Management for Professionals

Colin Chen

Practical Credit Risk and Capital Modeling, and Validation

CECL, Basel Capital, CCAR, and Credit Scoring with Examples

Springer

Management for Professionals

The Springer series "Management for Professionals" comprises high-level business and management books for executives, MBA students, and practice-oriented business researchers. The topics cover all themes relevant to businesses and the business ecosystem. The authors are experienced business professionals and renowned professors who combine scientific backgrounds, best practices, and entrepreneurial vision to provide powerful insights into achieving business excellence.

The Series is SCOPUS-indexed.

Colin Chen

Practical Credit Risk and Capital Modeling, and Validation

CECL, Basel Capital, CCAR, and Credit Scoring with Examples

Springer

Colin Chen
Data Science and Analytics Consultants
Bayside, NY, USA

ISSN 2192-8096 ISSN 2192-810X (electronic)
Management for Professionals
ISBN 978-3-031-52541-4 ISBN 978-3-031-52542-1 (eBook)
https://doi.org/10.1007/978-3-031-52542-1

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

"A book for all those who would like to know how financial institutions handle credit risks"

– Kathryn Zhao,
Global Head of Electronic Trading,
Cantor Fitzgerald

Preface

Credit risk assessment has been evolving quickly since the 2008 financial crisis both in accounting and capital regulation. On accounting, the incurred losses paradigm adopted by both the International Accounting Standards Board (IASB) and the Financial Accounting Standards Board (FASB) was one of the main critics for the failure of credit risk control in the financial crisis. A postcrisis review introduced the Current Expected Credit Loss (CECL) by FASB and the Internal Financial Report Standard 9 (or IFRS9) by IASB, both more focusing on the projected lifetime losses. On capital regulation, the US regulators introduced the Dodd-Frank stress testing and the annual Comprehensive Capital Analysis and Review (CCAR) procedures by the Dodd-Frank Wall Street Reform and Consumer Protection Act 2010 (Dodd-Frank Act, Pub. L. 111-203). Globally, Basel Committee on Banking Supervision (BCBS) proposed the Basel III accord (BIS, 2011). These regulations enhanced capital requirements both quantitatively and qualitatively.

Over the past few years, most banks, particularly large banks, have adopted these rules, and regulations and best practices start to converge. This book provides a comprehensive guide on credit risk modeling and validation for CECL/IFRS9, Basel capital, and CCAR. To complete the credit risk topic, the book also includes a chapter on credit underwiring and scoring, which addresses the credit risk at the front door of the lending business. The book is aimed at graduate students and practitioners. Examples are provided with data and programming protypes for learning practice by executing these examples with a specific programming language like Python, R, or SAS.

The three regulatory risk management frameworks have their designated purposes: the Allowance for Credit Loss (ACL) in accounting focuses on the lifetime Expected Credit Losses (ECL), the Capital assessment focuses on both ECL and Unexpected Credit Losses (UCL) in one year window, and the CCAR assessment focuses on the projected 9-quarter losses under different economic scenarios, especially the Severely Adverse (SA) scenario. Although they have different purposes, these assessments share similar data inputs and underlying loss forecasting models with different suitability. So, we first present the various data inputs and corresponding processes commonly for the three risk management frameworks, then we describe various common models used in credit risk assessments. Model suitability and model validation are discussed in each of the individual risk management frameworks. From a practical modeling perspective, such discussions would help modelers select the proper models to build from the beginning and improve model approval rate, as well as create foundational modeling tools to share among different applications.

The book's organization follows this logic. Chapter 1 presents a general review of the three risk assessment frameworks and the recent developments with each of them. We focus on their developing backgrounds and current usages in banks. First, we describe the newly developed ACL accounting rules, CECL for generally accepted accounting principles (GAAP) by FASB (2016) in the USA, and corresponding international ACL accounting rules IFRS9 by IASB (2014). While there are some differences between these rules, the major change as a response to the Financial Crisis is the allowance requirement with lifetime expected credit losses. So, we focus on the CECL modeling process. Then we introduce the capital management framework. We introduce the regulatory (equity) capital components for Basel IV, risk weighted asset (RWA), and the regulatory capital ratio and leverage ratio calculation. For RWA, we focus on the advanced approach, which bases on risk measures from models. We also introduce economic capital (EC), including its definition, relationship to RWA, and the simulation procedure commonly used for economic capital calculation, as well as some EC allocation approaches. Lastly, we introduce the stress test frameworks by focusing on the CCAR process due to the Dodd-Frank Act. We describe the two main components of the CCAR process, the stress test scenarios, and the stress test methodologies. In this chapter, we also introduce some different modeling traditions with these risk assessment frameworks. In the meantime, the three risk assessment frameworks have a dependence relation, for example, ACL affects the Basel regulatory capital calculation and they both affect the capital ratio calculation in CCAR. Such dependence will have impacts on new method adoption and implementation. We summarize some new developments driven by their interactions. In the last part of this chapter, we introduce credit underwiring and scoring. We briefly introduce credit underwriting processes with a focus on the model-based credit underwriting process, and then introduce a cyclical framework designed for the credit score modeling process.

We start from credit data and processing in Chap. 2. Given the structure difference, we classify credit transaction data as account level or cohort level. While most retail portfolios collect data in the more granular account level, wholesale portfolios more likely only have cohort level data available. Standard format and initial processing of these data are provided as examples. Segmentation and sampling are the two major data processing steps and have significant impact on model development and implementation, as well as model validation. They are required not only efficiently executable, but also theoretically sound. Statistical rigorousness should be embedded in final decision of these steps to properly defend the model. We describe some common segmentation methods. For sampling, besides the snapshot sampling, we introduce the Full Observation Stratified Sampling (FOSS) method for large panel data. We also describe the Synthetic Minority Over Sampling Technique

(SMOTE) popular in data science and point out potential issues if not used correctly. For microeconomic variables (MEVs), we focus on their classification and transformation. Finally, data integration and processing automation are discussed for efficiency.

Chapter 3 is the main source of various models used in this book. Aligning with the credit data processing, models are classified as account level models and cohort level models. Account level models are based on account level credit risk measures, Probability of Default (PD), Probability of Prepayment (PP), Loss Given Default (LGD), and Exposure at Default (EAD). For each account, the periodic losses over the observation unit (e.g., monthly or quarterly) are estimated as the combination of these measures over a specified projection window (12 months for Capital and 9 quarters for CCAR) or for the rest life of the account (CECL). These losses or their present values are summed up as the projected loss of the account. With each of these risk measures (PD, PP, LGD, EAD), various statistical and machine learning models are presented in this chapter. Especially for PD and PP models, besides they are modeled separately under the target event framework, we present two modeling frameworks, the competing risk framework and the multistate transition framework. For the competing risk framework, we demonstrate how binary generalized linear models can be used as unbiased approximation to the computationally expensive multinomial generalized liner models commonly used for competing risk modeling. For LGD models, we present a two-stage model with generalized linear models for the post-default resolution (gross loss or fully paid) probabilities and linear models for severities. For cohort level models, default rates, prepayment rate, or integrated loss rate are modeled as linear (or log-linear) and more often non-linear models linked to MEVs. For both account and cohort level models, variable selection is a major challenge, especially with model validation. We introduce the Adaptive and Exhaustive Variable Selection (AEVS) procedure as a general solution. It can be used for both linear (log-linear) and generalized liner models. For all these models, including some machines learning models as challengers, their advantages, disadvantages, and more importantly suitability and validation are discussed in each of the three credit risk management diagrams in the following chapters.

Chapter 4 focuses on the application of allowance for credit losses (ACL). After a brief introduction of the ACL process and platform, we describe some specific model data and processing for the CECL model application as example, including the design of a “reasonable” and “supportable” economic forecast. Development details on each component CECL models follow, especially the AEVS model section procedures for the PD model. These AEVS model selection procedures will also be used in the following chapters for other modeling areas. The multiple resolutions model is illustrated through the implementation for CECL LGD. Integration of these component models and loss forecasting results are presented. In this chapter, we also present a full framework for model validation and performance monitoring, including the main components commonly required in the framework and how to apply them in the CECL model validation procedure. Finally, we present some ACL methods with cohort level data as suggested by FASB (2016).

Chapter 5 focuses on the application of capital modeling. First, we introduce the regulatory capital and economic capital modeling processes. Then we focus on the capital model data processes. Since IRB is based on portfolio accounts, account level data are required. We describe the account level data, segmentation, sampling commonly used in capital modeling. We also describe some specific data requirements and processing for capital models, especially the economic cycles, downturn identification, and the through-the-cycle (TTC) concepts. We describe in detail how we create the reference data sets for PD, LGD models, since the risk measures based on these reference data sets are the inputs to the final RWA calculation under IRB. Model development details for the component models are presented using the mortgage reference data as example. Model integration and RWA calculation are illustrated through the internal risk rating (IRR) procedure based on TTC risk measures from the developed components models and the risk grid development process. Risk grid migration and monitoring is discussed. For economic capital, we present the details of the economic capital simulation and calculation, as well as some EC allocation methods. For capital models, we also present a full model validation and performance monitoring procedure, by applying a similar general procedure present in Chap. 4, but focusing on the sensitivity analysis on RWA and EC with the Vasicek distribution tail and MC-simulated loss distribution tail.

Chapter 6 is designated to the stress testing and CCAR applications. First, we introduce three popular stress test frameworks, regulatory stress test, systematic stress test, and reverse stress test. Given the recent trend of more popular uses of stress test results in risk management programs, we present some popular applications under each of these stress test frameworks, such as Stress Loss Usage (SLU), Risk Appetite (RA), Internal Capital Adequacy Assessment Process (ICAAP), CCAR, top risk identification, concentration risk management, and business risk limits management. Then we start on the model application in these stress test frameworks, by focusing on the regulatory stress test and CCAR. As a comprehensive capital analysis, CCAR uses both the expected losses (through equity capital projection) and the unexpected losses (through RWA as described in Chap. 5) to calculate the capital ratios under different scenarios. For the equity capital projection, it uses projected incomes through the Pre-provisioning Net Revenue (PPNR) and expected credit losses (ECL). In this book, we focus on the expected credit losses. We present all component models and their applications for a full list of credit products, including model segmentation and model methodology choice. To show details of the model development, we start from data processing in stress test, especially the macroeconomic variables (MEV) data under different scenarios. Using the sample mortgage data, we develop all component models, including a two-stage LGD model. Similar model integration for loss forecasting as for CECL in Chap. 4 was demonstrated. We complete this chapter by demonstrating how model validation and performance monitoring can be done for stress test models.

In Chap. 7, we present models for credit scoring and underwriting. Credit score models are popularly used for credit underwriting. First, we present a brief introduction to financial underwriting and the credit underwriting business and how risk rating, which includes credit scoring, is used in the underwriting process as a critical tool. Then we introduce how credit modeling is used in risk rating by focusing on the recent developments for score models and scorecards in credit scoring, as well as the impacts from regulations of fair lending. The details on the credit scoring modeling process are presented step by step on model population, segmentation, target definition, reject inference, score driver selection, fair lending testing, and building standard scorecards. We apply the modeling process on a research data set from small business lending by showing how to execute each of the steps. We summarize some of new trends on the scorecards' usage. Although score models are subject to less regulatory frameworks, to properly manage the risk of models critically important to a firm's business, we recommend and present a framework for comprehensive model validation and review, as well as a sound governance process for score models.

Finally, given the fast development of risk analysis on Environmental, Social, and Governance (ESG) aspects and the growing focus on sustainable finance from both the public and regulators, certainly I should mention the impact of ESG factors on credit risk. However, the investigation of ESG factors in credit risk has just started and mostly focuses on the wholesale credit underwriting business leading by the CRAs (Credit Rating Agencies), while this book is more focusing on the consumer credit underwriting.

Bayside, NY, USA Colin Chen

Acknowledgments

As for this first book in my career, my greatest thank is to my family. I would like to thank my parents who not only gave me life but also their continued unconditional love. I would like to thank my brother and sisters who support and encourage me. Of course, I would like to thank my beloved wife for her support and encouragement on this project, especially during the hard time of Covid-19 pandemic.

I would like to thank my teacher and professors – Mr. Yi Shen; Professor Yongguang Zhang and Professor Guoying Li at Academia China; Professor Thomas Sellke, Professor Shanti Gupta (Deceased), Professor Mary Ellen Bock (Emerita), and Professor George McCabe (Emeritus) at Purdue University.

I would like to thank my colleagues at SAS Institute Inc., Fannie Mae, Wells Fargo, Bank of America, JP Morgan Chase, and Citibank. Some notable names are Dr. Robert Rodriguez, Mr. Christopher Bailey, Dr. Jun Wang, Dr. Zack Li, Dr. William Kahn, Dr. Mark An, Dr. Mark Feldman, Dr. Cristian DeRitis, Dr. Henry Buist, Dr. Leonard Mills, Dr. Peter Abken, Dr. Taosong He, Dr. Patrick Gillen, Mr. Ming Fan, Ms. Jingjing Wang, Ms. Kathryn Zhao, and Ms. Barbara Ross.

I would like to thank my coauthors in previous publications – Dr. David Chock, Professor Xuming He, and Professor Keming Yu.

I would like to thank the five reviewers for this book and Ms. Nitza Jones for supporting and endorsing this project as the Springer Editor. I am beholden to Krishnakumar Pandurangan, Jill Balzano, Ben Ingraham, and Suhani Jain for their patience and affectionate help all along the publication process.

Contents

About the Author

Colin Chen is currently the Director and Founder of Data Science and Analytics Consultants, focusing on data science projects from financial and media industries. He has over 15 year's experiences in financial risk management. He worked at JPMorgan Chase as an Executive Director leading the Operational Risk modeling group and at Bank of America as a Director of Model Risk Management. He also worked at Wells Fargo and Fannie Mae on credit and market risk models with mortgages, credits cards, auto loans, mortgage servicing rights, and hedging. Colin also worked for the SAS institute for 10 years as a senior software developer. He was the 2013 Chair of the Section of Statistical Programmers and Analysts under the American Statistics Association. Colin holds a PhD in Statistics and a Master in Computer Science, both from Purdue University, and published over 20 papers in professional journals.

1 Introduction to Credit Risk and Capital Management Frameworks

Credit risk is the possibility of a loss resulting from the failure by a borrower, or more generally an obligor, to repay a loan or meet contractual obligations. In banking and financial industries, it refers to the risk that a lender may not receive the owed principal and interest, which results in an interruption of cash flows and increased costs for collection.

Lending businesses are significant portion of bank businesses. Consumer and commercial loans are critical asset on a bank's book. Assessing and managing the credit risk of these assets are critical for both profitability and business valuation. Over the time, banks have built up complex credit risk management systems, which consist of quantitative processes. On the center of these quantitative processes are various credit risk models built to meet either internal requests of risk management or external regulations. Although the statistical or mathematical theories behind these credit risk models are commonly known, building a model in production risk management systems is an iterative process starting from data processing to model selection and estimation, model output analysis and testing, model validation, and ongoing performance assessment, which decides whether the iterative process restarts. The entire modeling process is more of a practical execution rather than an application of some theories. A successful execution of this process needs solid execution of all these components under the Model Risk Management and Governance (MRMG). In the meantime, different quantitative risk management frameworks have different focuses and are executed differently. So, modeling process for each risk management framework is evaluated independently.

This chapter introduces such practical modeling executions for three credit risk and capital management frameworks, ACL, capital, and CCAR. These three frameworks cover the two most important regulatory areas, accounting and capital management. The introduction focuses on their backgrounds, usages, coverages, and recent developments. Details of modeling executions for each framework are in Chaps. 4, 5, and 6. In addition to this regulatory credit risk management, business also requests risk measurements, e.g., the credit scorecards by underwriting. Scorecard modeling is covered in Chap. 7.

C. Chen, *Practical Credit Risk and Capital Modeling, and Validation*, Management for Professionals, https://doi.org/10.1007/978-3-031-52542-1_1

Key Abbreviations and Symbols

ACL	Allowance for credit losses
CCAR	Comprehensive Capital Analysis and Review
CCR	Counterparty credit risk
CE	Credit exposure
CECL	Current expected credit loss
CVA	Credit valuation adjustment
DR	Default rate
EAD	Exposure at default
EC	Economic capital
EE	Expected exposure
EL	Expected loss
FASB	Financial Accounting Standards Board
IFRS	International Financial Reporting Standards
IASB	International Accounting Standards Board
ICAAP	Internal Capital Adequacy Assessment Process
LGD	Loss given default
MRMG	Model Risk Management and Governance
PD	Probability of default
PFE	Potential future exposure
RA	Risk appetite
RC	Regulatory capital
RWA	Risk-weighted asset
SEC	Securities and Exchange Commission
TSA	The standardized approach
UPB	Unpaid principal balance
VaR	Value at risk

1.1 Quantitative Credit Risk and Capital Management

By the definition, credit risk is the possibility of a loss resulting from a credit failure event. To assess credit risk, one can estimate this possibility, which is defined as the probability of default (PD). PD can be easily used to rank accounts, and this is the idea of credit scorecard, if the main purpose of the assessment is just relatively discrimination. The shortage of PD as a quantitative risk assessment for credit risk is that it doesn't measure the severity of the risk and depends heavily on the definition of default. To enhance this shortage, two more credit risk measures are introduced, loss given default (LGD) and exposure at default (EAD). LGD measures the loss percentage as regarding to the EAD, which is either the unpaid principal balance (UPB) or total credit draw at default. The production of these three quantities (PD *

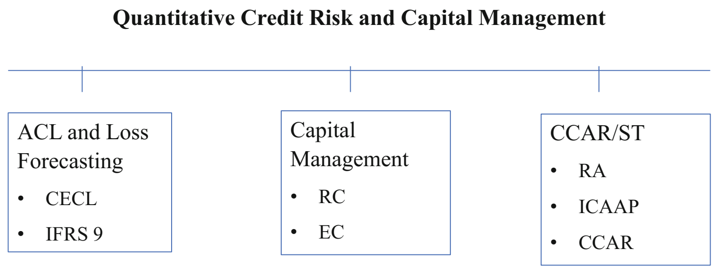


Fig. 1.1 Credit risk frameworks

LGD * EAD) is the expected loss (EL) for this account. For loss forecasting, models are built for PD, LGD, and EAD based on various risk factors.

The expected loss estimation through PD, LGD, and EAD is an account-level setup, which requires information related to each account. When account-level information is not available, we turn to the cohort or sub-portfolio aggregations. PD will be replaced with default rate (DR). DR can be estimated and predicted using various linear and nonlinear regression and time series models. If even the aggregated default information is not available, but the aggregated losses are available, with a proper scalar, the aggregated loss rates are computed. Loss rates can be estimated and predicted using regression and time series models. For loss forecasting, one can multiply the predicted loss rate with the predicted quantity of the scalar.

These loss forecasting methods are the core quantitative components of the three credit risk and capital management frameworks. For ACL with accounting, both account-level PD, LGD, and EAD models and cohort-level DR or loss rate are used depending on data availability. For Basel Capital, since the internal ratings-based (IRB) approach requires account-level information for account segmentation, account-level 1-year PD, LGD, and EAD are used. For CCAR, though one can use the account-level PD, LGD, and EAD to predict the nine-quarter losses, the loss rate is preferred due to its simplicity and the sensitivity requirement to macroeconomic factors. Figure 1.1 shows the three frameworks and their sub-applications. In the following sections, we introduce each of these credit risk and capital management frameworks.

1.2 ACL and Loss Forecasting

Credit risk has been managed through accounting rules. In the USA, the Securities and Exchange Commission (SEC) regulates that firms with listed securities report certain financial information, including the allowance for credit losses (ACL), quarterly (SEC 10-Q) and annually (SEC 10-K). For such financial information, SEC adopted accounting rules following the generally accepted accounting principles (GAAP) by FASB. The corresponding international accounting rules follow the International Financial Reporting Standards (IFRS) by IASB.

Financial assets (loans and contracts) subject to credit risk could be un-securitized (e.g., whole loans) or securitized (e.g., mortgage-backed securities or MBS). The un-securitized assets are in a financial institute's banking book, and their credit risks are managed through the allowance for loan and lease losses (ALLL) process. Securitized assets could be on either the financial institute's banking book or trading book. By FAS-115 (1993), they are classified into three segments:

- HTM – Debt securities that the enterprise has the positive intent and ability to hold to maturity are classified as held-to-maturity securities and reported at amortized cost.
- HFT – Debt securities that are bought and held principally for the purpose of selling them in the near term are classified as trading securities and reported at fair value, with unrealized gains and losses included in earnings.
- AFS – Debt securities not classified as either held-to-maturity securities or trading securities are classified as available-for-sale securities and reported at fair value, with unrealized gains and losses excluded from earnings and reported as other comprehensive income (OCI).

HTM securities are allocated in the banking book, while HFT securities are in the trading book. As for AFS securities, the Fundamental Review of the Trading Book (FRTB)[1] would have them allocated in the trading book, while the IFRS still prefers the separate category of fair value through OCI (FVTOCI). Under IFRS, accumulated OCI (or AOCI) is treated as other comprehensive equity (OCE) and can be reclassified as retained earnings (RE) when the asset is disposed. However, the new CECL rules in GAAP (FASB, 2016) would push credit losses from AFS debt securities with allowance aligning to HTM securities:

> 326-30-35-2 For individual debt securities classified as available-for-sale securities, an entity shall determine whether a decline in fair value below the amortized cost basis has resulted from a credit loss or other factors. An entity shall record impairment relating to credit losses through an allowance for credit losses. However, the allowance shall be limited by the amount that the fair value is less than the amortized cost basis. Impairment that has not been recorded through an allowance for credit losses shall be recorded through other comprehensive income, net of applicable taxes. An entity shall consider the guidance in paragraphs 326-30-35-6 and 326-30-55-1 through 55-4 when determining whether a credit loss exists.[2]

Expected credit loss assessment only applies to HTM and AFS/FVTOCI assets. For HFT, due to the short-term holding and main purpose of trading, ECL assessment is not needed.

---

[1] Fundamental review of the trading book: a revised market risk framework, BCBS 2013.

[2] Measurement of credit losses on financial instruments, an amendment of the FASB accounting standards codification, FASB ASU 2016-13, June, 2016 (referred to as "FASB, 2016" in this book).

For ACL in accounting, both IASB and FASB adopted the credit loss assessment approach using incurred losses before the financial crisis, though concerns have been expressed about delaying the recognition of credit losses and potential of balance sheet assets overstatement. The financial crisis highlighted this issue and forced a profound review of accounting standards. The reviews produced the International Financial Reporting Standard number 9 (IFRS9) (IASB, 2014)[3] and current expected credit loss (CECL) (FASB, 2016). IFRS9 went live in 2018 covering not only expected credit loss (ECL) rules, but also classification mechanics and hedge accounting. Hereafter, our focus is limited to ECLs. On the other hand, FASB's new credit impairment standard became effective for SEC filers for years beginning on or after December 15, 2019 (with early adoption permitted 1 year earlier), and 3 years later for other entities.[4]

The key change introduced by new accounting standards is a shift from a backward incurred losses perspective toward a forward-looking ECL representation. This change closes the methodology gap between ACL in GAAP accounting and the internal credit loss forecasting practices (e.g., ALLL). Making the forward-looking loss forecasting an accounting standard implies a deep review in terms of business interpretation, computational efficiency, and IT infrastructures, though some of these could be built upon existing loss forecast practices for some institutions. Furthermore, a convergence between finance and credit risk management both in methodology and view toward the forward-looking could be challenging. So, a holistic perspective is required in such a complex framework involving widespread competences to be aligned on a common goal.

While there are some differences between IFRS9 and CECL, the major change as a response to the financial crisis is the common allowance requirement with lifetime expected credit losses. So, we focus on CECL models while providing some comparisons between IFRS9 and CECL when significant differences exist.

1.2.1 CECL on Credit Risk

FASB ASU 2016-13 detailed the motivation of shifting from the incurred loss methodology to CECL for accounting on ACL underlined in the financial crisis:

> The global financial crisis underscored those concerns because users analyzed credit losses by utilizing forward-looking information to assess an entity's allowance for credit losses on the basis of their own expectations. Consequently, in the lead-up to the financial crisis, users were making estimates of expected credit losses and devaluing financial institutions before accounting losses were recognized, highlighting the different information needs of users from what was required by GAAP. Similarly, financial institutions expressed frustration during this period because they could not record credit losses that they were expecting but had not yet met the probable threshold.

[3] IASB, 2014. IFRS 9 Financial Instruments. July 2014. Technical report. International Accounting Standards Board.

[4] Due to COVID-19 pandemic, the original 1-year lag was extended to 3 years for other entities.

The objective of general-purpose financial reporting is to provide financial information about the reporting entity that is useful to existing and potential investors, lenders, and other creditors in making decisions about providing resources to the entity. In 2008, the Financial Accounting Standards Board (FASB) and the International Accounting Standards Board (IASB) established a Financial Crisis Advisory Group (FCAG) to advise the Boards on improvements to financial reporting in response to the financial crisis. The FCAG identified as a weakness in GAAP the delayed recognition of credit losses that results in the potential overstatements of assets. As a result, the FCAG recommended exploring more forward-looking alternatives to the incurred loss methodology. After several rounds of consultation starting from 2012, FASB issued the final GAAP update on ACL as described in FASB ASU 2016-13.

FASB ASU 2016-13 defines the CECL model as the entire updated GAAP accounting changes on ACL, not simply the concept of current expected credit losses. The main changes can be summarized in three areas:

CECL Changes for Assets Measured at Amortized Cost

- CECL requires estimate of full lifetime expected loss based on relevant information about past events, including historical experience, current conditions, and reasonable and supportable forecasts that affect the collectability of the reported amount.
- CECL eliminates the probable initial recognition threshold with the "incurred" loss method and, instead, reflects an entity's current estimate of all expected credit losses.
- CECL requires the disclosures of credit quality indicators further disaggregated by year of origination (or vintage), in addition to the disclosures of credit quality indicators in relation to the amortized cost of financing receivables, for better assessing changes in underwriting standards and credit quality trends in asset portfolios over time and the effect of those changes on credit losses.

CECL Changes for Available-for-Sale Debt Securities

- CECL requires an entity to determine whether a decline in fair value of an AFS security below its amortized cost basis has resulted from a credit loss or other factors and not use the length of time a security has been in an unrealized loss position to avoid recording a credit loss.
- CECL removes the requirements to consider the historical and implied volatility of the fair value of a security and recoveries or declines in fair value after the balance sheet date in determining whether a credit loss exists.
- CECL requires that full lifetime expected credit losses for an AFS security be estimated and presented as an allowance rather than as a write-down. The allowance enables an entity to record reversals of credit losses in current period as net income, which in turn should align the income statement recognition of credit losses with the reporting period in which changes occur.

- CECL limits the amount of the allowance for credit losses to the amount by which fair value is below amortized cost because the classification as available for sale is premised on an investment strategy that recognizes that the investment could be sold at fair value, if cash collection would result in the realization of an amount less than fair value.

CECL Changes for Purchased Credit Deteriorated (PCD) Assets

- CECL simplifies the loss allowance for PCD assets and requires similar treatment as other assets as at amortized cost or ASF.
- CECL requires the initial allowance for credit losses for PCD asset is added to the PCD purchase price rather than being reported as a credit loss expense. Only subsequent changes in the allowance for credit losses are recorded as a credit loss expense for these assets.
- CECL requires interest income for PCD assets should be recognized based on the effective interest rate, excluding the discount embedded in the purchase price that is attributable to the acquirer's assessment of credit losses at acquisition.

Among all these changes, the core is the forward-looking expected credit loss estimate. As always, under the non-prescriptive nature of the accounting principle, the Board does not specify a method for measuring expected credit losses and allows an entity to apply methods that reasonably reflect its expectations of the credit loss estimate. However, FASB ASU 2016-13 did provide several examples. Combining with the method we described in Sect. 1.1, the common CECL expected credit loss estimation methods are:

1. Loss rate
2. Vintage
3. Aging schedule
4. Discounted cash flow

Loss rate, aging schedule, and vintage methods use historical realized losses combining with some forward-looking adjustments. While these methods are simple and straightforward, the forward-looking adjustments are subjective and hard to justify. So, the discounted cash flow method based on PD, LGD, and EAD becomes the most popular method used in CECL.

The discounted cash flow needs economic forecast for the entire contractual life of the asset. While FASB requires such forecast should be based on current conditions, reasonable and supportable, it isn't expected that an entity will need to create an economic forecast over the entire contractual life of long-dated financial assets. Therefore, CECL allows an entity to revert to historical loss information that is reflective of the contractual term (considering the effect of prepayments) for periods that are beyond the time frame for which the entity is able to develop reasonable and supportable forecasts.

1.2.2 IFRS9 on Credit Risk

IFRS9 was developed in parallel with CECL by IASB addressing the same incurred losses issue identified by the joint Financial Crisis Advisory Group (FCAG) formed by IASB and FASB. FASB and IASB received different feedback on their respective proposed credit loss models. The IASB stakeholders strongly preferred an impairment model that utilizes a dual measurement approach, while US stakeholders strongly preferred the CECL lifetime expected loss measurement methodology proposed by FASB.

Some notable similarities and differences between the CECL model and IFRS9 are summarized below (abstracted from ASU 2016-13):

Similarities

(a) Both CECL and IFRS9 are considered to be expected credit loss models. The CECL model requires that the full amount of expected credit losses be recorded for all financial assets measured at amortized cost, whereas IFRS9 requires that an allowance for credit losses equal to the 12-month expected credit losses as defined in IFRS9 be recognized, until there is a significant increase in credit risk when lifetime expected credit losses are recognized.

(b) Under IFRS9, the full amount of expected credit losses is measured for financial assets that have experienced a significant increase in credit risk since initial recognition. For these assets, there may be similar measurements of expected credit losses under IFRS9 and CECL because, under both, an entity will measure credit losses over the expected life.

Differences

(a) Under CECL, there are different requirements based on the measurement attribute. Specifically, different considerations and indicators for impairment exist for available-for-sale debt securities. IFRS9 requires one credit loss approach for all financial assets (described as fair value through other comprehensive income assets under IFRS9), regardless of the measurement attribute.

(b) FASB acknowledges the time value of money is implicitly present in credit loss methodologies using amortized cost information, whereas IFRS9 requires an explicit consideration of the time value of money.

