CREDIT RISK ANALYTICS

MEASUREMENT TECHNIQUES, APPLICATIONS, *and* EXAMPLES *in* SAS®

+website

BART BAESENS
DANIEL RÖSCH
HARALD SCHEULE

WILEY

Contents

CHAPTER 1

Introduction to Credit Risk Analytics

Welcome to the first edition of *Credit Risk Analytics: Measurement Techniques, Applications, and Examples in SAS*.

This comprehensive guide to practical credit risk analytics provides a targeted training guide for risk professionals looking to efficiently build or validate in-house models for credit risk management. Combining theory with practice, this book walks you through the fundamentals of credit risk management and shows you how to implement these concepts using the SAS software, with helpful code provided. Coverage includes data analysis and preprocessing, credit scoring, probability of default (PD) and loss given default (LGD) estimation and forecasting, low default portfolios, Bayesian methods, correlation modeling and estimation, validation, implementation of prudential regulation, stress testing of existing modeling concepts, and more, to provide a one-stop tutorial and reference for credit risk analytics.

This book shows you how to:

- Understand the general concepts of credit risk management
- Validate and stress test existing models
- Access working examples based on both real and simulated data
- Learn useful code for implementing and validating models in SAS
- Exploit the capabilities of this high-powered package to create clean and accurate credit risk management models

WHY THIS BOOK IS TIMELY

Despite the high demand for in-house models, there is little comprehensive training available. Practitioners are often left to comb through piecemeal resources, executive training courses, and consultancies to cobble together the information they need. This book ends the search by providing a thorough, focused resource backed by expert guidance.

Current Challenges in Credit Risk Analytics

Commercial banks are typically large in size, and their fundamental business model continues to rely on financial intermediation by (1) raising finance through deposit taking, wholesale funding (e.g., corporate bonds and covered bonds), and shareholder capital, and (2) lending, which is a major source of credit risk.

Commercial bank loan portfolios consist to a large degree of mortgage loans, commercial real estate loans, and small and medium-sized enterprise (SME) company loans. SME loans are often backed by property collateral provided by the SME owners. The reliance of commercial bank loan portfolios on real estate is fundamental. Note that various types of mortgage loans exist. Examples are prime mortgages, subprime mortgages, reverse mortgages, home equity loans, home equity lines of credit (HELOCs), and interest-only loans, as well as variable, fixed-rate, and hybrid loans, to name a few.

Further loan categories include consumer loans (car loans, credit card loans, and student loans) and corporate loans. Loans to large companies also exist but compete with other funding solutions provided by capital markets (i.e., issuance of shares and corporate bonds).

Other sources of credit risk are fixed income securities (e.g., bank, corporate, and sovereign bonds), securitization investments, contingent credit exposures (loan commitments and guarantees), credit derivatives, and over-the-counter (OTC) derivatives.

Credit risk was at the heart of the global financial crisis (GFC) of 2007 to 2009 and is the focus of this book. Post GFC, prudential regulators have increased risk model requirements, and rigorous standards are being implemented globally, such as:

- Implementation of Basel III: The Basel rules concern capital increases in terms of quantity and quality, leverage ratios, liquidity ratios, and impact analysis. We will discuss the Basel rules in more detail later.
- Stress testing: Regulators require annual stress tests for all risk models.
- Consistency across financial institutions and instruments: Regulators are currently identifying areas where regulation is applied in inconsistent ways.
- Reinvigoration of financial markets (securitization): A number of markets, in particular the private (i.e., non-government-supported) securitization market, have declined in volume.
- Transparency: Central transaction repositories and collection of loan-level data mean more information is collected and made available to credit risk analysts.
- Increase of bank efficiency, competition, deregulation, and simplification: The precise measurement of credit risk is a central constituent in this process.

Risk model methodologies have advanced in many ways over recent years. Much of the original work was based in science where experiments typically abstracted from business cycles and were often applied within laboratory environments to ensure that the experiment was repetitive. Today, credit risk models are empirical and rely on historical data that includes severe economic downturns such as the GFC.

State-of-the-art credit risk models take into account the economic fundamentals of the data generating processes. For example, it is now common to include the life cycle of financial products from origination to payoff, default, or maturity while controlling for the current state of the economy. Another aspect is the efficient analysis of available information, which includes Bayesian modeling, nonparametric modeling, and frailty modeling. Risk models are extended to exploit observable and unobservable information in the most efficient ways.

