surprisingly, it has a number of weaknesses.

The recent financial crisis has put the Basel regulation of credit derivatives and securitized products under scrutiny. The Basel II enhancements and the Basel III reform try to fix the most important issues and strengthen the resilience of the banking sector. However, the reforms are still in the process of implementation with unclear outcomes, and the regulatory discussion continues. We believe that the ongoing research in the field of credit risk management, modeling, and general understanding of the topic, to which this book aims to contribute, is one way of avoiding at least some of the mistakes which were made in the past.

5.6 Counterparty Credit Risk

Counterparty credit risk (CCR) is a specific form of risk arising from the possibility of default of a counterparty before the maturity of a financial transaction. The products subject to counterparty credit risk are, in particular, over-the-counter (OTC) derivatives and securities financing transactions (e.g., repo operations). Exchange traded derivatives or derivatives settled with a central counterparty are only theoretically subject to a counterparty credit risk. Even an exchange or its clearinghouse (*Central Counterparty*—CCP) may go bankrupt. However, the probability of default of an OTC counterparty is generally much higher.

Let us consider a financial contract, for example, a forward or swap transaction, between a financial institution and a counterparty. If the counterparty defaults at time τ before the contract's maturity T then the institution may (but does not have to) suffer a loss depending on the legal set-up and the value of the transaction at the time of default. Under standard legal documentation (*The International Swaps and Derivatives Association*—ISDA Master Agreement) the market value f_τ (from the perspective of the institution) will be frozen at the time of default, remaining cash flows will be canceled, and the amount f_τ will be either payable by the institution to the counterparty, if negative, or by the counterparty to the institution, if positive. Hence, if $f_\tau \leq 0$, then the institution does not suffer any loss due to the counterparty's default, since the transaction is just closed-out before maturity for its market value. However, if $f_\tau > 0$ then the institution's exposure with respect to the counterparty will be probably paid-back only partially, or not at all. If l denotes the fractional Loss Given Default coefficient (LGD), then the loss will be $l \times f_\tau$. The loss could be much higher in a non-standard legal situation when the institution would be obliged to fulfill all the future payments under the contract, but the counterparty's payments would become part of bankruptcy claims. On the other hand, the loss on collateralized transactions, where the market value is secured by cash or other high quality collateral, may be completely eliminated in case of the counterparty's default. The situation becomes more complicated, provided all losses and profits with respect to a single defaulting counterparty can be mutually netted.

In any case, in market terms, the *Credit Valuation Adjustment* (*CVA*) can be defined as the difference between the market value of a transaction with respect to a theoretically risk-free counterparty and the market value of the identical transaction with respect to the specific risky counterparty. In other words, CVA can be specified as the theoretical cost of insurance against the counterparty credit losses, and therefore, under the risk neutral valuation principle, CVA is expressed as the expected discounted loss caused by a possible counterparty default event. Focusing on the standard set-up, we can write theoretically

$$CVA = E[\text{discounted CCR loss}] = E[e^{-r\tau}\max(f_\tau, 0) \times l \times I(\tau \leq T)], \quad (5.40)$$

Where $I(\tau \leq T)$ denotes the indicator function, i.e. it is 1 if $\tau \leq T$ and 0 otherwise, r is the risk-free rate in continuous compounding, and E is the expectation operator under the risk neutral probability measure. Given CVA, we can adjust the derivative market value as $f_d = f_{nd} - CVA$ where f_{nd} is the no-default (risk-free counterparty) valuation and f_d is the market value with respect to a given risky counterparty. It is clear that the CVA risk-neutral definition (5.40) involves several uncertainties: the time and probability of default and the derivative transaction value at the time of default. Even the loss given default, as well as the discount rate, depending on the time of default τ should be considered as stochastic. Moreover, all those variables might be mutually correlated. Therefore, CVA modeling is generally even more challenging than the valuation of complex derivatives without considering the counterparty credit risk.

Expected Exposure and Credit Valuation Adjustment
There are many approaches to CVA valuation. The most general method is based on a Monte Carlo simulation of future market factors, exposures, intensities of default, and times to default. On the other hand, a simplified practical "*add-on*" approach tries to express CVA as an expected exposure with respect to the counterparty, multiplied by the probability of counterparty default, and by the LGD. The expected exposure itself is, in general, estimated by a Monte Carlo simulation (that is usually much simpler compared to a full scale simulation of all the factors). For some products, the expected exposure can be calculated using the option valuation formula.

In order to come up with a realistic practical formula, one needs to make a number of simplifying assumptions. Firstly, let us assume independence between the time to default and exposure; moreover, assume that the loss rate l is constant, and that the discount rate $r(t)$ and the counterparty default (forward) intensity[6] $q(t)$ are deterministic functions of time. Dividing the time interval into subintervals $0 = t_0 < \ldots < t_m = T$, the simplified CVA expression (see e.g. Gregory 2010) can be written as:

$$CVA \doteq l \sum_{j=1}^{m} e^{-r(t_j) \times t_j} EE(t_j) q(t_j) \Delta t_j, \tag{5.41}$$

where $\Delta t_j = t_j - t_{j-1}$ and the *expected exposure* $EE(t) = E[\max(f_t, 0)]$ is defined independently of the other factors. The expected exposure can, in general, be estimated by the Monte Carlo simulation method. For some products, like forwards or swaps, it can be expressed in an analytic or semi-analytic form. On the other hand, if the derivative value f_t itself is not analytical, then we have to cope with a numerically difficult "double" Monte Carlo simulation, i.e. with simulations that are embedded in another Monte Carlo simulation.

Example 1 Let us consider a simple outstanding 1-year forward to buy a non-dividend stock for $K = 101$ with the current spot price $S_0 = 100$. Assume that the risk-free rate is $r = 1\%$, that the counterparty risk-neutral intensity of default $q = 4\%$, and the LGD $l = 60\%$, are constant. The forward value at time t is $f_t = S_t - e^{-r(T-t)}K$ conditional on the spot price S_t, and so the expected exposure

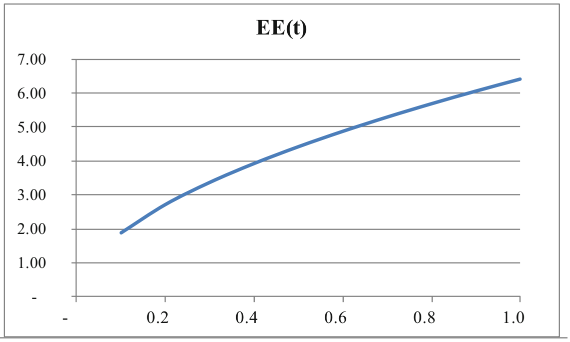
$$EE(t) = E\left[\max\left(S_t - e^{-r(T-t)}K, 0\right)\right] = c\left(t, e^{-r(T-t)}K\right)e^{rt}$$

is just the market value of the European call option with maturity t, and the strike price $e^{-r(T-t)}K$ that can be evaluated using the Black-Scholes formula and re-discounted forward to the time t (i.e., multiplied by the factor e^{rt}). Equivalently,

[6] That is $q(t)\Delta t$ is the probability of default over the period $[t, t + \Delta t]$.

Table 5.7 Long forward CVA calculation

t	EE(t)	CVA contr.
0.1	1.89	0.005
0.2	2.73	0.007
0.3	3.38	0.008
0.4	3.93	0.009
0.5	4.42	0.011
0.6	4.88	0.012
0.7	5.29	0.013
0.8	5.69	0.014
0.9	6.06	0.014
1.0	6.41	0.015
	CVA	0.107



the adjustment can be calculated by integrating the call option value from the time 0 to T

$$CVA = \int_0^T l \times c\left(t, e^{-r(T-t)}K\right) \times q \times dt \doteq \sum_{j=1}^{m} l \times q \times c\left(t_j, e^{-r\left(T-t_j\right)}K\right)\Delta t_j. \quad (5.42)$$

Table 5.7 shows a calculation according to (5.42) dividing the time to maturity into ten subintervals only. The resulting CVA presents more than 10 bps of the forward notional amount. We have assumed, for the sake of simplicity, that the interest rates and forward default intensities were constant. The calculation would remain essentially the same if we were given the term structure of interest rates and of the default intensities.□

The calculation of the expected exposure for option-like products turns out to be relatively simple, since the signs of the market values do not change. An option seller exposure would always be zero (once the premium was paid) while an option buyer exposure would always be just the option value. For a European option, the expected exposure would be, by the principle of iterated expectations, equal to the current option value re-discounted forward to the time t, i.e. $EE(t) = c_0 e^{rt}$.

The expected exposure of a forward transaction is an increasing function of time as illustrated above. However, for an outstanding interest rate swap, the expected exposure would be zero at the start date, provided the swap had been entered into under market conditions, and at maturity when all the cash flows have been settled; its maximum would be attained somewhere in the middle between the start date and the maturity (see Fig. 5.21 for an illustration). In fact, the expected exposure turns out to be just a swaption value that can be evaluated using a generalized Black-Scholes formula (see Witzany 2013a).