(c) The CECL model requires collective evaluation of credit losses when similar risk characteristics exist. IFRS9 states that the measurement of expected credit losses shall reflect a probability-weighted amount but particular measurement techniques are not prescribed. Therefore, IFRS9 allows collective evaluation of credit losses based on shared risk characteristics; however, unlike the CECL model, the probability-weighted outcomes must be considered.

(d) GAAP treats a concession provided to a troubled borrower to be a continuation of the original lending agreement. Differences exist for modifications of financial assets, and the concept of a troubled debt restructuring does not exist in IFRS9.

(e) Differences exist for purchased financial assets. IFRS9 also includes requirements for originated credit-impaired financial assets as well as purchased credit-impaired financial assets. GAAP does not contain provisions for originated impaired financial assets, and there are differences in the scope and measurement of expected credit losses for purchased financial assets.

(f) GAAP continues to permit the application of nonaccrual practices, whereas IFRS9 continues to preclude the use of nonaccrual practices. IFRS9 requires a net interest approach to be applied to the "Stage 3" assets, which represent individual assets that are credit impaired, whereas a gross interest approach is used otherwise.

(g) The discount rate utilized when a discounted cash flow approach is used under the CECL model is required to be the effective interest rate. IFRS9 provides that an entity also is permitted to use an approximation of the effective discount rate when discounting expected credit losses.

(h) The CECL model requires expected credit losses for unfunded commitments to reflect the full contractual period over which an entity is exposed to credit risk via a present obligation to extend credit. The CECL model does not require an allowance for expected credit losses beyond the contractual term or beyond the point in which a loan commitment may be unconditionally cancelled by the issuer. In contrast, for a financial asset that contains both a loan and an undrawn commitment component, IFRS9 states that an entity should measure expected credit losses over the period that an entity is exposed to credit risk and expected credit losses are not mitigated by credit risk management actions, even if that period extends beyond the maximum contractual period.

(i) The CECL model requires the amortized cost basis of financing receivables and net investment in leases to be disclosed by credit quality indicator, disaggregated by year of origination. This information is intended to help users understand the credit quality trends within the portfolio from period to period. IFRS9 requires an entity to disclose a reconciliation of the financial assets relating to the allowance for credit losses from the opening balance to the closing balance and requires explanations of how significant changes in the gross carrying amounts of financial assets during the period contributed to the changes in the allowance for credit losses.

1.2.3 Impact to Basel Capital and CCAR

CECL and IFRS9 impact credit loss allowance through provisions, which in turn impact an institution's actual capital and capital ratios under both Basel and CCAR. Sections 1.3 and 1.4 describe capital and capital ratios for Basel and CCAR, respectively. Figure 1.2 summarizes the relations between CECL/IFRS9 and Basel Capital and CCAR.

The CECL impact to Basel Capital is directly through allowance changes from the incurred loss methodology. In most cases, there is an increase in allowance when an institution switches to CECL from the incurred loss methodology. To relieve the impact of a sudden increase in regulatory capital requirement, regulators implemented a transition process with a 5-year transition period by allowing a

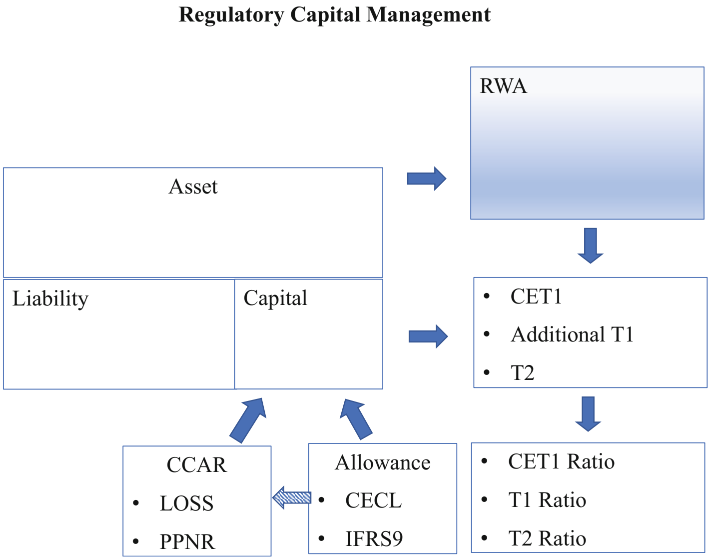


Fig. 1.2 Interactions among three risk frameworks

25% haircut of the after-tax provision for credit losses for the first 2 years of electing CECL. Then, starting from the third year, the CECL transition amount (defined as the difference between pre-CECL and post-CECL retained earnings at CECL adoption date) modified with the first 2-year transition,[5] which is fixed after the first 2 years of transition, is recognized in the retained earnings during the three remaining transition years by 75%, 50%, and 25%, respectively.[6] So, in year 6, the CECL transition is fully implemented.

To be consistent with the current regulatory capital rule for including the qualified allowance into Tier 2 capital calculation, the adjusted allowance for credit losses (AACL) is defined as those allowances that have been charged against earnings or retained earnings, which would include credit loss allowances related to financial assets, except allowances for PCD assets and AFS debt securities. AACL would be eligible for inclusion in an institution's Tier 2 capital subject to the current limit for including ALLL in Tier 2 capital under the capital rule. AACL transition amount and modified AACL transition amount are defined similarly as CECL transition.

[5]Defined as the CECL transition amount (also called the day-one impact of CECL) plus the difference between retained earnings reported in the most recent regulatory report and retained earnings as of the beginning of the fiscal year that the banking organization adopts CECL. It is the cumulative CECL transition amount after the first 2 years of transition.

[6]Regulatory Capital Rule: Revised Transition of the Current Expected Credit Losses Methodology for Allowances Federal Register Vol. 85, No. 190, September 30, 2020.

Table 1.1 CECL transitional amounts to apply to regulatory capital components during the final 3 years of the 2020 CECL transition

	Year 3	Year 4	Year 5
Increase retained earnings and average total consolidated assets by the following percentages of the modified CECL transitional amount	75%	50%	25%
Decrease temporary difference DTAs by the following percentages of the DTA transitional amount			
Decrease AACL by the following percentages of the modified AACL transitional amount			

Temporary difference on tax deferred asset (TDA) due to CECL adoption is defined as TDA transition amount and calculated similarly as CECL and AACL transition amounts.

Table 1.1 from the 2020 regulatory CECL transition rule summarizes the regulators' solution to the CECL impact on regulatory capital components.

Similarly, CECL will impact CCAR through CCAR components under both the baseline and stressed scenarios. CECL would impact CCAR through the allowance for credit losses, but not the pre-provision net revenue (PPNR).

The allowance will be projected under different CCAR scenarios. Here we have 2 sets of loss forecast models – CCAR that focuses on the 9–13 quarters of loss forecast under different scenarios and CECL that focuses on the lifetime of loss forecast under a reasonable and supportable scenario. The first question is should we reconcile the two sets of loss forecast models, since, in most cases, the two sets of models were developed differently. One way may be extending CCAR model ability to project lifetime losses. This will keep the CCAR modeling process integrity and its purpose for stress testing. However, given the CCAR purpose of forecasting relatively short-term losses, such model extension may not be feasible. The other direction is recalibrating the CECL model to adapt to both baseline and stressed scenarios beyond the reasonable and supportable scenario. This could lead to complicated model redeveloping.

Nevertheless, a "reasonable and supportable" economic forecast beyond the CCAR 9–13 quarters window would have to be created under different scenarios to adopt CECL. Due to all these considerations, the regulators delayed the incorporation of CECL into CCAR after 2022.

Although integrating CECL into CCAR is complex, however, based on the properties of these two loss forecasting processes, we can foresee some results of the integration:

- Integrating CECL in CCAR will generally lead to increases in capital requirements.
- CECL pulls forward loan loss provisions and leads to a greater decline in peak-to-trough regulatory capital ratios under stress.

- The front-loading of loan loss provisions would vary across portfolios and assumptions on scenario design and would have a greater impact on banks with a higher concentration of retail portfolios and portfolios with longer tenors.

For instance, the unemployment rate peak and the house price trough, which are key drivers of expected losses on retail portfolios, are typically reached later in the scenario. Therefore, the earlier recognition of losses under CECL has a larger impact on retail portfolios. In summary, incorporating CECL in CCAR will increase the stringency of stress tests, especially for banks that hold loans with longer maturities.

1.3 Capital Management

The idea behind capital management is to require institutions to hold enough capital to cover both expected and unexpected risks. For credit risks, since the allowance for credit losses in accounting has been created to cover the expected credit risks, capital is used to cover the unexpected risks. The time window of the loss measurement is limited to 1 year (12 months) instead of lifetime as required in CECL. The unexpected credit loss is defined as the difference between a least possible loss (or high quantile of the loss distribution) and the expected credit loss as shown in Fig. 1.3 with a typical credit loss distribution – large unexpected losses have low probabilities, while smaller losses tend to have high probabilities. It should be emphasized that the expected credit loss under the capital management framework is for a 1-year window measurement and noted as EL in Fig. 1.3 to distinguish with expected credit loss under CECL (noted as ECL).

So, we need to figure out the portfolio credit loss distribution in order to estimate the unexpected credit loss. There are different approaches to approximate the loss

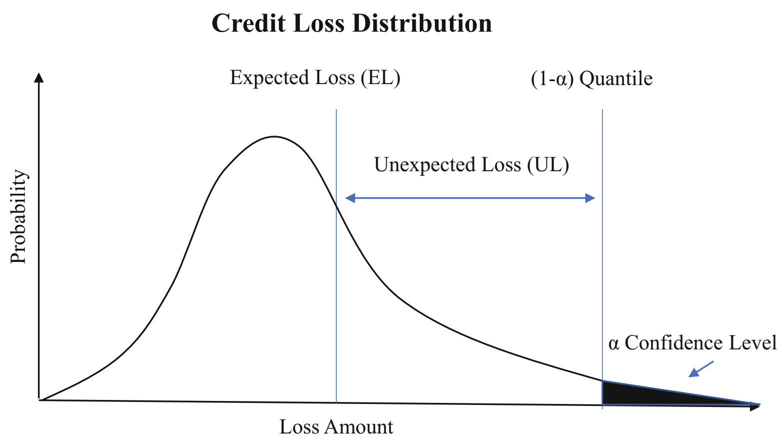


Fig. 1.3 Expected and unexpected losses. Note: Here we define *α confidence level* as the probability of loss amount exceeding the (1-*α*) *quantile*

distribution. The internal ratings-based (IRB) approach, which was introduced in Basel II, becomes the standard for regulatory purpose, and the economic capital approach is adopted for internal use by institutions.

1.3.1 Regulatory Capital

The regulatory (Basel) capital management framework adopts the risk-weighted asset (RWA) approach since Basel II, which assesses an institution's total asset with risks. These risks include credit risks, market risks, and operational risks, and an institution's total RWA is the sum of RWAs under each of these risks:

$$\mathrm{RWA}_{\mathrm{Total}} = \mathrm{RWA}_{\mathrm{Credit}} + \mathrm{RWA}_{\mathrm{Market}} + \mathrm{RWA}_{\mathrm{OpRisk}} \tag{1.1}$$

Basel II requires that an institution's regulatory capital from its own funding is no less than 8% of the total RWA:

$$\text{Regulatory Capital} \geq 8\% * \mathrm{RWA}_{\mathrm{Total}} \tag{1.2}$$

The logic using RWA instead of assets in regulatory capital requirement is taking into consideration the impact of risk. Assuming that a risky asset has a 100% weight, a $100.00 investment on such asset involves a minimum regulatory capital of $8.00. On the other hand, a risk-free asset having a 0% weight does not need any capital exceeding expected losses requirement, which has already been included within own funds by means of provisions.

Figure 1.2 in Sect. 1.2.3 helps in understanding the relationship between assets, liabilities, and regulatory capital. Assets are classified according to their risk profile in order to compute the RWA. The latter may be lower or greater than the total asset value. Regulatory capital components, on the other hand, are ranked in line with their capability to absorb losses, such as CET1, T1, and T2. Common equity Tier 1 capital is made up by instruments like common shares, retained earnings, and so on. Debt-like instruments with some degree of subordination are included as additional Tier 1 and Tier 2 components. It is worth noting that ECL estimates affect regulatory capital by means of provisions. RWA is computed under the capital management framework and is not affected by ECL in accounting. In Sect. 1.3.4, we described how CECL impacts the provisions and regulatory capital and a transition procedure introduced by the regulator to relieve the sudden change of regulatory capital for impacted institutions.

So, the capital ratio is calculated as:

$$\text{Capital Ratio} = \frac{\text{Regulatory Capital}}{\mathrm{RWA}_{\mathrm{Total}}} \tag{1.3}$$

With different levels of regulatory capital, different level of capital ratios can be computed. As a response to the financial crisis, capital requirements are enhanced by

Table 1.2 Basel Committee on Banking Supervision (BCBS) minimum capital requirements' roadmap. Dates referred to as January 1

	2016 (%)	2017 (%)	2018 (%)	2019 (%)	2020 (%)
Minimum common equity capital ratio	4.500	4.500	4.500	4.500	4.500
Capital conservation buffer	0.625	1.250	1.875	2.500	2.500
Minimum common equity plus capital conservation buffer	5.125	5.750	6.375	7.000	7.000
Minimum Tier 1 capital	6.000	6.000	6.000	6.000	6.000
Minimum total capital	8.000	8.000	8.000	8.000	8.000
Minimum total capital plus conservation buffer	8.625	9.250	9.875	10.500	10.500

gradual introduction of capital conservation buffer as shown in Table 1.2 for the transition periods.

In the following, we introduce more details on how RWA is computed using the IRB approach and how provisions based on accounting ECL impact the regulatory capital ratios.

There are different methods to compute RWA, and IRB becomes the most popular methods over the time. In general, the IRB approach converts unexpected credit loss estimates for asset portfolios into RWAs by multiplying the reverse of 8% (12.5) to these estimates. The unexpected credit loss estimates are obtained by an asymptotic single risk factor (ASRF) model (or Merton-like model) based on firm's asset value returns, which infers the probability of default (PD) distribution, combined with the loss given default (LGD) and exposure at default (EAD) models.

The IRB approach is based on the assumption that accounts with similar risk profiles follow the same loss distribution and an internal risk rating is used as a proxy of the risk profile. There are different internal risk ratings, but the most popular one is based on the through-the-cycle (TTC) probability of default (PD). Accounts with the same rating form a grid, and the average actual default rate in this grid is used as the default probability for all accounts in this grid.

Within each grid g, the normalized asset return ($y_{i,\,g,\,t}$) of i_{th} account over time t is driven by a single common factor X and an idiosyncratic noise component ($\epsilon_{i,\,g,\,t}$) as follows:

$$y_{i,g,t} = \sqrt{\rho_g} X + \sqrt{1 - \rho_g}\ \epsilon_{i,g,t} \tag{1.4}$$

where ρ_g is the correlation between asset returns and the common factor X and X and $\epsilon_{i,\,g,\,t}$ are independent and identically distributed $N(0, 1)$. Therefore, $y_{i,\,g,\,t}$ has a standardized normal distribution.

A default event happens when the return of the i_{th} account over time t passes a threshold as defined by the following binary variable:

$$D_{i,g,t}^{\text{def}} = \begin{cases} 1, & \text{for } y_{i,g,t} \le \Phi^{-1}\left(\text{PD}_{i,g,t}\right), \\ 0, & \text{Otherwise.} \end{cases} \tag{1.5}$$

where Φ^{-1} is the inversion of the standard normal cumulative distribution function (CDF) and $\text{PD}_{i,g,t}$ is the default probability threshold. The conditional default probability given a common factor scenario X_s can be computed as:

$$\begin{aligned} P\left(D_{i,g,t}^{\text{def}} = 1 | X = X_s\right) &= P\left(y_{i,g,t} \le \Phi^{-1}\left(\text{PD}_{i,g,t}\right) \mid X = X_s\right) \\ &= P\left(\sqrt{\rho_g} X + \sqrt{1-\rho_g}\, \epsilon_{i,g,t} \le \Phi^{-1}\left(\text{PD}_{i,g,t}\right) \mid X = X_s\right) \\ &= P\left(\epsilon_{i,g,t} \le \frac{\Phi^{-1}\left(\text{PD}_{i,g,t}\right) - \sqrt{\rho_g} X}{\sqrt{1-\rho_g}} \mid X = X_s\right) \\ &= \Phi\left(\frac{\Phi^{-1}\left(\text{PD}_{i,g,t}\right) - \sqrt{\rho_g} X_s}{\sqrt{1-\rho_g}}\right) \end{aligned} \tag{1.6}$$

The above Merton (1974)[7] single asset (or account) model was extended to a portfolio model by Vasicek (2002)[8] and Gordy (2003)[9]. The portfolio (or grid as used in IRB) is assumed infinitely granular, and each debtor in the portfolio is independent from each other. Under these assumptions and the conditional default probability in Eq. (1.6), the marginal probability of default (MPD) distribution function of the portfolio (grid) is a monotonic function of the risk factor $-X$. Given $-X$ also has a standard normal distribution, the quantiles of the marginal probability of default distribution of the portfolio (grid) correspond to the normal quantiles from $-X$. So, the ASRF assumption informs the IRB formula used to assess capital requirements (BIS, 2006)[10] as follows:

$$\text{MPD}_{i,g,t}(1-\alpha) = \Phi\left(\frac{\Phi^{-1}\left(\text{PD}_{i,g,t}\right) + \sqrt{\rho_g}\, \Phi^{-1}(1-\alpha)}{\sqrt{1-\rho_g}}\right) \tag{1.7}$$

[7] Merton, R., 1974. On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance 29*(2), 449–470.

[8] Vasicek, O., 2002. Loan portfolio value. *Risk 15*(2), 160–162.

[9] Gordy, M., 2003. A risk-factor foundation for risk-based capital rules. *Journal of Financial Intermediation 12*, 199–232.

[10] BIS, 2006. Basel II International Convergence of Capital Measurement and Capital Standards: A Revised Framework. BIS, Basel.

$$\mathrm{CL}_{1-\alpha} = \sum_{g=1}^{G} \sum_{i=1}^{N_g} \mathrm{EAD}_{i,g,t} * \mathrm{LGD}_{i,g,t} * \Phi\left(\frac{\Phi^{-1}\left(\mathrm{PD}_{i,g,t}\right) + \sqrt{\rho_g}\,\Phi^{-1}(1-\alpha)}{\sqrt{1-\rho_g}}\right) \tag{1.8}$$

where $1 - \alpha$ is the quantile level (α is the confidence level) and $\mathrm{MPD}_{i,g,t}(1-\alpha)$ is the $1-\alpha$ quantile of the marginal probability of default distribution of the portfolio (grid) and $\mathrm{CL}_{1-\alpha}$ is the credit loss quantile of the total portfolio that includes G grids and N_g account within grid g. $\mathrm{EAD}_{i,g,t}$ and $\mathrm{LGD}_{i,g,t}$ are exposure at default and loss given default of the i_{th} account.

If $\mathrm{PD}_{i,g,t}$ and $\mathrm{LGD}_{i,g,t}$ are replaced by the grid-specific probability of default and grid-specific loss given default, PD_g and LGD_g, and the grid total exposure at default is EAD_g for the 1-year window, assuming the portfolio has a single asset type, thus a common correlation parameter ρ, then the unexpected credit loss at 0.999th quantile (0.1% confidence level) can be computed by:

$$\mathrm{UL}_{0.999} = \sum_{g=1}^{G} \mathrm{EAD}_g * \mathrm{LGD}_g * \left(\Phi\left(\frac{\Phi^{-1}\left(\mathrm{PD}_g\right) + \sqrt{\rho}\,\Phi^{-1}(0.999)}{\sqrt{1-\rho}}\right) - \mathrm{PD}_g\right) \tag{1.9}$$

This is the most popular formula used in the regulatory RWA calculation with the IRB approach. Regulatory RWAs are computed by applying the reciprocal of 8% on Eq. (1.9) together with an additional 1.06 overall adjustment, as listed below:

$$\begin{aligned}\mathrm{RWA} = 12.5 * 1.06 * \sum_{g=1}^{G} \mathrm{EAD}_g * \mathrm{LGD}_g \\ * \left(\Phi\left(\frac{\Phi^{-1}\left(\mathrm{PD}_g\right) + \sqrt{\rho}\,\Phi^{-1}(0.999)}{\sqrt{1-\rho}}\right) - \mathrm{PD}_g\right) * \mathrm{adj}(M)\end{aligned} \tag{1.10}$$

where $\mathrm{adj}(M)$ represents an account-specific maturity adjustment. It is worth noting that Eq. (1.10) is used for the non-defaulted portfolio only. Additionally, a distinction holds between foundation and advanced IRB approaches. According to the foundation approach, a bank is authorized to use internally estimated PDs, whereas all other parameters are given. In contrast, an advanced IRB bank relies on internally estimated PDs, LGDs, and EADs.

In Eq. (1.9), the following part:

$$\mathrm{EL} = \sum_{g=1}^{G} \mathrm{EAD}_g * \mathrm{LGD}_g * \mathrm{PD}_g \tag{1.11}$$

is the expected loss under the capital management framework. It is compared to the loss provision to decide provision shortfall or excess provision (see Figs. 1.4 and 1.5). These quantities will impact the regulatory capital amount calculation, which will be detailed in the following.

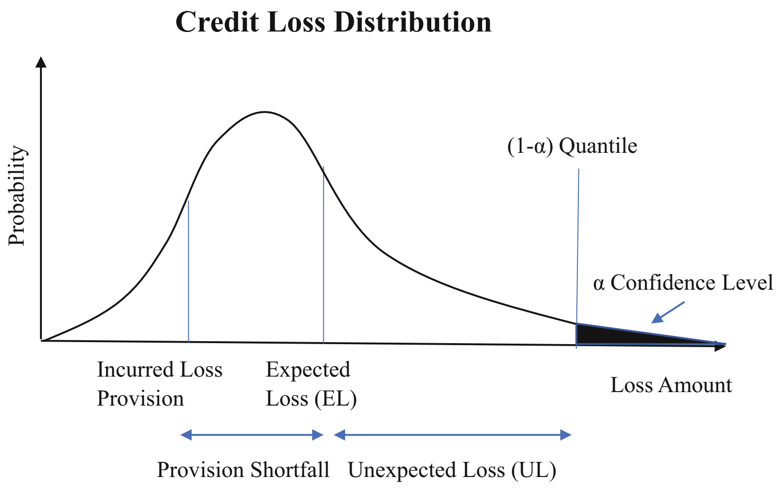


Fig. 1.4 Provision shortfall

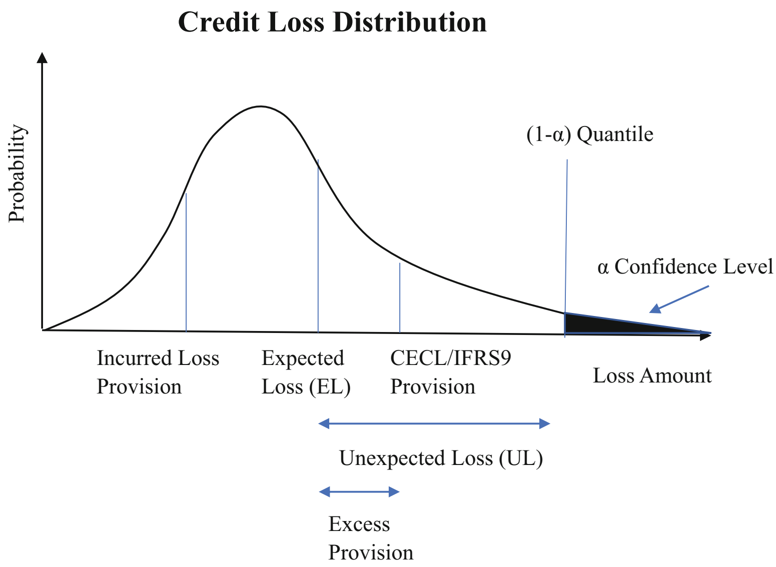


Fig. 1.5 Excess provision

Once the regulatory RWA is obtained, the next step is to calculate the regulatory capital amount. We need some background on regulatory capital identification.

Regulatory capital consists of the following components:

- *Tier 1 – common equity (CET1).* Common equity Tier 1 comprises an institution's core capital and includes primarily common shares and disclosed reserves (retained earnings). It also includes stock surpluses resulting from the issue of common shares, common shares issued by subsidiaries and held by third parties, and accumulated other comprehensive income (AOCI).

- *Tier 1 – additional capital (AT1).* This additional layer of the Tier 1 capital consists of instruments paying discretionary dividends, having neither a maturity date nor an incentive to redeem, for example, contingent convertible (CoCos) or hybrid security, which has a perpetual term and can be converted into equity when a trigger event occurs. An event that causes a security to be converted to equity occurs when CET1 capital falls below a certain threshold. Innovative hybrid capital instruments are phased out in Basel III because of their fixed distribution percentage, non-loss absorption capabilities, and incentive to redeem through features like step-up clauses. As a general rule, loss-absorbing preferred stocks qualify for inclusion in additional Tier 1 capital. However, clauses, covenants, and restrictions that make these shares more debit-like may cause them not to be acceptable for Tier 1 capital.
- *Tier 2 capital (T2).* Tier 2 consists of the bank's supplementary capital including undisclosed reserves, revaluation reserves, general and excess provisions, subordinated term debt, and hybrid capital instruments. Tier 2 capital is split into upper and lower levels. Upper-level Tier 2 capital consists of securities that are perpetual (no maturity date), for example, cumulative perpetual preferred stock, revaluation reserves, and fixed asset investments. Lower-level Tier 2 capital consists of subordinated debt with minimum original term of 5 years or more and is generally inexpensive for a bank to issue.

Following the 2008 financial crisis, a deep reform modified regulatory capital definitions and ratios and has been reflected in the transition from Basel II to Basel III. In Basel III, the definition of Tier 1 capital is moving toward the definition of tangible common equity. Tier 1 capital (referred to as going concern capital) must consist primarily of common equity and retained earnings and exclude goodwill and other intangible assets. Non-equity Tier 1 must be subordinated and have discretionary dividends/coupons with no incentive to redeem in times of stress. Hybrid capital instruments are limited in Tier 1. Innovative hybrid instruments, such as step-up instruments, cumulative preferred stock, and trust preferred stock, are excluded or grandfathered from Tier 1. Non-innovative hybrid instruments can be included in AT1 with conditions.

As mentioned above, regulatory capital is impacted by accounting provisions. The difference between the IRB expected loss (EL shown in Figs. 1.4 and 1.5) and accounting provision is called provision shortfall or provision excess if negative. Under Basel III, the provision shortfall is required to be deducted from CET1 capital; however, the provision excess is only allowed to be added back to Tier 2 capital subjected to a cap. Since the adoption of CECL is likely triggering provision excess (Fig. 1.5), there is a call for creating symmetry around the loss absorption capabilities of the allowance and capital within the CET1 capital.[11] Such asymmetry

[11] Capital and the allowance for credit losses – an *opportunity exists to level the playing field among international banks*, Deloitte, 2018.

Table 1.3 Example of provision impact on regulatory capital

	Eligible capital		Required capital
	Before CECL	CECL	
Accounting provisions for financial instruments	150	250	
Basel regulatory expected loss	(200)	(200)	
CET1 before provisions for financial instruments	1000	1000	
Provisions for financial instruments	(150)	(250)	
CET1 before regulatory adjustments due to provisions	850	750	
Regulatory adjustments due to provisions	(50)	0	
CET1	800	750	4.5% of RWA
Additional Tier 1	100	100	
Tier 1	900	850	6% of RWA
Tier 2 before regulatory adjustments due to provisions	100	100	
Regulatory adjustments due to provisions	0	50	
Tier 2	100	150	
Tier 1 + Tier 2	1000	1000	8% of RWA

treatment in regulatory capital has been existing with unrealized gains (Tier 2) and unrealized losses (Tier 1 through provision on allowance) for certain securities. Table 1.3 presents an example for the impact of regulatory capital by provisions.

In Table 1.3, we assume the accounting provisions before and after the adoption of CECL are \$150 and \$250, respectively, and the regulatory expected loss is \$200. So, the incurred loss method results in \$50 provision shortfall, and the CECL results in \$50 provision excess. We assume the CET1 is \$1000, additional T1 is \$100, and T2 is also \$100 (we will explain how these are computed from the balance sheet later). Provision shortfall will reduce the post-provision CET1 to \$800 for the incurred loss method, while the provision excess will directly reduce the post-provision CET1 to \$750 for CECL. Such impact will not be changed for final T1 by adding AT1 for the two methods, till the provision excess makes up to the T2. So, provision excess will mainly impact CET1 and T1.

In the following, we will introduce how CET1, T1, and T2 are accounted in Basel III. As the regulatory capital in Basel III has been shifted more on CET1, we add more details on CET1 requirements in the following.

CET1

Basel III has tightened the criteria governing instrument's inclusion in CET1. For an instrument to be included in CET1, the predominant form of Tier 1 capital, it must meet all of the following criteria:

1. It represents the most subordinated claim in liquidation of the bank.
2. It is entitled to a claim of the residual assets that is proportional to its share of issued capital, after all senior claims have been repaid in liquidation (i.e., it has an unlimited and variable claim, not a fixed or capped claim).
3. Principal is perpetual and never repaid outside of liquidation (setting aside discretionary repurchases or other means of effectively reducing capital in a discretionary manner that is allowable under national law).
4. The bank does nothing to create an expectation at issuance that the instrument will be bought back, redeemed, or cancelled, nor do the statutory or contractual terms provide any feature which might give rise to such an expectation. This criterion does not oppose banks being market makers in their own shares.
5. Distributions (i.e., dividends and coupons) are paid out of distributable items (retained earnings included). The level of distributions is not in any way tied or linked to the amount paid in at issuance, and they are not subject to a cap (except to the extent that a bank is unable to pay distributions that exceed the level of distributable items).
6. There are no circumstances under which the distributions are obligatory. Non-payment is therefore not an event of default.
7. Distributions are paid only after all legal and contractual obligations have been met and payments on more senior capital instruments have been made. This means that there are no preferential distributions, including in respect of other elements classified as the highest-quality issued capital.
8. It is the issued capital that takes the first and proportionately greatest share of any losses as they occur. Within the highest-quality capital, each instrument absorbs losses on a going concern basis proportionately with all the others.
9. The paid-in amount is recognized as equity capital (i.e., not recognized as a liability) for determining balance sheet insolvency.
10. The paid-in amount is classified as equity under the relevant accounting standards.
11. It is directly issued and paid up, and the bank cannot directly or indirectly have funded the purchase of the instrument.
12. The paid-in amount is neither secured nor covered by a guarantee of the issuer or related entity or subject to any other arrangement that legally or economically enhances the seniority of the claim. Related entities are the issuer's subsidiaries, the parent undertaking of the issuer or its subsidiaries, the parent financial holding company or its subsidiaries, the mixed activity holding company or its subsidiaries, the mixed financial holding company and its subsidiaries, or any undertaking that has close links with the previous entities.
13. It is only issued with the approval of the owners of the issuing bank, either given directly by the owners or, if permitted by applicable law, by the Board of Directors or by other persons duly authorized by the owners.
14. It is clearly and separately disclosed on the bank's balance sheet.