Despite all these advancements, a word of caution is in order. All empirical risk models remain subject to model risk as we continue to rely on assumptions and the historical data that we observe. For example, it is quite common to obtain *R*-squared values of 20 percent for linear LGD and exposure at default (EAD) models. As the *R*-squared measures the fraction of the observed variation that is explained by the model, these numbers suggest that there is a considerable amount of variation that these models do not explain. Providing more precise models will keep us busy for years to come!

A Book on Credit Risk Analytics in SAS

In our academic research, we work with a number of software packages such as C++, EViews, Matlab, Python, SAS, and Stata. Similar to real languages (e.g., Dutch and German), being proficient in one package allows for quick proficiency in other packages.

In our dealings with credit risk analysts, their financial institutions, and their regulators, we realized that in the banking industry SAS is a statistical software package that has come to be the preferred software for credit risk modeling due to its functionality and ability to process large amounts of data. A key consideration in the industry for using SAS is its quality assurance, standardization, and scalability. We will discuss this point in the next chapter in more detail.

Most documentation available for statistical software packages has been developed for scientific use, and examples usually relate to repeatable experiments in medicine, physics, and mathematics. Credit risk analytics is multidisciplinary and incorporates finance, econometrics, and law. Training material in this area is very limited, as much of the empirical work has been triggered by the digitalization and emergence of big data combined with recent econometric advances. Credit risk analytics requires the consideration of interactions with the economy and regulatory settings, which are both dynamic and often nonrepeatable experiments. We learned a great deal from existing literature but continuously reached limits that we had to overcome. We have collected much of this research in this text to show you how to implement this into your own risk architecture.

Structure of the Book

This book contains 15 chapters. We deliberately focused on the challenges in the commercial banking industry and on the analysis of credit risk of loans and loan portfolios.

Following the introduction in the first chapter, the book features three chapters on the preparation stages for credit risk analytics. The second chapter introduces Base SAS, which allows you to explicitly program or code the various data steps and models, and SAS Enterprise Miner, which provides a graphical user interface (GUI) for users that aim to extract information from data without having to rely on programming. The third chapter introduces how basic statistics can be computed in SAS, and provides a rigorous statistical explanation about the necessary assumptions and interpretations. The fourth chapter describes how data can be preprocessed using SAS.

Next, we have included five chapters that look into the most modeled parameter of credit risk analytics: the probability of default (PD). The fifth chapter develops linear scores that approximate the default probabilities without the constraints of probability measures to be bounded between zero and one. Credit scores are often provided by external appraisers to measure default behavior. Examples are real estate indexes, bureau scores, collateral scores, and economic indicators. The sixth chapter discusses methodologies to convert scores and other pieces of information into default probabilities by using discrete-time hazard models. Discrete-time methods are relatively simple, and their estimation is robust and has become a standard in credit risk analytics. The seventh chapter builds further on this and estimates default probabilities using continuous-time hazard models. These models explicitly model the life cycle of a borrower and do not assume that observations for a given borrower are independent over time, which discrete-time hazard models often do. The eighth chapter discusses the estimation of default probabilities for low default portfolios, which is a particular concern for small portfolios in relation to large and/or specialized loans.

In the next section, we consider other important credit risk measures. In Chapter 9, we estimate default and asset correlations. We compute credit portfolio default rates and credit portfolio loss distributions using analytical and Monte Carlo simulation–based approaches, and show the reader how correlations can be estimated using internal data. The tenth chapter presents marginal loss given default (LGD) models and LGD models that condition on the selecting default event. The eleventh chapter discusses exposure at default (EAD) models, which are similar in structure to LGD models.

In the last part of the book, we discuss capstone modeling strategies that relate to the various models built in prior sections. Chapter 12 discusses Bayesian models, which allow the analyst to base the model estimation on the data set and prior information. The priors may stem from experts or information collected outside the analyzed system. We show how to implement Bayesian methods and where they might be most useful. Chapter 13 reviews concepts of model validation along with regulatory requirements, and Chapter 14 discusses stress testing of credit risk models by building credit risk measures conditional on stress tests of the macroeconomy, idiosyncratic information, or parameter uncertainty. Chapter 15 concludes the book.

The companion website (www.creditriskanalytics.net) offers examples of both real and simulated credit portfolio data to help you more easily implement the concepts discussed.