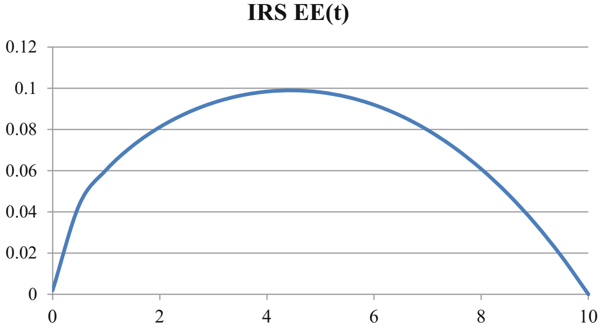


Fig. 5.21 Expected exposure of a 10-year interest rate swap

Further simplification can be achieved by introducing the *expected positive exposure* defined as an average of $EE(t)$ over time:

$$EPE = \frac{1}{T}\int_0^T EE(t)dt \doteq \frac{1}{T}\sum_{j=1}^m EE(t_j)\cdot \Delta t_j. \tag{5.43}$$

If we assume that the intensity of default is constant or independent of the expected exposure then CVA can be approximately written as

$$CVA \doteq \bar{q} \times l \times EPE \times A(0) \tag{5.44}$$

Where $\bar{q}$ is the average intensity of default and

$$A(0) = \sum_{j=1}^m e^{-r(t_j)\times t_j}\Delta t_j \tag{5.45}$$

is the risk-free annuity value.

Example 2 The expected positive exposure of the forward contract from Example 1 can be estimated as an average of values from the second column in Table 5.7 $EPE \doteq 4.47$. Similarly, the annuity (5.45) can be estimated as an average of the discount factors applied: $A(0) \doteq 0.99$. Therefore, we have a simple calculation

$$CVA \doteq 0.04 \times 0.6 \times 4.47 \times 0.99 \doteq 0.107$$

which gives us exactly the same result as in Table 5.7.□

The concept of EPE is also useful in connection with *Credit Default Swap* (CDS) quotations. Let us assume for the sake of simplicity that the CDS payoff is settled in cash and that default can take place only in times $t_1, \ldots, t_m$ when the spread is being paid. If $S(t)$ denotes the reference entity (risk-neutral) survival probability function[7]

[7] $S(t) = \Pr[\tau > t]$ is defined as the probability that default does not take place until time t.

then the spread must satisfy the classical insurance equivalence relation (see also Sect. 5.1):

$$\sum_{j=1}^{m} e^{-r(t_j)\times t_j} X \times \Delta t_j \times S(t_j) = \sum_{j=1}^{m} e^{-r(t_j)\times t_j} l \times \Delta S(t_j), \tag{5.46}$$

where $\Delta S(t_j) = S(t_j) - S(t_{j-1})$ is the probability of default during the time interval $[t_{j-1}, t_j)$, i.e. $\Delta S(t_j) = q(t_j)\Delta t_j$ using the concept of default intensity. The left-hand side of (5.46) corresponds to the expected discounted premium income, while the right-hand side corresponds to the expected discounted payoff.

Now, if X is a market CDS spread quotation for the reference entity being equal to our counterparty and for the maturity T, and if we replace $EE(t_j)$ by the constant EPE in (5.41), then

$$\begin{aligned} CVA \doteq EPE \sum_{j=1}^{m} e^{-r(t_j)\times t_j} \times l \times \Delta S(t_j) = \\ = EPE \sum_{j=1}^{m} e^{-r(t_j)\times t_j} X \times \Delta t_j \times S(t_j) = X \times EPE \times A_{CDS}(0). \end{aligned} \tag{5.47}$$

Consequently, CVA can be simply approximated as the CDS spread times EPE times the risky (CDS) annuity

$$A_{CDS}(0) = \sum_{j=1}^{m} e^{-r(t_j)\times t_j} \Delta t_j \times S(t_j).$$

For swap-like products, the CVA is often expressed as a spread X_{CVA} that can be added to the periodical fixed or float payment. Since the payments are terminated in the case of default, we need to solve the equation

$$CVA = X_{CVA} \times L \times \sum_{j=1}^{m} e^{-r(t_j)\times t_j} \Delta t_j S(t_j), \tag{5.48}$$

where L is the swap notional. Therefore, combining (5.47) and (5.48) we get a nice and simple approximation:

$$X_{CVA} \doteq X_{CDS} \frac{EPE}{L}. \tag{5.49}$$

Example 3 Let us consider a 10-year 1 billion CZK notional interest rate swap where we should receive the fixed coupon with the expected exposure profile as shown in Fig. 5.21. Without considering the counterparty credit risk the IRS market rate would be 4%. The expected positive exposure (based on 40% volatility of the IRS rates) is 71 million CZK. Assume that the CDS spread quoted for our counterparty equals 250 bps. According to (5.49) the CVA spread comes out at approximately 18 bps. Consequently, the adjusted fix payment paid by the counterparty should be 4.18%, which significantly differs from the 4% rate without the CVA adjustment.□

The decompositions (5.44) or (5.47) allow certain stressing of the CVA. Besides the probability of default (intensity or the CDS spread), the exposure can be stressed by introducing the *potential future exposure* (PFE) as a quantile of the average exposure (depending on paths from 0 to T). The Basel II internal approach also introduces the notion of *effective EE* that is defined as EE but with the additional requirement of being non-decreasing. *Effective EPE* is then defined as the average of effective EE. This approach is related rather to a netted portfolio of transactions with respect to a counterparty, where maturing transactions are expected to be replaced with new ones.

Collateralization and Netting

The counterparty credit risk of an OTC derivative can be mitigated in a way similar to the margin mechanism applied to exchange traded products if the OTC counterparties agree to post collateral, usually in cash, covering the derivative market value. In practice, this can be achieved by signing the *Credit Support Annex* (CSA) of the ISDA Master Agreement. The collateralization can be two-way or one-way. For example, a bank would require a corporate counterparty to post collateral covering the exposure $\max(f_t, 0)$, but no collateral would be sent by the bank if f_t became negative. Two-way collateralization has recently become quite normal between banking counterparties. According to the ISDA Margin Survey 2015 (ISDA 2015), the use of collateral has indeed become extensive: in 2014 89% of non-cleared fixed-income derivatives and 97% of non-cleared credit derivatives were collateralized with a CSA agreement, while more than 80% of derivatives portfolios with more than 2500 trades are reconciled daily, and the use of cash or government securities accounts for over 90% of total collateral. These numbers have had an increasing trend over the last years (for more details see also Baran 2016).

If the collaterals were recalculated and posted on a continuous basis, then the CCR would be virtually eliminated. In practice, there is a standard remargining period or a minimum threshold, and so there might be a residual counterparty risk. For example, if the remargining period was 1 day, then there still should be a CVA corresponding to the 1 day horizon during which the margin does not necessarily cover the market value in the case of unexpected market volatility.

Another way to mitigate CCR with respect to a counterparty is a netting agreement allowing one to net the positive and negative market values of different derivative contracts in the case of the counterparty's default, i.e. the exposure needs to be defined and monitored on the portfolio basis as $E(t) = \max(V(t), 0)$ where $V(t) = \sum f_i(t)$ is the sum of the outstanding transactions' market values with respect to the single counterparty. For a more complex portfolio, there is little chance to find a precise analytical formula for the expected netted exposure. Nevertheless, if we can assume that the portfolio linearly depends on market factors and that $V(t)$ approximately follows a generalized Wiener process, i.e. if

$$V(t) = V(0) + \mu t + \sigma Z\sqrt{t} \text{ where, } Z \sim N(0,1)$$

then we can find relatively easily analytical formulas for EE, EPE, or PFE (Gregory 2010). In particular, if $V(0) = 0$ and $\mu = 0$, then $EE(t) = \sigma\sqrt{t}\varphi(0)$ and

$$EPE = \frac{\sigma\varphi(0)}{T}\int_0^T \sqrt{t}dt = \frac{2}{3}\sigma\varphi(0)\sqrt{T} \doteq 0.27\sigma\sqrt{T}.$$

For more complex non-linear portfolios, where we cannot assume normality, a Monte Carlo simulation needs to be used. Notice that the problem is technically similar to the VaR estimation. We need to model future exposure probability distribution, focusing in this case on the positive rather than on negative values of the portfolio. However, the time dimension makes the task even more challenging.

Wrong Way Risk

We have already emphasized that the simplified CVA formula (5.41) is based on the assumption that the exposure and the event of default are independent. This formula should not be used if there is evidence of a link between default, or intensity of default, and of the exposure. The formula can, in fact, be easily fixed defining the expected exposure conditional on default $EE^*(t) = E\left[\max(f_t, 0)\middle|t = \tau\right]$. Then the analogous formula

$$CVA \doteq l\sum_{j=1}^{m} e^{-r(t_j)\times t_j} EE^*(t_j) \cdot q(t_j) \cdot \Delta t_j \tag{5.50}$$

becomes consistent with (5.40). In terms of causality, there is usually a common driver for the exposure and the event of default, and so we cannot say that one event causes the other or vice versa. For example, if a company is sensitive to currency devaluation and if the exchange rate impacts the exposure then there could be either a *wrong way risk*, $EE^*(t) > EE(t)$, with the exposure increasing in the case of devaluation, or a *right way risk*, $EE^*(t) < EE(t)$, with the exposure going down in the case of revaluation. Both wrong way and right way risks exist similarly for interest rate products, but in the case of CDS exposures the risk is almost always in the wrong way direction from the perspective of the credit protection buyer. If the systematic credit risk increases, e.g. during a financial crisis, then the CDS exposure goes up and the counterparty credit risk generally increases as well.