These criteria make CET1-qualified instruments fully equivalent to ordinary shares in terms of their capital quality as regards loss absorption and do not possess features that could cause the condition of the bank to be weakened as a going concern during periods of market stress. For example, when a bank issues non-voting ordinary shares, to be included in CET1, they must be identical to voting common shares of the issuing bank in all respects except the absence of voting rights.

The total qualified CET1 capital under Basel is the sum of the following components:

- Capital instruments (e.g., common stocks) and related share premium accounts
- Retained earnings, prior to the inclusion of any interim net profit or losses
- Independently reviewed interim profits, net of any foreseeable charge or dividend
- Accumulated other comprehensive income
- Other reserves
- Funds for general banking risk
- CET1-eligible minority interests
- *Less* regulatory adjustments (provision shortfall)

These elements (except the regulatory adjustments), making up the shareholders' equity of a bank balance sheet, represent capital invested in the bank by third-party investors and capital generated and not returned to these investors since its constitution. These elements can be grouped into the following four sections.

Capital Section

The capital section groups transactions with owners acting in their capacity as owners, comprising the following:

Ordinary shares (or common stock) represent the basic ownership interest in the reporting entity. It is the residual corporate interest that bears the ultimate risk of loss, as it is subordinate to all other instruments issued by the bank. A reporting entity may have more than one class of ordinary shares.

Share premium (or additional paid-in capital or APIC) is the excess amount paid by an investor over the par value of a stock issue. In addition, non-stock-related contributions from an investor, such as cash or property, are normally reflected in share premium.

Treasury shares (or own shares or treasury stock) are created when a reporting entity reacquires its own ordinary shares. If the treasury shares are not constructively or actually retired upon its reacquisition, the bank presents it on the shareholders' equity of the balance sheet as a reduction from capital.

Other equity instruments are instruments other than ordinary shares that from an accounting perspective are classified as equity instruments as opposed to liabilities. In a bank, AT1 instruments are included in this section of shareholders' equity.

Earnings Section

The earnings section groups the following items:

Retained earnings represent the earned capital of the bank. Earned capital is the capital that develops and builds up over time from profitable operations. It consists of all undistributed income that remains invested in the bank.

Profit or loss (or net income) represents the bank's net earnings generated during the reporting period. This statement mainly comprises net interest income after provisions for credit losses, net non-interest income, and taxes.

Dividends are distributions to shareholders and owners of other equity instruments (such as equity-classified AT1 instruments). Cash dividends declared are generally reported as a deduction from retained earnings until the retained earnings account is exhausted. In the absence of retained earnings, cash dividends should generally be charged to share premium. Stock dividends declared are generally shown as a deduction from retained earnings and added to common stock and share premium.

Reserves Section: Other Comprehensive Income (OCI)

Other comprehensive income (OCI) mainly represents items that are temporarily recorded in the bank's balance sheet. While a large proportion of the items in OCI will be subsequently reclassified to profit or loss, some items (e.g., remeasurement of gains/losses related to defined benefit pension plans) are reclassified to other sections of shareholders' equity.

Non-controlling Interests Section

Issuances of a subsidiary's ordinary shares and other equity instruments (e.g., AT1 instruments) to third parties are treated as non-controlling interests (i.e., minority interests). Non-controlling interests are shown in the consolidated balance sheet as a separate component of equity, which is distinct from the group's shareholders' equity.

As an example, Table 1.4 shows HSBC's fully loaded CET1 capital as of December 31, 2015.

Table 1.4 Example of HSBC's CET1 capital on December 31, 2015

Item amount	USD (mm)
Ordinary shares and related share premium	20,858
Retained earnings	143,976
Other comprehensive income (and other reserves)	−453
Eligible minority interests	3519
Independently reviewed interim net profits net of any foreseeable charge or dividend	−3717
Regulatory adjustments	−33,320
Total CET1 capital	130,863

T1

Tier 1 capital comprises CET1 and AT1 capital. AT1 is the sum of the following components:

- Eligible AT1 instruments (e.g., CoCos and qualified hybrid securities) and share premium (related to eligible AT1 instruments)
- Qualified AT1 minority interests
- *Less* direct and indirect holdings by the bank of own AT1 instruments
- *Less* reciprocal cross-holdings of AT1 instruments designed to inflate the AT1 funds artificially
- *Less* the applicable amount of direct, indirect, and synthetic holdings by the bank of the AT1 instruments of financial sector entities where the bank does not have a significant investment in those entities
- *Less* direct, indirect, and synthetic holdings by the bank of the AT1 instruments of financial sector entities where the bank has a significant investment in those entities
- *Less* qualifying Tier 2 deductions that exceed the Tier 2 capital of the bank
- *Less* foreseeable tax charges relating to AT1 items

T2

Tier 2 capital consists of the sum of the following elements:

- Capital instruments and subordinated loans issued by the bank that meet the criteria for inclusion in Tier 2 capital and are not included in Tier 1 capital (these instruments shall not qualify as CET1 or AT1 items).
- Share premium (stock surplus) resulting from the issue of instruments included in Tier 2 capital.
- Eligible Tier 2 minority interests.
- For banks calculating RWAs in accordance with the standardized approach to credit risk, general provisions can be included in Tier 2 capital subject to the limit of 1.25% of RWAs calculated in accordance with the standardized approach.
- For banks calculating RWAs in accordance with the internal ratings-based (IRB) approach to credit risk and where the total expected loss amount is less than the total eligible provisions, banks may recognize the difference in Tier 2 capital up to a maximum of 0.6% of credit RWAs calculated in accordance with the IRB approach.
- *Less* direct and indirect holdings by the bank of own Tier 2 instruments, including own Tier 2 instruments that the bank could be obliged to purchase as a result of contractual obligations.
- *Less* reciprocal cross-holdings of Tier 2 instruments of financial sector entities designed to inflate artificially the bank's Tier 2 funds.
- *Less* the applicable amount of direct, indirect, and synthetic holdings by the bank of the Tier 2 instruments of financial sector entities where the bank does not have a significant investment in those entities.

- *Less* direct, indirect, and synthetic holdings by the bank of the Tier 2 instruments of financial sector entities where the bank has a significant investment in those entities, excluding underwriting positions held for fewer than five working days.

The calculation of CET1, T1, and T2 regulatory capital and corresponding capital ratio could become complex if an institution has complex asset holding structures and accounting.

Leverage Ratio

For capital measures, besides the regulatory capital ratios described above, Basel III introduced the leverage ratio[12] (LR), which is defined as the ratio of capital over total leverage exposure. The LR, which does not distinguish between assets based on risk, is conceived as a backstop to risk-weighted capital requirements. The idea behind the leverage ratio is to assess excessive leverage exposures, which could be a significant risk itself in stress regardless of their risks as assets. The United States implemented the Basel III LR as the supplementary leverage ratio (SLR):

$$\mathrm{SLR} = \frac{\text{Regulatory Capital}}{\text{Total Leverage Exposure}} \tag{1.12}$$

Regulatory capital is defined as the Tier 1 capital. In contrast, the financial accounting balance sheet is used as a starting point to measure the leverage exposure. Specific provisions and valuation adjustments may be deducted from the exposure to which they relate. As a general principle, collateral, guarantees, and purchased credit risk mitigation may not be deducted from exposures. A bank's total leverage exposure measure is the sum of the items listed below[13]:

- *On-balance sheet exposures.* All balance sheet assets are included according to their accounting measurement. On-balance sheet derivatives, collateral, and covenants for securities finance transactions, different from those described below, are also included.
- *Derivative exposures.* For derivatives, two types of exposures are considered. On the one hand, exposure may arise from the instrument underlying the derivative. On the other hand, a counterpart credit risk exposure is taken into account. All in all, derivatives are measured with the use of the accounting exposure that reflects the fair value of the contract. An add-on for potential future exposure is also used to ensure a consistent conversion to a loan equivalent amount.

[12] BIS, 2014b. Revised Basel III leverage ratio framework and disclosure requirements. Bank for International Settlements, Basel.

[13] To ease strains in the Treasury market resulting from the coronavirus and increase banking organizations' ability to provide credit to households and businesses, on April 1, 2020, Fed temporally changed its supplementary leverage ratio rule. The change would exclude US Treasury securities and deposits at Federal Reserve Banks from the calculation of total leverage exposures for holding companies. This change expired on March 31, 2021.

- *Securities finance transactions.* Secured lending and borrowing is an important source of leverage. Repurchase agreements and securities finance are included by use of the accounting measure of exposure. Regulatory netting rules are applied.
- *Off-balance sheet items.* Off-balance sheet items including commitments, letters of credit, failed transactions, and unsettled securities are subject to a uniform 100% credit conversion factor. The only exception is that any commitments that are unconditionally cancelable by the bank at any time without prior notice may have a credit conversion factor of 10%.

LR was introduced by the Basel Committee in 2010 as a measure to control bank excessive leverages exposures, which demonstrated a significant risk during the financial crisis, especially with derivatives. It was finalized in January 2014 and the US regulators finalized the SLR in September 2014.

Basel III requires a minimum SLR of 3% for large financial institutions with more than $250 billion in total consolidated assets. In addition, banking organizations that are subject to Category I standards, which are the global systemically important bank holding companies (US GSIBs), as well as their depository institution subsidiaries, are subject to enhanced supplementary leverage ratio (eSLR) standards. The eSLR standards require each US GSIB to maintain a supplementary leverage ratio above 5% to avoid limitations on the firm's distributions. The US regulators also require each of insured depository institutions to maintain a supplementary leverage ratio of at least 6% to be deemed "well capitalized" under the prompt corrective action framework of each agency.

1.3.2 Economic Capital

The concept of economic capital differs from regulatory capital in the sense that regulatory capital is the mandatory capital the regulators require to be maintained, while economic capital is the best estimate of required capital that financial institutions use internally to manage their own risk and to allocate the cost of maintaining regulatory capital among different units within the organization.

As shown in Fig. 1.6, economic capital (EC) directly uses the unexpected credit loss of a portfolio of assets, which is the difference between the $(1-\alpha)$ quantile and the expected value of the credit loss distribution of the portfolio, as a measure of credit risk instead of transferring that into RWA to measure the credit risk of the portfolio and requiring minimum regulatory capital ratios as in the regulatory capital management. So, EC, as well as some other credit risk measures such as the expected shortfall (ES), more focuses on the loss distribution of a portfolio, especially loss distribution under the full picture of economic scenarios.

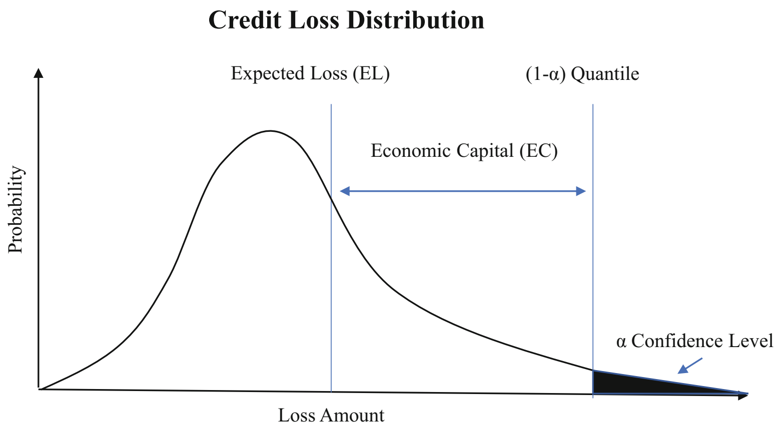


Fig. 1.6 Economic capital

So, let's come back to the portfolio loss distribution. Assume there are N accounts/assets within the portfolio and L_i, $i = 1, , , N$ are the random variables representing the loss from each account over a window of time (12 months) from the reporting date. The portfolio loss is the sum of all account losses:

$$L = \sum_{i=1}^{N} L_i \tag{1.13}$$

Let F_p be the loss distribution function of the portfolio; EC is defined as:

$$\mathrm{EC}(1-\alpha) = \mathrm{VaR}_{F_p}(1-\alpha) - \mathrm{EL} \tag{1.14}$$

where α is the confidence level such that loss has α probability exceeding the $(1 - \alpha)$ quantile or $(1 - \alpha)$ value at risk $\mathrm{VaR}_{F_p}(1-\alpha)$ of the loss distribution L or, equivalently, $(1 - \alpha)$ is the confidence level such that loss has $(1 - \alpha)$ probability not exceeding the $(1 - \alpha)$ value at risk $\mathrm{VaR}_{F_p}(1-\alpha)$ of the loss distribution F_p. EL is the expected portfolio loss under this loss distribution:

$$\mathrm{EL} = E\,(L) \tag{1.15}$$

Value at risk is a cutoff value used in the economic capital definition, and it does not provide information on what happens when loss exceeds this cutoff value. The expected shortfall is defined as the expected loss given that loss exceeds the cutoff value:

$$\begin{aligned}
\mathrm{ES}(1-\alpha) &= E\left(L \middle| L > \mathrm{VaR}_{F_p}(1-\alpha)\right) \\
&= \frac{\int_{\mathrm{VaR}_{F_p}(1-\alpha)}^{\infty} L f(L)\mathrm{dL}}{\int_{\mathrm{VaR}_{F_p}(1-\alpha)}^{\infty} f(L)\mathrm{dL}} \\
&= \frac{1}{\alpha}\int_{\mathrm{VaR}_{F_p}(1-\alpha)}^{\infty} L f(L)\mathrm{dL} \\
&= \frac{1}{\alpha}\int_{(1-\alpha)}^{1} \mathrm{VaR}_{F_p}(r)\mathrm{dr}
\end{aligned} \tag{1.16}$$

So, expected shortfall is some weighted average of those VaR values beyond $\mathrm{VaR}_{F_p}(1-\alpha)$. ES is a commentary risk measure to EC. It indicates the average loss given a default event, i.e., when the economic capital is not sufficient to absorb the losses. Expected shortfall takes a conditional average. As such, it is a more stable estimate than VaR measures. Therefore, ES is often preferred over VaR for capital allocation. Expected shortfall has subadditivity (ES of two portfolios combined not exceeding the sum of individual portfolio ESs), while EC does not.

Though ES is a better measure of the loss distribution tail compared to EC, it is less intuitive and more complex in computation, which limits its use in credit risk management. However, it is more popularly used in market risk management.

All these measures depend on the portfolio loss distribution F_p, which needs to be estimated from the account-level information. There are various models that can be used to estimate the portfolio loss distribution. Largely these models can be categorized into analytical models and simulation-based models.

Analytical models make simplifying assumptions on the loss distributions of the asset/account classes. Exposures are grouped into homogeneous asset/account classes on which the loss distributions are calculated and afterward aggregated to the full portfolio level. Given the assumptions made, the results are obtained from the analytical expressions allowing fast computation. The Vasicek one-factor model we described in Sect. 1.3.1 for regulatory capital management is the most popular analytical model used to estimate the portfolio loss distribution. A disadvantage of these models is the restrictive assumptions that have to be made in order to obtain closed-form solutions from analytical expressions.

Simulation-based models aim to approximate the true portfolio distribution by empirical distributions from a large number of Monte Carlo simulations. Because the portfolio losses are obtained from simulations, one does not have to rely upon the stringent assumptions in analytical models. The main disadvantages are the high computation time and the volatility of the results at high confidence levels.

Whether analytical or simulation-based models, to estimate the portfolio loss distribution, we need to start from the portfolio loss generating process. Given that the portfolio loss is the sum of individual account losses as in Eq. (1.13), we start from the individual account loss. Under the PD/LGD/EAD framework, the individual account loss can be decomposed as:

$$L_i = \mathrm{EAD}_i * \mathrm{LGD}_i * I\ (D_i = 1), i = 1 \ldots, N \tag{1.17}$$

where EAD_i and LGD_i are the exposure at default and loss given default and D_i is the binary default event for i_{th} account with probability $\mathrm{PD}_i = P(D_i = 1)$. This is a binary loss distribution model forming the base of the portfolio loss distribution. It follows the real credit loss generating process. Portfolio credit loss observed over a fixed period of time (12 months) consists of credit losses resulting from individual account default events, which lead to nonperforming accounts with cash flow disruptions.

As an example of analytical model, we can use normality approximation based on central limit theory.

If the individual losses L_i are assumed independent, which means that accounts are not related on the event of loss, following Eq. (1.13), the central limit theory guarantees that the portfolio loss distribution can be approximated by a normal distribution with mean EL and standard deviation:

$$\sigma_L = \left(\sum_{i=1}^{N} \sigma_{L_i}^2\right)^{1/2} = \left(\sum_{i=1}^{N} \left[\mathrm{EAD}_i * \mathrm{LGD}_i * \sqrt{\mathrm{PD}_i * (1 - \mathrm{PD}_i)}\right]^2\right)^{1/2} \tag{1.18}$$

when the number of accounts N in the portfolio is large enough.

The above result can be relaxed with correlation ρ_{ij} between the default events D_i and D_j:

$$\sigma_L = \left(\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{EAD}_i * \mathrm{LGD}_i * \mathrm{EAD}_j * \mathrm{LGD}_j * \rho_{ij} * \sqrt{\mathrm{PD}_i * (1 - \mathrm{PD}_i) * \mathrm{PD}_j * \left(1 - \mathrm{PD}_j\right)}\right)^{1/2} \tag{1.19}$$

Under the normality, EC can be approximated by applying a standard normal VaR multiplier on σ_L:

$$\mathrm{EC} = \mathrm{VaR}_{\Phi}(1 - \alpha) * \sigma_L - \mathrm{EL} \tag{1.20}$$

where $\mathrm{VaR}_{\Phi}(1 - \alpha)$ is the $(1 - \alpha)$ quantile of the standard normal distribution.

Though the normality approximation results in a simple analytical EC estimate, it is highly unreliable due to the much heavier tail of the real portfolio losses than the normal distribution.

Due to the hardness to defend the strict assumptions in the analytical models for portfolio loss distribution estimation, simulation-based models have become the main trend in practical EC estimation processes.

Following the loss generating process described in Eq. (1.17), account-level risk parameters EAD, LGD, and PD are estimated based on risk factors including macroeconomic variables. Unlike the marginal TTC estimates of these risk parameters used in the regulatory capital estimation, EC uses PIT (point in time) estimates of these risk parameters conditioning on risk factors. The PIT estimates are conditional on the risk factors observed at the reporting date. It is assumed that,

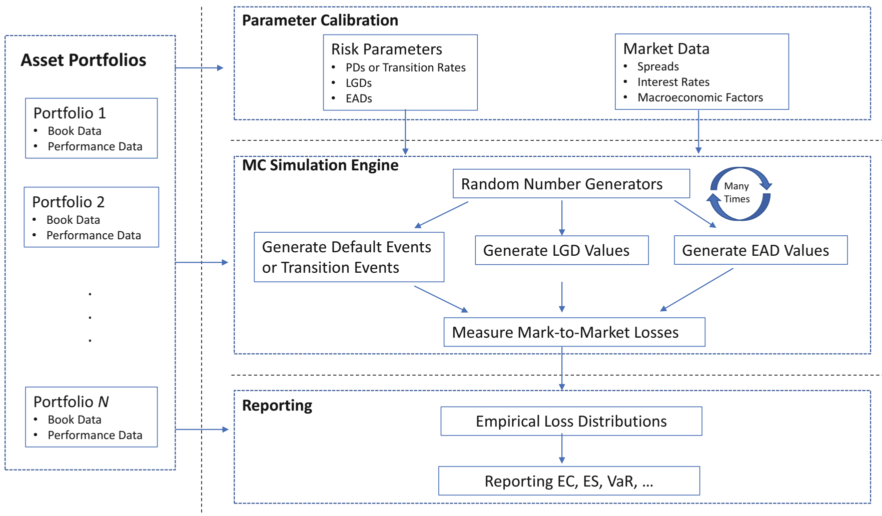


Fig. 1.7 Simulation-based EC process

conditioning on risk factors X, the default events of individual accounts in the portfolio are independent with the estimated probability:

$$\mathrm{PD}_i(x) = P(D_i = 1 | X = x) \tag{1.21}$$

Similarly, the other two risk parameters EAD and LGD are also assumed following some distributions (e.g., beta or generalized gamma) conditioning on risk factors X. Conditioning on risk factors, these three risk parameters are assumed mutually independent. Under this model framework, default events and losses correlations among individual accounts are implied by the dependence of individual account risk parameters on the common risk factors, as well as the correlation among PD, LGD, and EAD. More specifically, all these correlations are embedded in the PD, LGD, and EAD models estimated from the historical data including macroeconomic variables.

The EC described above is a conditional version based on the risk factors observed at the reporting date. By shocking the risk factors or designing stressed risk factor scenarios, we can carry out EC sensitivity and stressing analysis under the simulation-based model. By simulating the correlated risk factors, we can also obtain marginal EC. We will describe correlated risk factor simulation methods based on t-copula in Chap. 5. Figure 1.7 presents a flowchart of the simulation-based EC implementation.

Such implementation can easily be extended to also incorporate multiple asset classes, even multiple risk types such as market risk, interest risk, credit risk, operational risk, and insurance risk.

With multiple portfolios, ECs can be estimated individually for each portfolio, or a single EC for the entire pool of all portfolios combined can be estimated. Except for some extreme cases (e.g., extremely heavy tailed and highly correlated loss distributions among portfolios), the single pooled EC is less than the sum of the ECs of individual portfolios. This is the case for all VaR-based measures. The difference is called diversification benefit (DB):

$$\mathrm{DB} = \sum_{i=1}^{N} \mathrm{EC}(P_i) - \mathrm{EC}(P_1 + P_2 +, \ldots, +P_N) \tag{1.22}$$

To measure the impact of an individual portfolio to the combined portfolios on EC, the marginal EC of an individual portfolio is defined as:

$$\mathrm{MEC}(P_i) = \mathrm{EC}(P_1 + P_2 +, \ldots, +P_N) - \mathrm{EC}(P_1 + P_2 +, \ldots, +P_{i-1} + P_{i+1} +, \ldots, +P_N) \tag{1.23}$$

which measures additional EC increased if the portfolio is added into a pool of portfolios.

The diversification benefit should be allocated back to each portfolio inversely proportional to their marginal EC contributions, and capital allocation for the i_{th} portfolio given the pooled EC should be:

$$\mathrm{EC}_{\mathrm{Allocated}}(P_i) = \mathrm{EC}(P_i) - \frac{\mathrm{MEC}(P_i)^{-1}}{\sum_{j=1}^{N} \mathrm{MEC}(P_j)^{-1}} \mathrm{DB} \tag{1.24}$$

This allocation method can be used for any sub-level of entities from the combined entity, for example, from firm to line of businesses (LOB) EC or from firm to legal entities.

When there are a large number of portfolios in the combined entity (e.g., $N > 1000$), the computation for the marginal EC for all portfolios could be expensive due to a large number of simulations required in the EC calculation. In such case, the above capital allocation method may not be feasible, and some alternative approaches could be used. One such method is based on the EC contribution (vs. the diversification benefit above). For the contribution-based EC allocation method, first a portfolio independent grid is designed to cover all portfolios (e.g., a grid based on risk rating and maturity tenor in wholesale credit). Each portfolio's exposure is mapped into the cells of this grid. Then each cell is considered as a portfolio, and we can focus on these much smaller numbers of "portfolios." These cell portfolios are first considered as diversified, and normal intensity coefficients are calculated for these cells based on their EC contributions (measured by loss volatility), and then concentration intensity coefficients are calibrated based on an exponential EC growth rate with the exposure when the exposure exceeds a threshold. The combined diversified and concentration ECs are set to the total entity EC with proper scaling. Then, the allocated EC for each original portfolio is the combined diversified EC and concentration EC proportional to their exposures mapped to each cell. Refer to Sect. 5.4.5 for more details.

Allocated economic capital and their use in risk/reward ratios reveal which business lines an institution should pursue to make the best use of the risk/reward tradeoff. Performance measures that use economic capital include return on risk-adjusted capital (RORAC), risk-adjusted return on capital (RAROC), and economic value added (EVA). Business units that perform better on measures like these can receive more of the firm's capital in order to optimize risk.

1.4 Stress Test and CCAR

Besides the accounting rule changes, a significant post-financial crisis (2008) regulatory change for the financial industry risk management is on stress testing. This is legislated through the Dodd-Frank Act[14] in the United States and a series of stress testing rules required by the European regulators.

In the United States, the stress testing emphasizes on capital management. The Comprehensive Capital Analysis and Review (CCAR) was introduced as a key component of the Dodd-Frank Act and is an annual exercise conducted by the Federal Reserve on the largest bank holding companies operating in the United States (FRB, 2020a).[15] CCAR consists of two components, the quantitative part based on the Dodd-Frank stress test results (FRB, 2020b)[16] and a qualitative assessment of firms' capital plans. In addition, capital surcharges for the stress capital buffer (SCB) and global systemically important banks (GSIBs) are assessed for each participating bank from 2020 as a simplification of the minimal capital requirements under both normal and stress scenarios.

For the quantitative component, banks are required to project losses, incomes, and regulatory capital ratios under both the FRB-supplied macroeconomic scenarios and banks' own designed scenarios with their own methodologies for nine consecutive quarters. Based on the banks' input data and balance sheet information, FRB carries out its own assessment of these quantities with its own methodologies and publishes the results for all participating banks as the exam results. The purpose of these quantitative measurements under different scenarios, especially the adverse and severely adverse scenarios, is to provide a forward-looking assessment of capital adequacy under a range of stressful scenarios as a key input to a firm's capital plan. For purposes of the qualitative assessment, the Federal Reserve assesses the strength of the firm's capital planning practices, including the firm's ability to identify,

[14] The Dodd-Frank Wall Street Reform and Consumer Protection Act (commonly referred to as the Dodd-Frank Act) is a US federal law that was enacted on July 21, 2010. The law overhauled financial regulation in the aftermath of the Great Recession, and it made changes affecting all federal financial regulatory agencies and almost every part of the nation's financial services industry.

[15] FRB, 2020. Comprehensive Capital Analysis and Review 2020 Summary Instructions March 2020. Board of Governors of the Federal Reserve System, Washington, DC.

[16] FRB, 2020. Dodd-Frank Act Stress Test 2020: Supervisory Stress Test Methodology. March 2020. Board of Governors of the Federal Reserve System, Washington, DC.

measure, and determine the appropriate amount of capital for its risks, and controls and governance supporting capital planning. The qualitative assessment is informed by a review of the materials each firm provides in support of its annual capital plan submission. In addition, the Board's qualitative assessment incorporates supervisory assessments that are undertaken throughout the year. The qualitative assessments serve as the basis for the Federal Reserve's decision to object to certain firms' capital plans.

In European Union (EU), several types of stress test are carried out. Every 2 years, the European Banking Authority (EBA) carries out EU-wide stress tests in cooperation with the European Central Bank (ECB), the European Systemic Risk Board (ESRB), and the national supervisory authorities. The ECB together with the EBA deployed this EU-wide stress testing exercise back in 2011, and then the 2-year cycle started from 2014 and up to 2020[17] (EBA, 2020).[18] In addition to the EU-wide EBA stress tests, EU law requires the ECB to carry out stress tests on its supervised banks at least once per year through the Supervisory Review and Evaluation Process (SREP). ECB also exercises other stress tests in addition to the SREP process. The Bank of England in 2014 aligned with the Federal Reserve Board in committing to an annual stress testing process for major UK banks (BOE, 2015).[19] Other regulators across the world followed the same approach by introducing a stress testing assessment on a regular basis.

The two main components for stress testing are stress scenarios and stress methodologies. Design of stress scenarios and selection of stress methodologies are essential steps for stress testing. In the following, we will focus on the CCAR exercise on these steps.

1.4.1 Scenarios for Stress Testing

A stress scenario is not a forecast of macroeconomic and financial conditions. On the contrary, it is a coherent set of conditions designed specifically to assess the resilience of banks when facing a deterioration in global economic conditions. In the United States, FRB defined the Scenario Design Framework[20] through the Dodd-Frank Act stress test rules. In addition, CCAR participating firms are also required to develop their own stress scenarios as part of the stress testing results in the CCAR submission.

---

[17] In view of the extraordinary circumstances relating to the coronavirus (COVID-19) pandemic, the EBA has decided to postpone the 2020 EU-wide stress test until 2021. The ECB supports this decision and has postponed its own 2020 SREP stress test until 2021 as well.

[18] EBA, 2020. 2020 EU-wide stress test. Methodological note. European Banking Authority, London.

[19] BOE, 2015. Stress testing the UK banking system: key elements of the 2015 stress test. Bank of England Publications, London.

[20] 12 C.F.R. Appendix A to Part 252 – Policy Statement on the Scenario Design Framework for Stress Testing. Feb 28, 2019.

FRB Scenario Design Framework

The stress test rules define scenarios as "those sets of conditions that affect the U.S. economy or the financial condition of a company that the Board annually determines are appropriate for use in stress tests, including, but not limited to, baseline, adverse, and severely adverse scenarios." The stress test rules define baseline scenario as a "set of conditions that affect the U.S. economy or the financial condition of a company and that reflect the consensus views of the economic and financial outlook." The stress test rules define adverse scenario as a "set of conditions that affect the U.S. economy or the financial condition of a company that are more adverse than those associated with the baseline scenario and may include trading or other additional components." The stress test rules define severely adverse scenario as a "set of conditions that affect the U.S. economy or the financial condition of a company and that overall are more severe than those associated with the adverse scenario and may include trading or other additional components." With these general guidelines, the FRB Scenario Design Framework consists of two components for stress conditions, the macroeconomic scenarios and the market shocks. While the macroeconomic scenarios focus on the general stress factors, market shocks focus on instantaneous events, which immediately affect the market value of the companies' trading assets and liabilities. So, market shocks apply only to companies with significant trading activity. FRB design the three scenarios – baseline, adverse, and severely adverse scenarios – for these two components correspondingly. Instead of technique details, only general guidelines are provided on how these scenarios are created in the FRB Scenario Design Framework. We summarize these guidelines in the following.