THE CURRENT REGULATORY REGIME: BASEL REGULATIONS

We take a closer look at the Basel I, Basel II, and Basel III Capital Accords. These are regulatory guidelines that were introduced in order for financial institutions to appropriately determine their provisions and capital buffers to protect against various risk exposures. One important type of risk is credit risk, and in this section we discuss the impact of these accords on the development of PD, LGD, and EAD credit risk models. The Basel regulations underly many aspects of credit risk analytics, and we will come back to the various issues in later chapters.

Regulatory versus Economic Capital

Banks receive cash inflow from various sources. The first important sources are bank deposits like savings accounts, term accounts, and so on. In return, the depositors receive a fixed or variable interest payment. Another source is the shareholders or investors who buy shares, which gives them an ownership in the bank. If the firm makes a profit, then a percentage can be paid to the shareholders as dividends. Both savings money and shareholder capital are essential elements of a bank's funding. On the asset side, a bank will use the money obtained to make various investments. A first investment, and part of a key banking activity, is lending. Banks will lend money to obligors so that they can finance the purchase of a house or a car, study, or go traveling. Other investments could be buying various market securities such as bonds or stocks.

Note that these investments always have a risk associated with them. Obligors could default and not pay back the loan, and markets could collapse and decrease the value of securities. Given the societal impact of banks in any economic system, they need to be well protected against the risks they are exposed to. Bank insolvency or failure should be avoided at all times, and the risks that banks take on their asset side should be compensated by appropriate liabilities to safeguard their depositors. These people should be guaranteed to always get their savings money back whenever they want it. Hence, a bank should have enough shareholder capital as a buffer against losses. In fact, we could include retained earnings and reserves and look at equity or capital instead. In other words, a well-capitalized bank has a sufficient amount of equity to protect itself against its various risks. Thus, there should be a direct relationship between risk and equity.

Usually, this relationship is quantified in two steps. First, the amount of risk on the asset side is quantified by a specific risk number. This number is then plugged into a formula that precisely calculates the corresponding equity and thus capital required. There are two views on defining both this risk number and the formula to be used.

The first view is a regulatory view whereby regulations such as Basel I, Basel II, and Basel III have been introduced to precisely define how to calculate the risk number and what formula to use. Regulatory capital is then the amount of capital a bank should have according to a regulation. However, if there were no regulations, banks would still be cognizant of the fact that they require equity capital for protection. In this case, they would use their own risk modeling methodologies to calculate a risk number and use their own formulas to calculate the buffer capital. This leads us to the concept of economic capital, which is the amount of capital a bank has based on its internal modeling strategy and policy. The actual capital is then the amount of capital a bank actually holds and is the higher of the economic capital and the regulatory capital. For example, Bank of America reports at the end of 2015 a ratio of total capital to risk-weighted assets using advanced approaches of 13.2 percent and a current regulatory minimum capital of 8 percent (this number will increase as Basel III is fully phased in). Therefore, the capital buffer is currently 5.2 percent.

Note that various types of capital exist, depending upon their loss-absorbing capacity. Tier 1 capital typically consists of common stock, preferred stock, and retained earnings. Tier 2 capital is of somewhat less quality and is made up of subordinated loans, revaluation reserves, undisclosed reserves, and general provisions. The Basel II Capital Accord also included Tier 3 capital, which consists of short-term subordinated debt, but, as we will discuss later, this has been abandoned in the more recent Basel III Capital Accord.

Basel I

The Basel Accords have been put forward by the Basel Committee on Banking Supervision. This committee was founded in 1974 by the G10 central banks. Nowadays, it counts 27 members. They meet regularly at the Bank for International Settlements (BIS) in Basel, Switzerland.

The first accord introduced was the Basel I Capital Accord, in 1988. As already mentioned, the aim was to set up regulatory minimum capital requirements in order to ensure that banks are able, at all times, to return depositors' funds. The Basel I Accord predominantly focused on credit risk and introduced the idea of the capital or Cooke ratio, which is the ratio of the available buffer capital and the risk-weighted assets. It put a lower limit on this ratio of 8 percent; in other words, the capital should be greater than 8 percent of the risk-weighted assets. We have been asked where this number comes from and speculate that it was an industry average at the time of implementation of the first Basel Accord. Changing the capital requirement by only a few percentage points is a challenging undertaking for large banks and takes many years. The capital could consist of both Tier 1 and Tier 2 capital, as discussed earlier.