One way to solve the conditional expected exposure analytically, in some cases, is to use the standard Gaussian copula model where the time to default $\tau = S^{-1}(\Phi(X))$ is driven by a normally distributed variable $X \sim N(0,1)$ transformed using the inverse survival function. Similarly, let us assume that the derivative or portfolio value $V(t) = G(Z)$ at time t is driven by a normal variable $Z \sim N(0,1)$ through an increasing function G. Now, the exposure-default correlation can be captured by the correlation ρ between the normal variables X and Z. Since a high value of X is translated into a low value of τ, as the survival function is decreasing, a positive correlation $\rho > 0$ corresponds to the wrong way risk while a negative correlation $\rho < 0$ corresponds to the right way risk. The conditional expected exposure can then be written as

$$EE^*(t) = E\left[G\left(\sqrt{1-\rho^2}Y + \rho X\right)^+ \middle| X = \Phi^{-1}(S(t))\right] = \\ = \int_{-\infty}^{\infty} G\left(\sqrt{1-\rho^2}y + \rho\Phi^{-1}(S(t))\right)^+ \varphi(y)dy, \tag{5.51}$$

Decomposing $Z = \sqrt{1-\rho^2}Y + \rho X$ where $X, Y \sim N(0,1)$ are independent. For example, if $V(t) = \mu t + \sigma\sqrt{t}Z$ follows just the generalized Wiener process then

$$EE^*(t) = \int_{-a/b}^{\infty} (a + by)\varphi(y)dy = a\Phi\left(\frac{a}{b}\right) + b\varphi\left(\frac{a}{b}\right),$$

where $a = \mu t + \rho\sigma\sqrt{t}\Phi^{-1}(S(t))$ and $b = \sqrt{1-\rho^2}\sigma\sqrt{t}$.

The principle (5.51) can be used to express the expected loss of a forward or to price an option with the wrong way risk by a semi analytic formula (Gregory 2010). In a similar fashion, Černý and Witzany (2015) obtained and tested a semi-analytical formula to price the CVA of interest rate swaps with the wrong way risk. In general, a Monte Carlo simulation for underlying market factors and the counterparty time of default needs to be run (see, e.g., Brigo and Pallavicini 2008).

Example 4 Let us consider the outstanding 1-year forward from Example 1 and let us recalculate the CVA with the wrong way risk Gaussian correlation $\rho = 0.5$. The stock price is lognormally distributed and can be written as

$$S_t(Z) = S_0 \exp\left(\left(r - \sigma^2/2\right)t + \sigma\sqrt{t}Z\right).$$

If we assume for the sake of simplicity that the default density $q = 4\%$ is constant over the 1-year horizon then the survival function is linear, i.e. $S(t) = 1 - qt$. Therefore, in line with (5.51) the conditional expected exposure can be written as

$$EE^*(t) = \int_{-\infty}^{\infty} \left(S_t\left(\sqrt{1-\rho^2}y + \rho\Phi^{-1}(S(t))\right) - e^{-r(T-t)}K\right)^+ \varphi(y)dy$$

This integral could be solved analytically, as mentioned above, but we can also evaluate it numerically for $t = 0.1, 0.2, \ldots, 1$ similarly to Table 5.7. The conditional $EE^*(t)$ values, their average $EPE^* = 11.21$, and the corresponding $CVA = 0.282$ come out more than twice as high than without the wrong way risk!□

Bilateral Counterparty Credit Risk

So far, we have assumed that there is a default-free institution that has a counterparty with a positive probability of default. In reality, both counterparties may default. If the institution defaults at time *t* and if the outstanding derivative market value $f_t < 0$ is negative then the counterparty will lose and the institution will "save" the amount $-l_I \times f_t$ where l_I is the institution's LGD ratio. In this sense, the institution has an "option to default" with a potential positive payoff. *Bilateral credit valuation adjustment* (BCVA, sometimes also denoted BVA or TVA—*total valuation adjustment*) takes into account both effects of the potential loss due to the counterparty's default and the potential "profit" due to the institution's own default. Let τ_I denote the institution's time of default and τ_C the counterparty's time of default. Then the BCVA can be decomposed into two parts[8]

$$BCVA = CVA_C - CVA_I$$

where CVA_C covers the counterparty's default, provided that the institution has not defaulted sooner

$$CVA_C = E\left[e^{-r\tau_C}\max\left(f_{\tau_C}, 0\right) \times l_C \times I(\tau_C \leq T \& \tau_C < \tau_I)\right],$$

and analogously

$$CVA_I = E\left[e^{-r\tau_I}\max\left(-f_{\tau_I}, 0\right) \times l_I \times I(\tau_I \leq T \& \tau_I < \tau_C)\right].$$

The CVA_I is also sometimes called the *Debit Valuation Adjustment* (DVA). If we assume that the institution and the counterparty cannot both default before *T* (or that the probability of this event is negligible) then CVA_C and CVA_I are just the "one-way" CVAs we have discussed so far from the opposite perspectives. The probability of joint default, i.e. of $\tau_C, \tau_I \leq T$ is negligible if the defaults are independent and their probabilities are low. Otherwise, this possibility needs to be taken into account in the context of a correlation model.

The advantage of the concept of BCVA is that it makes derivatives valuation symmetric again. Note that with the "one-way" CVA, the institution's market value is $f_I = f_{nd} - CVA_C$ while the counterparty's market value is $f_C = -f_{nd} - CVA_I \neq -f_I$. With the bilateral adjustment we have $f_I = f_{nd} - BCVA$ and $f_C = -f_{nd} + BCVA = -f_I$.

[8] Implicitly assuming that $\Pr[\tau_C = \tau_I] = 0$.

Fig. 5.22 Germany 5Y CDS spread (Source: Thomson Reuters)

The accounting of CVA, DVA and BCVA has gradually become a market standard and has been mandatory since 1/2013 (IFRS 13). On the other hand, it should be noted that BCVA has the strange effect that the deterioration of the institution's own credit quality is translated into its accounting profit. This can be compared to marking down liabilities due to own credit downgrade. During the crisis, this has indeed happened, when a large investment bank reported a 1 billion USD profit based on this effect. Such a situation is not acceptable to auditors and regulators who prefer conservative accounting principles and tend to require banks to account just the CVA rather than BCVA.

What Are the Risk-Free Rates?

With the emergence of the ever-present counterparty credit risk, the markets started to reconsider the classical approach to the construction of the risk-free rates from the government bond yields or from interest rate swap rates. Figure 5.22 shows the German government 5-year CDS spread development (approximating the government bond spreads over the risk free rate) that went as high as 100 bps during 2011. The risk of the German government is considered to be almost minimal compared to other countries where the CDS spreads even went up to hundreds of basis points.

This market reality leads to a preference for a zero coupon curve built from interest rate swap rates where the counterparty risk is much smaller (the expected exposure is usually just a fraction of the swap notional amount). A fixed IRS rate is the cost of the rolled-over financing of a reference rated (e.g., AA) bank, where the bank can be periodically replaced by another one in the case of its credit deterioration. Nevertheless, we have to keep in mind that even a high rated entity can default during the reset time horizon (3M or 6M), and so short rates do incorporate certain

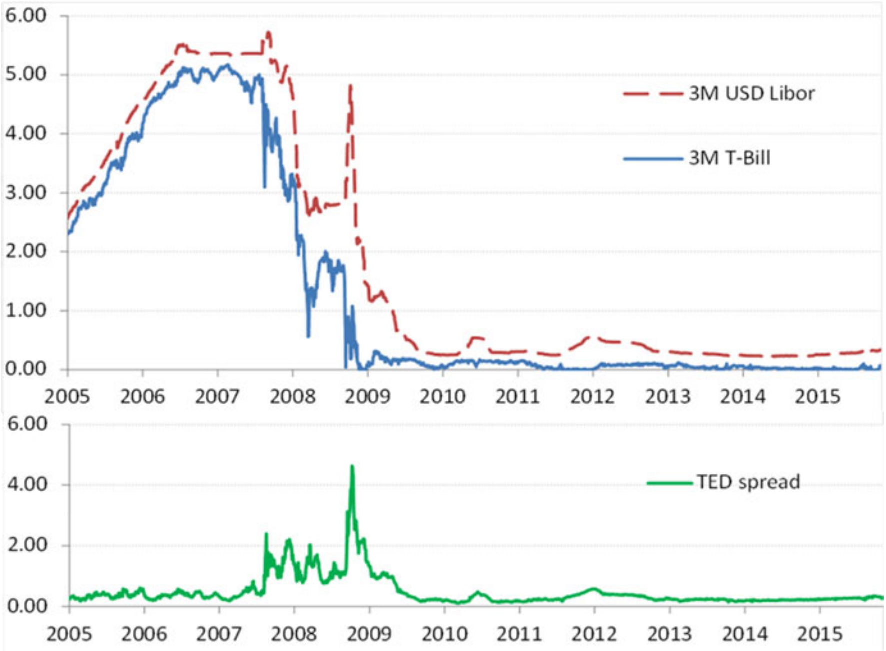


Fig. 5.23 The spread between the 3 M USD Libor and 3 M US T-bill rates (Source: Bloomberg)

credit spreads that are, consequently, reflected in the IRS rates. Figure 5.23 shows the development of the 3M USD Libor and the US Treasury 3M bill rate (which is close to the ideal risk free rate). The spread between the two rates went over 400 bps and still remains relatively high.