For macroeconomic scenarios, the three scenarios are generated according to the following guidelines:

- *Baseline* – The baseline scenario will be developed around a macroeconomic projection that captures the prevailing views of private sector forecasters (e.g., Blue Chip Consensus Forecasts and the Survey of Professional Forecasters), government agencies, and other public sector organizations (e.g., the International Monetary Fund and the Organization for Economic Co-operation and Development) near the beginning of the annual stress test cycle. The baseline scenario is designed to represent a consensus expectation of certain economic variables over the time period of the tests, and it is not the Board's internal forecast for those economic variables.
- *Severely Adverse* – The Board intends to use a recession approach to develop the severely adverse scenario. In the recession approach, the Board will specify the future paths of variables to reflect conditions that characterize post-war US recessions, generating either a typical or specific recreation of a post-war US recession. The Board chose this approach because it has observed that the conditions that typically occur in recessions such as increasing unemployment, declining asset prices, and contracting loan demand can put significant stress on companies' balance sheets. This stress can occur through a variety of channels, including higher loss provisions due to increased delinquencies and defaults;

losses on trading positions through sharp moves in market prices; and lower bank income through reduced loan originations. For these reasons, the Board believes that the paths of economic and financial variables in the severely adverse scenario should, at a minimum, resemble the paths of those variables observed during a recession.

- *Adverse* – The adverse scenario can be developed in a number of different ways, and the selected approach will depend on a number of factors, including how the Board intends to use the results of the adverse scenario. Generally, the Board believes that the companies should consider multiple adverse scenarios for their internal capital planning purposes, and likewise, it is appropriate that the Board consider more than one adverse scenario to assess a company's ability to withstand stress. Accordingly, the Board does not identify a single approach for specifying the adverse scenario. Rather, the adverse scenario will be formulated according to one of the possibilities (scaling, probabilistic, stable). The Board may vary the approach it uses for the adverse scenario each year so that the results of the scenario provide the most value to supervisors, in light of current condition of the economy and the financial services industry.

For the severely adverse macroeconomic scenarios based on historical recessions, FRB first creates the path for the unemployment rate, which feathers a jump from 3% to 5% points from the initial point of testing over the recession period. FRB calibrates this jump either based on the limit level of at least 10% (e.g., at an economic expansion period) or a lower end of the change (e.g., at an early economic recovery period). Once the unemployment rate path is created, all other variables in the severely adverse scenario will be specified to be consistent with the increase in the unemployment rate. The approach for specifying the paths of these other variables in the scenario will be a combination of (1) how economic models suggest that these variables should evolve given the path of the unemployment rate, (2) how these variables have typically evolved in the past US recessions, and (3) and evaluation of these and other factors.

For the adverse macroeconomic scenarios, in addition to the three methods – scaling (scale the differences between severely adverse and baseline (by half or two-thirds)), probabilistic (take a quantile of the macroeconomic variable distribution developed from the baseline), and stable (take stable value of macroeconomic variables) – FRB gives flexible guidelines including adding or reducing macroeconomic variables from the severely adverse scenarios and using firms' internal designed stress scenarios.

FRB also issues guidelines for market shocks under the three scenarios:

- *Baseline* – No shocks by definition.
- *Severely Adverse* – The general market practice for stressing a trading portfolio is to specify market shocks either in terms of extreme moves in observable, broad market indicators and risk factors or directly as large changes to the mark-to-market values of financial instruments. These moves can be specified either in relative terms or absolute terms. Supplying values of risk factors after a "shock" is

roughly equivalent to the macroeconomic scenarios, which supply values for a set of economic and financial variables; however, trading stress testing differs from macroeconomic stress testing in several critical ways.

- *Adverse* – The market shock component included in the adverse scenario will feature risk factor movements that are generally less significant than the market shock component of the severely adverse scenario. However, the adverse market shock may also feature risk factor shocks that are substantively different from those included in the severely adverse scenario, in order to provide useful information to supervisors. As in the case of the macroeconomic scenario, the market shock component in the adverse scenario can be developed in a number of different ways.

For the severely adverse market shocks, FRB prefers the risk factors method. The market shock component for the severely adverse scenario will incorporate key elements of market developments during the second half of 2008 but also incorporate observations from other periods or price and rate movements in certain markets that the Board deems to be plausible though such movements may not have been observed historically. Over time, the Board also expects to rely less on market events of the second half of 2008 and more on hypothetical events or other historical episodes to develop the market shock. Once broad market scenarios are agreed upon, specific risk factor groups will be targeted as the source of the trading stress.

Statistical Scenario Framework

Different from the regulatory scenario design framework, statistical scenario framework explores the full joint distribution of macroeconomic variables based on statistical models instead of focusing on few special scenarios. Loss distributions are derived from the joint distribution of the macroeconomic variables. Stressed loss is defined as a high quantile of the loss distribution or the expected shortfall similar as in the economic capital framework.

The statistical scenarios are commonly used under the economic capital framework as described in Sect. 1.3.2, which can be used as a bottom-up stress testing scheme with integrated risks. Based on integrated capital and liquidity thresholds, a reverse stress testing also can be designed based on the statistical scenarios. We will give more details on reverse stress testing in Chap. 6.

In the following, we briefly introduce common models used to generate statistical scenarios.

The first model popularly used to generate statistical scenarios for macroeconomic variable is the vector autoregression (VAR). Assuming X_t is the vector of macroeconomic variable at time t, the VAR model with order p is:

$$X_t = \boldsymbol{a} + B_1 X_{t-1} + B_2 X_{t-2} + \ldots + B_p X_{t-p} + \varepsilon_t \tag{1.25}$$

where $\boldsymbol{a}$ is the constant p-vector, B_i, $i = 1, \ldots, p$ are p by p coefficient matrices, and $\varepsilon_t \sim N(0, \sigma^2 I)$ is the standard normal random vector. When all coefficient matrices B_i,

$i = 1, \ldots, p$ are diagonal, the VAR model can be represented as p one-dimension autoregression models AR(p):

$$x_{i,t} = a_i + b_{i,1}x_{i,t-1} + b_{i,2}x_{i,t-2} + \ldots + b_{i,p}x_{i,t-p} + \epsilon_{i,t} \tag{1.26}$$

Using the lag operator L, this can be rewritten as:

$$\left(1 - b_{i,1}L - b_{i,2}L^2 - \ \ldots \ - b_{i,p}L^p\right)x_{i,t} = \lambda_i(L)x_{i,t} = a_i + \epsilon_{i,t} \tag{1.27}$$

The AR(p) model is called stationary if the root of the polynomial $\lambda_i(L)$ falls outside of the unit circle. Time series from a stationary model has bounded variance. If one (or more) root falls on the unit circle, one can use difference (more higher-order difference) to transfer to a stationary model:

$$\lambda_i^*(L)(1 - L)x_{i,t} = \lambda_i^*(L)\Delta x_{i,t} = a_i + \epsilon_{i,t} \tag{1.28}$$

with the difference $\Delta\ x_{i,\ t} = x_{i,\ t} - x_{i,\ t\ -\ 1}$ to obtain the error correction model:

$$\Delta x_{i,t} = a_i + b_{i,1}^*\Delta x_{i,t-1}b_{i,2}^*\Delta x_{i,t-2} + \ldots b_{i,p}^*\Delta x_{i,t-p} + \epsilon_{i,t} \tag{1.29}$$

So, we assume the VAR model (1.25) is stationary. A time trend can be added to the time series, and then the AR model is called trend stationary. The coefficients of the AR model can be estimated from historical data using least squares. Then the estimated model can be used to project the future values of the series based on the current observed values. The stationary requirement of the AR model makes sure that the projection is meaningful and not going to infinity certainly.

A VAR model as in Eq. (1.25) or more generally a vector error correction (VEC) model under the normality assumption is popularly used to generate macroeconomic scenarios for economic capital based on loss distribution analysis and stress testing. However, the normality assumes a very thinner tail of the distribution and does not provide sufficient variability presented by the macroeconomic variables. An alternative method to generate the macroeconomic scenarios is using the t-copula. This can be done in the following steps:

1. Predict each individual macroeconomic variable according to its AR model with time trend.
2. At each time point, use a t-copula to create the joint distribution of all macroeconomic variables, and then simulate from this joint distribution.
3. Add the predicted values at each time point to the simulated random sample from the t-copula and use the combined as the simulated paths of the macroeconomic variables.

In summary, the simulated macroeconomic variable paths are $\widehat{x}_{i,t} + \widehat{\epsilon}_{i,t,g}$, $g = 1, \ldots, G$, where

$$\widehat{x}_{i,t} = \widehat{a}_{\mathrm{i}} + \widehat{b}_{i,1}\widehat{x}_{i,t-1} + \widehat{b}_{i,2}\widehat{x}_{i,t-2} + \ldots + \widehat{b}_{i,p}\widehat{x}_{i,t-p} \tag{1.30}$$

and $\widehat{\epsilon}_{i,t,g} \sim C_p(\mathrm{df}, \Sigma_t)$ is a p-dimension t-copula with degree of freedom df and correlation Σ_t.

The correlation matrix can be estimated using historical data and may be adjusted at different time points. The degree of freedom df is a judgmental call, and the common values are 2–7 with the larger the number, the thinner the tail of the copula.

1.4.2 Stress Test Methodologies

As the quantitative component of CCAR, the Dodd-Frank stress test is required to estimate the effect of supervisory scenarios on the regulatory capital ratios of firms participating in the supervisory stress test by projecting net income and other components of regulatory capital for each firm over a nine-quarter projection horizon. Projected net income, adjusted for the effect of taxes, is combined with noncommon capital action assumptions and other components of regulatory capital to produce post-stress capital ratios.

The Federal Reserve's approach to modeling post-stress capital ratios generally follows US generally accepted accounting principles (GAAP) and the regulatory capital framework. Figure 1.8 illustrates the framework used to calculate changes in net income and regulatory capital.

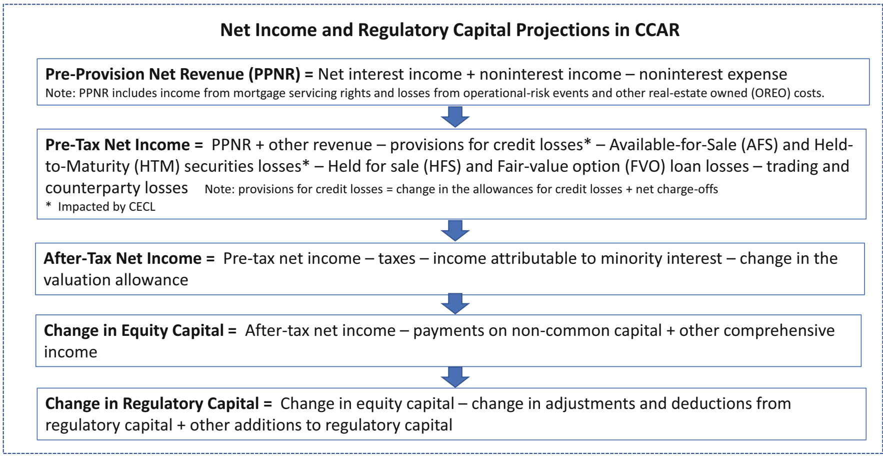


Fig. 1.8 CCAR process

The Federal Reserve calculates projected pre-tax net income for the firms subject to the supervisory stress test by combining projections of revenue, expenses, provisions for credit losses, and other losses, including:

1. PPNR
2. Provisions for credit losses
3. Credit losses on investment securities in the available-for-sale (AFS) and held-to-maturity (HTM) portfolios
4. Losses on loans held for sale (HFS) or for investment and measured under the fair value option (FVO)
5. Losses on market risk exposures, credit valuation adjustment (CVA), and incremental default risk (IDR) for firms subject to the global market shock
6. Losses from a default of the largest counterparty for firms with substantial trading, processing, or custodial operations

While projecting each of these items under different scenarios requires detailed treatment, we focus on the credit losses, which are covered in Items 2 and 3.

Provisions for credit losses of accrual loan portfolios in a quarter cover both the loss allowance changes and net charge-offs (or recorded loss) in this quarter. The loss allowance is defined as the sum of future four quarters expected credit losses before the adoption of CECL (which is defined as the lifetime expected credit loss at the current date.) So, in CCAR, 13 quarters of losses are required to be projected under different scenarios. Over the nine quarters, projection of provisions in each quarter includes the projected credit loss in this quarter and the changes of the next four quarters projected credit loss sum. Once CECL is adopted, the provisions over the nine quarters will be simplified as the lifetime expected credit loss at the beginning of each quarter.

Investment securities in the available-for-sale (AFS) and held-to-maturity portfolios could be subject to credit losses, for example, corporate debt, municipal debt, collateralized loan obligation (CLO), and collateralized debt obligation (CDO) securities. Quarterly loss projections are required similarly as for accrual portfolios in the provisions for the current quarter charge-offs and the allowance changes. In addition, for these security portfolios, unrealized gains or losses due to interest rate and other risk factors over the nine-quarter horizon are required to be recognized in capital for some firms through other comprehensive income (OCI) as described in Sects. 1.2 and 1.3.

For the quarterly loss projection, there are two approaches commonly used, the expected credit loss projection based on PD, LGD, and EAD at account level and the charge-off projection at portfolio level. These models take in specific macroeconomic factors as direct inputs, and stressing effects are represented by the model outputs and the final projected losses under different scenarios. Chapter 6 presents more details on these approaches.

Once the regulatory capital is projected in each of the nine quarters under a scenario, capital ratios under this scenario can be computed as the ratio of the regulatory capital and the risk-weighted asset (RWA) based on its asset structure in the quarter. The Federal Reserve generally projects that a firm takes actions to maintain its current level of assets, including its securities, trading assets, and loans, over the projection horizon. The Federal Reserve assumes that a firm's risk-weighted assets (RWAs) remain unchanged over the projection horizon except for changes primarily related to items subject to deduction from regulatory capital or due to changes to the Board's regulations.

The regulatory stress test framework described above is mainly used for CCAR. Based on this framework, firm-specific stress tests can be built for some firm-specific purposes. For example, a single risk factor shock can be tested across different scenarios to see its impact under different scenarios. Rather than assuming a constant RWA, firm can also change the asset structure to test the impact on capital ratios across different scenarios.

Besides the regulatory stress test framework, the economic capital framework described in Sect. 1.3.2 is another framework that can be used for stress testing. Combined with the statistical scenarios designed in Sect. 1.4.1, the economic capital framework can provide full distribution of incomes, losses, liquidity, and capital. Unlike the regulatory stress test framework, which focuses on capital and leverage, the economic capital framework is more flexible to be able to test various risks, e.g., concentration risk and liquidity risk, and can perform reverse stress testing. We will share more details on using economic capital framework for stress testing in Chap. 6.

1.5 Underwriting and Credit Scoring

While the credit model applications in the previous three chapters are closer to the inhouse credit risk management, credit underwriting and scoring are for the front door credit lending business. As we know if the doorkeeper does a good job to turn the bad apples away, the party in the bar may avoid some violence and people in the bar can have an enjoyable night. Credit underwriting works as the doorkeeper for the credit lending business and tries to discriminate the crowd with some superintelligence (called credit intelligence), which including the credit scoring techniques, to rein the potential risks with those pass the door.

Credit underwriting can go back to ancient time as an act of selection in sharing or giving with certain expected return. Over the long history, some insights and experiences have been collected, for example, the 5 Cs (character, capacity, collateral, capital, and conditions), which have been commonly used by credit managers from modern time credit underwriting until now. Most of these factors are hard to measure (especially the first two), while partial information could be obtained through in-person interviews and other relationship channels. Such manual underwriting process could not meet the large volume of credit demands with the economy growth. This became a bottleneck in the underwriting business, especially for consumer lending, until the modern computing power enabled the execution of statistical models from the second half of the last century. A credit underwriting process developed on predictive statistical models becomes mature. Here we briefly introduce such a full credit underwriting process and share more details in Chap. 7.

Risk rating is the core of a model-based credit underwriting process and can be executed through various techniques, among which credit scoring based on odds estimated by the logistic regression model becomes the most popular one. We developed a full credit score modeling process from model population, segmentation, target definition, reject inference, score driver selection, and fair lending testing to the final step of building standard scorecards. In this section, we briefly introduce this modeling process and leave the details for each of these steps to Chap. 7, including our recommendations on some popular methods for data and model as well as some warnings on risks of direct adoption of some techniques from machine learning.

1.5.1 Credit Underwriting

Credit underwriting is one major financial underwriting besides insurance underwriting. Financial underwriting is the process of customer information collection and gathering, customer segmentation and risk classification, and decision-making and actions related to the sale of the financial product. So, financial underwriting is largely a customer assessment and selection process for specified financial products. These customers have been collected by the marketing operation, and some preselection process has been executed, for example, the popular 4 Rs – response, risk, retention, and revenue. However, the real customer assessment starts from the underwriting process.

For credit underwriting, many underwriting processes have been developed in modern time, from the early relationship-based underwriting, to experience-based underwriting focusing on some basic customer characteristics (e.g., collateral and capacity), and to the expanded experience-based underwriting with a full set of customer characteristics (e.g., the five Cs plus commitment, competency, etc.). Some of these underwriting processes still exist in underdeveloped economies or underground lending. However, all these underwriting processes are largely manual and could not meet the requirement on cost and efficiency to process a large volume of applicants, especially in the consumer lending area, as the economy grew from the second half of the last century. Regulations also have become more and more demanding on the transparency of the credit underwriting process and issued several fair lending laws. In the meantime, computing power enabled the execution of statistical models, which has the power to do customer classification and discrimination based on public data in a much more "transparent and unbiased" way rather than the opaqued manual underwriting process by a risk manager, which is often considered hard to interpret and has the suspension of human biases. So, credit underwriting has converged to the model-based automated process, which turns out to be the most successful application area for predictive statistical models.

While there are variant of designs for a credit underwriting process based on statistical models, we present an underwriting process in Fig. 1.9, which can generally fit for all credit products. The process also takes into account the current

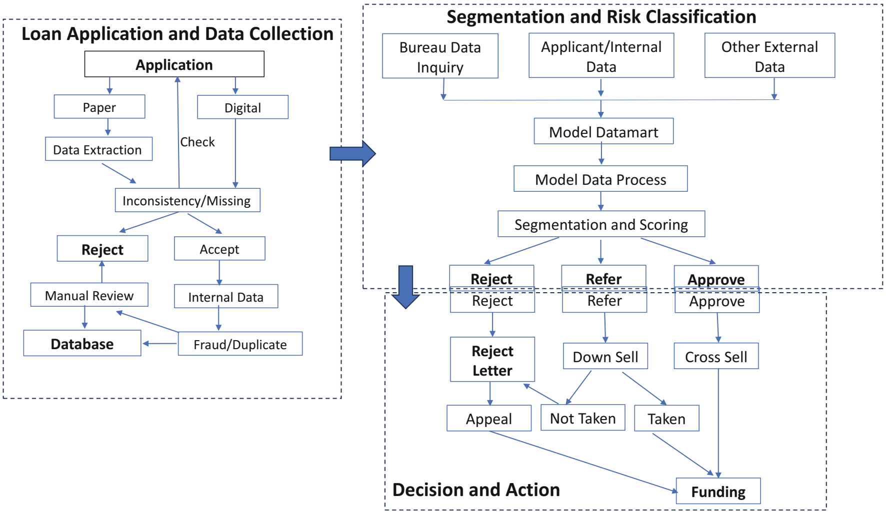


Fig. 1.9 Credit underwriting process

consumer lending practices implemented by most large banks and FinTech companies.

We split the credit underwriting process into three stages:

- *Loan Application and Data Collection*: Covers the information and data collection from the application as well as the application sanity and fraud checks
- *Segmentation and Risk Classification*: Covers the data ETL, model data processing, and underwriting segmentation and scoring as the core of the underwriting process
- *Decision and Action*: Covers underwriting decision processing based on segmentation and scoring results, as well as the following actions taken

In Chap. 7, we present more details on each of these three stages, their interactions, as well as impacts from regulations on these stages.

As the core of the credit underwriting process, segmentation and risk classification deserve further diving in, especially the quantitative tools used in this stage. While segmentation could depend on factors including products and line of businesses, risk classification more depends on quantitative measurements from credit models, which is the focus of this book. We introduce the credit modeling components by the order of risk rating, credit scoring, and scorecards sequentially following the implementation order of these quantitative tools:

- *Risk Rating*: is the base measurement for risk classification, including risk grades for business risk rating (or corporate rating) by a rating agency (e.g., Moody, Fitch, S&P) and risk scores for consumer risk rating (or consumer credit score) by credit reporting agency (e.g., Experian, Equifax, and Transunion) or lending businesses for specific credit products using statistical models, which is our focus

- *Credit Scoring*: is the modeling process by using predictive statistical models to assess the credit worthiness of a consumer and translate the model outputs into an easily interpretable measurement framework like scorecard for the credit underwriting process
- *Scorecard*: is the score system using integers between 1 and 999 to interpret the credit worthiness estimated from the credit model based on odds of repayment risks

More details on each of these three components and the relationships in the implementation of these components are discussed in Chap. 7.

1.5.2 Credit Scoring

Credit scoring is the modeling process for score models and scorecards, which take in measurable consumer characteristics (called score drivers) as input and produce an output measure for the consumer rating on repayment risk.

Among the different types of models for risk rating, the probability (or odds)-based models are selected for credit scoring due to several reasons. First, creditworthiness has been considered to be the proper quantity to evaluate a consumer, and over the history, creditworthiness has been defined as a function (most popularly, the linear function) of consumer characteristics. However, we don't know this function, and we can only observe the consumer's repayment risk through the observed repayment behavior, like a failure event to meet the required payment. To link the creditworthiness with the observed repayment behavior, the logit link (or log odds) between the risk event probability and creditworthiness becomes a natural selection, and the link function can be easily estimated through the statistical maximum likelihood method. Second, scores are considered continuous quantity although they are preferred to be rounded to be integers. Mapping from the model output to credit scores is preferred to be from a continuous quantity to a continuous quantity, while probability and odds naturally meet this requirement. The FICO-type mapping described in Sect. 7.2.7 has become the most popular mapping from odds to scores. Finally, the probability (or odds)-based models are easy to interpret on the dynamics of scores, both for the overall odds changes and the impact of individual score driver. This is one of the critical reasons why scorecards are so popular on risk rating. In addition, the probability model based on the logit link or logistics regression model has good stability over large data samples with overlapped observations as discussed in Sect. 3.1.1.1, which are commonly true for credit scoring data.

Risk rating directly based on classification models, like those models from machine learning, has been a hot topic. However, the lack of stability and interpretability makes these ML-based models hard to be accepted as a credit scoring tool for the daily business run. A better use of these models is for the complementary data pattern exploration and consumer risk mining, which can be combined with the

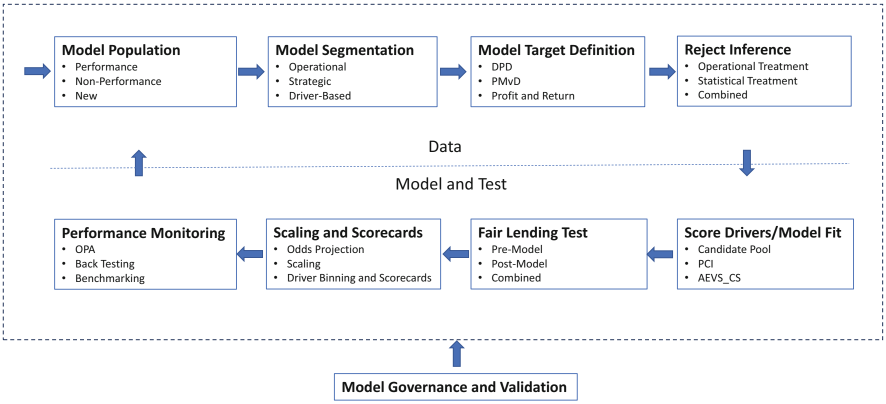


Fig. 1.10 Credit score modeling process

primary credit scoring system for additional selection and filtering on targeted population with alternative data. Caution should be taken for such exercises to make sure regulations (like those on fair lending) are not violated.

We focus on the primary credit scoring system. We design the full step-by-step modeling process in a cyclical framework as shown in Fig. 1.10.

We present a brief summary for each of these steps and would refer to Sect. 7.2 for details.

- *Model Population*: defines the data population according to the modeling purpose, including performance population and non-performance population based on the availability of payment performance observation and a new population for testing purpose. Model population identification and definition should be the start of the score modeling process, and population stability based on PSI should be checked to make sure model populations don't change significantly.
- *Model Segmentation*: describes the three types of factors for model segmentation in credit scoring – operational, strategic, and driver-based. Operational factors, like product type and line of business, are the most commonly used segmentation factors as in other credit model application. Strategic factors based on business strategic are more often for credit scoring due to the business requirement with credit underwriting. Driver-based segmentation is mainly for granularity to meet score model requirement, e.g., monotone scorecards.
- *Model Target Definition*: introduces various model targets for a score model, which could be based on different measures on payment performance, profitability, and returns. However, there are some basic properties the model target should hold to be a valid model target.
- *Reject Inference*: deals with the non-performance population as a unique but essential model data treatment in credit score modeling. Borrowing the techniques from missing data treatment in modeling, we included both the operational treatment and statistical treatment based on extra data and information

availability, model assumptions, as well as decisions by modeler and model sponsor.

These steps generally focus on the data part of the credit score modeling, while the following steps more focus on the model fit, testing, and performance monitoring:

- *Score Drivers and Model Fit*: defines the Prohibited Correlation Index (PCI) for individual and a group of score drivers as a measure of the correlation with the prohibited base and implements it in the Adaptive and Exhaustive Variable Selection process designed for Credit Scoring (AEVS_CS) based on the general AEVS.
- *Fair Lending Testing*: defines the PCI-based fair lending test (FLT), which could be implemented in the pre-model variable selection (AEVS_CS) process or in the post-model testing based on specified dataset or a combination of both.
- *Target Prediction, Scaling, and Scorecards*: describes the model target prediction based on odds and its relationship with creditworthiness, the properties a scorecard should have, and the FICO-type score scaling process.
- *Performance Monitoring*: This step is adopted from model validation (Sect. 7.6.5) to complete the credit score modeling cycle. The monitoring could include on-going performance monitoring (OPA) designed for the specific score model based on criteria and tests (e.g., PSI tests), a general back-testing procedure, and some benchmarks. Breaks of the monitoring rules could lead to the requirement of model update and redevelopment.

Using the small business lending data as an example, we implement all these modeling steps in Chap. 7 (Sects. 7.3 and 7.4). A full model validation procedure for score models is presented in Sect. 7.6, including the demonstration of issues with the SMOTE sampling method in credit score modeling and alternative score models using WoE, as well as score driver binning for scorecards.

Credit Data and Processing

2

Credit data refer to any data related to a credit product during its lifetime. Credit data vary with different products, for example, retail and wholesale loans have different types of credit data. Credit data also depends on availability by aggregation levels. Account-level data have more granular information for each account compared to aggregated cohort-level data. In the time horizon, credit data can be classified as origination data and transaction data. While the origination data describe all characteristics at the credit product origination, the transaction data record the periodic (e.g., daily, weekly, monthly) status changes of the credit product. So, the origination data are static, and the transaction data are dynamic. In the content horizon, origination data includes feathers describing borrower/obligor characteristics, product characteristics, and collateral characteristics if the product is secured by some underlying asset. For transaction data, it could consist of the full transaction history since origination, or transactions from a booking date, or just a snapshot of the booking at some specific dates. Figure 2.1 shows the structures of credit data by these different dimensions.

Credit data have greater availability compared to other risk data, especially for retail products, like mortgage, credit cards, auto loans, small business loans, and other retail products. This is largely due to the stable servicing regardless if the product is held to maturity or securitized and sold by the originator. Although data are collected and processed differently at different firms, some common practices both in data processing and modeling converge. In this chapter, we will describe some formats of credit data commonly used in the financial industry on credit modeling and analytics.

Credit data are essential inputs for credit modeling and analytics. Different types of data can produce different types of model estimates, as well as different models require different formats of data. In this chapter, we will focus on the steps of data processing for preparing model-ready data, which include segmentation, sampling, transformation, and integration with other data, for example, the macroeconomic data. The models will be discussed in the following chapter.

C. Chen, *Practical Credit Risk and Capital Modeling, and Validation*, Management for Professionals, https://doi.org/10.1007/978-3-031-52542-1_2

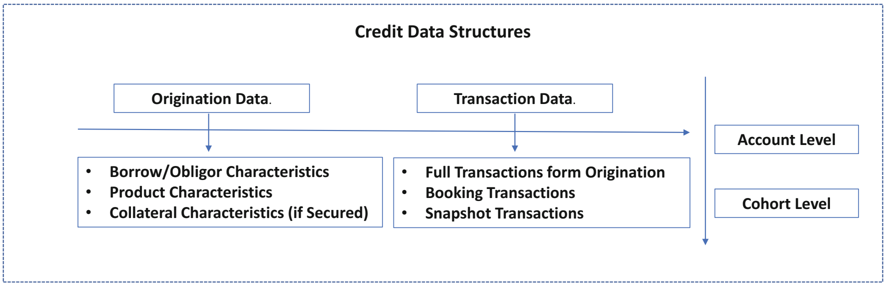


Fig. 2.1 Credit data by structures

Key Abbreviations and Symbols

ACL	Allowance for credit losses
AI	Artificial intelligence
CCAR	Comprehensive Capital Analysis and Review
CECL	Current expected credit loss
CPI	Consumer Price Index
CSD	Cross-sectional data
DFS	Dealer financial services
DPI	Disposable personal income
DTI	Debt to income
DR	Default rate
EAD	Exposure at default
EC	Economic capital
EDF	Empirical distribution function
ETL	Extraction, transferring, and loading
FOSS	Full observation stratified sampling
GDP	Gross domestic product
IRB	Internal ratings-based
IV	Information value
LGD	Loss given default
LTV	Loan to value
OS	Observation sampling
PD	Probability of default
PIT	Point in time
PTI	Payment to income
RC	Regulatory capital
SMOTE	Synthetic Minority Oversampling Techniques
SQL	Sequential Query Language
TTC	Through-the-cycle
UPB	Unpaid principal balance
VaR	Value at risk
WAC	Weighted average coupon
WALA	Weighted average loan age
WoE	Weight of Evidence

2.1 Account-Level Data

Credit data in account level record information related to the account once the account is set up for the credit product, for example, a mortgage, a credit card, an auto loan in retail lending, or a commercial and industry loan in the wholesale sector. The information includes the periodic transactions such as payments and status updates, as well as all information in the application/origination. The information can be used to monitor the account performance and provide early warning, which might lead to some lender action to mitigate the potential loss. Such information can also be used to forecast the obligor's future behavior at the beginning of the loan application and assist the lending to price the product. We classify such information into transaction data and origination data and discuss more in details in the following sections.