In terms of credit risk, the Basel I Capital Accord introduced fixed risk weights dependent on the exposure class. For cash exposures, the risk weight was 0 percent, for mortgages 50 percent, and for other commercial exposures 100 percent. As an example, consider a mortgage of \$100. Applying the risk weight of 50 percent, the risk-weighted assets (RWA) then become \$50. This is the risk number we referred to earlier. We will now transform this into required capital using the formula that regulatory minimum capital is 8 percent of the risk-weighted assets. This gives us a required capital amount of $4. So, to summarize, our $100 mortgage should be financed by least $4 of equity to cover potential credit losses.

Although it was definitely a good step toward better risk management, the Basel I Accord faced some important drawbacks. First, the solvency of the debtor was not properly taken into account since the risk weights depended only on the exposure class and not on the obligor or product characteristics. There was insufficient recognition of collateral guarantees to mitigate credit risk. It also offered various opportunities for regulatory arbitrage by making optimal use of loopholes in the regulation to minimize capital. Finally, it considered only credit risk, not operational or market risk.

Basel II

To address the shortcomings of the Basel I Capital Accord, the Basel II Capital Accord was introduced. It consists of three key pillars: Pillar 1 covers the minimal capital requirement, Pillar 2 the supervisory review process, and Pillar 3 market discipline and disclosure. (See Exhibit 1.1.)

Under Pillar 1, three different types of risk are included. Credit risk is the risk faced when lending money to obligors. Operational risk is defined as the risk of direct or indirect loss resulting from inadequate or failed internal processes, people, and systems, or from external events. Popular examples here are fraud, damage to physical assets, and system failures. Market risk is the risk due to adverse market movements faced by a bank's market position via cash or derivative products. Popular examples here are equity risk, currency risk, commodity risk, and interest rate risk. In this

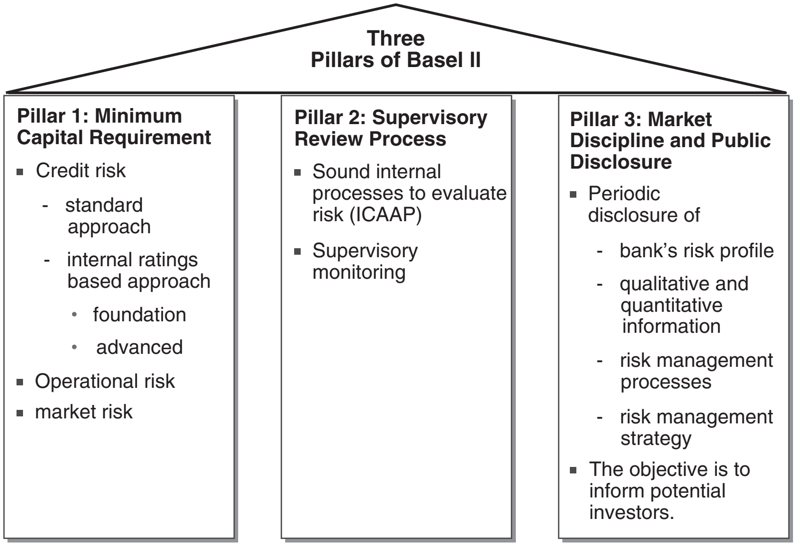

Exhibit 1.1 Pillars of the Basel II/III Regulation

book, we will closely look at credit risk. The Basel II Capital Accord foresees three ways to model credit risk: the standard approach, the foundation internal ratings based approach, and the advanced internal ratings based approach. All boil down to building quantitative models for measuring credit risk.

All quantitative models built under Pillar 1 need to be reviewed by overseeing supervisors. This is discussed in Pillar 2. Key activities to be undertaken are the introduction of sound processes to evaluate risk, such as the internal capital adequacy assessment process (ICAAP) and supervisory monitoring.

Finally, once all quantitative risk models have been approved, they can be disclosed to the market. This is covered by Pillar 3. Here, a bank will periodically disclose its risk profile, and provide qualitative and quantitative information about its risk management processes and strategies to the market. The objective is to inform the investors and convince them that the bank has a sound and solid risk management strategy, which it hopes will result in a favorable rating, in order for the bank to attract funds at lower rates.