Another argument against the IRS rates is the emergence of significant basis swap spreads. Figure 5.24 shows that the markets have recently perceived the 6M Euribor credit premium to be at least 10 bps higher than the 3M Euribor financing premium. This phenomenon indicates that, in spite of the still relatively high rating of the financial market counterparties, the intensities of default are viewed as non-negligible and increasing in time (see Fig. 5.16). Practically, it means that the zero coupon curve based on swaps with 6M float payments would differ from the curve based on swaps with 3M periodicity, and so on. Using the swaps, there would be a multitude of risk-free curves, completely changing the paradigm of one single risk-free curve. The situation would become even more complicated considering cross-currency swap basis spreads (see Baran and Witzany 2013; Baran 2016).

The current solution generally accepted by the market is to use the *Over-Night Index Swap* (OIS) rates in order to construct the risk free curve, since the 1-day horizon risk (default intensity) is considered minimal (Fig. 5.16). An OIS is similar to plain vanilla IRS with the difference that the float rate is calculated daily (every business day) as an official overnight rate and, in order to simplify the settlement, compounded over longer periods into an *Overnight Index Average* (ONIA), or the *Effective Federal Funds Rate* for USD, the *Euro Overnight Index Average*

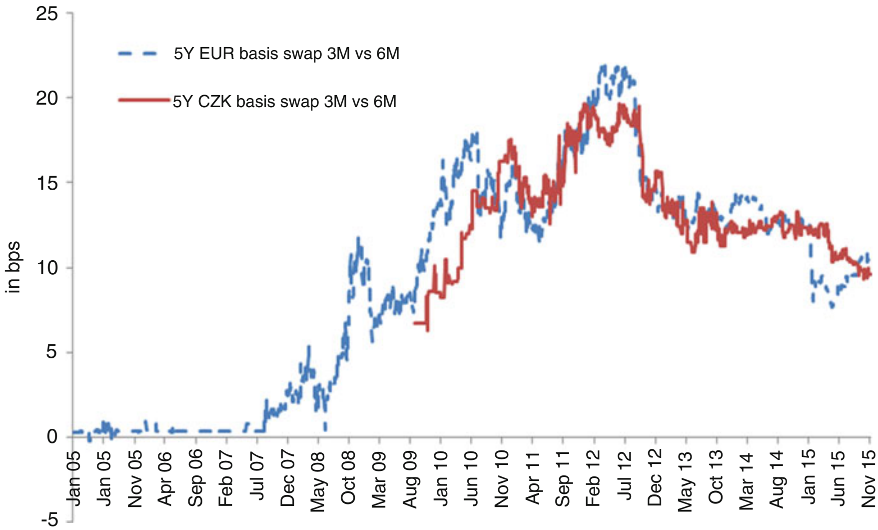


Fig. 5.24 The 5Y basis spreads for 3 M/6 M Euribor and 3 M/6 M Pribor (Source: Bloomberg)

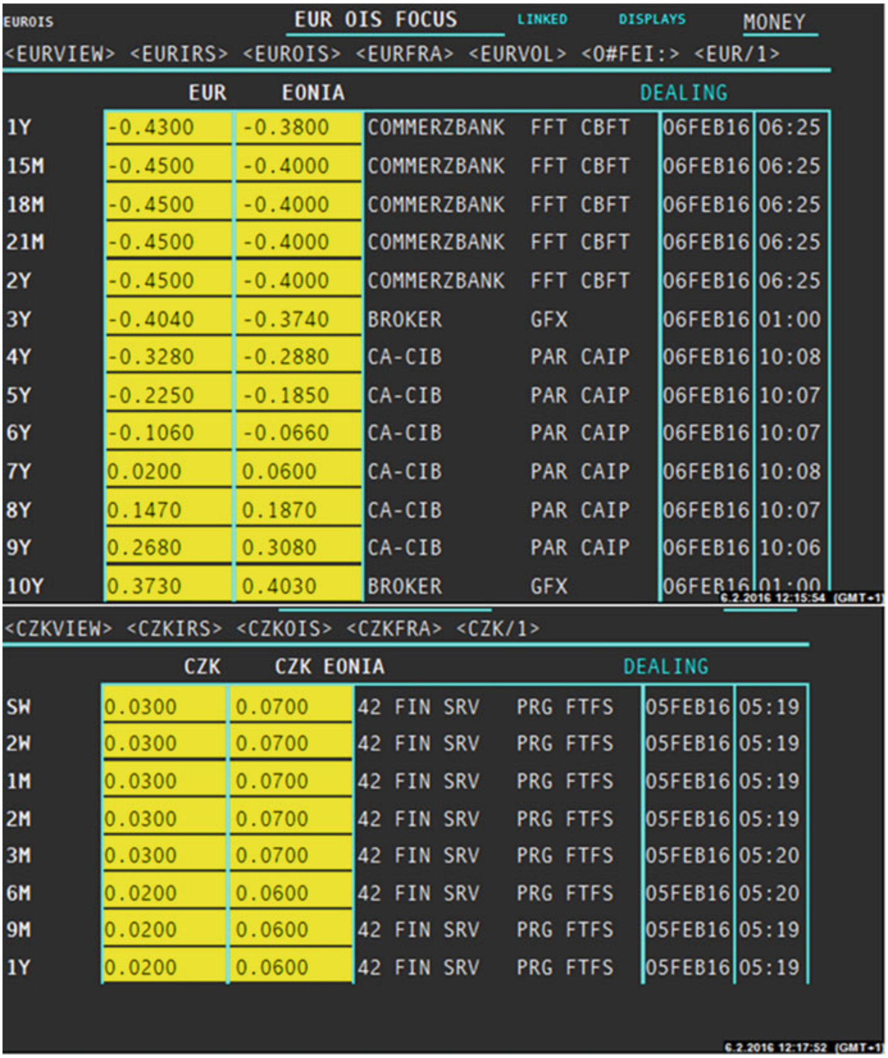
(EONIA) for EUR, the *Sterling Overnight Index Average* (SONIA) for GBP, the *Czech Overnight Index Average* (CZEONIA) for CZK, etc. The compounding is based on daily (business day) O/N deposit periods

$$r = \left(\prod_{t=1}^{n_b} \left(1 + \frac{r_t \times n_t}{360} \right) - 1 \right) \frac{360}{n},$$

Where r_t is the O/N rate, n_t the number of calendar days in the O/N period (normally 1 day, but it can also be 3 days for a weekend), n_b the number of business days in the compounding period (e.g. 3 or 6 months), and n is the total number of days. OIS swaps tend to be short lived, often only 3 months or less. For swaps of 1 year or less, there is only a single payment at maturity defined as the difference between the fixed rate and the compounded OIS rate. For longer swaps the payments are made quarterly or annually.

A fixed OIS rate again represents the cost of the rolled-over financing of a reference rated bank, where the bank can be replaced by another one in the case of credit deterioration. In this case, the roll-over periods are only one business day and the probability of a full default of a reference rated entity (e.g., AA rated) during one single business day is considered almost negligible (there is usually a sequence of downgrades before an AA bank ends up in default). Indeed, Fig. 5.26 shows that the spread between the 3M USD Libor and the 3M OIS rates approximates well the TED[9] spread (compare to Fig. 5.23). Therefore, if there is

[9]The TED spread is the difference between the interest rates on interbank loans (USD Libor) and on short-term U.S. government debt (“T-bills”).



EUROIS EUR OIS FOCUS LINKED DISPLAYS MONEY

<EURVIEW> <EURIRS> <EUROIS> <EURFRA> <EURVOL> <0#FEI:> <EUR/1>

	EUR	EONIA	DEALING			
1Y	-0.4300	-0.3800	COMMERZBANK	FFT CBFT	06FEB16	06:25
15M	-0.4500	-0.4000	COMMERZBANK	FFT CBFT	06FEB16	06:25
18M	-0.4500	-0.4000	COMMERZBANK	FFT CBFT	06FEB16	06:25
21M	-0.4500	-0.4000	COMMERZBANK	FFT CBFT	06FEB16	06:25
2Y	-0.4500	-0.4000	COMMERZBANK	FFT CBFT	06FEB16	06:25
3Y	-0.4040	-0.3740	BROKER	GFX	06FEB16	01:00
4Y	-0.3280	-0.2880	CA-CIB	PAR CAIP	06FEB16	10:08
5Y	-0.2250	-0.1850	CA-CIB	PAR CAIP	06FEB16	10:07
6Y	-0.1060	-0.0660	CA-CIB	PAR CAIP	06FEB16	10:07
7Y	0.0200	0.0600	CA-CIB	PAR CAIP	06FEB16	10:08
8Y	0.1470	0.1870	CA-CIB	PAR CAIP	06FEB16	10:07
9Y	0.2680	0.3080	CA-CIB	PAR CAIP	06FEB16	10:06
10Y	0.3730	0.4030	BROKER	GFX	06FEB16	01:00

6.2.2016 12:15:54 (GMT+1)

<CZKVIEW> <CZKIRS> <CZKOIS> <CZKFRA> <CZK/1>

	CZK	CZK EONIA	DEALING			
SW	0.0300	0.0700	42 FIN SRV	PRG FTFS	05FEB16	05:19
2W	0.0300	0.0700	42 FIN SRV	PRG FTFS	05FEB16	05:19
1M	0.0300	0.0700	42 FIN SRV	PRG FTFS	05FEB16	05:19
2M	0.0300	0.0700	42 FIN SRV	PRG FTFS	05FEB16	05:19
3M	0.0300	0.0700	42 FIN SRV	PRG FTFS	05FEB16	05:20
6M	0.0200	0.0600	42 FIN SRV	PRG FTFS	05FEB16	05:20
9M	0.0200	0.0600	42 FIN SRV	PRG FTFS	05FEB16	05:19
1Y	0.0200	0.0600	42 FIN SRV	PRG FTFS	05FEB16	05:19

6.2.2016 12:17:52 (GMT+1)

Fig. 5.25 EUR and CZK OIS quotations (Source: Thomson Reuters, 5.2.2016)

a liquid market for the OIS swaps (which is the case for USD and EUR) then the rates can be used to construct an almost ideal risk-free zero coupon curve. For currencies with a limited or no OIS market, an approximation needs to be used (see Baran and Witzany 2013, for a discussion).