2.1.1 Account Transaction Data

Transaction data record the periodic activities of the account and its status updates. Credit product has periodic payment processing daily, weekly, biweekly, or monthly according to product characteristic. For example, mortgage, auto loan, and credit card most often are paid monthly. However, for small business lending, due to its short-term property, the payment could be daily, weekly, or biweekly. Transaction data also record the account status – current, past due, or default.

Table 2.1 shows an example of small business lending transaction data.

Figure 2.2 shows the first 20 rows of the transaction data set. The transaction data include the unique account identifier (AccountID) and funding information that includes the funding date (FDate), quarter (FQuarter), and amount as the funded amount outstanding (FAO) at the loan origination of the transaction date (PDate), which transaction is also marked as "Purchase" with the transaction type attribute. Pay frequency (PayFreq) indicates this loan is scheduled as daily payment, and gross payment (GrossPay) is the daily payment, and its cumulation is marked by the cum collection (CumColl) attribute. The outstanding gross receivable (OutGrossRe) is the difference between the total gross receivable and the cum collection. Similarly, the funded amount outstanding is the difference between the total funded amount and the cumulative funded amount collection (CumFAC), and income outstanding (IncOut) is the difference of the total interest income and cumulative income/interest collection (CumIncColl). DPD is the number of days post due, which is calculated based on the scheduled receivable and cumulative actual payments. Deal status marks the account status as active, repaid, or written-off.

Account transaction data record the loan amortization and cash flow processes. For longer contract term products, such as mortgage and auto loans, portfolios consisting these products will be on the banking book for some time. Either 1-year or lifetime loss forecast for such portfolios needs to base on the current conditions at forecasting time as well as conditions in the future. To mimic the banking book

Table 2.1 Small business lending transaction data description

Field	Description
AccountID	Unique (internal) account identifier; each contract is corresponding to one account
Funding quarter	Quarter in which funds are received by customer
Funding date	Date funds are received by customer
Payment date	Date of transaction
Transaction type	Type of the transaction (e.g., purchase, payment, early payoff, partial payoff, recovery, charge-back)
Payment frequency	Repayment frequency (e.g., daily, weekly, bi-monthly, monthly)
Gross payment	Gross payment received on transaction date
Cum collections	Cumulative repayment amount as of payment date
Outstanding gross receivable	Total remaining receivable
Cum funded amount collections	Total portion of funded amount repaid as of payment date
Funded amount outstanding	Portion of funded amount outstanding as of payment date
Cum income collections	Total income/interest paid as of payment date
Income outstanding	Portion of income outstanding as of payment date
DPD	Days past due
Deal status	Status as of transaction date (active, repaid, write-off)

AccountID	FQuarter	FDate	PDate	TransType	PayFreq	GrossPay	CumColl	OutGrossRe	CumFAC	FAO	CumIncColl	IncOut	DPD	DealStatus
2291	2016 Q1	2016-01-07	2016-01-07	Purchase	Daily	-25000	0	30500	0	25000	0	5500	0	Active
2291	2016 Q1	2016-01-07	2016-01-08	Payment	Daily	181.55	181.55	30318.45	148.81	24851.19	32.74	5467.26	0	Active
2291	2016 Q1	2016-01-07	2016-01-11	Payment	Daily	181.55	363.1	30136.9	297.62	24702.38	65.48	5434.52	0	Active
2291	2016 Q1	2016-01-07	2016-01-12	Payment	Daily	181.55	544.65	29955.35	446.43	24553.57	98.22	5401.78	0	Active
2291	2016 Q1	2016-01-07	2016-01-13	Payment	Daily	181.55	726.2	29773.8	595.25	24404.75	130.95	5369.05	0	Active
2291	2016 Q1	2016-01-07	2016-01-14	Payment	Daily	181.55	907.75	29592.25	744.06	24255.94	163.69	5336.31	0	Active
2291	2016 Q1	2016-01-07	2016-01-15	Payment	Daily	181.55	1089.3	29410.7	892.87	24107.13	196.43	5303.57	0	Active
2291	2016 Q1	2016-01-07	2016-01-19	Payment	Daily	181.55	1270.85	29229.15	1041.68	23958.32	229.17	5270.83	0	Active
2291	2016 Q1	2016-01-07	2016-01-20	Payment	Daily	181.55	1452.4	29047.6	1190.49	23809.51	261.91	5238.09	0	Active
2291	2016 Q1	2016-01-07	2016-01-21	Payment	Daily	181.55	1633.95	28866.05	1339.3	23660.7	294.65	5205.35	0	Active
2291	2016 Q1	2016-01-07	2016-01-22	Payment	Daily	181.55	1815.5	28684.5	1488.11	23511.89	327.39	5172.61	0	Active
2291	2016 Q1	2016-01-07	2016-01-25	Payment	Daily	181.55	1997.05	28502.95	1636.93	23363.07	360.12	5139.88	1	Active
2291	2016 Q1	2016-01-07	2016-01-26	Payment	Daily	181.55	2178.6	28321.4	1785.74	23214.26	392.86	5107.14	0	Active
2291	2016 Q1	2016-01-07	2016-01-27	Payment	Daily	181.55	2360.15	28139.85	1934.55	23065.45	425.6	5074.4	0	Active
2291	2016 Q1	2016-01-07	2016-01-28	Payment	Daily	181.55	2541.7	27958.3	2083.36	22916.64	458.34	5041.66	0	Active
2291	2016 Q1	2016-01-07	2016-01-29	Payment	Daily	181.55	2723.25	27776.75	2232.17	22767.83	491.08	5008.92	0	Active
2291	2016 Q1	2016-01-07	2016-02-01	Payment	Daily	181.55	2904.8	27595.2	2380.98	22619.02	523.82	4976.18	0	Active
2291	2016 Q1	2016-01-07	2016-02-02	Payment	Daily	181.55	3086.35	27413.65	2529.8	22470.2	556.55	4943.45	0	Active
2291	2016 Q1	2016-01-07	2016-02-03	Payment	Daily	181.55	3267.9	27232.1	2678.61	22321.39	589.29	4910.71	0	Active
2291	2016 Q1	2016-01-07	2016-02-04	Payment	Daily	181.55	3449.45	27050.55	2827.42	22172.58	622.03	4877.97	0	Active

Fig. 2.2 Small business lending transaction data sample

portfolios at forecasting time, samples based on observed transaction paths in the book history can be drawn. In Sect. 2.4, we will discuss different methods for the transaction data sampling and compare their advantages and disadvantages.

For short contract term products, for example, loans less than a year, 1-year and lifetime loss forecasts are the same. A simplification of loss forecast for such products is just treating the entire portfolio as new origination and forecasting lifetime loss for all accounts in the portfolio at the forecasting time. For such short-term products, loss forecasting starting from origination will depend more on the account and loan characteristics, which we will describe in the next section.

2.1.2 Account and Product Characteristics Data

For credit products, data are collected at application and origination. These data can be classified into two categories – data related to the account profile and data related to the credit product. Together these data provide important information for underwriting.

Table 2.2 shows an example of components for these data in small business lending. For account profile, besides account ID, there are attributes covering the

Table 2.2 Small business lending origination data description

Field	Description
Account characteristics	
AccountID	Unique (internal) account identifier; each contract is corresponding to one account
Deal type	Identifies if customer has been funded previously (e.g., new, returning)
Lead source	Source of merchant acquisition
Sub-channel	Detailed marketing channel
Industry	Industry of business/operation
Sub-industry	Detailed industry description
State	US state/location
City	City in which business resides
Personal credit score	Reported personal credit score
Months in business	Time (months) since the business was incepted
Annual revenue	Annual revenue of business
Product characteristics	
Payment type	Method by which repayment occurs (e.g., ACH/credit card split)
Funded date	Date funds are received by customer
Expected term	Anticipated duration (in months) for customer repayments
Funded amount	Gross amount received by customer
Factor rate	Factor rate used to calculate repurchase amount
Initial receivable amount	Gross amount due from customer
Discount	Discount offered for early payoff
Adjusted receivable amount	Total amount received from early payoffs with an applied discount
Frequency	Repayment frequency (e.g., daily, weekly, bi-monthly, monthly)
Expected payment	Expected periodic payment amount
Expected no. of payments	Anticipated total number of payments to be made

account channel, account industry, geographic information, personal credit score, business life length, and revenue. Product characteristics cover the attributes related to the product, payment type and frequency, funding information (date and amount), contract term, factor and gross amount due, discount and adjusted receivable, expected periodic payment, and total number of payments. For small business lending, instead of coupon rate, a factor is popularly used to calculate the total gross amount due. For underwriting, these drivers will determine creditworthiness of an origination, and thus a decision of approval or rejection can be made. We will discuss creditworthiness and scorecard in Chap. 7.

2.1.3 Collateral Characteristics Data

For some credit products, for example, mortgage, the underlying property is the collateral. The characteristics of the collateral will affect the performance of the credit product. For example, if the property has high market price, the mortgage unlikely becomes default since the borrower has the option to sell the property. The following is an example of collateral characteristics for mortgage products:

- Property type (condo, single family)
- Property risk rating (H, M, L)
- Property state
- Occupancy status (owner, non-owner)

Property type and occupancy status are known at the origination and are popularly used in the underwriting and default models. Most of these collateral characteristics are also used in the LGD model to project the recoveries after default.

2.2 Cohort-Level Data

When account-level data are not available, cohort-level data, which consist of some aggregated accounts, may be collected. For example, mortgages based on vintage (origination time) may only collect aggregated data by vintages instead of by each account. In such cases, individual account performance is not recorded, but the charge-off/payoff is recorded. Loss rate or payoff rate will be the targets of performance monitoring during transactions. Besides the data availability, using cohort can reduce the data size significantly when data size becomes an issue for an application, and the aggregation in cohorts will not loss too much information with respect to the application, for example, mortgage prepayment modeling for portfolios with a large number of accounts.

Table 2.3 Mortgage cohort transaction data description

Field	Description
Cohort ID	Unique cohort identifier
Booking date	Starting transaction date observed on book
Transaction date	Date of transaction observed (end of the month)
Payment method	Monthly pay, pickup-a-pay, prepay penalty
Monthly scheduled payment	Total scheduled monthly payment
Interest payment	Monthly scheduled interest payment
Principal payment	Monthly scheduled principal payment
Monthly payment	Total payment received monthly
Monthly interest	Interest received monthly
Monthly principal	Principal received monthly
Cum principal payment	Cumulative principal payment at end of month
Unpaid principal balance (UPB)	Portion of funded amount outstanding at beginning of month
Scheduled UPB	Scheduled UPB at beginning of month
Weighted average loan age (WALA)	Scheduled UPB weighted average loan age
Weighted average coupon (WAC)	Scheduled UPB weighted average coupon
Interest and fees	Monthly unpaid interest and fees
Cum interest and fees	Cumulative unpaid interest and fees
Charge-off	Monthly charge-off
Prepayment	Monthly prepaid principal

2.2.1 Cohort Transaction Data

Cohort transaction data are aggregated based on individual accounts. So, some transaction attributes shown in Table 2.1 will become the cohort aggregated attributes, for example, monthly scheduled and actual payments are total payment for the cohort instead of for individual account. This logic applies similarly to the interest and principal payments for the mortgage cohort shown in Table 2.3. Unpaid principal balance (UPB) is calculated at the beginning of the month for the cohort, while charge-off and payoff are all calculated at the end of the month for the cohort. So, the rate of charge-off over UPB and the rate of prepayment over UPB are the main performance measures of the cohort. Scheduled UPB is the UPB at the beginning of the month by scheduled amortization. The weighted average loan age (WALA) and weighted average coupon (WAC) are all based on the scheduled UPB instead of observed UPB. These quantities are commonly used in the mortgage prepayment model. Note that the payment method attribute in Table 2.3 is for the entire cohort instead of individual account as in Table 2.1.

Table 2.4 Mortgage cohort origination data description

Field	Description
Cohort account characteristics	
Cohort ID	Unique (internal) cohort identifier
Channel	Marketing channel
State	US state/location
City	City in which business resides
Average FICO score	Cohort average FICO score
Average DTI	Cohort average debt to income ratio
Average PTI	Cohort average payment to income ratio
Cohort product characteristics	
Origination UPB	Cohort total origination loan amount
Origination WAC	Cohort UPB weighted average coupon at origination
Average coupon	Average coupon rate
Average loan size	Average loan size at origination
Average LTV	Average loan to value ratio at origination
Weighted LTV	Origination UPB/origination value

2.2.2 Cohort Mean Characteristics Data

Similar to the transaction data for cohort, the characteristics at origination are aggregated in average for the entire cohort. Some characteristics with the cohort are constant, and they are treated as cohort characteristics, for example, in Table 2.4, channel, state, and city are cohort characteristics with a unique constant value. Other account characteristics, for example, FICO, DTI (debt to income ratio), and PTI (payment to income ratio) are all averages of the cohort.

For product characteristics, origination UPB is the sum of all origination loan amount, origination WAC is the UPB weighted average coupon at origination, and weighted LTV is the house value weighted loan to value ratio at origination. The rest of the cohort product characteristics are simple arithmetic averages – average coupon, average loan size, and average LTV at origination. These attributes at origination should not be mixed up with similar attributes in the transaction data (e.g., WAC).

2.3 Segmentation and Grouping

Segmentation is an essential step in credit modeling. First, proper segmentation can reduce modeling complexity. A comprehensive segmentation analysis should dive into the risk generating process and fully understand the financial products and its risks. Only accounts with similar risk profiles should be grouped into a common segment. Although credit risks are defined to be caused by the default, path to default, and severity of losses due to default, even the difference in default definition among credit products that leads to nature segmentation and modeling based on such segmentation will significantly reduce modeling complexity. Second, proper segmentation can improve the model accuracy. Accounts in a segment sharing common risk profile will largely share the common risk factors, and selecting proper risk factors is the key to model accuracy.

Finally, due to the large scale of credit portfolios, especially with retail products, proper segmentation can scale down the modeling process in a divide-conquer approach and significantly improve the modeling efficiency.

In the following sections, we will describe some commonly used segmentation methods for various credit products. We focus on the ACL/loss forecast and CCAR model segmentation, since capital models have relatively simpler segmentation than these two suits of models.

2.3.1 Common Segmentation Methods

Segmentation for credit accounts of portfolios is commonly done in the following dimensions: business units and legal entities, line of business, product types, collateral types, and other account, product, and collateral characteristics if deemed proper. In the following, we present some commonly used segmentation for both retail and wholesale credit products.

Mortgage

Segmentation for residential mortgage modeling is commonly based on loan product types and collateral types. More granular segmentation can be carried out based on the property location and mortgage performance at forecast date, for example, current or delinquent. Table 2.5 shows a model segmentation topology of a firm's mortgage portfolio.

Table 2.5 Mortgage model segmentation

First/second lien	Target			
	D90	Default	Resolution	Prepayment
First lien	FL current to D90	FL current to default		FL 15-yr conforming, Alt-A, and jumbo
				FL 30-yr conforming jumbo
				FL 30-yr Alt-A
				FL 30-yr government
	FL delinquent to D90	FL delinquent to default		FL 3/1 hybrid
				FL 5/1 and 7/1 hybrid
				FL subprime
				FL pay option
Second lien	SL current to D90	SL current to default		SL fixed
	SL delinquent to D90	SL delinquent to default		SL HELOC
First and second lien			FL and SL default to resolution	FL and SL default to prepayment

By product type, the portfolio is first segmented by first lien and second lien, since different collateral privileges have impacts on all aspects of loan behaviors from early default warning (defined as 90 days delinquent or serious delinquent D90) to default (defined as 180 days delinquent), as well as early payoff (prepayment). Within the first lien, early warning and default are segmented based on the present loan status – current or delinquent – which certainly impacts the transition to these defined terminal targets. For prepayment, first lien has much more granular segmentation based on various deeper level product types – especially loan term and risk – which have been demonstrated essential in loan payoff speed. Besides default and prepayment, a post-default transition model is also built to estimate the default to resolution dynamics. This is usually called the second-stage transition model for loss forecast and can provide more accurate loss forecast in magnitude and timing given such post-default behavior observations. Once the loan is default, the post-default behavior is less relevant to lien status, so a combined model is built. We will discuss such models in more details in Chap. 3.

Credit Cards

Different from mortgages, credit cards modeling segmentation depends more on the charge-off types and payment behaviors. This is largely due to lack of collateral for the credit cards, and any account facing payment risk and delinquency in certain degree (defined as 180 days delinquent) will result in direct charge-off. Table 2.6 shows model segmentation topology of a firm's US credit card portfolio.

Based on the charge-off types, defaults are segmented into contractual, settlement, bankruptcy, fraud, and deceased. Within the contractual charge-off, defaults are further segmented by account types defined at the forecast date based on the account status and payment behaviors. “Closed,” “Delinquent,” and “Inactive” are account status at forecast date, while “Revolver” and “Transactor” are all current account with different payment behaviors (paid in full or by minimum). The “New” segment is defined as all account with less than 6 months from opening. A single payoff model applies to the entire portfolio.

Table 2.6 Credit card model segmentation

	Target	
Charge-off types	Default	Prepayment
Contractual charge-off	Closed	Payoff and account closure
	Delinquent	
	Inactive	
	New	
	Revolver	
	Transactor	
Settlement charge-off	Settlement	
Bankruptcy charge-off	Bankruptcy	
Fraud charge-off	Fraud	
Deceased charge-off	Deceased	

Table 2.7 Dealer financial services model segmentation

	Target		
Product types	Default	Prepayment	Recovery
Auto loans	Current	Current	Open
	Delinquent	Delinquent	Closed
Marine, motorcycle, and RV	Current	Current	Open
	Delinquent	Delinquent	Closed
Aircraft	Current	Current	Open
	Delinquent	Delinquent	Closed

Auto Loans and Other Dealer Financial Services

For auto loans and other dealer financial services (DFS) loans, which include marine, motorcycle, and recreation vehicles (RV) and aircraft, the definition of default is different from mortgage or credit(s). Instead of 180 days in delinquency, default of DFS is defined as 120 days in delinquency. The Basel-compliant Final Rule retail definition of default is:

Specifically, revolving retail exposures and residential mortgage exposures would be in default at 180 days past due; other retail exposures would be in default at 120 days past due. In addition, a retail exposure would be in default if the bank had taken a full or partial charge-off or write-down of principal on the exposure for credit-related reasons. Such an exposure would remain in default until the bank had reasonable assurance of repayment and performance for all contractual principal and interest payments on the exposure.[1]

Table 2.7 shows the modeling segmentation for DFS portfolio. It is relatively simple and only based on the loan status at the forecasting date.

Small Business

Some small business lending portfolios are similar to credit cards. For such portfolios, modeling segmentation is similar to that of credit cards. Small business lending portfolios consist of short- to medium-term loans instead of line of credits that can be segmented based on characteristics of clients and loan products, for example, channel, industry, business location, and business life length, as well as payment types and payment frequency.

Wholesale

For wholesale credit portfolios, besides line of business, obligor, and loan product characteristics, credit risk profiles are classified using either internal risk ratings or external risk ratings. These risk ratings are based on the concept of TTC (through-the-cycle) and are less sensitive to the economic dynamics, so it is preferred for modeling segmentation for wholesale credit portfolios.

[1] OCC, FRB, DTS, OTS, and FDIC. Risk-Based Capital Standards: Advanced Capital Adequacy Framework Basel II; Final Rule. Federal Register Vol. 72, No. 235 (December 7 2007), 69288–69336.

Table 2.8 Wholesale portfolio risk ratings group segmentation

Risk ratings group	Description
Large corporation – financial (LC-F)	Bank and issnsurance belong to similar industries and have low default rates
Large corporation – non-financial (LC-NF)	Remaining large corporations are a diverse set of industries with average to good credit quality. Retail is assigned to large corporation system due to industry similarity and high credit quality
Middle market (MM)	Middle markets are small in the number of obligors and average to good credit quality
Consumer and small business banking (CSBB)	CSBB has different customers and exposure amount than all other categories
Business banking (BB)	BB has high obligor count and high historical default rates
Individuals	Individual accounts use FICO scores for risk ratings
Real estate (RE)	Real estate has a differentiated client type and industry

Obligors in different industries and sub-industries use different risk rating systems. Table 2.8 shows a risk rating system group segmentation for one big bank's wholesale portfolio. Each obligor is mapped into one of these seven risk rating groups. For example, restaurant, agriculture, and healthcare are mapped into middle market; utilities, media, large public and private corporations, telecom and technology, and money managers are mapped into LC-NF.

Due to sample size limitation, wholesale credit modeling segmentation is more portfolio dependent. Large portfolios could have more granular segmentation, while small portfolios may suffer from limited sample size with granular segmentation. A balance between sample size limitation and segmentation granularity should always be considered for wholesale credit modeling.

2.3.2 Segmentation Stability

As part of the model validation, segmentation stability is one measurement of the model performance in the ongoing performance assessment (OPA) process. The segmentation stability is defined as homogeneity of risk profiles within segments. As for the credit risk modeling segmentation stability, it is expected that accounts should not be significantly different in default probability within the same segment. Besides some qualitative assessments, a homogeneity test can be used to assess the segmentation stability.

The homogeneity test is a two-sample t-test to test if the default probabilities from two populations are statistically the same or not. Assume there are d_1 defaults among n_1 accounts for the first population and there are d_2 defaults among n_2 accounts from the second population. The two-sample t-test tests the hypotheses:

$$H_0 : p_1 = p_2 \ \text{ vs } \ H_1 : p_1 \neq p_2 \tag{2.1}$$

where p_1, p_2 are the default probabilities for the two segments. Let $\widehat{p}_1 = \frac{d_1}{n_1}$ and $\widehat{p}_2 = \frac{d_2}{n_2}$, under the null hypothesis H_0, and assume equal variance, $T = \frac{\widehat{p_1 - p_2}}{\widehat{\sigma}(\widehat{p_1} - \widehat{p_2})}$, asymptotically a t-distribution with a degree of freedom $n_1 + n_2 - 2$, where:

$$\widehat{\sigma}(\widehat{p}_1 - \widehat{p}_2) = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \ \text{ and } \ s_p = \sqrt{\frac{n_1\,\widehat{p}_1(1-\widehat{p}_1) + n_2\,\widehat{p}_2(1-\widehat{p}_2)}{n_1 + n_2 - 2}} \tag{2.2}$$

Using this two-sample t-test, the segmentation stability test can be carried out with two steps:

- Merging test – Testing any two segments to see if they pass the homogeneity test; if they do, review the possibility of merging the two segments.
- Split test – Within each segment, check if a sub-segment has over 50% population; if it does, then test the homogeneity of this sub-segment with the rest of the population within this segment, and if the t-test fails, then review the possibility of splitting out this sub-segment.

Iteratively executing these two steps can achieve an initial segmentation, which could be further validated qualitatively. Once the segmentation is built, continuous checking with these two tests can monitor the segmentation stability.

The segmentation stability test described above can be extended to other risk modeling where the risk profile is characterized by a loss distribution. The two-sample test can be replaced by tests of the homogeneity of two distributions based on empirical distribution functions (EDF), for example, Kolmogorov-Smirnov (KS), Anderson-Darling (AD), Cramer-Von Mises (CVM), and other EDF tests.[2]

Cluster analysis as a data mining tool is also popularly used in segmentation analysis. It can base on multiple factors to group population without a target measurement, and thus its interpretation could be difficult. So, it is often used as an exploratory tool for initial segmentation analysis for populations with similar characteristics, for example, the initial segmentation for a mortgage portfolio based on the estimated TTC (through-the-cycle) 1-year default probability in the internal ratings-based (IRB) capital process. Such initial segments are refined based on additional qualitative and quantitative rules to form the final IRB grids of the portfolio. The portfolio populations are mapped into these grids based on their estimated TTC 1-year default probability at a snapshot. Population migration among grids within the portfolio over a series of snapshots is closely monitored to check the segmentation stability for the IRB capital process.

[2] Stephens, M. A. (1986). Tests based on EDF statistics, in Goodness-of-fit Techniques (eds, R. B. D'Agostino and M. A. Stephens), Ch. 4. New York: Marcel Dekker.

2.4 Account Data Sampling

In credit modeling, data sampling is essential, especially for retail portfolios, which usually have a large number of accounts and records of transactions. In summary, there are three main advantages for data sampling in credit modeling. First, data sampling can improve the computing efficiency in modeling process without loss of much accuracy of modeling results. Computing power has increased exponentially in the recent decade and enables a "big data" era. However, even with today's computer power, some credit modeling running over multimillion data records with expensive optimization processes and a large number of repeated simulations is still a bottleneck, and efficient modeling requires reducing "non-informative" processing, for example, some of the maximum likelihood estimation (MLE) processing. Second, proper data sampling is required by some statistical methods and machine learning algorithms in credit modeling. For example, credit events like default are much less frequent in the transaction data. Bootstrapping the original transaction data without proper stratified sampling could significantly underestimate the default likelihood and cause the failure of some statistical estimates.

Lastly, if the available data collected are known unbalanced or unrepresentative with respect to the entire population, oversampling or undersampling techniques are needed to form the proper modeling data. Although this happens less with credit modeling since data for the full portfolio are usually collected, drawing conclusions with less representative characteristics of factors could be risky if sampling was not done properly.

2.4.1 Population Representation and Sampling

Before doing any sampling, one should understand on what population the model or analysis is going to make references. Without the target population in mind, modeling results could be misleading. For example, using oversampling the default loans and undersampling the non-default ones in a mortgage portfolio, one can prepare a modeling data set with defaults and non-defaults splitting in half. Such modeling data set clearly deviates from the original portfolio population, and models built on it without proper stratified weights will be biased toward heavy default penalty. Such models will produce downward biased results if applied to the original portfolio.

One exception is that the data collected do not represent the target population on which to make references. Under such situation, sampling techniques may be used to recover the target population. Models then can be built on newly constructed sample data set as a representation of the target population. However, such assumption of the constructed sample representing the target population needs to be checked carefully.

In credit performance modeling, data are commonly collected for the full portfolio. Models are built on the target portfolio and are used to track and project the portfolio performance. However, in credit underwriting, the in-house portfolio population may be different from the application population. Using in-house population to approximate the application population needs careful review including the homogeneity testing with respect to underwriting risk factors. A simple oversampling the defaults (or undersampling the non-defaults) to form some synthetic data is not sufficient and may lead to significant biases.

In the following, we will introduce several sampling methods that are commonly used to form modeling data to approximate the target population.

2.4.2 Snapshot Sampling

Snapshot sampling is commonly used in building loss forecast models. Forecasting performance for all loans/accounts of a portfolio on the book at a forecasting date is taking a snapshot of the portfolio book at the forecasting date to form a target population. Given the dynamics of a portfolio, some loans/accounts existing on early snapshots could be non-existing at later snapshots due to a credit event (default or prepay). To approximate the target population at a forecast date, snapshots at historical book dates are used to form a pool of loan and observation date combinations. In the following, we show an example of such approximation with mortgage loans.

Let the cross section of loans be represented as $i = 1, \ldots, N$. Simply put, there are N loans available in the book history to build a model. Each of those loans is observed monthly over time and may or may not be still active today. Let loan age be indexed naturally as $0, 1, \ldots, t, \ldots, T_i$ where t is an arbitrary loan age and T_i is the total number of months loan i was observed and t_{ij} is the i th loan age observed at time j.

Let $\omega(i, t)$ represent all information known about loan i at the beginning of age t. The population of all loan-month information sets observed from time $T = 0$ to time $T = J$ can be represented as an unbalanced matrix:

$$
\Omega = \left(\begin{array}{|c|c|c|c|c|c|c|}
\hline
 & \cdots & \omega(1, t_{1j}) & \cdots & \cdots & & \omega(1, t_{1J}) \\
\hline
\vdots & & \vdots & & & & \\
\hline
\omega(i, t_{i0}) & \cdots & \omega(i, t_{ij}) & \cdots & \cdots & \cdots & \omega(i, t_{iJ}) \\
\hline
\vdots & & \vdots & & & & \\
\hline
\omega(N, t_{N0}) & \cdots & \omega(N, t_{Nj}) & \cdots & \cdots & \omega(N, T_N)) & \\
\hline
\end{array}\right)
$$

$$
\xrightarrow{\quad T = 0 \qquad\qquad T = j \qquad\qquad \ldots \qquad\qquad T = J \quad}
$$

To approximate the portfolio population at $T = J + 1$, loans observed at $T = 0, \ldots, J$ with their observed characteristics can be pooled to form the model data.

Such approximation is based on the assumption that the portfolio risk profile is relatively stable without any sudden changes. The portfolio risk profile consists of the structure of individual loan risk profiles. So, the assumption requires that the structure of individual loan risk profiles should be stable. To mitigate the model risk with this assumption due to the portfolio dynamics with credit events, one can use relatively recent snapshots to form the model data.

However, for regulatory capital modeling, the concept of through-the-cycle (TTC) requires model reference data to cover several economic cycles, and using only recent snapshots would not meet this requirement. So, the reference data for regulatory capital modeling take in all snapshots (usually quarterly), and loans at different snapshots are assumed conditionally independent even with the same loan (a critical assumption in models using snapshot sampling). These loans at different snapshots are grouped into grids based on their estimated TTC probability of defaults (TTC PDs), which defines the portfolio risk profile. Then, at a specific snapshot date, loans are mapped into these grids based on their estimated TTC PDs. Unexpected credit losses are projected based on the grid observed PDs and assumed correlation for different products (e.g., mortgage, credit cards) and are summed up by grids. To monitor the stability of the portfolio risk profile, migration of loans among grids between continuous snapshots is observed by the average observed PDs of the mapped loans in each grid. Crossing of these observed average PDs over grids indicates stability risk, and less granular grids may be used to mitigate this risk, which more likely increases the capital.