Basel III

The Basel III Capital Accord was introduced as a direct result of the GFC. It builds upon the Basel II Accord, but aims to further strengthen global capital standards. Its key attention point is a closer focus on tangible equity capital since this is the component with the greatest loss-absorbing capacity. It reduces the reliance on models developed internally by the bank and ratings obtained from external rating agencies. It also places a greater emphasis on stress testing. (See Exhibit 1.2.)

For important banks, it stresses the need to have a loss-absorbing capacity beyond common standards. It puts a greater focus on Tier 1 capital consisting of shares and retained earnings by abolishing the Tier 3 capital introduced in Basel II, as it was deemed of insufficient quality to absorb losses. A key novelty is that it introduces a risk-insensitive leverage ratio as a backstop to address model risk. It also includes some facilities to deal with procyclicality, whereby due to a too cyclical nature of capital, economic downturns are further amplified. The Basel III Accord also introduces

	Basel II	**Basel III**
Common Tier 1 capital ratio (common equity = shareholders' equity + retained earnings)	2% * RWA	4.5% * RWA
Tier 1 capital ratio	4% * RWA	6% * RWA
Tier 2 capital ratio	4% * RWA	2% * RWA
Capital conservation buffer (common equity)	—	2.5% * RWA
Countercyclical buffer	—	0%–2.5% * RWA

Note: RWA = risk-weighted assets.

Exhibit 1.2 Basel III: Capital Requirements a liquidity coverage and net stable funding ratio to satisfy liquidity requirements. We will not discuss those further, as our focus is largely on credit risk. The new Basel III standards took effect on January 1, 2013, and for the most part will become fully effective by January 2019. Compared to the Basel II guidelines, the Basel III Accord has no major impact on the credit risk models themselves. It does, however, introduce additional capital buffers, as we will discuss in what follows.

The Tier 1 capital ratio was 4 percent of the risk-weighted assets (RWA) in the Basel II Capital Accord. It was increased to 6 percent in Basel III. The common Tier 1 capital ratio whereby common Tier 1 capital consists of common equity, which is common stock and retained earnings, but no preferred stock, was 2 percent of the risk-weighted assets in Basel II and is 4.5 percent of the risk-weighted assets in Basel III. A new capital conservation buffer is introduced that is set to 2.5 percent of the risk-weighted assets to be covered by common equity. Also, a countercyclical capital buffer is added, ranging between 0 and 2.5 percent of the risk-weighted assets.

As already mentioned, a non-risk-based leverage ratio is introduced that should be at least 3 percent of the assets and covered by Tier 1 capital. Very important to note here is that we look at the assets and not risk-weighted assets, as with the previous ratios. The assets also include off-balance-sheet exposures and derivatives. The idea here is to add this ratio as a supplementary safety measure on top of the risk-based ratios.

Basel III includes (relative to Basel II) the capital conservation buffer, the countercyclical capital buffer, and, if relevant, an additional capital ratio for systemically important banks.

Basel Approaches to Credit Risk Modeling

In what follows, we will discuss how credit risk can be modeled according to the Basel II and III Capital Accords. Basically, there are three approaches available, as already discussed: the standardized approach, the foundation internal ratings based approach, and the advanced internal ratings based approach. The approaches differ in terms of their sophistication and level of flexibility related to using internally estimated risk numbers.

Standardized Approach

Let us first discuss the standardized approach. For nonretail exposures, this approach relies on external credit assessment institutions (ECAIs) to provide credit ratings. Popular examples of ECAIs are Moody's, Standard & Poor's, and Fitch. Given the crucial impact of these ECAIs, the Basel Accords have introduced eligibility criteria such as objectivity, independence, transparency, and disclosure that need to be fulfilled in order to be officially recognized as an ECAI. The ratings provided by the ECAIs will then be mapped to risk weights provided in the accords. Risk weights are provided for sovereigns, banks, corporates, and other exposures. The capital itself is then calculated as 8 percent of the risk-weighted assets.

For retail, the risk weight is 75 percent for nonmortgage exposures and 35 percent for mortgage exposures. Remember, in Basel I the risk weight for mortgages was higher at 50 percent. For corporates, the risk weights vary from 20 percent for AAA-rated exposures to 150 percent for exposures rated B or lower. For sovereigns, the risk weights vary from 0 percent for AAA-rated countries to 150 percent for countries rated B or lower. For loans already in default, the risk weight can go up to 150 percent. Note that the European Banking Authority (EBA) has introduced mapping schemes to transform an ECAI's credit ratings to credit quality steps, which can then be further mapped to risk weights using the European capital directive. Let us illustrate this with an example.