Figure 5.25 shows an example of EUR and CZK OIS quotes. The EUR OIS market turns out to be quite liquid, with quotes going from 1M up to 50Y, but the CZK OIS market unfortunately provides quotes only up to 1 year. The quotes can be compared with Euribor, Pribor, or IRS rates that, indeed, turn out to be at least 20–40 bps higher than the respective OIS rates.

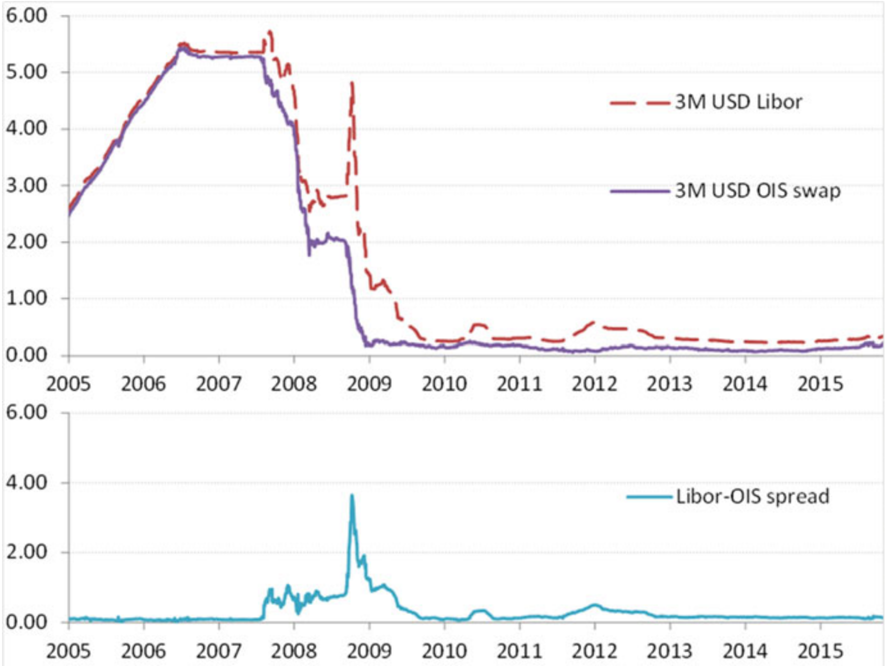


Fig. 5.26 The spread between 3 M USD Libor and 3 M USD OIS rates (Source: Bloomberg)

It should be pointed out that the "true" risk-free rates are a necessary input of derivative valuation models (Hull and White 2012a). Derivatives should be first valued, assuming there are no defaults, and then adjusted for the (bilateral) counterparty credit risk $f = f_{nd} - BCVA$. Generally, it is not correct to use the interest rate reflecting the counterparty's cost of financing as an input of the derivative valuation model. For example, for a long European call option, it is possible to write $f = e^{-r_C T}\hat{E}\left[(S_T - K)^+\right]$, where r_C is the counterparty's cost of financing (assuming no credit risk on the part of the institution), but the risk-neutral expected value is based on the assumption that the drift of S_t is the risk free rate r_0 and not the risky rate. In this particular case, the no-default option value can in fact be adjusted as $f = e^{-(r_C - r_0)T} f_{nd}$ using the counterparty's credit spread. Nevertheless, this formula is not applicable to derivatives like forwards or swaps, where the cash flows can be both positive and negative.

The same discussion applies to collateralized derivative transactions. If there is a two-way continuous collateralization, then the discounting rate can be effectively replaced by the rate r_M accrued on the margin account, yet the drift of the asset prices is still the risk-free rate (in the risk neutral world). In this case, we can use the multiplicative adjustment $f = e^{-(r_M - r_0)T} f_{nd}$ for all types of derivatives. If $r_M > r_0$ then the collateral interest brings an advantage to the counterparty receiving a positive payoff, and vice versa if $r_M < r_0$.

Basel III CVA Capital Charge

The accounting of CVA and DVA has definitely improved both awareness and the management of the counterparty credit risk. For large institutions the total P/L impact of the seemingly small adjustments might be in billions of USD. The new BCVA accounting practice has at the same time highlighted the existence of a new price risk category: movements of counterparty credit risk (and of own credit risk), and of exposures, causing changes in the total BCVA with a positive or negative impact on P/L. BCBS (2010) notes that while the Basel II standard covers the risk of a counterparty default, it does not address similar CVA risk, which during the financial crisis was a greater source of losses than those arising from outright defaults. Therefore, the Basel III regulation (BCBS 2010) introduces a new CVA capital charge to cover the risk of mark-to-market losses on CVA for the OTC derivatives. Banks are not required to calculate this capital charge for transactions with respect to a *central counterparty* (CCP) or securities financing transactions. Note that the regulator does not, in contrast to IFRS 13, consider the bilateral BCVA, which could increase the total market value due to institutions' own credit deterioration.

In principle, the regulator wants banks, in the Internal Market Model (IMM) approach, to calculate the VaR of their portfolio market value incorporating the credit value adjustments

$$MV = \sum_i \left(f_{nd,i} - CVA_i\right)$$

depending not only on the market factors but also on the credit spreads of their individual counterparties. More precisely speaking, the CVA capital charge should be calculated separately from the pure market risk capital charge, i.e. considering counterparty defaults as the only source of losses (but at the same time simulating future exposures depending on the underlying market factors).

Firstly, BCBS (2010) explicitly requires banks to use a simplified formula similar to (5.41) for the purpose of the capital charge calculation:

$$\begin{aligned} CVA = {} & LGD_{MKT} \times \\ & \sum_{i=1}^{T} \max\left(0, \exp\left(-\frac{s_{i-1}\cdot t_{i-1}}{LGD_{MKT}}\right) - \exp\left(-\frac{s_i \cdot t_i}{LGD_{MKT}}\right)\right) \cdot \\ & \left(\frac{EE_{i-1}\cdot D_{i-1} + EE_i \cdot D_i}{2}\right) \end{aligned} \tag{5.52}$$

where s_i is the credit spread corresponding to maturity t_i. The banks should use market CDS spreads whenever available. If a CDS spread is not available, the banks are supposed to use a proxy. Similarly, LGD_{MKT} is the loss given default of the counterparty based on a market instrument, if available, and on a proxy otherwise. Finally, EE_i is the expected exposure and D_i the discount factor, both corresponding to the revaluation time t_i. Note that the term

$$\exp\left(-\frac{s_i \cdot t_i}{LGD_{MKT}}\right)$$

estimates the survival probability until time t_i implied by the loss given default LGD_{MKT} and the spread s_i (and an implicit assumption of a constant intensity of default). Therefore, the first factor in the summation approximates the probability of the counterparty's default between t_{i-1} and t_i, while the second factor stands for an average discounted exposure during this time interval. According to the regulation, any internal model, based on direct CVA revaluation, or on spread sensitivities, must be based on (5.52). Likewise, the Basel III market risk capital charge and the CVA capital charge must be calculated as the sum of non-stressed and stressed VaR components. The non-stressed VaR components use the expected exposures and spread variations corresponding to normal market conditions, while the stressed VaR must use stressed future exposures and spread variations corresponding to a financial crisis.

Looking at the IMM requirements, it is not surprising that the majority of banks will opt, regarding the total capital charge, for a relatively simple standardized formula where the capital charge is calculated as a percentage of each exposure depending on the counterparty's rating, transaction maturity, and possible (counterparty) credit risk protection. The individual capital charges are then aggregated on a portfolio level:

$$K = 2.33\sqrt{\begin{array}{l}\left(\sum_i 0.5w_i\left(M_i EAD_i^{total} - M_i^{hedge}B_i\right) - \sum_{ind} w_{ind}M_{ind}B_{ind}\right)^2 + \\ \sum_i 0.75w_i^2\left(M_i EAD_i^{total} - M_i^{hedge}B_i\right)^2\end{array}} \tag{5.53}$$

where w_i is the weight corresponding to counterparty i based on its rating and Table 5.8 (if there is no external rating, then the bank must, subject to regulatory approval, map its internal rating to an external rating); EAD_i^{total} is the total exposure with respect to the counterparty i (with or without netting) including the effect of collateral; M_i is the effective weighted maturity (duration) of the transaction with respect to the counterparty i; B_i is the notional of purchased single-name CDS hedges (with respect to the reference entity i) with maturity M_i^{hedge}; and, finally, B_{ind} is the notional of purchased index CDS hedge with maturity M_{ind}.