Although snapshot sampling has a critical issue – repeatedly using the same loan at different snapshots could introduce correlations and violate the independence assumption in some models, especially models using panel data – in practice, model risk related to this assumption is low as weak conditional independence can be achieved with sufficient risk factors. In addition, due to its simplicity in concept and implementation, snapshot sampling is popularly used in loss forecast and capital models.

2.4.3 Observation Sampling

Instead of snapshotting the portfolio book history, observation sampling (OS) is based on all observations in the portfolio book history, which could cut from some portfolio booking date. If the portfolio book history covers all loans/accounts transaction history from origination, this data set is usually called Complete Panel Data from Origination (CPDO). An origination data (OD), which only takes the loans/accounts information at the origination date plus some performance measurements, can be extracted from the CPDO. The origination data is also a cross-sectional data (CSD), which is defined as data that only include loans/accounts information at a single date point in lieu of transactional information in the panel data. If the complete panel data only cover transaction information from a booking date for some accounts as we have shown in Sect. 2.4.2 for the mortgage portfolio, we call such panel data complete panel data from booking (CPDB). In credit modeling, CPDB is the most common panel data due to most portfolios that include purchased loans without early transaction information from origination.

For cross-sectional data (CSD), observation sampling is on the loans/accounts level and is usually based on either the performance target (e.g., default indicator) or risk factors (e.g., product types) or combined. Since the original data represent the model population, stratified sampling should be used to avoid bias. A standard application is building underwriting models on the origination data (OD), where credit events (defaults) are oversampled and non-events are undersampled with stratified weights. There is an argument on whether the in-house portfolio population can represent the application population. One answer could be that the unbiased model built on the in-house portfolio population can be extrapolated to the application population assuming that the underwriting process does no cherry-picking. An alternative is expanding the in-house data with external data to better approximate the application population. However, simply modifying the stratified weights or not using the stratified weights is a risky practice due to difficulty to justify, although a single risk multiplier based on management judgment can be used. We will describe more details in Chap. 7 for credit scoring and underwriting.

For panel data (CPDO or CPDB), observation sampling is based on how observations are extracted from the observed transaction history. A full exploded observation sampling takes every observation in the loan/account transaction history and produces a quadratic number of records for the full panel (called exploded panel or EP in some modeling practices) data. Referring to the snapshot sampling, this is equivalent to the full snapshot at every booking date of the portfolio. However, the difference from the snapshot sampling is that instead of taking the vertical snapshot of the portfolio, observations are tracked on the horizontal transaction level. The full exploded observation sample could be assumed an approximation to the portfolio population at a forecast date. However, due to its explosive sample size and frequent repeated use of observations on the same loan, it is not recommended to directly use the EP data in loss forecast models. Some downsize sampling techniques should be used. In the next section, we introduce one sound method – full observation stratified sampling (or FOSS) – which can both downsize the EP and avoid correlations caused due to the frequent repeated use of same loan observations.

2.4.4 Full Observation Stratified Sampling (FOSS)

The full observation stratified sampling (FOSS) method is a stratified sampling technique applied on the panel data. It first segments the full observations of each loan's transaction history based on loan performance or other measures of the observations and defines the eligible sets for each segment. Then, for each segment, stratified sampling is applied on the eligible sets of each loan. The obtained stratified samples can be further downsized by stratified sampling on the cross-section loan level.

As an extension of the mortgage panel data in Sect. 2.4.2, let the cross section of loans be represented as $i = 1, \ldots, N$ and each of those loans is observed periodically (e.g., monthly or quarterly) over time and may or may not be still active today. Let loan age be indexed naturally as $0, 1, \ldots, t, \ldots, T_i$ where t is an arbitrary loan age and

T_i is the total number of periods loan i was observed and t_{ij} is the i th loan age observed at period j.

Again, let $\omega(i, t)$ represent all information known about loan i at the beginning of age t. The population of all loan information sets observed from origination can be represented as an unbalanced matrix:

$$\Omega = \left(\begin{array}{|c|c|c|c|c|c|c|}
\hline
\omega(1,0) & \cdots & \omega(1,t) & \cdots & \omega(1, T_1) & & \\
\hline
\vdots & & \vdots & & & & \\
\hline
\omega(i,0) & \cdots & \omega(i,t) & \cdots & \cdots & \cdots & \omega(i,T_i) \\
\hline
\vdots & & \vdots & & & & \\
\hline
\omega(N,0) & \cdots & \omega(N,t) & \cdots & \cdots & \omega(N,T_N)) & \\
\hline
\end{array}\right)$$

where a row represents the periodic time series data for a given loan and a column represents data across loans at a given age t. Note that, for notation simplicity, we do not align the observations with the calendar time anymore, though that is needed when merging with other data, especially the macroeconomic data.

More generally, on the portfolio book, loans could be observed at its any age point, and the general unbalanced matrix could be:

$$\Omega` = \left(\begin{array}{|c|c|c|c|c|c|c|}
\hline
\omega(1,t_{10}) & \cdots & \omega(1,t_{1t}) & \cdots & \omega(1, t_{1T_1}) & & \\
\hline
\vdots & & \vdots & & & & \\
\hline
\omega(i,t_{i0}) & \cdots & \omega(i,t_{it}) & \cdots & \cdots & \cdots & \omega(i,t_{iT_i}) \\
\hline
\vdots & & \vdots & & & & \\
\hline
\omega(N,t_{N0}) & \cdots & \omega(N,,t_{Nt}) & \cdots & \cdots & \omega(N,t_{NT_N}) & \\
\hline
\end{array}\right)$$

Segments can be decided by loan performance, like delinquency status (including "New," "Current," "30DPD," "60DPD," "90DPD," etc.). Denote these loan status segments as $d = 1, \ldots, 5$. On the transaction history, any given loan could have the same instance at different observations, especially "Current." We can extract loan observations with the same status.

Let Ω^d be all loan observations of Ω having loan status d. Trivially, $\Omega = \cup_{\mathrm{d}} \Omega^d$. A representative Ω^d can be depicted as:

$$\Omega^{d} = \left(\begin{array}{|c|c|c|c|c|c|c|}
\hline
 & \omega(1,\underline{t}_1) & \cdots & & \omega(1,\bar{t}_1) & & \\
\hline
 & & \vdots & & \vdots & & \\
\hline
\omega(j,\underline{t}_j) & \cdots & \cdots & \omega(j,t) & \cdots & \cdots & \omega(j,\bar{t}_j) \\
\hline
 & \vdots & \vdots & & \vdots & & \\
\hline
 & & \omega(J,\underline{t}_J) & \cdots & \cdots & \omega(J,\bar{t}_J) & \\
\hline
\end{array}\right)$$

where $j = 1, \ldots, J$ loans have at least one instance of the d status, $\underline{t_j}$ is the earliest occurrence of a d status for loan j, and $\overline{t_j}$ is the latest occurrence of a d status for loan j. As illustrated, there is no systematic timing relationship among the $\{\underline{t_j}\}$ across loans or among the $\left\{\overline{t_j}\right\}$ across loans, though $\underline{t_j} \leq \overline{t_j}$.

As panel data, each loan information set $\omega(j,t)$ can be treated as a hypothetical forecast date in history. Time goes forward from each fixed forecast date. Some variables, such as loan age and macroeconomic variables, can be updated after the fixed snapshot date, whereas other variables remain static. Denote these spawned transaction sets as $\omega(j,t + k;t)$ for $k \geq 0$ representing a panel set at age $t + k$ spawned from the root loan observation $\omega(j,t)$, where for notational symmetry $\omega(j,t;t) = \omega(j, t)$. Denote the spawned panel set from the root observation $\omega(j,t)$ as Ψ where:

$$\Psi(\omega(j,t)) = \left(\begin{array}{|c|}
\hline
\omega(j,t;t) \\
\hline
\omega(j,t+1;t) \\
\hline
\omega(j,t+2;t) \\
\hline
\vdots \\
\hline
\omega(j,T_j;t) \\
\hline
\end{array}\right)$$

To avoid using a loan outcome observed at its end date T_j more than once, we can randomly pick exactly one forecast date τ_j for each loan j represented as a row of Ω^d. That is, for each loan $j \in \{1, \ldots, J\}$, draw an observation as forecast date $\tau_j \in \left\{\underline{t_j}, \ldots, \overline{t_j}\right\}$ with probability:

$$1/ \mid \Omega^{d}(j, .) \mid \tag{2.3}$$

where $|v|$ is the cardinality of v.

Across all loans, the chosen dates τ_j spawn a panel data set, which can be depicted as (suppressing the extraneous loan index j from $\omega(j,\bullet)$):

$$
\Psi^{\mathrm{d}}(\tau) = \begin{pmatrix}
 & \omega(\tau_1;\tau_1) & \ldots & \omega(T_1;\tau_1) & & \\
 & \vdots & & \vdots & & \\
\omega(\tau_j;\tau_j) & \ldots & \ldots & \ldots & \ldots & \omega(T_j;\tau_j) \\
\vdots & & & & & \vdots \\
 & & \omega(\tau_J;\tau_J) & \ldots & \omega(T_J;\tau_J) &
\end{pmatrix}
$$

where the rows of this matrix are simply the transpose of $\Psi(\omega(j,t))$. It follows that the weights w_t to use in model estimation for each spawned panel observations are the reciprocal of the draw probability for the row:

$$w_t\left[\omega\left(\tau_j + k;\tau_j\right)\right] = \mid \Omega^d(j,.) \mid, \text{for all } k = 0, \ldots, T_j - \tau_j. \tag{2.4}$$

We can repeat the sampling and weighting exercise for each separately modeled status d. A generalization to sample fewer than J loans is useful when the cardinality of Ω^d is large, as will often be the case for current loans $d = 1$. Suppose the conditional probability of focus is the target default. Partition $\{1, \ldots, J\}$ into loans that defaulted or not so that loans $j' \in \{1, \ldots, J'\}$ defaulted and loans $j^* \in \{1, \ldots, J^*\}$ didn't default. Partition the rows of Ω^d into $\Omega^d(J')$ and $\Omega^d(J^*)$ with the obvious meaning. Among non-defaults, draw a sample size of $n \leq J^*$ picking each loan j^* with probability:

$$n* \mid \Omega^d(j^*,.) \mid / \mid \Omega^d(J^*) \mid \tag{2.5}$$

instead of n/J^* as under uniform random sampling. This design better represents snapshot portfolios through calendar time. Choose one random starting forecast date τ_j as before for J' defaults and the n non-defaults. For the defaults $j' \in \{1, \ldots, J'\}$, the panel data weights to be used in model estimation are:

$$w_t\left[\omega\left(\tau_{j'} + k;\tau_{j'}\right)\right] = \left| \Omega^d(j',.)\right|, \text{for all } k = 0, \ldots, T_j - \tau_j. \tag{2.6}$$

Now consider the sampled non-defaults $j^* \in \{1, \ldots, n\} \subseteq \{1, \ldots, J^*\}$. Under our chosen sample rules for picking the rows and columns of $\Omega^d(J^*)$, the probability of drawing any observation as the forecast date equals:

$$n* \mid \Omega^d(j^*,.) \mid / \mid \Omega^d(J^*) \mid *1/ \mid \Omega^d(j^*,.) \mid \tag{2.7}$$

which simplifies the reciprocal panel data weights to:

$$w_t\left[\omega\left(\tau_{j*} + k;\tau_{j*}\right)\right] = \left| \Omega^d(J^*)\right|/n, \quad \text{for all } j^*, k = 0, \ldots, T_{j*} - \tau_{j*}. \tag{2.8}$$

After pooling all loan-months generated from both defaults and non-defaults, it is appropriate to normalize the loan-month weights so that the sum of normalized weights equals the number of loan-months used in the model estimation.

If one also would sample down the defaults, the same method applies as with the non-defaults described above. FOSS implementation can be found in Chap. 4.

2.4.5 Synthetic Minority Oversampling Techniques (SMOTE)

Different from the sampling methods we described above, the Synthetic Minority Oversampling Techniques[3] (SMOTE) creates synthetic observations from the observed samples. SMOTE first identifies the minority classes (e.g., default in credit modeling). Then, based on the observed minority class samples, SMOTE creates more synthetic examples of the class rather than by oversampling with replacement. There are different methods to create such synthetic examples. SMOTE generates synthetic examples in a less application-specific manner, by operating in "feature space" rather than "data space." The minority class is oversampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbors. Depending upon the amount of oversampling required, neighbors from the k nearest neighbors are randomly chosen.

The SMOTE implementation is straightforward. If we set the number of nearest neighbors (or synthetic sampling directions) $k = 5$ and the amount of oversampling needed is 200%, only two neighbors from the five nearest neighbors are chosen, and one sample is generated in the direction of each. Synthetic samples are generated in the following way: take the difference between the feature vector (sample) under consideration and its nearest neighbor. Multiply this difference by a random number between 0 and 1, and add it to the feature vector under consideration. This causes the selection of a random point along the line segment between two specific features. This approach effectively forces the decision region of the minority class to become more general.

SMOTE has been proved efficient in machine learning applications by increasing the accuracy of classification and model fitting. In credit modeling, when observed data samples do not cover some segments, e.g., in underwriting, SMOTE can make up these missing segments and improve the accuracy of model fitting. However, some cautions should be mentioned in applying SMOTE in credit modeling. First of all, SMOTE adds synthetic loans/accounts, so it is important to check such synthetic loans/accounts are potentially feasible. While SMOTE can be used for feature spaces with both continuous and nominal factors, their combinations with non-monomial factors after SMOTE should be checked for feasibility. Second, SMOTE changes the population structure by adding more synthetic loans/accounts. To make the sample

[3] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP: SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research 16 (2002) 321–357.

population after SMOTE comparable to the target population, proper weights should be adopted based on the SMOTE oversampling parameters. Without such weights, model fits will be biased. More details can be found in Chap. 7.

2.5 Macroeconomic Data

Macroeconomic factors are essential in credit modeling due to their impact to the credit performance. While there are a large number of macroeconomic factors, we focus on the ones included in the CCAR practice due to coverage, systemic classification, and processing of the data by regulators and agencies (CCAR pool). More importantly, scenario forecasts on these macroeconomic factors are provided annually by regulators though the forecasts are hypothetic. Still, these hypothetic forecasts can be used as reference when "real" forecasts are required in practical applications, for example, CECL. There are alternatives for creating macroeconomic scenarios as we described in Sect. 1.4, and the CCAR pool could also be enriched by individual institutions when developing their own macroeconomic scenarios.

2.5.1 MEV Classification

In CCAR, FRB classifies the domestic macroeconomic variables (MEV) into three classes – economic activity, financial conditions, and interest rates.

- *Measures of economic activity (six variables)*: quarterly percent changes (at an annual rate) in real and nominal gross domestic product (GDP), real and nominal disposable personal income (DPI), the Consumer Price Index (CPI) for all urban consumers, and the level of the unemployment rate of the civilian non-institutional population aged 16 years and over
- *Measures of financial conditions (four variables)*: indexes of house prices, commercial real estate prices, equity prices, and stock market volatility
- *Measures of interest rates (six variables)*: the rate on 3-month Treasury securities; the yield for 5-year Treasury securities; the yield for 10-year Treasury securities; the yield for 10-year BBB corporate securities; the interest rate associated with conforming, conventional, 30-year fixed-rate mortgages; and the prime rate

The variables describing international economic conditions in each scenario include three variables in four countries or country blocs:

- *The three variables for each country or country bloc*: quarterly percent changes (at an annual rate) in real GDP and in consumer price indexes or local equivalent and the level of the US dollar exchange rate
- *Four countries or country blocs*: the euro area (the 19 European Union member states that have adopted the euro as their common currency); the United

Kingdom; developing Asia (the nominal GDP-weighted aggregate of China, India, South Korea, Hong Kong Special Administrative Region, and Taiwan); and Japan

While these MEVs do not cover the full pack of macroeconomic variables, they are considered as representatives of each category in CCAR modeling.

2.5.2 MEV Forecasting

A key component of MEV data is the MEV forecasting data. Since there are different purposes for MEV forecasting in different applications, there is no unique way to generate MEV forecasting data. In the following, we describe MEV forecasting data generation for the three credit loss forecasting applications – CECL, CCAR, and capital management.

MEV Forecasting with CECL

CECL requires estimate of full lifetime expected loss based on relevant information about past events and historical experience, current conditions, and reasonable and supportable forecasts that affect the collectability of the reported amount as described in Sect. 1.2.1. While past events and current conditions are straightforward, the "reasonable and supportable" forecasts affecting collectability of credit products are a little wide. Model-based approaches (either loss rate models or PD/LGD/EAD models) narrow such "reasonable and supportable" factors to MEVs and largely assume other factors as fixed or having known variations in the future. So, in the following, we focus on how to create reasonable and supportable MEV forecasting data in CECL.

There are two methods to create MEV forecasts, which are recommended by regulators. The first method is the mean reversion method. It assumes MEVs can have reasonable and supportable forecasts only over a sub-period from the forecasting date instead of the lifetime of the credit product, and beyond that period, forecasts are reverted to historical averages as suggested in ASU 2016-13. Figure 2.3 presents an illustrative example of such a reversion for an MEV forecasting.

MEVs have a strong time series style. So traditionally, time series models such as autoregression (AR) and error correction models based on historical data have been used for forecasting. These models have demonstrated forecasting sufficiency for relatively short periods and could fail for long term due to bias convolution; thus, other information including management judgments are often added for long-term forecasting. However, such extra information could be subject to uncertainty due to the long term. So very often, businesses are able to build models and carry out forecasts for these MEVs for a specified period of time (e.g., 1–5 years) as shown in Fig. 2.3 with the three dashed color lines – blue for the baseline, red for a more severe scenario (S2), and green for a less severe scenario (S1). One can notice that, for the baseline scenario, the MEV model forecast has been reverted to the historical average at the end of the forecasting period. This is often a recommended design of

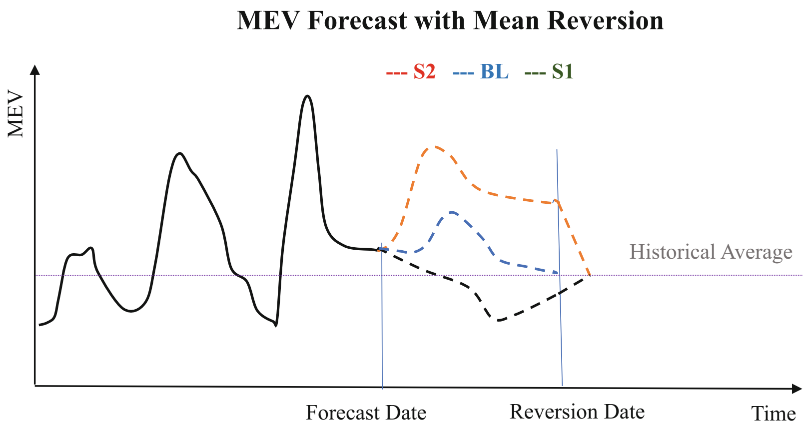


Fig. 2.3 CECL MEV forecast

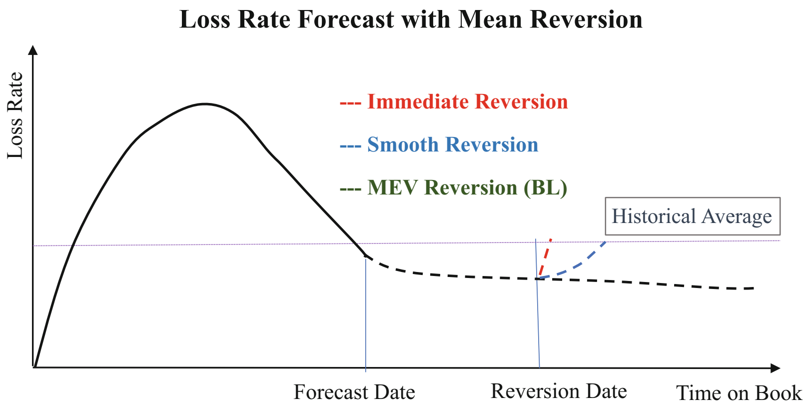


Fig. 2.4 CECL loss rate forecast

the model for a smoothing transition to the historical average. For the other two scenarios (S1 and S2), the smoothing transition to the historical average is delayed. Such design is certainly more "reasonable and supportable" than the sudden jumping to the historical averages simultaneously.

MEV mean reversion is considered as input reversion for loss forecast models. It requires the model fits the MEV sufficiently well in the sense that extreme values outside the popular range of MEV values should have stable impacts to the model. One good example is the mortgage prepayment models with the primary mortgage rate being revered to the long-term average from the secondary market rate. The S-curve used in the prepayment models for the rate incentive controls the rate extreme variations well. For models do not control such extreme variations well, the input reversion is not recommended. Instead, one can consider the output reversion – directly revert the loss rate (or PD). Figure 2.4 presents a mean reversion for the loss rate.

To stabilize the loss forecast results, instead of using the MEV mean reversion beyond the forecasting period (being reasonable and supportable), one can directly revert the loss forecast model results to the historical averages. The output reversion ignores the MEV impact beyond the reversion date, as well as other risk factors. So, unless the portfolio runs out beyond the reversion date, otherwise any significant risk factor impact beyond reversion date should be reviewed before adopting the output reversion.

Besides the mean reversion method, the second method for MEV forecasting is based on modeling the joint distribution of all MEVs at each future time without any mean reversion. As described in Sect. 1.4.1, at each forecast time t, the MEV conditional distribution is approximated by the simulated paths $\widehat{x}_{i,t} + \widehat{\epsilon}_{i,t,g}, g = 1, \ldots, G,$ where $\widehat{x}_{i,t} = \widehat{a}_{\mathrm{i}} + \widehat{b}_{i,1}\widehat{x}_{i,t-1} + \widehat{b}_{i,2}\widehat{x}_{i,t-2} + \ldots + \widehat{b}_{i,p}\widehat{x}_{i,t-p}$ and $\widehat{\epsilon}_{i,t,g} \sim C_p(\mathrm{df}, \boldsymbol{\Sigma}_t)$, a p-dimension t-copula with degree of freedom df and correlation $\boldsymbol{\Sigma}_t$. Since $\widehat{x}_{i,t}$ is assumed a stable (or trend stable) time series, the idiosyncratic variation from $\widehat{\epsilon}_{i,t,g}$ could be significant. Various MEV forecasts could be generated from these conditional distributions, such as the median, quantiles, and some weighted averages. From these forecasts, some reasonable and supportable MEV forecasts could be selected. Such forecasts have the attraction of continuity and could be reasonable and supportable for some economic trends if such trends are deemed acceptable in the future.

One of such reasonable and supportable MEV forecasts could be a forward-looking through-the-cycle (TTC) estimate. Different from the regulatory capital TTC, the forward-looking TTC incorporates the historical cycles with current forward-looking. A solid justification of the forward-looking, such as the lifetime MEV PIT (point in time) conditional distribution we described above, should be in place first, and then some weighted averages can be taken as the forward-looking TTC based on these forecasts.

MEV Scenarios with CCAR

For CCAR, as described in Sect. 1.4.1, FRB creates and distributes the MEV scenario data each year in February to CCAR participants. As pointed out by the regulators, the MEV scenario data should not be treated as forecasting, but three scenarios were designed for the purpose of stress testing (baseline, adverse, and severely adverse). The baseline scenario MEVs are created based on available consensus views of the macroeconomic outlook by professional forecasters, government agencies, and other public sector organizations as of the beginning of the annual stress test cycle. The severely adverse scenario MEVs are created based on the economic and financial conditions that reflect the conditions of post-war US recessions. The adverse scenario MEVs are somehow between the baseline and the severely adverse scenarios.[4] FRB only provides scenario data for the 28 MEVs listed in Sect. 2.5.1 and provide direction and intensity for other MEV variables which

[4] 12 C.F.R. Appendix A to Part 252 – Policy Statement on the Scenario Design Framework for Stress Testing. Feb 28, 2019.

participant institutes can develop for their own uses. FRB also provides the historical values for these 28 MEVs in its distributions.

The 28 MEVs in the FRB distributions can be the basis for developing other MEVs. Various methods can be used. For model-based methods, based on historical data, one can use simple linear regression or more fine-tuned autoregression models with exogenous variables selected from the given MEVs. Together, the MEVs from both the regulator distributions and the proprietary MEV scenario generating processes form a rich pool of MEV scenario data.

MEV Forecasting with Capital Measurements

For regulatory capital management, the main purpose is to rank the accounts within the portfolio according to risk profiles. So, MEV impacts are averaged out from the risk profiles by using the through-the-cycle (TTC) MEV estimates. The 1-year-ahead PD is estimated based on the current observations, and MEV forecasts are not needed. All reference data for PD, LGD, and EAD are observed historical data.

For economic capital management, as described in Sect. 1.3.2, conditional EC is based on PIT risk parameters estimated at the current observed MEVs, while marginal EC requires full distributions of MEVs, which can be approximated by empirical MEV distributions based on all observed values of MEVs.

2.6 Data Processing, Automation, and Data Models

Data are essential inputs for models. Data processing is the first step for any modeling process and usually takes over half of the total resources in the modeling project. So, it is critical to have the data and process architectures well designed before building a model. In the previous sections of this chapter, we present the data architectures and some data processes for credit risk models, especially for the three credit risk model frameworks, ACL, capital, and CCAR. In this section, we make an effort to connect these components by adding more process components to form a more connected process architecture, which we define as data model for credit risk.

Figure 2.5 summarizes the credit risk data model architecture.

Our data model architecture starts from planning and searching of all available data sources. As most of the data sources are created for some specific purposes by line of businesses, it is not necessary that such data sources are created with the intention of modeling. So, while a targeted searching of data sources within the origination is helpful, more investigation and studies should be built on the communication with line of businesses and related parties. Be prepared that data sources could be in various formats and types. Data in large volumes, like the retail transactions, are likely stored in a well-maintained database. For such data, standard extraction, transferring, and loading (ETL) tools (e.g., SQL for various relational databases) are efficient to process data once data standards are clearly defined. Special distributed data are likely in spreadsheet or text format. For such data, proprietary tools need to be developed. To reduce the workload of processing such data, some common standards should be defined and mutually agreed upon if the

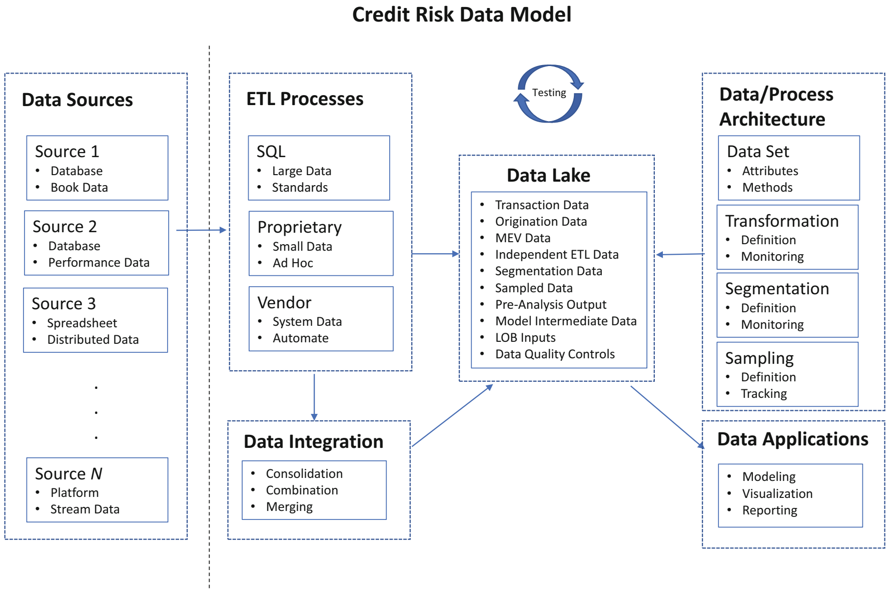


Fig. 2.5 Data model

data distribution will be repeated. With the development of smart data collection using artificial intelligence (AI), data generated from automated systems will become more popular. Such data are likely stored in certain platforms and can be smoothly transferred between systems. While building proprietary tools to process data distributed in some basic formats may not be totally eliminated in credit risk data model, such processes should be automated as much as possible to increase efficiency and control errors.

Once we have all the sources of data loaded, we can build a data lake for credit risk analysis and modeling. In the following sections, we discuss some of these processes to connect separate data processes and automate the data model architecture.

2.6.1 Data Integration

The data integration in our data model is the following-up step once data are all loaded from different sources. To be simple, we define the data integration as the procedures of consolidation, combination, and merging. These procedures should be executed in consulting with the data and process architectures in the data model design.

Data consolidation is the procedure of data quality check and reconciliation. With different ETL tools and different operating systems, data attribute types and formats may not be consistent. For example, spreadsheet data may change data attribute types during data transferring between different operating systems. So, it is critical to check the consistency of data attributes. If such consistency can't be controlled during the transferring and loading, then data type conversion should be physically done during data consolidation. When data from different sources have overlaps, data reconciliation should be executed with reasonable logics and be documented.

Data combination is a data integration procedure that simply joins two or more individual data sets/files or some specific data structures into one. It is often used to purposely split data blocks for large data files with the assumption that these data blocks are homogeneous. However, it is safe to check the data block homogeneity before simple combinations. Combination for heterogeneous data blocks should be limited for specific purposes (e.g., linked storages).

Data merging is a more demanded data integration procedure. It has higher requirements for the data sets/files to be merged. Data merging needs aligned common key (or keys) among the merged data. Although multiple keys can be used with data merging, it is recommended to use single key with multiple merging steps to better track the merging process, especially when the common key is not matched well among the merged files and missing values may need some special treatments.

2.6.2 Data Transformation

Data transformation converts the original data or attributes into new data or attributes or adds new data or attributes based on the original ones. There are various kinds of data transformation. For credit risk data, we only focus on the data records and attributes. So, we classify data transformation as records transformation and attributes transformation.

For credit risk data, a popular data records transformation is expanding the book observation data into panel data. Very often book records only include credit events (default, delinquency, payoff, and resolution) instead of the full transaction history for saving of storage space. To expand such concise data into full transaction panel data for credit modeling and analysis, periodic records need to be added. This is the same as the transformation of time-to-event data to panel data.

For data attributes transformation, there are several types of transformation based on the characteristics of the attributes and purposes of analysis. In the following, we discuss a list of popular data attributes transformation in credit analysis and modeling:

Lags

Lags are commonly used on macroeconomic and financial risk factors in credit risk modeling, since these risk factors have delayed effects on credit risk events. For example, market interest rate always has a lagged effect on borrower's refinance decisions. Both periodic and cumulative lags can be used.