Assume we have a corporate exposure of $1 million. It is unsecured with a maturity of five years, and Standard & Poor's assigns an AA rating to it. Using the European directive, an AA rating corresponds to a credit quality step of 1, which, according to Article 122, will map to a risk weight of 20 percent. The risk-weighted assets thus become 20 percent out of $1 million, or $0.2 million. The regulatory minimum capital can then be calculated as 8 percent thereof or thus $0.016 million. The standardized approach also provides facilities for credit risk mitigation in case of collateralized loans.

Although the standardized approach looks simple and appealing at first sight, it suffers from inconsistencies between ratings of different ECAIs with the accompanying danger of banks' cherry-picking the ECAIs. It also has problems in terms of coverage of various types of exposures. For example, retail exposures are discriminated only in terms of mortgage or nonmortgage. A more detailed categorization is highly desirable. Ideally, every obligor should have his or her own risk profile, whereby not only default risk is considered, but also loss and exposure risk as measured by LGD and EAD.

Internal Ratings Based (IRB) Approach

The internal ratings based (IRB) approach is a more sophisticated approach for quantifying credit risk. It relies on four key risk parameters, which we will introduce first. The PD is the probability of default of an obligor over a one-year period. It is expressed as a decimal and when converted to percentage ranges between 0 and 100 percent. The EAD is the exposure at default and is the amount outstanding. It is measured in currency terms. The LGD is the loss given default or the ratio of the loss on an exposure due to default of an obligor on the amount outstanding (EAD). It is also expressed as a decimal and ranges between 0 and 100 percent.

The PD, LGD, and EAD parameters can now be used to calculate the expected loss (EL) which becomes PD * LGD * EAD. Suppose the EAD is $10,000 and the LGD equals 20 percent. This means that upon default 20 percent out of $10,000 will be lost (= $2,000). The probability of losing this amount equals the probability of default let's say 1 percent. The expected loss then becomes $20. These risk parameters are used in the IRB approach to quantify credit risk.

Basically, there are two subapproaches of the IRB approach, the foundation IRB approach and the advanced IRB approach. In the foundation IRB approach, the PD is estimated internally by the bank, while the LGD and EAD are either prescribed in

	PD	LGD	EAD	Asset Correlations
Foundation approach	Internal estimate	Regulator's estimate	Regulator's estimate	Regulator's estimate
Advanced approach	Internal estimate	Internal estimate	Internal estimate	Regulator's estimate

Exhibit 1.3 Basel Foundation and Advanced IRB Approach

the Basel Accord or provided by the local regulator. In the advanced IRB approach, all three risk parameters, PD, LGD, and EAD, can be estimated internally by the bank itself. Furthermore, regulators provide asset correlations, which measure the degree to which the asset values underlying the credit exposures are correlated. In this setting, the asset correlations are either constant or a monotone function that is declining with increasing PDs. (See Exhibit 1.3.)

A distinction is made between the following types of exposure classes: corporate, retail, central governments (sovereigns) and central banks, institutions, equity exposures, securitization positions, and other non-credit-obligation assets. The foundation IRB approach is typically not permitted for retail exposures. Hence, for retail exposures, you can choose either the standard or the advanced IRB approach. Once the PD, LGD, and EAD are known, risk weight functions provided in the Basel Accord or directive can be used to calculate the regulatory capital.

We will describe this process in more detail in Chapter 14 on stress testing, but, in essence, capital is set to equal unexpected losses (ULs). Unexpected losses are the difference between the credit value at risk (VaR) and the expected losses (ELs). The reason for this is embedded in the accounting regime for credit risk. Expected losses are provisioned for, and provisions are losses in the profit and loss (P&L) statement and hence already netted with the equity account. As a result, the capital of a bank should cover losses that exceed provisions, and these losses are called unexpected losses. The credit value at risk (VaR) is computed in a similar way as expected losses, with the distinction that PD, LGD, and EAD are stressed to reflect an economic downturn:

- PD is stressed via the concept of a worst-case default rate given a virtual macroeconomic shock based on a confidence level of 99.9 percent and a sensitivity to the macroeconomy that is based on the asset correlation.
- LGD is based on an economic downturn.
- EAD is based on an economic downturn.