To get some intuition on the formula, note that w_i looks like a regulatory estimate of a "standard" annualized CVA change as a percentage of the duration and hedge adjusted exposure $M_i EAD_i^{total} - M_i^{hedge}B_i$. Therefore, the first part under the square root of (5.53) corresponds to an estimate of the total portfolio CVA standard deviation assuming that the counterparties are perfectly correlated, but allowing for a systematic CDS index hedge, while the second part under the square

Table 5.8 Regulatory CVA weights (BCBS 2010, par. 104)

Rating	Weight $w_i(\%)$
AAA	0.7
AA	0.7
A	0.8
BBB	1.0
BB	2.0
B	3
CCC	10.0

root corresponds to the standard deviation estimate assuming the counterparties are independent. The two estimates correspond to decomposition into single-factor systematic and idiosyncratic factors with the weight $\rho^2 = 0.5^2 = 0.25$ for the systematic variance and the complementary weight $1 - \rho^2 = 0.75$ for the idiosyncratic variance. Therefore, the square root stands for a conservative portfolio CVA standard deviation estimate and the multiplier 2.33 is just the standard normal 99% quantile. So, indeed, the result of (5.53) appears to estimate the 99% CVA VaR in a 1-year horizon (for a more detailed explanation see, for example, Pykhtin 2012).

FVA, KVA, MVA, and Other XVAs

Besides the CVA, DVA, and BCVA that have become more-or-less standard accounting and regulatory concepts, there are other more controversial valuation adjustments like FVA, LVA, KVA, or MVA, altogether denoted as XVAs (Gregory 2015).

To explain the reasoning behind the *Funding Value Adjustment* (FVA), let us consider as an example a non-collateralized derivative position with a positive market value, e.g. a long option. The derivative position is an asset that has been acquired by paying a premium, and the cost is funded internally by a rate corresponding to the institution's cost of financing. On the other hand, the market value accrues only the risk-free rate used for discounting of the derivatives expected cash flow, e.g. the OIS rate. Consequently, there is a difference between the institution's financing rate and the risk-free (OIS) rate, which calls for an additional valuation adjustment. The same, but positive, effect is applicable if a derivative market value is negative, i.e. a liability. In this case, the interest cost of the liability equals the OIS rate, but the interest revenue is the funding rate. The funding cost spread FS_C and the funding benefit spread FS_B could be generally different, and so we should calculate separately the two $FVA = FCA + FBA$ components, i.e. the *Funding Cost Adjustment* (FCA) and the *funding benefit adjustment* (FBA). Mathematically,

$$FVA = E\left[\int_0^T e^{-rt}\max(f_t, 0)\cdot FS_C(t)\cdot S(t)dt\right] - E\left[\int_0^T e^{-rt}\max(-f_t, 0)\cdot FS_B(t)\cdot S(t)dt\right], \quad (5.54)$$

Where FS_C is the funding (cost) spread on the asset side, FS_B is the funding (benefit) spread on the liability side, and $S(t)$ the transaction survival probability, i.e. the joint survival probability for both counterparties.

The adjusted derivative market value should now be

$$f_0^{adj} = f_0^{nd} - CVA + DVA - FVA.$$

It should be noted that the concept of FVA remains controversial. According to Hull and White (2012b), FVA should not be reflected in the valuation of derivatives—the standard valuation argument says that derivative value should be equal to the risk-neutral expectation of the cash-flows discounted by the risk-free rate, not by any entity specific funding rate. Moreover, if the funding costs are applied, then the valuation of derivatives will be asymmetric and arbitrage opportunities will exist. In spite of the continuing academic discussion, according to a market survey (Gregory 2015), the majority of large global banks do account for FVA with a total impact in billions of USD.

Regarding practical calculations, the usual assumption is that the exposure and funding spreads are independent. Moreover, depending on close-out assumptions (Gregory 2015), the survival probability can be neglected. Then, after a standard discretization, we have the following relatively simple formula

$$FVA = \sum_{j=1}^{m} e^{-rt_j}EE(t_j)\cdot \overline{FS}_C(t_j)\Delta t_j - \sum_{j=1}^{m} e^{-rt_j}ENE(t_j)\cdot \overline{FS}_B(t_j)\Delta t_j \quad (5.55)$$

Where $ENE(t)$ is the expected negative exposure, $\overline{FS}_C(t)$ the expected (or forward) funding cost spread, and $\overline{FS}_B(t)$ the expected funding benefit spread.

However, defining the FVA component more precisely, it has become obvious that there is an overlap with the concept of DVA. The institution's funding spread should be theoretically equal to the product of the default probability and the institution's LGD, i.e. $\overline{FS}_B(t_j) = l\cdot q(t_j)$, hence

$$FBA = -\sum_{j=1}^{m} e^{-rt_j} ENE(t_j) \cdot \overline{FS}_B(t_j)\Delta t_j = -l\sum_{j=1}^{m} e^{-rt_j} ENE(t_j) \cdot q(t_j)\Delta t_j = DVA.$$

Note that this argument does not apply to CVA and FCA, since CVA uses the counterparty's default probabilities, while the FCA institution's funding spread depends on its own funding probabilities. One simple solution of this finding is to apply either FBA or DVA, i.e. the total adjustment would be either the *"CVA and symmetric funding"* $CVA + FVA$ or *"bilateral CVA and asymmetric funding"* $BCVA + FCA$. According to Gregory (2015), the market practice prefers the asymmetric funding approach.

Another consistent solution to the CVA/FVA overlap is to define the funding benefit spread only as the liquidity spread above the risk-free rate, plus a standard institution's credit spread, i.e. we should define rather the liquidity (benefit) spread as the difference between the real market spread and the theoretical spread: $LS_B(t_j) = MS_B(t_j) - l \cdot q(t_j)$. Some authors (Gregory 2015; Hull and White 2012b, c; Crépey et al. 2014) argue that the same principle applies to FCA on the asset side. The reasoning is that the asset quality influences an institution's overall credit quality, and so the cost of funding should depend on the asset credit risk. For example, an investment into treasury bonds has no (negative) contribution to the institution's credit quality, and so should be funded by the risk-free rate, possibly plus an institution-specific liquidity spread. Therefore, we obtain an alternative definition of FVA, denoted rather as *LVA—Liquidity Valuation Adjustment* (Crépey et al. 2014), where the funding spread in (5.54) is replaced by the liquidity spread $LS_B(t_j) = MS_B(t_j) - l \cdot q(t_j)$ only. In market terms, the liquidity spread can be estimated as the difference between the institution's bond yield spread and the CDS spread. It seems that the concept of LVA resolves the academic controversy, and is accepted, for example, by Hull and White (2014).

Another XVA to be mentioned is the *Margin Valuation Adjustment—MVA*. While FVA is related to uncollateralized transactions, MVA arises due to standard overcollateralization requirements, mainly due to initial margin posting. Organized exchange derivatives positions or OTC positions cleared by a CCP (central counterparty) involve initial margin and maintenance margin management. The requirement is defined not to collateralize actual counterparty's losses, but in order to cover potential losses over a short-time horizon (1–10 days) and on a high confidence level (e.g. 99%). The excess margin balance usually exists even for bilateral transactions depending on the margin mechanism. In any case, the margin balance earns a return R_{IM} that will be at most equal to the OIS rate, and its financing at the same time represents a funding cost FC that will be larger than the OIS rate. Therefore, the MVA can be defined mathematically as follows:

$$MVA = E\left[\int_0^T e^{-rt} IM(t) \cdot (FC(t) - R_{IM}(t)) \cdot S(t)dt\right]$$

Where $IM(t)$ is the initial margin balance and $S(t)$ the (joint) survival probability. As above, the calculations can be simplified by discretizing the time interval and assuming independence between the margin balance and the funding spread.

Finally, let us look the *Capital Valuation Adjustment—KVA*, which is supposed to reflect the cost of regulatory capital related to derivative transactions. Traditionally, there has been the market risk capital requirement, calculated for different product portfolios and market factors. The capital requirement is significant for proprietary positions but can be neglected for back-to-back operations. Another component is the classical default CCR capital requirement defined as the RWA times 8%. The derivative exposures can be calculated according to Basel II rules by several methods: CEM—current exposure method, SM—standardized exposure method, and IMM—internal model method. The new Basel III component is the CVA charge described above. The total capital requirement $C(t)$ that needs to be calculated on a (derivative) portfolio level again represents a cost, in this case the cost of capital $CC(t)$ The cost of capital should be considered rather as the spread between the required capital return and the risk-free rate (since the capital per se can be invested into risk-free assets). Thus, the KVA mathematical definition is

$$KVA = E\left[\int_0^T e^{-rt} C(t) \cdot CC(t) \cdot S(t)dt\right].$$

As usual the formula can be discretized and based on the expected capital requirements $EC(t)$ and other relevant expected future parameters.

While CVA and (more or less) FVA have become accounting standards, MVA and KVA are used rather for reporting and monitoring. There is an ongoing debate regarding the consequences of their accounting, possible side-effects and overlaps. In any case, the debate around FVA mentioned above should be resolved first of all, before the institutions start to account for the other XVAs.