Functions

Function transformation, for example, the logarithm transformation on positive attributes, is used to explore the best functional relationships between the target and risk factors. More complex function transformation like the spline and other nonlinear functions are also often explored. It is well known that in mortgage prepayment models, the relationship between interest rate incentive and prepayment probability is like S-curve. Interaction between risk factors can be considered as more complex function transformation, especially between risk factors in different types, for example, continuous and categorical risk factors. Software is commonly needed for handling complex interactions.

Types

Data attribute's type may be transferred. For example, text data can be decomposed into multiple binary attributes based on the Bag of Words (BOW) by marking if the word in the BOW shows in the text content. In this way, we can transfer the text attribute to a bag of binary attributes.

The most popular attribute type transformation is converting risk factors from continuous to categorical as in the underwriting and credit score models. This is the bucketing process. For credit risk models with binary or multinomial targets, the bucketing process or the more general binning process can be applied on both continuous and categorical risk factors, which we describe in the next section.

2.6.3 Data Binning and WoE

Binning can be applied on both continuous and categorical attributes. The idea of binning is grouping accounts with close attributes into a single risk profile for risk classification purpose. There are various ways to do binning. For credit models with binary or multinomial targets, the most popular method is the Weight of Evidence (WoE). To be simple, we focus on the binary target – event or non-event.

Assume the risk factor x is binned into J buckets and define the following parameters:

r_T^{NET}: total number of non-events
r_T^{ET}: total number of events
r_j^{NET}: total number of non-events in j th bucket
r_j^{ET}: total number of events in j th bucket
$r_{\min}^{\mathrm{NET}}$: minimum number of non-events per bucket
$r_{\min}^{\mathrm{ET}}$: minimum number of events per bucket
$r_{\max}^{\mathrm{NET}}$: maximum number of non-events per bucket
$r_{\max}^{\mathrm{ET}}$: maximum number of events per bucket
$b_{\min}$: minimum number of buckets
$b_{\max}$: maximum number of buckets

For this binning, the information value (IV) with respect to this risk factor x is:

$$\mathrm{IV}(x) = \sum\nolimits_{j=1}^{J} \left(\frac{r_j^{\mathrm{NET}}}{r_T^{\mathrm{NET}}} - \frac{r_j^{\mathrm{ET}}}{r_T^{\mathrm{ET}}} \right) \log \left(\frac{r_j^{\mathrm{NET}}}{r_T^{\mathrm{NET}}} \Big/ \frac{r_j^{\mathrm{ET}}}{r_T^{\mathrm{ET}}} \right) \tag{2.9}$$

An optimal binning (J^*) maximizes IV(x) with restraints on $r_{\min}^{\mathrm{NET}}$, $r_{\min}^{\mathrm{ET}}$, $r_{\max}^{\mathrm{NET}}$, $r_{\max}^{\mathrm{ET}}$, $b_{\min}$, and $b_{\max}$. Such optimal binning can be solved either through recursive splitting and joining algorithms or a more efficient mathematic programming design.

The corresponding items

$$\mathrm{WOE}_j(x) = \log \left(\frac{r_j^{\mathrm{NET}}}{r_T^{\mathrm{NET}}} \Big/ \frac{r_j^{\mathrm{ET}}}{r_T^{\mathrm{ET}}} \right), j = 1, \ldots, J^* \tag{2.10}$$

are the Weight of Evidence (WoE) with the optimal binning. So, the original risk factor x is converted to a numeric variable WoE(x) with only J^* values.

WoE transformation (also called WoE scaling) is popular in credit risk modeling due to some attractive features. WoE mitigates the need to do extensive outlier analysis, as extreme values just end up in the first and last bins and thus won't ruin the fit of a model. Since bins are used, we can create an extra bin for NA's, so no treatment of missing values is necessary either. Then, for modeling the nonlinear effects, WoE simply summarizes them into those weights. Finally, another advantage is that fewer parameters are needed for categorical variables, as a categorical level is transformed to a numerical WoE value and thus only a single parameter has to be estimated. So, WoE is also used for parameter reduction when a large number of categorical risk factors with many levels are in the model.

However, one should pay attention to some of the details of WoE in practice and not blindly use it for everything. The WoE transformation is a supervised binning method. As with any binning, it requires quite some pre-setups as we show above with the optimal binning. In the meantime, some commercial or open-source software implementations may be suboptimal. For example, the number of bins with the optimal binning may be set to a range too narrow to have an optimal solution. Most often, the suboptimal solution by some software may be just a local stop point in the splitting and joining recursive process, and some management judgments are still required.

WoE has also been criticized for its way of handling nonlinear effects. For a nonlinear effect, WoE essentially uses a linear function of the WoE weights to approximate the nonlinear relation by assuming the nonlinear relation is summarized by these WoE weights. Such assumption needs to be checked, and if it fails, one may need to reverse the WoE transformation and use the function transformation discussed in Sect. 2.6.2 to catch the nonlinear effect more efficiently.

2.6.4 Data Lake and Business Analysis

In our data model design, we have a central component called data lake. The data lake includes all loaded data from various sources in the designed data architectures as well as some processed data from data processing and preliminary data analysis. Such data are not as wild as the data distributed in various sources and have transparency to be ready for some business analysis. The data files in the data lake can be considered having standard formats and being stable. The data lake has several functions.

Storage for All "Original" Data

Data lake stores all loaned data from various sources in the format of well-designed architectures. These data files are taken as original data and the starting point for business analysis. They are relatively stable in the sense that only periodic updates are needed to add more data or modify some of the current data. Between these updates, business can base on these data for preliminary analysis, reporting, model development, performance monitoring, and various testing purposes. So, the data lake bears a time stamp as a going-on data storage function.

Business Analysis

Data lake provides both the inputs for business analysis, which includes data visualization, preliminary analysis (e.g., bivariate analysis), and data transformation, segmentation, and sampling, as well as the data storage place for the outputs of these analysis and even more business intelligence applications.

Data lake can be used to fill the data gaps between different line of businesses and parties. Very often, risk-related data are spread across different parts of an organization with complex access requirements, especially true for large organizations. An integrated, adaptive, efficient, transparent, but with proper security and quality control risk data system is the base for a new generation of risk management. The data lake concept developed here can meet such requirements. Data systems consist of well-designed data lakes that can be built on more open environments, like public or proprietary clouds, and better shared across all participants with AI (artificial intelligent)-style management and communication.

2.6.5 Processing Automation and AI

The data model we described in Fig. 2.5 is connected with many data processes. Automation of these processes is the key to have an efficient data system. Automation starts from the design of the data model. For each process within the data model, first identify the parts that currently are designed to be executed manually. Then, consider if automation is possible. If not, then consider to redesign the process. When redesigning the process, one should have the AI thoughts in place. For example, triggers may be created to turn on the automation execution.

While AI can enhance the automation within a data model, solid tests must also accompany these automated processes to make sure the automation is achieved not in sacrifice of accuracy. Also controls need to be set up to check the process execution. While data quality controls can be placed in the ETL stages, other controls are also needed in the data processing and analysis stages. Such controls could provide early error discovery and ongoing system monitoring. The combination of automation and AI is the future of credit risk data systems.

3 Credit Modeling Techniques

Credit risk modeling techniques become mature over more than a half century of developments. While modeling for credit risk could be traced back much earlier, theoretical affirmation of statistical models, for example, the multinomial logit model as a special case of the more general conditional logit model, was first provided about a half century ago (McFadden, 1974)[1] using the random utility maximization paradigm. Since then, statistical models like the generalized linear models (GLM) have become the most popular selection in modeling credit risks, though machine learning models start to challenge that dominance in some areas in recent years. Figure 3.1 outlines the structure of various models discussed in this chapter.

Aligning with the credit data processing in Chap. 2, models are classified as account-level models and cohort-level models. Account-level models are based on account-level credit risk measures, probability of default (PD), probability of payoff (PP), loss given default (LGD), and exposure at default (EAD). For each account, the periodic losses over the observation unit (e.g., monthly, quarterly, or annually) are estimated as the combination of these measures over a specified projection window (1 year for capital and nine quarters for CCAR) or for the rest life of the account (CECL). These losses or their present values are summed up as the projected loss of the account.

With each of these risk measures (PD, PP, LGD, EAD), various statistical and machine learning models are presented in this chapter. For PD and PP, besides they are modeled separately under the target event framework, we present two frameworks under which they are simultaneously modeled, the competing risk framework and the multistate transition framework. For the competing risk framework, we demonstrate how binary generalized linear models can be used as unbiased approximation to the computationally expensive multinomial logit models

[1] McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, (ed.), Frontiers in Econometrics, New York: Academic Press, 105–142.

C. Chen, *Practical Credit Risk and Capital Modeling, and Validation*, Management for Professionals, https://doi.org/10.1007/978-3-031-52542-1_3

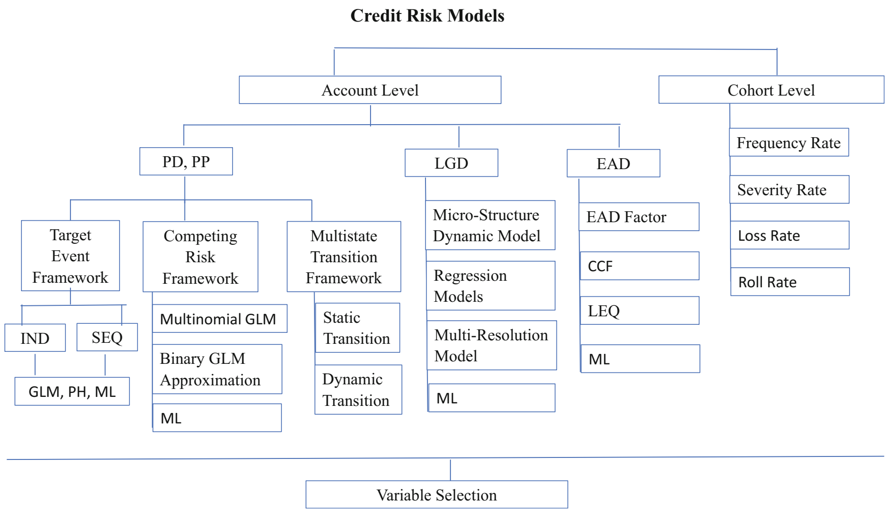


Fig. 3.1 Model coverage

commonly used for competing risk modeling. For LGD models, we present a dynamic resolution model with generalized linear models for the post-default dynamic resolution (gross loss or fully paid) probabilities and linear models for severities at resolution. This is also called micro-structure LGD model with loss timing and forward-looking. In addition, we also present the commonly used regression models for LGD, which directly model the LGD ratio as a function of information at default. A more granular LGD model would explore multiple post-default resolutions with their predicted probabilities and severities at default. For EAD models, we focus on some commonly used methods for products with credit commitments.

Machine learning models are also discussed. These techniques including tree-based classifiers and other enhanced approaches are ideally suited for credit risk analysis. The extraordinary speedup in computing, coupled with significant algorithmic theoretical advances, has created a renaissance in computational credit risk modeling, especially in line of business risk management practices.

For cohort-level models, default rates, payoff rate, and integrated loss rate are modeled as linear (or loglinear) and more often nonlinear models linked to MEVs. Lastly, we also include the cohort-level roll rate model, which corresponds to the multistate transition model for account level.

For both account- and cohort-level models, variable selection is a major challenge, especially with model validation. We introduce the adaptive and exhaustive variable selection (AEVS) process as a general solution. It can be used for both linear (loglinear) and generalized linear models.

It should be pointed out that the general models presented in this chapter need to be tuned to fit into the specific credit risk management diagrams. For example, for RWA in the regulatory capital management, the 1-year PD model is fitted to cross-sectional data from snapshots instead of panel data as in the loss forecast models used in ACL or CCAR. For all these models, including some machine learning models as challengers, their advantages, disadvantages, and more importantly suitability and validation are discussed in each of the three credit risk management diagrams in the following chapters.

Key Abbreviations and Symbols

ACL	Allowance for credit losses
AEVS	Adaptive and exhaustive variable selection
AI	Artificial intelligence
CCAR	Comprehensive Capital Analysis and Review
CECL	Current expected credit loss
CSD	Cross-sectional data
DR	Default rate
EAD	Exposure at default
EC	Economic capital
GLM	Generalized linear model
IRB	Internal ratings-based
IV	Information value
LGD	Loss given default
LTV	Loan to value
PD	Probability of default
PH	Proportional hazard
PIT	Point in time
PTI	Payment to income
RC	Regulatory capital
RWA	Risk-weighted asset
SMM	Simple monthly mortality
SMOTE	Synthetic Minority Oversampling Techniques
TTC	Through-the-cycle
WAC	Weighted average coupon
WALA	Weighted average loan age
WOE	Weight of Evidence

3.1 Account-Level Models

When account-level data are available, we can build account-level risk models (PD, PP, LGD, and EAD). While the model inputs, especially the credit product performance, are required at the account level, the model outputs can be aggregated at any higher (e.g., cohort) level. So, account-level models can fulfill applications with higher-level risk aggregations. In the meantime, dividing the loss modeling into the

PD/PP, LGD, and EAD structure follows the credit loss generating process and shares more insights on the source of the credit risk and how to efficiently control it.

With the account-level models, under the target event framework, a single target risk event (default or payoff) is modeled either by not considering other risk events or after removing other specific risk events. When multiple risk events are modeled simultaneously, we cover two popular credit risk modeling frameworks – the competing risk framework (CRF) and the multistate transition framework (MTF). The CRF models are risk factor-based models, which usually are designed for long-term loss forecasting, while the MTF models are transition models, which are more suitable for short-term loss forecasting. Though there are modified MTF models that use risk factors to predict the key transition rates, such dynamic MTF models may be used for long-term loss forecasting with frequent performance monitoring.

3.1.1 Target Event Framework

When a specific credit risk event (default or payoff) is the target, a binary outcome can be formulated by defining the target event as the risk event and all others as non-event. So, the target risk event is modeled independently. The target event framework fits the capital and credit scoring modeling paradigms, where default is the only concern and payoff is classified into non-default.

The target risk event can also be modeled after removing other risk events, for example, default can be modeled after removing payoff event. This is equivalent to modeling the default event in condition of non-payoff and is a conditional risk event model.

Combining the independent risk event model and the conditional risk event model, one can form a sequential risk model framework by first modeling one target risk event independently and treating all other risk events as non-events and then removing this risk event and independently modeling the second target risk event and treating the rest of risk events as non-events.

Sequential risk model framework has been popularly used in retail consumer products, especially with mortgages by the argument that borrowers always consider payoff before default. However, there were evidences that borrowers may execute strategic default before considering payoff, especially when house price was deeply down and home equity became negative as seen during last housing crash. In such case, a competing risk framework may be more reasonable if both risk events need to be modeled.

3.1.1.1 Generalized Linear Models

For binary target, generalized linear model has been the most popular model used, especially the binary logit model. The target variable for a risk event (e.g., default) is defined as:

Table 3.1 Generalized linear models

Distribution	CDF	Link
Logistic	$F(x'\beta) = \Lambda(x'\beta) = \frac{\exp(x'\beta)}{1+\exp(x'\beta)}$	$g(P) = \Lambda^{-1}(P) = \log\left(\frac{P}{1-P}\right)$
Normal	$F(x'\beta) = \Phi(x'\beta)$	$g(P) = \Phi^{-1}(P)$
Gompertz	$F(x'\beta) = G(x'\beta) = (1 - \exp(-\exp(x'\beta))$	$g(P) = G^{-1}(P) = \log(-\log(1-P))$

$$Y = \begin{cases} 1, & \text{Event} \\ 0, & \text{Non-Event} \end{cases} \tag{3.1}$$

Event Probability and Link Function

Given the risk factor vector $X = x$ in m dimension (for simplicity, we assume x includes the constant factor), the generalized linear model assumes that the probability of the risk event is a function of the linear predictor $x'\boldsymbol{\beta}$:

$$P(Y = 1|X = x) = F(x'\beta) \tag{3.2}$$

where β is the m-dimension parameter vector, F is a cumulative distribution function (CDF), and its inverse $g = F^{-1}$ that links the event probability to the linear predictor is called link function in the generalized linear model. If F is the logistic CDF:

$$F(x'\beta) = \Lambda(x'\beta) = \frac{\exp(x'\beta)}{1 + \exp(x'\beta)} \tag{3.3}$$

The link function is the logit function:

$$g(P) = \log\left(\frac{P}{1-P}\right) \tag{3.4}$$

which is the logarithm of the risk event probability odds defined as $P/(1 - P)$. Other commonly used CDF and link function pairs in GLM for binary target are normal CDF and probit link and extreme value (or Gompertz) CDF and gompit link. Table 3.1 summarizes these GLM models. Note that Φ is the standard normal distribution function, Λ is the standard logistic distribution, and G is the standard Gompertz distribution. Logistic and normal distributions are similar symmetric distributions, while Gompertz is asymmetric and biased to the left and less sensitive to the large values. For credit risk models, logistic distribution is the most popularly used one due to the proportional odds property and computing efficiency.

Proportional Odds

Assume we have only one standardized risk factor x. We would like to know what is the odds ratio if x increases 1 unit, for example, from 0 to 1, 1 to 2, and so on. Under logistic distribution, the log of the odds ratio is:

$$\log\left(\frac{P_1}{1-P_1}\right) - \log\left(\frac{P_0}{1-P_0}\right) = \beta_0 + \beta_1(x+1) - \beta_0 - \beta_1 x = \beta_1 \tag{3.5}$$

which is exactly the coefficient itself, and the odds ratio is $\exp(\beta_1)$ and doesn't depend on the risk factor anymore. So, we say that the log odds is proportional to the change of any risk factor in the model. The proportional odds property is attractive in interpretation of the risk factor impacts if the risk factor has such a multiplicative relation to the risk event probability.

Maximum Likelihood Estimate

The model parameters can be estimated through maximum likelihood. Assume the observed outcomes and risk factors are $\{(y_{\mathrm{i}}, x_{\mathrm{i}}), i = 1, \ldots, N\}$. For each observation (y_{i}, x_i), the binomial likelihood conditioning on $\mathrm{x_i}$ is:

$$L_i(\beta) = P_i^{y_i}(1-P_i)^{1-y_i} \tag{3.6}$$

where $P_i = P(Y=1|X=x_i) = \exp(x_i'\beta)/(1+\exp(x_i'\beta))$ is the conditional risk event probability $i = 1, \ldots, N$. Assume these risk events are **conditionally independent**, and then we can maximize the combined likelihood $L(\beta) = \prod_{i=1}^N L_i(\beta)$ (or equivalently its logarithm $l(\beta) = \log(L)$) with respect to the model parameters β:

$$\max_\beta\left(L(\beta) = \prod_{i=1}^N L_i(\beta)\right) = \max_\beta \prod_{i=1}^N \left[\frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}\right]^{y_i}\left[\frac{1}{1+\exp(x_i'\beta)}\right]^{1-y_i} \tag{3.7}$$

$$\Leftrightarrow \max_\beta l(\beta) = \max_\beta \sum_{i=1}^N \left\{y_i x_i'\beta - \log\left(1+\exp(x_i'\beta)\right)\right\} \tag{3.8}$$

It turns out that the loglikelihood function $l(\beta)$ is a very well-behaved concave function for almost any observations $\{(y_{\mathrm{i}}, x_i), i = 1, \ldots, N\}$ except some very rare cases with separation.[2] So, with overlapped observations, the maximum likelihood estimate $\widehat{\beta}_{\mathrm{ML}}$ is finite and unique and can be found efficiently. This adds an advantage for binary logit model in credit risk modeling with large data.

[2]Definition of observations separation and overlap:

- Complete separation: $\exists$a vector b that correctly allocates all observations to their response groups, i.e.:

$$\begin{cases} b'x_i > 0 & y_i = 1 \\ b'x_i < 0 & y_i = 0 \end{cases}$$

- Quasi-complete Separation: $\exists$a vector b such that:

$$\begin{cases} b'x_i \geq 0 & y_i = 1 \\ b'x_i \leq 0 & y_i = 0 \end{cases}$$

  with equality holds for at least one point
- Overlap: Neither complete separation nor quasi-complete separation.

Under the following very general conditions:

- Y_{i} are independent condition on $X_i = x_i$.
- $X = [x_1, \ldots, x_N]'$ is full rank and bounded.
- $\frac{1}{N}X'X$ converges to a positive definite matrix.

$\widehat{\beta}_{\mathrm{ML}}$ is consistent, asymptotically normal, and efficient. In summary, the MLE of the binary logit model has the following properties:

- $l(\beta)$ is globally concave and MLE $\widehat{\beta}_{\mathrm{ML}}$ is unique (and finite with no data separation).
- MLE $\widehat{\beta}_{\mathrm{ML}}$ satisfies the following first-order condition:

$$\frac{\partial}{\partial\beta} l\left(\widehat{\beta}\right) = \sum\nolimits_{i=1}^{N} x_i \left[y_i - \Lambda\left(x_i'\widehat{\beta}\right)\right] = 0 \tag{3.9}$$

  which mimics the normal equations in linear regression (see Appendix 1 for linear models).
- If the logistic regression has an intercept, the above equation shows that $N^{-1}\sum_i y_i = N^{-1}\sum_i \Lambda\left(x_i'\widehat{\beta}\right)$, that is, the sample mean of response equals to the mean of predicted response $\widehat{\pi}_i = \frac{1}{N}\sum_{i=1}^{N} F\left(x_i'\widehat{\beta}\right)$.
- If the logistic regression includes categorical covariates, the same normal equation implies that the mean response equals to the mean of predicted for each bucket of the covariate.
- Under very *general conditions*, $\widehat{\beta}_{\mathrm{ML}}$ is consistent, asymptotically normal, and the most efficient.

Goodness-of-Fit

Generalized linear models are fitted using historical data. To make sure that the model fits the data well, various goodness-of-fit (GOF) measures can be used. The following is a list of these GOF measures:

- R^2 and Adjusted $\overline{R}^2$: $R^2 = 1 - \left\{\frac{L(0)}{L\left(\widehat{\beta}\right)}\right\}^{\frac{2}{N}}, R^2_{\max} = 1 - \{L(0)\}^{\frac{2}{N}}, \overline{R}^2 = \frac{R^2}{R^2_{\max}}$, where $L(0)$ is the intercept-only likelihood and $L\left(\widehat{\beta}\right)$ is the likelihood of the specified model with MLE. The larger these values, the better the model fit.
- $-2\log\left(L\left(\widehat{\beta}\right)\right)$, AIC, SIC: $\mathrm{AIC} = -2\log\left(L\left(\widehat{\beta}\right)\right) + 2p, \mathrm{SIC} = -2\log\left(L\left(\widehat{\beta}\right)\right) + p\log(N)$, where $L\left(\widehat{\beta}\right)$ is the likelihood of the specified model with MLE. The lower these values, the better the model fit.

- Rank Correlations: are based on ranking a pair of observations according to the combination of the observed responses (1 or 0 as event or non-event) and the model predicted event probabilities (scores). A pair of observations with different observed responses (1 or 0 in the binary case) is said to be concordant if the observation with the lower response value (0) has a lower predicted score than the observation with the higher response value (1). If the observation with the lower response value (0) has a higher predicted score than the observation with the higher response value (1), then the pair is discordant. If the pair is neither concordant nor discordant, it is a tie (which means the same predicted scores for the pair). The following rank corrections are popularly used:

$$N - \#\text{of obs.} t - \#\text{of pairs}, n_c - \#\text{of concordant}, n_d - \#\text{of discordant}$$

$$c\ (\text{Concordance Index}) = \frac{n_c + 0.5(t - n_c - n_d)}{t}$$

$$\text{Somers}'\ \text{D}\ (\text{Gini}) = \frac{n_c - n_d}{t}$$

$$\text{Goodman} - \text{Kruskal Gamma} = \frac{n_c - n_d}{n_c + n_d}$$

$$\text{Kendall}'\text{s Tau} - \text{a} = \frac{n_c - n_d}{0.5N(N-1)}$$

  Note that, without ties, $t = n_c + n_d$ and Somers' D (Gini) $= 2c - 1$. The higher these values, the better the fit.
- Receiver Operating Characteristic Curves (ROC): as the most popular graphic GOF measure of the binary logit model. It is the plot of *sensitivity* vs. 1 − *specificity* defined as:

$$\text{POS}(z) = \sum_{i \in \Omega_1} I(\widehat{\pi}_i \geq z), \text{FAL\_POS}(z) = \sum_{i \in \Omega_0} I(\widehat{\pi}_i \geq z)$$

$$\text{sensitivity} = \text{POS}(z)/n_1, 1 - \text{specificity} = \text{FAL\_POS}(z)/n_0,$$

  where Ω_1 is the group with n_1 events and Ω_0 is the group with n_0 non-events. The area under the ROC curve (also called AUC) as z changes from 1 to 0 is the Concordance Index c.

Variable Specification and Selection

The classical variable selection methods for linear and generalized linear models are forward, backward, and stepwise based on goodness-of-fit statistics (usually likelihood-based). However, a model that fits the historical data well may not be a good predictor. A balance between fitting and prediction can be achieved through some statistical criteria like cross-validation. Model variable specification and selection need to consider economic and finance theories for interpretation, as well as business inputs to reflect the reality and practice. In addition, when multiple models are available, we need to defend why the selected models are optimal or suboptimal under some conditions. In Sect. 3.3, we present an adaptive and exhaustive variable selection (AEVS) procedure to address these issues for both linear and generalized linear models.

3.1.1.2 Proportional Hazard Models

Different from the generalized linear models, which directly specify the event probability function, the proportional hazard models derive the event probability through the specification of the hazard function of event time. Although there is a one-to-one mapping between the event probability distribution and the event-time hazard, the motivation and logic of the specifications are different. Modeling of the hazard of a risk event (or failure) originated from reliability theory and biostatistics for continuous observations with measured durations between events or censored observation times. So, the data for such duration analysis are repeated measures on the same object or called longitudinal data. For credit risk modeling, such longitudinal data is called panel data, since the event-time duration measures can be easily converted to periodic (monthly, quarterly, or annually) panel observations and vice versa. The exchangeability between the event-time measures and panel observations provides an information equivalence from input data aspect between generalized linear models and hazard models and is a base that these two types of models can have close estimates of risk event probability in credit modeling.

The proportional hazard model assumes that the hazard function of the risk event time consists of a baseline hazard function and some multipliers in the form of parametric risk factor function. The parametric risk factor multiplier leads to the proportional risk factor contribution to the hazard similar to the proportional contribution to the odds in the binary logit model. The baseline hazard function can take both parametric and nonparametric forms. The parametric baseline hazard connects the proportional hazard model to the accelerated failure time models, while the nonparametric baseline connects it to the Kaplan-Meier nonparametric survival estimator. Because of the dependence of the event probability on the event-time hazard function, the baseline hazard function needs to be estimated from observed event times, which should be measured using a proper starting point such as the initial book date, origination date, or some other date that the hazard of the risk event has a fundamental relation. Once the event-time hazard function is estimated, it can be converted to the event probability.

Event Probability and Hazard Function

Assume the risk event time T measured with a proper starting point has a cumulative distribution function (CDF) $F(t)$ and a probability density function (PDF) $f(t)$ on $(0, \infty)$. So, together with the survival function $S(t)$, they have the following relationships:

$$\text{CDF} : F(t) = P(T < t)$$

$$\text{PDF}: f(t) = \partial F/\partial t$$

$$\text{Survival Function}: S(t) = 1 - F(t) = P(T \geq t)$$

The hazard function of the risk event time T is defined as the ratio of PDF and survival function $S(t)$:

$$h(t) = f(t)/S(t) \tag{3.10}$$

A more intuitive definition of the hazard is the conditional risk event rate given that the account has survived (free from the risk event) before time t:

$$h(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P[T \in [t, t + \Delta t) | T \geq t] = -d\ \log(S(t))/dt \tag{3.11}$$

The relation between the hazard function and the PDF is one-to-one:

$$h(t) = \frac{f(t)}{\int_t^{\infty} f(u)du} \tag{3.12}$$

$$f(t) = h(t)S(t) = h(t)\exp\left\{-\int_0^t h(u)du\right\} \tag{3.13}$$

and

$$S(t) = \exp\left\{-\int_0^t h(u)du\right\} \tag{3.14}$$

So, the risk event probability for a given period of Δt can be estimated from the estimated hazard function $\widehat{h}$ as:

$$P(Y = 1 | t, \Delta t) = \widehat{S}(t) - \widehat{S}(t + \Delta t) = \exp\left\{-\int_0^t \widehat{h}(u)du\right\} - \exp\left\{-\int_0^{t+\Delta t} \widehat{h}(u)du\right\} \tag{3.15}$$

where:

$$Y = \begin{cases} 1, & \text{Event during } [t, t + \Delta t) \\ 0, & \text{Non-Event during } [t, t + \Delta t) \end{cases} \tag{3.16}$$

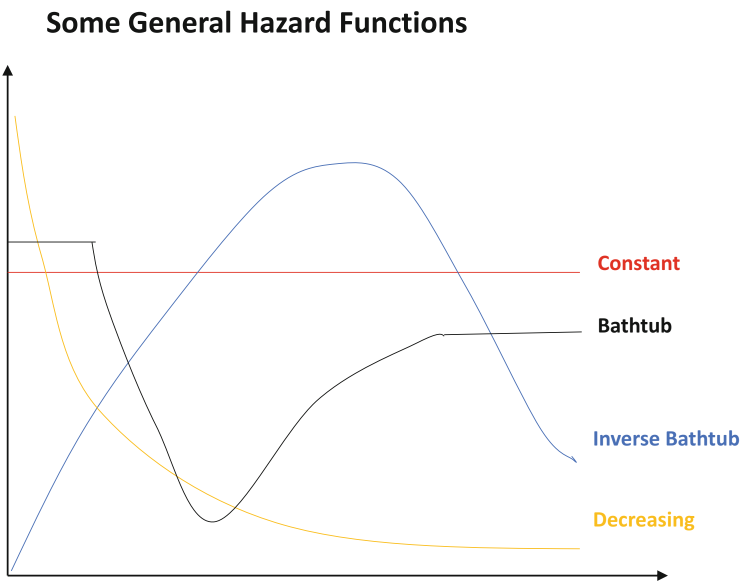


Fig. 3.2 Hazard functions

As examples, the following are two parametric specifications:

$$T \sim \text{Exponential}: f(t, \lambda) = \lambda \exp(-\lambda t), h(t) = f(t)/S(t) = \lambda$$

$$T \sim \text{Weibull}: f(t, \lambda) = \lambda\alpha(\lambda t)^{\alpha-1} \exp(-(\lambda t)^{\alpha}), h(t) = f(t)/S(t) = \alpha\lambda^{\alpha}t^{\alpha-1}$$

Figure 3.2 shows more general patterns of hazard function.