We will provide more specific details of the Basel regulations in the next chapters. These chapters include modeling default probabilities, loss given default, exposure at default, and validation as well as stress testing.

INTRODUCTION TO OUR DATA SETS

We have made four data sets available for student use via the book's companion website (www.creditriskanalytics.net). Exhibit 1.4 shows the four data sets and their applications.

Data Set HMEQ

The data set HMEQ reports characteristics and delinquency information for 5,960 home equity loans. A home equity loan is a loan where the obligor uses the equity of his or her home as the underlying collateral. The data set has the following characteristics:

- BAD: 1 = applicant defaulted on loan or seriously delinquent; 0 = applicant paid loan
- LOAN: Amount of the loan request
- MORTDUE: Amount due on existing mortgage
- VALUE: Value of current property
- REASON: DebtCon = debt consolidation; HomeImp = home improvement
- JOB: Occupational categories
- YOJ: Years at present job
- DEROG: Number of major derogatory reports

Data Set	Risk Segment	Chapter
HMEQ	Retail	5 Credit scoring
Mortgage	Retail	4 Exploratory data analysis 6 Probabilities of default (PD): discrete time hazard models 7 Probabilities of default: continuous time hazard models and practical implications 8 Low default portfolios 9 Default correlations and credit portfolio risk 11 Exposure at default (EAD) and adverse selection 12 Bayesian methods 13 Model validation 14 Stress testing
LGD	Corporate	10 Loss given default (LGD) and recovery rates
Ratings	Corporate	5 Credit scoring

Exhibit 1.4 Data Set Usage in This Book

- DELINQ: Number of delinquent credit lines
- CLAGE: Age of oldest credit line in months
- NINQ: Number of recent credit inquiries
- CLNO: Number of credit lines
- DEBTINC: Debt-to-income ratio

Data Set Mortgage

The data set mortgage is in panel form and reports origination and performance observations for 50,000 residential U.S. mortgage borrowers over 60 periods. The periods have been deidentified. As in the real world, loans may originate before the start of the observation period (this is an issue where loans are transferred between banks and investors as in securitization). The loan observations may thus be censored as the loans mature or borrowers refinance. The data set is a randomized selection of mortgage-loan-level data collected from the portfolios underlying U.S. residential mortgage-backed securities (RMBS) securitization portfolios and provided by International Financial Research (www.internationalfinancialresearch.org). Key variables include:

- id: Borrower ID
- time: Time stamp of observation
- orig_time: Time stamp for origination
- first_time: Time stamp for first observation
- mat_time: Time stamp for maturity
- balance_time: Outstanding balance at observation time
- LTV_time: Loan-to-value ratio at observation time, in %
- interest_rate_time: Interest rate at observation time, in %
- hpi_time: House price index at observation time, base year = 100
- gdp_time: Gross domestic product (GDP) growth at observation time, in %
- uer_time: Unemployment rate at observation time, in %
- REtype_CO_orig_time: Real estate type condominium = 1, otherwise = 0
- REtype_PU_orig_time: Real estate type planned urban development = 1, otherwise = 0
- REtype_SF_orig_time: Single-family home = 1, otherwise = 0
- investor_orig_time: Investor borrower = 1, otherwise = 0
- balance_orig_time: Outstanding balance at origination time
- FICO_orig_time: FICO score at origination time, in %
- LTV_orig_time: Loan-to-value ratio at origination time, in %
- Interest_Rate_orig_time: Interest rate at origination time, in %

- hpi_orig_time: House price index at origination time, base year = 100
- default_time: Default observation at observation time
- payoff_time: Payoff observation at observation time
- status_time: Default (1), payoff (2), and nondefault/nonpayoff (0) observation at observation time

Data Set LGD

The data set has been kindly provided by a European bank and has been slightly modified and anonymized. It includes 2,545 observations on loans and LGDs. Key variables are:

- LTV: Loan-to-value ratio, in %
- Recovery_rate: Recovery rate, in %
- lgd_time: Loss rate given default (LGD), in %
- y_logistic: Logistic transformation of the LGD
- lnrr: Natural logarithm of the recovery rate
- Y_probit: Probit transformation of the LGD
- purpose1: Indicator variable for the purpose of the loan; 1 = renting purpose, 0 = other
- event: Indicator variable for a default or cure event; 1 = event, 0 = no event