Conclusion 6

The art and science of credit risk pricing, measurements, and management have been on a long journey and made a significant progress during recent decades. The classic credit risk management question as to whether a loan application should be approved or not, and, possibly, under which conditions, used to be approached by experts based on their experience and analytical skills. Advances in mathematics, statistics, and computer power have brought new sophisticated, automated methods that either support or completely replace skilled credit analysts. The growth of banking portfolios, in particular in the area of consumer and mortgage financing, has highlighted the issue of portfolio credit risk that is not driven by losses on any single exposure, but rather by losses due to higher than expected default rates. As we have seen in Chap. 4, portfolio modeling involves not only the estimation of individual default probabilities, but also the concept of default correlation, which remains challenging even today. The development of credit risk management standards in financial institutions has gone hand in hand with changes in the Basel regulation, which aims to set basic risk management standards and define regulatory capital based on the risk undertaken. The concepts of rating, PD, LGD, EAD, or expected and unexpected loss, were used by many advanced credit risk managers before Basel II, but since the introduction of the new regulation, these concepts have really become standard and widespread. Since the nineties we have seen rapid growth of the derivative markets. Counterparty credit risk used to be handled in a relatively simple way through limits on counterparty exposures based on various simple equivalents. Recently, in particular after the financial crisis, the issue of counterparty credit has become much more complicated with the advance of the many different valuation adjustments (XVAs) discussed in Sect. 5.6. Last but not least, we should mention the credit derivatives and credit derivative-linked securities, such as CDOs, which started to be traded actively at the beginning of the previous decade. Their pricing and risk management, seriously challenged by the

J. Witzany, *Credit Risk Management*, DOI 10.1007/978-3-319-49800-3_6

financial crisis, still pose a real challenge to financial engineers, as shown in Chap. 5. We hope that this text provides not only an overview of credit risk pricing, measurement, and management methods, but also that it will contribute to the ongoing research in this area.

Index

A
Accuracy ratio (AR), 22
Adjustment
  capital valuation (KVA), 239
  credit valuation, bilateral (BCVA), 228
  debit valuation (DVA), 228
  funding benefit (FBA), 236
  funding cost (FCA), 236
  funding value (FVA), 236
  liquidity valuation (LVA), 238
  margin valuation (MVA), 238
  total valuation (TVA), 228
  value (XVA), 236
Affine form, 207
Algorithm
  back-propagation, 81
  Newton–Raphson, 43, 72
Altman's Z-Score, 38
Analysis, univariate, 45
Annuity
  risk-free, 223
  risky, 224
Approach
  add-on, 221
  advanced internal rating based (IRBA), 17, 110
  advanced measurement(AMA), 16
  basic indicator, 16
  cohort, 105
  comprehensive, 112
  extrapolation, 63
  fixed time horizon, 105
  foundation internal rating based (IRBF), 16, 110
  internal assessment (IAA), 219
  internal market model (IMM), 234
  internal model (IMA), 15
  internal rating based (IRB), 13, 110, 219
  momentum, 102
  nearest neighbor, 83
  PD-weighted, 104
  pool level, 107
  portfolio invariant, 16
  rating based (RBA), 219
  simple, 112
  standardized (SA), 13, 110
  substitution, 217
  supervisory formula (SFA), 219
  Value at Risk, 15
  variable time, 105
Area under the curve (AUC), 22
Arrears, payments in, 162
Artificial neural networks, 37
Asset
  backed security (ABS), 174
  correlation, 117
  return, standardized, 123, 126
  swap, 164
  volatility, 126, 142
Asset-backed security (ABS)
  pass-through, 174
  tranche, 175
  waterfall, 175
Augmentation, 62

B
Bagging, 88
Bclear platform, 171
Beta, asset, 149
Bond
  junior subordinated, 122
  senior, 122
Breslow-Crowley estimator, 72

J. Witzany, *Credit Risk Management*, DOI 10.1007/978-3-319-49800-3

Buffer
conservation, 157
countercyclical, 157

C
Canonical maximum likelihood (CML), 197
Capital
adequacy ratio (CAR), 16, 110
CVA charge, 217, 234
economic, 118, 120, 157
market line, 117
regulatory, requirement, 16
risk adjusted return (RAROC), 120
CDX NA IG, 170
Chinese wall, 8
Choleski decomposition, 197
Classification
accuracy (ACC), 29
coarse, 52
system, 8
tree, 75
Collateralization
continuous, 225
minimum threshold, 225
one-way, 225
real estate, 131
two-way, 225
Collateralized Debt Obligations (CDO)
cash flow, 176
hybrid, 176
market value, 176
synthetic, 176
unfunded synthetic, 178
waterfall, 182
Committee
Basel, on Banking Supervision (BCBS), 13
New Product, 12
rating, 35
Compensator, 211
Confidence interval, 130
Confusion matrix, 29
Contingent claim, 146
Continuous compounding, 207
Contract, forward, 8
Conversion
factor (CF), 16, 102
factor, credit (CCF), 102
Copula
Archimedean, 195
Archimedean, generator, 196
bivariate, 193
Clayton's, 196
double-t, 191
Frank's, 196
Fréchet family of, 195
Fréchet mixture, 195
function, 193
Gaussian, 182, 226
Gumbel's, 196
implied, 198
independent, 194
logistic, 191
Marshall-Olkin, 195
maximum, 194
minimum, 194
multi-variate, 198
product, 194
Student t, 195
Correctness of categorical predictions, 29
Correlation
asset, 121
base, 188
compound, 188
default, recovery rate, 131
implied, 188
interest rate and default intensity, 215
matrix, 47
rating migration, 121
regulatory, 155
skew, 190
Cost of errors, 31
Counterparty
central (CCP), 219, 234
credit risk, bilateral, 228
credit risk (CCR), 219
Credit
bureau, 8
conversion factor (CCF), 112
default swap (CDS), 160, 223
enhancement level (L), 219
event, 161
linked note, 12
multi-name derivatives, 179
protection, 217
protection buyer, 162
protection seller, 163
risk mitigation, 217
spread, 121, 163
support annex (CSA), 225
valuation adjustment (CVA), 3, 10, 220
valuation adjustment, capital charge, 234
Value at Risk, 153
Credit default swaps (CDS)
add-up, 172

basket, 172
basket, analytical valuation, 184
first-to-default, 173
forward, 179
index, 170
multi-name, 163
nth-to-default, 173
option, 179
protection, 164
second-to-default, 173
single-name, 163
spread, 168
valuation, 169
CreditMetrics, 120
CreditPortfolioView, 139
CreditRisk+, 133
Cross-validation, 54
Cumulative accuracy profile (CAP), 22
Cut-off score, 29

D
Default
correlation, 33, 173
cross, 144
definition of, 20
distance to (DD), 144, 208
observation cohorts, 20
Poisson arrival of, 210
rate, 20
rate, unexpected, 154
threshold, 148
time to, 153
Derivative
collateralized, 233
credit, 2
financial, 159
over-the-counter (OTC), 219
payoff, 159
Discriminant analysis, 39
Distribution
beta, 131
binomial, 34, 134, 185
chi-squared, 32, 45
empirical, 34
gamma, 134, 135
generalized-Poisson, 203
hyperbolic, 192
logistic, 41
loss, 203
marginal, 193
multivariate normal, 153
negative binomial, 136
normal, 34
normal inverse, 143
normal inverse gaussian (NIG), 192
Poisson, 134
risk neutral loss, 181
Student t, 191

E
Economic downturn, 115, 156
Effective
federal funds rate, 230
maturity (M), 16
number of exposures (N), 219
Efficient Frontier, 117
Eligible protection provider, 217
Entropy index, 76
Equation
Chapman-Kolmogorov, 92
stochastic differential, 125
Error
function, squared, 81
standard, 43
type I, 29
type II, 29
Euclidean metrics, 83
Exercise price, 143
Expected
default frequency (EDF), 142, 144, 148
risk-neutral value, 233
shortfall, 139
tranche loss (ETL), 200
value, conditional, 206
Expert systems, 38
Exposure
corporate, 113
counterparty risk equivalent, 10
credit, 9
at default (EAD), 16, 96, 102
expected conditional, 226
expected (EE), 221
expected effective, 225
expected positive (EPE), 223
potential future, 225
re-securitized, 217
retail, 115
External Credit Assessment Institutions (ECAIs), 111
External rating agency
Moody's, 19
Standard & Poor's, 19

F
Factor
  debtor specific, 127
  idiosyncratic, 121, 184
  systematic, 127, 184
Fannie Mae, 175
Fee, upfront, 180
Financial
  crisis, 160
  ratio, 45
Formula
  Basel II, 153
  Black-Scholes-Merton, 126, 143
  CVA regulatory, 235
  regulatory, 16, 113
  Vasicek's, 154
Forward transaction, 9, 222
Fréchet bounds, 194
Freddie Mac, 175
Function
  activation, 80
  baseline hazard, 71
  characteristic, 203
  cumulative hazard, 69, 184
  Excel, BINOMDIST, 33
  Excel, NORMSINV, 119
  2-increasing, 193
  indicator, 220
  link, 41, 108
  logistic, 80
  logit, 108
  moment-generating, 191
  probability generating, 134
  tangent hyperbolic, 80
  threshold, 80

G
Geometric Brownian motion, 125, 142
Gini
  coefficient, 22
  index, 76
Ginnie Mae, 175
Goodness-of-fit, 22
Gradient, 81
Gross carrying amount, 102

H
Hessian, 43
Hypothesis
  null, 43
  testing, 28

I
Impurity measure, 76
Index, country, 127
Inequality, Fréchet–Hoeffding, 194
Information value (IV), 45, 49, 51
Installment, calendar, 109
Insurance equivalence relation, 224
Intensity
  of default, 69, 165, 221, 224
  of default, forward, 209
  Poisson, 202
  risk neutral, 167
  of transition, 93
Internal cost of funds, 94
International Swap and Derivatives Association (ISDA), 163
  framework documentation, 163, 220
iTraxx Europe, 170