Data Set Ratings

The ratings data set is an anonymized data set with corporate ratings where the ratings have been numerically encoded (1 = AAA, and so on). It has the following variables:

- COMMEQTA: Common equity to total assets
- LLPLOANS: Loan loss provision to total loans
- COSTTOINCOME: Operating costs to operating income
- ROE: Return on equity
- LIQASSTA: Liquid assets to total assets
- SIZE: Natural logarithm of total assets

HOUSEKEEPING

We are planning to regularly update this book in the future and need your help. Please forward any feedback, errata, extensions, or topics that you would be interested in seeing covered in the next edition to us:

- Bart Baesens: bart.baesens@kuleuven.be
- Daniel Rösch: daniel.roesch@ur.de
- Harald Scheule: harald@scheule.com

Also check the book website for further details: www.creditriskanalytics.net.

Furthermore, we have generated a set of teaching slides that we are happy to share with university lecturers. Check the website or e-mail us if you are interested in obtaining the material.

We hope you have as much fun reading this book as we had writing it. Without further ado, let's get started and explore credit risk analytics.

Bart Baesens, Daniel Rösch, and Harry Scheule
September 2016

About the Authors

Bart Baesens

Bart Baesens is a professor at KU Leuven (Belgium) and a lecturer at the University of Southampton (United Kingdom). He has done extensive research on big data and analytics, credit risk modeling, customer relationship management, and fraud detection. His findings have been published in well-known international journals and presented at top-level international conferences. He is the author of various books, including *Analytics in a Big Data World* (see http://goo.gl/kggtJp) and *Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques* (see http://goo.gl/P1cYqe). He also offers e-learning courses on credit risk modeling (see http://goo.gl/cmC2So) and advanced analytics in a big data world (see https://goo.gl/2xA19U). His research is summarized at www.dataminingapps.com. He regularly tutors, advises, and provides consulting support to international firms with respect to their big data, analytics, and credit risk management strategy.

Daniel Rösch

Daniel Rösch is a Professor of Business and Management and holds the chair in Statistics and Risk Management at the University of Regensburg (Germany). Prior to joining the University of Regensburg in 2013, he was Professor of Finance and Director of the Institute of Banking and Finance at Leibniz University of Hannover from 2007 to 2013. He earned a PhD (Dr. rer. pol.) in 1998 for work on empirical asset pricing. From 2006 to 2011 he was visiting researcher at the University of Melbourne. Since 2011 he has been visiting professor at the University of Technology in Sydney. His research interests cover banking, quantitative financial risk management, credit risk, asset pricing, and empirical statistical and econometric methods and models. He has published numerous papers in leading international journals, earned several awards and honors, and regularly presents at major international conferences.

Rösch's service in the profession has included his roles as president of the German Finance Association, co-founder and member of the board of directors of the Hannover Center of Finance, and deputy managing director of the work group Finance and Financial Institutions of the Operations Research Society. He currently serves on the editorial board of the *Journal of Risk Model Validation*. Professor Rösch has worked with financial institutions and supervisory bodies such as Deutsche Bundesbank in joint research projects. Among others, his work has been funded by Deutsche Forschungsgemeinschaft, the Thyssen Krupp Foundation, the Frankfurt Institute for Finance and Regulation, the Melbourne Centre for Financial Studies, and the Australian Centre for International Finance and Regulation. In 2014 the German Handelsblatt ranked him among the top 10 percent of German-speaking researchers in business and management.

Harald Scheule

Harald "Harry" Scheule is Associate Professor of Finance at the University of Technology, Sydney, and a regional director of the Global Association of Risk Professionals. His expertise is in the areas of asset pricing, banking, credit and liquidity risk, home equity release, house prices in distress, insurance, mortgages, prudential regulation, securities evaluation, and structured finance

Scheule's award-winning research has been widely cited and published in leading journals. He currently serves on the editorial board of the *Journal of Risk Model Validation*. He is author or editor of various books.

Harry has worked with prudential regulators of financial institutions and undertaken consulting work for a wide range of financial institutions and service providers in Asia, Australia, Europe, and North America. These institutions have applied his work to improve their risk management practices, comply with regulations, and transfer financial risks.