K
Kendall's tau, 194
Kernel trick, 85
KMV Portfolio Manager, 142
Knock-Out criteria, 61
Kurtosis, 41

L
Leg
  default, 188
  premium, 188
Lemma, Ito's, 143, 207
Likelihood, 42
  partial, 72
Limit
  concentration, 118
  credit, 9
  exposure, 6
Linear programming, 83, 84
Loan
  impaired, 8
  loss provisions, 91
  past due, 112
Lognormal property, 143
Lorenz curve, 23
Loss
  expected, best estimate (BEEL), 115
  expected (EL), 15, 95, 118
  given default, exposure weighted (EWALGD), 219
  given default (LGD), 16, 220
  mark-to-market, 234

unexpected (UL), 15, 118
unexpected, modeling, 180

M

Margin
conservative, 110
of conservatism, 114, 156
countercyclical, 110
hard, 84
initial, 238
maintenance, 238
soft, 85
Markov
chain, 92
property testing, 92
Martingale, 206
Matrix
Cholesky, 129
transition, 90
transition intensity, 93
transition probability, 90
Maturity
adjustment, 155
mismatches, 217
Mean
default-weighted, 107
weighted, 107
Measure
forward risk-neutral, 205
real-world probability, 205
Method
current exposure (CEM), 239
fuzzy, 62
gradient-descent, 81
hard cut-off, 63
inference for the margins (IFM), 197
maximum likelihood estimation (MLE), 41
momentum, 82
Model
accelerated failure time, 71
affine intensity, 216
auto-regressive, 140
capital asset pricing (CAPM), 117, 149
CIR with jumps, 214
Cox, Ingersoll, and Ross (CIR), 208, 212
Cox proportional hazard, 71
ensemble, 86
exponential, 70, 94
first-passage, 208
generalized Poisson loss (GPL), 202
Heath, Jarrow, and Morton (HJM), 204, 212
Ho-Lee, 208
homogenous Markov chain, 90
Hull-White, 208
internal, 155
internal market (IMM), 234
KMV, 126
large homogenous portfolio (LHP), 187
LIBOR-market, 204
logit, 40
loglogistic, 71
lognormal, 71
Markov chain, 89
Merton's structural, 123
multi-factor, macroeconomic, 139
non-parametric, 73
one-factor Gaussian copula, 153
parametric, 73
perfect, 24
probit, 40
random, 24
reduced form, 133, 204, 210
risk, 190
short-rate, 204
standard market, 185
structural, 38, 41, 204
structural stochastic, 208
two-systematic-factor, 131
Vasicek's, 153, 204, 210
Weibull, 71
Money-market account, 205
Monte Carlo simulation, 34, 119, 181
Mortgage-backed security (MBS), 174

N

Netting agreement, 225
Neural network, 79
deep, 80
multi-layer, 80
single-layer, 80
Neuron
hidden, 80
output, 80
Numeraire, 205
NYSE Euronext, 171

O

Observation, weighted, 62
Odds
good-bad, 43
log good-bad, 43
population good-bad, 49
Office

Office (*cont.*)
back, 11
front, 11
middle, 11
Officer
chief executive, 6
chief finance (CFO), 5
chief risk (CRO), 5
Option
American, 126
barrier, 208
European call, 125, 143
European put, 125
Overnight Index Average (ONIA), 230
Czech (CZEONIA), 231
Euro (EONIA), 230
Sterling (SONIA), 231
Over the counter (OTC), 159

P
Parameter
scale, 135
shape, 135
Partial differential equation, Black-Scholes-Merton, 207
Pillar
market discipline, 110
minimum capital requirements, 110
supervisory review process, 110
Point
attachment, 186
detachment, 186
Point in time (PIT), 65, 111
Portfolio
credit risk, 117
granularity, 154
invariant, 154
large homogenous (LHP), 199
market, 149
simulation, 129
theory, 117
Precision, 29
Present value
contingent cash flows, 146
risk adjusted, 122
Principle, equivalence, 109
Prioritization scheme
fast, 183
uniform, 183
Probability
conditional, of default, 187
of default (PD), 16
PIT, of default, 66
rating migration, 121
risk neutral, 146, 163
survival, 108, 223
transition, 90, 122
TTC, of default, 66
Process
basic affine, 212
compound Poisson, 209
continuous time Markov, 93
credit assessment, 19
default-counting, 202
doubly stochastic, 210
mean-reverting with jumps, 212
Poisson, 202
rating, 19
recalibration, 65
supervisory review, 17
Wiener, generalized, 226
Pro-cyclicality, 156
Provisions, 8
eligible, 116
individual, 116
portfolio, 116
Pruning, 75
P-value, 44

Q
Quadratic programming, 85
Quantile, 118
Quantitative study, 155

R
Random forest, 86
Rate
agency, 35
analytical, 34–37
calibration, 64
discount, 109
effective interest (EIR), 101
external, 16, 111
false alarm, 24
forward, 122
grades, 66
hazard, 69, 165
hit, 24
instantaneous interest, 205
Kaplan-Meier hazard, 73
model, automated, 37
OIS, 231

qualitative, 63
quantitative, 63
recovery, 97, 122, 156
risk-free interest, 121, 164, 229
scale, 19
shadow, 37, 67
symbols, 35
system, 20
system, discrimination power of, 22
zero coupon, 122
Ratio
leverage, 217
Sharpe's, 149
Receiver operating characteristic curve, 22
Reference
date, 103
data set (RDS), 104
entity, 161
Regression
linear, 39
logistic, 40
OLS, without constant, 108
ordinary least squares (OLS), 40
Regulation
Basel II, 12, 110, 153, 218
Basel 2.5, 218
Basel III, 11, 13, 110, 157
Capital Adequacy Directive (CAD), 13
credit-risk securitization framework, 13
Market Risk Amendment, 15
Solvency II, 155
The first Capital Accord, 13
The New Capital Accord, 13
Reject
bias, 61
inference, 61
Remargining period, 225
Repo operation, 219
Reweighting, 62
Risk
adjustment, 149
counterparty credit (CCR), 9
credit, 2
management, 5
market, 13
operational, 13
premium (RP), 94, 108, 149
price of, 148
right way, 226
settlement, 9
weight, 16, 111
weighted assets (RWA), 16, 110
wrong way, 226
Risk-neutral
probability of default, 144
process, 148
valuation principle, 220
world, 142, 205
Risky annuity, 186

S
Sample
development, 44
in, 44
out, 44
testing, 27
training, 27
validation, 44
Score
behavioral, 7
cut-off, 96
log-odds, 43
naïve Bayes, 50
rating, 43
scale, 20
Sector index, 121
Securities Industry and Financial Markets Association (SIFMA), 176
Securitization, 217
Selection
backward, 44
forward, 44
forward stepwise, 45
of variables, 44
Separation
hyperplane, 84
of powers, 5
Settlement
financial, 161
physical, 161
Signal
correct, 22
wrong, 22
Simulation, inverse survival function, 211
Single tranche trading, 178
Sinking-fund, 183
Somers' D, 22
Spearman's rho, 194
Special Purpose Vehicle (SPV), 174
Spread
basis swap, 230
bond, 229
CDS, 229
credit, up-front, 181
cross-currency swap basis, 230

Spread (*cont.*)
funding benefit, 236
funding cost, 236
TED, 231
Standards
International Accounting (IAS), 101
International Financial Reporting (IFRS), 101
State, absorbing, 90
Statistic
CS, 32, 48, 77
F-value, 39
Hosmer-Lemeshow, 32
Kendall's tau, 26
Kolmogorov-Smirnov, 26, 75
Kruskall-Wallis, 26
T, 28
Wald, 43
Statistical package, 28
Stress testing, 157
Supervisory formula (SF), 219
Support vector, 85
machines (SVM), 38, 83
Survival
analysis, 67
function, 69
Kaplan-Meier, function, 73
Swap
credit default (CDS), 223
interest rate, 222
over-night index (OIS), 230

T
Tail dependency, 195
Taylor expansion, 136
Test
binomial, 33
Hosmer-Lemeshow, 32
Theorem
Bayes, 49
Sklar's, 193
Theory, Markowitz portfolio, 117
Thickness (T), 219
Through the cycle (TTC), 65, 111, 114
Total return
payer, 169
receiver, 169
swap (TRS), 169
Tranche
equity, 175
mezzanine, 175
payoff, 180
senior, 175
subordinated, 176
Transformation
Fourier, 191, 203
Gil-Pelaez, 191
Laplace, 191
quantile-to-quantile, 132, 153, 196
Trees, binary classification, 77

V
Validation
out-of-sample, 27, 47
of PD estimates, 32
sample, 32
Value
forward, 122
negative predictive, 30
positive predictive, 29
at Risk, conditional, 139
at Risk, non-stressed, 235
at Risk, stressed, 235
at Risk (VaR), 118
Variable
categorical, 40, 47
continuous, 40
default, 39
dummy, 40, 47
explanatory, 38
ordinal, 40
short list, 45
uniform, 195

W
Weight of evidence (WoE), 49
WoEization, 53
Workout, 7

Z
Zero coupon curve
OIS, 230
risk-free, 232
swap, 229