Proportional Hazard

When incorporating risk factors $X = x$, the proportional hazard (PH) model specifies the hazard of the risk event as:

$$h(t|x) = \lambda(t) \exp(x'\beta) \tag{3.17}$$

In this form, the exponential function of the risk factor linear combination is a multiplier to the baseline hazard function $\lambda(t)$. Like the odds ratio with the binary logit model, the logarithm of the hazard ratio with one unit change for any of the risk factors is the risk factor coefficient in the PH model. For simplicity, assuming only one risk factor x:

$$\log(h(t|x+1)/h(t|x)) = \beta \tag{3.18}$$

So, the logarithm hazard change is proportional to the risk factor change, which is exactly the same as the logarithm odds change that is proportional to the risk factor change as in the binary logit model.

Baseline Hazard Specification
As we mentioned early, the baseline hazard function can be specified in a parametric or a nonparametric form. We will show how the parametric baseline hazard connects the proportional hazard model to the accelerated failure time models, as well as how the nonparametric baseline hazard connects it to the Kaplan-Meier nonparametric survival estimator.

Define the cumulative baseline hazard function $\Lambda(t) = \int_0^t \lambda(u)du$, and then we have the following facts under the proportional hazard specification of (3.17).

Lemma 3.1 If the event time T has the proportional hazard as specified in (3.17) condition on risk factors $X = x$, then given $X = x$, the quantity:

$$\log \Lambda(T) + x'\beta \tag{3.19}$$

has an extreme value distribution.

Combining (3.14) and (3.17), we have:

$$\begin{aligned} S(t) &= \exp\left\{ -\int_0^t h(u)du \right\} = \exp\{-\exp(x'\beta)\int_0^t \lambda(u)du\} \\ &= \exp\{-\exp(x'\beta)\Lambda(t)\} \end{aligned}$$

So, $\log\Lambda(T) + x'\beta = \log[-\log S(T)]$. Condition on $X = x$, $\log[-\log S(T)]$ is an extreme distribution, since $S(T)$ is a uniform distribution (just as $F(T)$) and $[-\log S(T)]$ is an exponential distribution and the logarithm of an exponential distribution is an extreme distribution.

So, we can scale it to a standard extreme distribution ϵ with the scale σ:

$$\log \Lambda(T) + x'\beta = \sigma\epsilon \tag{3.20}$$

If baseline hazard has a parametric form, for example, $\lambda(t) = \alpha\, t^{\alpha-1}$, as with the Weibull distribution, then $\log\Lambda(T) + x'\beta = \log T^{\alpha} + x'\beta = \sigma\epsilon$, and let $T_0 = \exp(\epsilon)$ be a standard exponential event-time variable, and then:

$$T^{\alpha} = \exp(-x'\beta)T_0^{\sigma} \tag{3.21}$$

which is exactly the accelerated failure time model. (3.20) is also useful in testing the proportional hazard specification.

As we know that while the accelerated failure time models are popular in reliability and survival analysis applications, they are usually not suitable for credit risk modeling due to the too restrictive parametric form resulting in poor model fit. So, as long as the estimation of the event probability is in concern, semiparametric or nonparametric baseline hazard should be considered.

Assume $t_1 < t_2 < \ldots < t_N$ are observed event times. In credit risk modeling, these times are periodic times in month, quarter, or year. With $t_0 = 0$ and $t_{N+1} = \infty$, the baseline hazard can be specified as piecewise constant with $\alpha_0 = 1$ and α_i, $i = 1, \ldots, N$ being parameters:

$$\lambda(t) = 1 - \alpha_i, t \in (t_{i-1}, t_i] \tag{3.22}$$

With the piecewise constant specification, the baseline hazard function can be estimated through maximum likelihood of product limit (Kalbfleisch & Prentice, 1980),[3] which we will show later.

The piecewise constant specification can be generalized to piecewise linear or more general spline functions. For all these specifications, the knots do not need to be the observed event times. To reduce the number of parameters to be estimated, reduced number of knots can be selected.

For the nonparametric method, we do not specify a format for the hazard function. Without risk factors considered, given the risk event times, the nonparametric maximum likelihood estimates of the survival function are given by Kaplan-Meier nonparametric survival estimator (Meeker & Escobar, 1998).[4] When risk factors are considered, combining with partial maximum likelihood estimates of the risk factor parameters with the empirical cumulative hazard function estimates through (3.14), the survival function $S(t)$ can be estimated. The following section explains these estimates.

Model Estimation

The model estimation for the proportional hazard model is based on both the baseline hazard specification and the time-event data. As we explained in the previous section, both parametric and nonparametric formats can be used for the baseline hazard. As for the event-time data, there could be censoring and ties, which need being handled differently from a complete event time. Since specifying the event-time hazard function is equivalent to specifying the survival function (thus the event probability in any period), maximum likelihood estimation is the most efficient approach for proportional hazard model.

For a set of complete event-time observations $\{(t_i, x_i), i = 1, \ldots, N\}$, under the proportional hazard specification and conditional independence of event times, we have the loglikelihood:

$$\begin{aligned} h(t|x) &= \lambda(t, \alpha)e^{x'\beta} \\ \Rightarrow f(t|x) &= \lambda(t, \alpha)e^{x'\beta} \exp\left(-e^{x'\beta}\Lambda(t, \alpha)\right) \\ \Rightarrow l(\alpha, \beta) = \log(L) &= \sum_{i=1}^{N} \left[\log(\lambda(t_i, \alpha)) + x_i'\beta - e^{x_i'\beta}\Lambda(t_i, \alpha)\right] \end{aligned} \tag{3.23}$$

[3] Kalbfleisch, J. D. and Prentice, R. L. (1980), The Statistical Analysis of Failure Time Data, New York: John Wiley & Sons.

[4] Meeker, W. Q. and Escobar, L. A. (1998), Statistical Methods for Reliability Data, New York: John Wiley & Sons.

For most parametric specification of the baseline hazard $\lambda(t, \alpha)$, the loglikelihood function $l(\alpha, \beta)$ is globally concave, and a unique MLE of (α, β) can be found by a simple gradient algorithm.

For a set of event-time observations $\{(t_i, x_i), i = 1, \ldots, N\}$ with censoring, we have the loglikelihood:

$$\begin{aligned} l(\alpha,\beta) = & \sum \log(f(t_i,\alpha,\beta)) && \text{(non-censoring)} \\ & + \sum \log(S(t_i,\alpha,\beta)) && \text{(right censoring)} \\ & + \sum \log(F(t_i,\alpha,\beta)) && \text{(left censoring)} \\ & + \sum \log\left(F\left(t_i^U,\alpha,\beta\right) - F\left(t_i^L,\alpha,\beta\right)\right) && \text{(interval censoring)} \end{aligned} \tag{3.24}$$

Global concave of $l(\alpha, \beta)$ and a unique MLE of (α, β) can't be guaranteed anymore. However, for most common parametric specifications of the baseline hazard used in accelerated failure time models, global optimization algorithms can find the MLE of (α, β), which is usually the best local MLE for the given data.

Instead of specifying the full parametric form of the hazard function, the semiparametric PH model can skip specifying the baseline hazard and directly estimate the risk factors' effects using partial likelihood. Then if needed as in the credit risk modeling for the event probability, the baseline hazard is estimated in a second step. Since the estimates of the risk factors' effects using partial likelihood are consistent, the baseline hazard estimates in the second step based on the first step estimates are also consistent.

One of the advantages of the PH model is the independent estimation of the risk factors' effects using partial likelihood from the baseline hazard. Assume that the ordered risk event times without censoring and ties are $t_1 < t_2 < \ldots < t_N$, and then under the proportional hazard of (3.17), the conditional probability that risk event 1 occurs at t_1 given that any of the N observations could occur at t_1 is:

$$\frac{h(t_1|x_1)}{\sum_{i=1}^{N} h(t_1|x_i)} = \frac{\exp\left(x_1'\beta\right)}{\sum_{i=1}^{N} \exp\left(x_i'\beta\right)} \tag{3.25}$$

This is the contribution to the likelihood from the first observation; for the ith-ordered observation, the contribution is:

$$\frac{h(t_i|x_i)}{\sum_{k=i}^{N} h(t_i|x_k)} = \frac{\exp\left(x_i'\beta\right)}{\sum_{k=i}^{N} \exp\left(x_k'\beta\right)} \tag{3.26}$$

Let R_i be the set of individuals still exposed to risk of fail at t_i, and this can be easily extended to cases with censoring and ties ($j \in D_i$ and $|D_i| = d_i$) assuming conditional independence of ties:

$$\prod_{j \in D_i} \frac{h(t_i|x_j)}{\sum_{k \in R_i} h(t_i|x_k)} = \frac{\prod\limits_{j \in D_i} \exp\left(x_j'\beta\right)}{\left[\sum_{k \in R_i} \exp\left(x_k'\beta\right)\right]^{d_i}} \tag{3.27}$$

So, assuming the conditional independence of ordered event times and ties, the log partial likelihood is:

$$L(\beta) = \prod_{i=1}^{N} \frac{\beta' \sum_{j \in D_i} x_j}{\left[\sum_{k \in R_i} \exp\left(x_k'\beta\right)\right]^{d_i}} \tag{3.28}$$

This is the Breslow likelihood and its logarithm is:

$$l(\beta) = \sum_{i=1}^{N} \left\{ \beta' \sum_{j \in D_i} x_j - d_i \log\left[\sum_{k \in R_i} \exp\left(x_k'\beta\right)\right] \right\} \tag{3.29}$$

If there is a weight on each observation, the weighted Breslow loglikelihood is:

$$l_w(\beta) = \sum_{i=1}^{N} \left\{ \beta' \sum_{j \in D_i} w_j x_j - \left(\sum_{j \in D_i} w_j\right) \log\left[\sum_{k \in R_i} \exp\left(x_k'\beta\right)\right] \right\} \tag{3.30}$$

The Efron likelihood is a bias correction version of the Breslow likelihood:

$$L_{\text{Efron}}(\beta) = \prod_{i=1}^{N} \frac{\beta' \sum_{j \in D_i} x_j}{\prod\limits_{j=1}^{d_i} \left(\sum_{k \in R_i} \exp\left(x_k'\beta\right) - \frac{j-1}{d_i} \sum_{k \in D_i} \exp\left(x_k'\beta\right)\right)} \tag{3.31}$$

In summary, the partial likelihood has the following properties:

- The partial likelihood does not involve the baseline hazard.
- The partial likelihood does not depend on the magnitudes of event times (duration), but only the rank orders of event times.
- Censoring observations only contributes to the denominators through the exposure risk sets.
- Ties are treated as individuals.
- It can be proved that the maximizers of the partial likelihood are consistent and asymptotically normal.

Once the maximum partial likelihood estimate (MPLE) $\widehat{\beta}$ is obtained, the baseline hazard function $\lambda(t)$ can be estimated through two methods according to the baseline hazard specifications.

For the piecewise constant specification in (3.22), let D_i be the set of individual accounts with event times at t_i and C_i be the individual accounts censored between (t_i, t_{i+1}) and R_i be the set of individual accounts still exposed to risk at t_i, and then the product limit estimates maximize the likelihood function:

$$L_{\mathrm{pl}} = \prod\nolimits_{i=0}^{N} \left\{ \prod_{j \in D_i} \left(S_0(t_i)^{\exp\left(x_j' \widehat{\beta}\right)} - S_0(t_i + 0)^{\exp\left(x_j' \widehat{\beta}\right)} \right) \prod_{k \in C_i} S_0(t_i + 0)^{\exp\left(x_k' \widehat{\beta}\right)} \right\} \tag{3.32}$$

where $S_0(t)$ is the survival function corresponding to the baseline hazard function and $\widehat{\beta}$ is the MPLE. Consider the discrete hazards $(1 - \alpha_i)$ over $(t_{i-1}, t_i]$.So, $S_0(t_i) = S_0(t_{i-1} + 0) = \prod_{j=1}^{i-1} \alpha_j$, and the simplified likelihood of α_i, $i = 1, \ldots, N$ $(\alpha_0 = 1)$ is:

$$L_{\mathrm{pl}}(\alpha_i) = \prod_{j \in D_i} \left(1 - \alpha_i^{\exp\left(x_j' \widehat{\beta}\right)} \right) \prod_{k \in R_i - D_i} \alpha_i^{\exp\left(x_k' \widehat{\beta}\right)} \tag{3.33}$$

Given $\widehat{\beta},$ the MLE $\widehat{\alpha}_{\mathrm{i}}$ solves:

$$\sum_{j \in D_i} \frac{\exp\left(x_j' \widehat{\beta}\right)}{1 - \widehat{\alpha}_i^{\exp\left(x_j' \beta\right)}} \sum_{k \in R_i} \exp\left(x_k' \widehat{\beta}\right) \tag{3.34}$$

So, $\widehat{S}_0(t) = \widehat{S}_0(t_{i-1} + 0) = \prod_{j=0}^{i-1} \widehat{\alpha}_j$ for $t \in (t_{i-1}, t_i], i = 1, \ldots, N,$ and $\widehat{S}(t, x) = \left[\widehat{S}_0(t)\right]^{\exp\left(x' \widehat{\beta}\right)}$.

The estimated piecewise constant baseline hazard can be smoothed using piecewise linear or smooth spline interpolations.

The second method is directly estimating the cumulative hazard function using the empirical cumulative hazard function, which is corresponding to the Kaplan-Meier nonparametric survival function estimates.

Let $N_i(s)$ be event counting process for the ith account and $y_i(s)$ denotes its risk exposure status, and the empirical cumulative hazard function is:

$$\widehat{\Lambda}(t, x) = \sum_{i=1}^{N} \int_0^t \frac{dN_i(s)}{\sum_{j=1}^{N} y_j(s) \exp\left(\widehat{\beta}' \left(x_j - x\right)\right)} \tag{3.35}$$

And the survival function is estimated according to (3.14), $\widehat{S}(t, x) = \exp\left(- \widehat{\Lambda}(t, x)\right)$.

Without risk factors and using the first-order approximation of the exponential function:

$$\log\left(\frac{\widehat{S}(t_k)}{\widehat{S}(t_{k+1})}\right) = \widehat{\Lambda}(t_{k+1}) - \widehat{\Lambda}(t_k) = \sum_{i=1}^{N}\int_{t_k}^{t_{k+1}} \frac{dN_i(s)}{\sum_{j=1}^{N} y_j(s)} = \frac{1}{N-k}$$

$$\widehat{S}(t_k) = \widehat{S}(t_{k+1})\exp\left(\frac{1}{N-k}\right) \approx \widehat{S}(t_{k+1})\frac{N-k+1}{N-k}$$

So as N is large, we have the Kaplan-Meier nonparametric survival function estimates with $\widehat{S}(t_0) = 1$:

$$\widehat{S}(t_{k+1}) = \widehat{S}(t_k)\frac{N-k}{N-k+1} \tag{3.36}$$

The connection between the empirical cumulative hazard estimates of (3.35) without risk factors and the Kaplan-Meier nonparametric survival estimates of (3.36) can claim that the cumulative hazard-based event probability estimates based on (3.35) are a combination of the Kaplan-Meier nonparametric survival function estimates and the MPLE of risk factor effects.

While both the PL and the nonparametric event probability estimates are popularly used in PH models, they have different performance. Computationally, the nonparametric method is more efficient for not solving multiple equations as required by PL method. For the estimated event probability, the nonparametric is less sensitive to data, while the PL method can be fine-tuned based on the selection of the knot points. In credit risk, the PL method is preferred due to its flexibility to adjust the accuracy requirements.

Goodness-of-Fit

Goodness-of-fit measures are the same as the binary generalized linear models presented in Sect. 3.1.1.1, since the event probability is the target. The only difference is that likelihood-based GOF measures should use the full likelihood for parametric models and partial likelihood for the semiparametric models.

3.1.1.3 Machine Learning Models

Although the classical statistical models, like generalized linear models and proportional hazard models discussed above, are still the most popular models used in credit risk management, especially in regulatory areas, in recent years, a growing attention has been devoted to machine learning procedures. Originated from the data-driven statistical thinking from a group of elite statisticians and combined with significant breakthroughs in computing speed, some difficult practical problems can be solved by brute force algorithms on very large data. Due to the lack of elegance in theory as shown by classical statistical models, these algorithm-based procedures were not recognized as models and were treated as some preliminary indicative data analysis. However, the free from restrictions as those imposed on the classical statistical models leads to powerful insight analysis of large and complex data on which the classical models usually have difficulty to extract useful information to make decisions.

Such advantages were rediscovered in computer science. Machine learning (ML) in computer science literature refers to a set of algorithms specifically designed to tackle computationally intensive pattern recognition problems based on extremely large data sets. In recent years, these techniques have achieved great successes in artificial intelligence (AI). These techniques including tree-based classifiers and other enhanced ML approaches are ideally suited for credit risk analysis. The extraordinary speedup in computing, coupled with significant algorithmic theoretical advances, has created a renaissance in computational credit risk modeling.

With their increased popularities in risk modeling, an urgent question is how to manage the risks of using these ML procedures in different areas of risk modeling. Since the regulatory guidance of SR 11-7, which made the model risk an independent tier of risk from credit, market, and operational risks, the model risk management framework in financial applications has become mature. At least based on the following informal definition of model, most ML procedures are models:

A model refers to a quantitative method, approach, or procedure that applies statistical, economic, financial or mathematical theories, techniques and assumptions to process input data into quantitative estimates, which are for uncertain quantities with varying precision and accuracy and can be translated into business information.

A model consists three components: the input and sources component, which includes the input data and assumptions; the processing component, which transforms the inputs into quantitative estimates; and the output and reporting component, which translates the quantitative estimates into business information.

The hesitance to include ML procedures into model is largely due to the second component – the ML processing component – which is criticized as black box lacking understandable descriptions not mentioned to compare with the explicit mathematical formula as presented in classical statistical models. Progresses have been made to shed lights on such black box by making the complex algorithms more explicit and trackable. In addition, the lack of stability and statistical inference with ML procedures has also started to improve with more and more efforts to enhance the ML procedures with sensitivity analysis and performance monitoring. There is also the argument that the black box does not need to be transparent to human as long as the machine can "understand" it and track the input and output relation as a general nonparametric function.

A more practical reason to not include ML procedures into the model class is that the standard requirements for model risk management could make some of the ML procedures hard to be fully implemented. For example, model selection with ML procedures is somewhat arbitrary and lack of solid defense when challenged with alternatives. In Sect. 3.3, we present more discussion on ML model selection.

Even with all these challenges, due to their powerful insight discovery ability on large and complex data, some ML procedures should be classified as models and managed with proper level of model risks. Both the regulators and practitioners have been actively pursuing a balance between accepting ML procedures into the

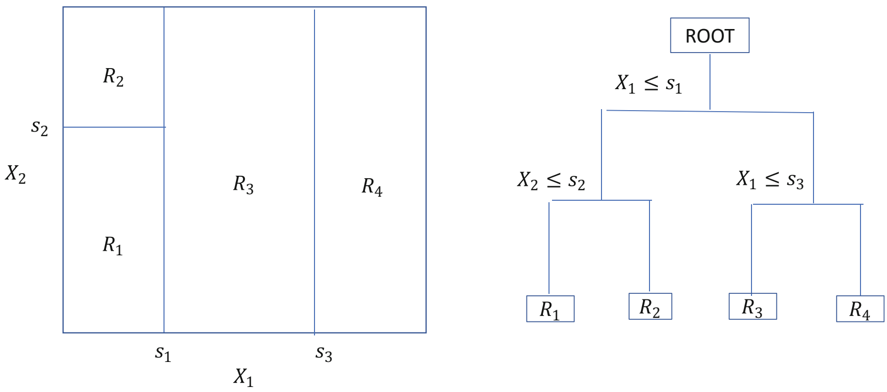


Fig. 3.3 Binary splitting

regulatory model class and proper model risk management. We hope such a balance could be achieved in the future in credit risk modeling.

For credit risk modeling with binary targets, we will focus on the tree-based, bagging, and boosting ML models. We consider them as models at least based on our informal model definition given early.

Classification and Regression Tree (CART)

Tree-based methods are considered nonparametric as they do not specify function formats between the target and factors; instead they partition the factor space into a set of rectangles and then fit a simple model (like a constant) in each one. They are conceptually simple yet powerful. Tree-based classification and regression are the most popular ML models in various applications.

Let's first consider a *regression tree model*. To be simple, we assume there are two factors X_1 and X_2 standardized to unit and the target Y is continuous with N observations $\{x_j, y_j\}$, $x_j = (x_{j1}, x_{j2})$, $j = 1, \ldots, N$. Without specification on any relationship between the target and factors, the regression tree model starts from binary splitting using one factor X_1, as shown in the left side of Fig. 3.3, at a selected point s_1, and then recursively does such binary splitting on the subregions based on new selected factors and points, X_2 at s_2 and X_1 at s_3 as shown in Fig. 3.3. After these binary splits, the entire factor space is petitioned into four subregions, and then a constant in each subregion is used as the estimate of the target, which means the regression tree model fits a function:

$$f(X_1, X_2) = \sum_{i=1}^{4} c_i \, I\{(X_1, X_2) \in R_i\} \tag{3.37}$$

which is the representation of the tree model structure shown in the right side of Fig. 3.3. If we choose minimization of the sum of squares $\sum_{j=1}^{N} \left(y_j - f\left(x_{j1}, x_{j2}\right)\right)^2$ as our criterion, it is easy to see that the best c_i is just the average of y_j in region R_i, $c_i = \text{ave}(y_j | (x_{j1}, x_{j2}) \in R_i)$. This can be generalized to K-factors $X = (X_1, \ldots, X_K)$ and M partitioned subregions:

$$f(X) = \sum_{i=1}^{M} c_i \, I\{X \in R_i\} \tag{3.38}$$

with the criterion $\sum_{j=1}^{N} \left(y_j - f(x_j)\right)^2$ minimized at $\widehat{c}_i = \text{ave}\left(y_j | x_j \in R_i\right)$. However, we have the following list of questions remained to answer:

1. How to decide which factor to use as a splitter and the split point?
2. How to decide when to stop the splitting and the tree size?

The first question can be answered by a greedy algorithm. Starting from all data, for a pair of factor and split point (k, s) combination, the factor space is partitioned into two half-plans:

$$R_1(k, s) = \{X | X_k \leq s\} \text{ and } R_2(k, s) = \{X | X_k > s\} \tag{3.39}$$

Then we seek the splitting factor k and split point s that solve:

$$\min_{(k,s)} \left\{ \min_{c_1} \sum_{x_j \in R_1(k,s)} \left(y_j - c_1\right)^2 + \min_{c_2} \sum_{x_j \in R_2(k,s)} \left(y_j - c_2\right)^2 \right\} \tag{3.40}$$

For any chosen factor k and split point s, the minimization is achieved at:

$$\widehat{c}_1 = \text{ave}\left(y_j | \mathrm{x}_j \in R_1(k, s)\right) \text{ and } \widehat{c}_2 = \text{ave}\left(y_j | \mathrm{x}_j \in R_2(k, s)\right) \tag{3.41}$$

For each splitting factor, the determination of the split point s can be done by scanning through all of the input points. So, determination of the best pair (k, s) is feasible. Having found the best split, we partition the data into the two resulting regions and repeat the splitting process on each of the two regions. This process can be repeated on all of the resulting regions. Then, we need to answer the second question with a stopping rule.

Clearly a very large tree might overfit the data, while a small tree might not capture the important structure. Tree size is a tuning parameter governing the model's complexity, and the optimal tree size should be adaptively chosen from the data. One approach would be to split tree nodes only if the decrease in sum-of-squares due to the split exceeds some threshold. This strategy is too short-sighted, however, since a seemingly worthless split might lead to a very good split below it.

The preferred strategy is to grow a large tree T_0, stopping the splitting process only when some minimum node size (say 5) is reached. Then this large tree is pruned using cost complexity pruning, which is defined as follows:

Define a subtree $T \subset T_0$ to be any tree that can be obtained by pruning T_0, that is, collapsing any number of its internal (non-terminal) nodes. For any node m in a tree T, its impurity is defined as:

$$Q_m(T) = \frac{1}{\mid R_m \mid} \sum_{x_i \in R_m} (y_i - \widehat{c}_m)^2 \tag{3.42}$$

and the cost complexity criterion is defined as:

$$C_\alpha(T) = \sum_{m=1}^{|T|} \sum_{x_i \in R_m} (y_i - \widehat{c}_m)^2 + \alpha|T| = \sum_{m=1}^{|T|} |R_m|\, Q_m(T) + \alpha|T| \tag{3.43}$$

where $|T|$ denotes the tree size (number of terminal nodes or final regions), R_m denotes the mth region and its size is $|R_m|$ and target mean is $\widehat{c}_m$, and α is a prespecified constant to penalize the tree complexity.

The idea is to find, for each α, the subtree $T_\alpha \subset T_0$ to minimize $C_\alpha(T)$. The tuning parameter $\alpha \geq 0$ governs the tradeoff between tree size and its goodness-of-fit to the data. Large values of α result in smaller trees T_α and conversely for smaller values of α. As the notation suggests, with $\alpha = 0$, the solution is the full tree T_0. For each $\alpha \geq 0$, one can show that there is a unique smallest subtree T_α that minimizes $C_\alpha(T)$. This can be done using the *weakest link pruning*: successively collapse the internal node that produces the smallest per-node increase in the sum of squares and continue until we produce the single-node (root) tree. This gives a finite sequence of subtrees, and one can show this sequence must contain T_α. See Breiman et al. (1984)[5] for details.

Like all other penalized optimization, the penalizing constant α can be estimated using cross-validation based on the input data and that makes the tree method self-complete as a statistical model.

For *classification tree model*, if the target is a classification outcome taking classes 1, 2,. . ., L, the only changes needed in the tree algorithm pertain to the criteria for splitting nodes and pruning the tree. For regression, we used the squared-error node impurity measure $Q_m(T)$, but this is not suitable for classification. In a node m, representing a region R_m, let

$$\widehat{p}_{ml} = \frac{1}{\mid R_m \mid} \sum_{x_i \in R_m} I\{y_i = l\} \tag{3.44}$$

be the proportion of l class observations. We classify the observations in node m to class $l(m) = \text{argmax}_l\, \widehat{p}_{ml}$, the majority class in node m. So, for classification tree, we can use the following node impurity measures $Q_m(T)$:

[5] Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984). Classification and Regression Trees, Wadsworth, New York.

$$\text{Misclassification error}: \frac{1}{\mid R_m \mid}\sum\nolimits_{x_i \in R_m} I\{y_i \neq l(m)\} = 1 - \widehat{p}_{ml(m)}$$
$$\text{Gini index}: \sum\nolimits_{l \neq l'} \widehat{p}_{ml}\widehat{p}_{ml'} = \sum\nolimits_{l=1}^{L} \widehat{p}_{ml}(1 - \widehat{p}_{ml}) \tag{3.45}$$
$$\text{Cross} - \text{entropy}: \ -\sum\nolimits_{l=1}^{L} \widehat{p}_{ml} \log(\widehat{p}_{ml})$$

For binary classification, if p is the proportion of the event, these three measures are $1 - \max(p, 1 - p)$, $2p(1 - p)$, and $-p \log p - (1 - p) \log(1 - p)$, respectively. All three are similar, but cross-entropy and the Gini index are differentiable and hence computationally easier for numerical optimization.

With these node impurity measures, the cost complexity for classification tree can be computed according to (3.43) similarly as for the regression tree and thus the tree growing and pruning. Since the Gini and entropy criteria more likely select purified nodes than the misclassification criterion, they are preferred in the tree growing, while the misclassification criterion is preferred in the tree pruning.

The key advantages of using trees can be summarized as follows:

- Easy to explain. Trees can be displayed graphically and can be easily interpreted even by nonexpert.
- Handle continuous and categorical factors. Categorical factors can be used without the need to create dummy variables. When categorical factors are ordinal – having ordered values – the splitting is the same as for continuous factors. When a categorical factor has q possible unordered values, there are $2^{(q-1)} - 1$ possible partitions of the q values into two groups, and the computations become prohibitive for large q. However, with a 0–1 outcome, this computation simplifies. We order the factor classes according to the proportion falling in outcome class 1. Then we split this factor as if it were an ordered factor. One can show this gives the optimal split, in terms of cross-entropy or Gini index, among all possible $2^{(q-1)} - 1$ splits. This result also holds for a quantitative outcome and square error loss – the categories are ordered by increasing mean of the outcome.
- Handle missing values for factors. One way is treating the missing values as a separate class if the factor is categorical. The more general method for factors with missing values is using surrogate splitters. When deciding the best (primary) splitter from the factors, only use observations without missing values for that factor and keep a list of all possible splitters by different factors (called surrogate splitters) ordered by their performance in mimicking the primary splitter. Then when constructing the tree and carrying out prediction, use the highest correlated surrogate splitter when the primary splitter has missing values.
- Mirror human decision-making. Trees are in many cases easier to grasp than regression, given that their mechanics are usually aligned with business decision-making.

In terms of disadvantages, tree models are not smoothing and have hard time to fit additive models. Also, trees do not usually have the same level of predictive accuracy of other methods, such as regression. The most critical disadvantage is the instability of the tree model. Some small data variation could lead to large variation in both the tree structure and the tree outcome. Such instability is due to the cumulative error from the recursive splitting process. Nevertheless, some enhancements, for instance, bagging, random forest, and boosting, improve their predictive performance, as detailed in the following section.

Bagging, Boosting, and Random Forest

Similar to tree models, most algorithm-based data-driven models have the disadvantage of instability. One approach to alleviate such instability is “smoothing” the model outputs based on some kind of average.

The bootstrap approach by Efron (1979)[6] is a general statistical approach for the estimation of sampling distribution of any statistics by using random sampling with replacement. It falls in the broader resampling methods. In statistical inference, it is popularly used as a way of assessing the accuracy of a parameter estimate or a prediction. Bootstrap can also be used to improve the estimate or prediction through bootstrap aggregation (or bagging), which averages such estimate or prediction over a collection of bootstrap samples, thereby reducing its variance. Bagging follows a simple procedure:

- Training Data Generation. A prespecified number of B bootstrapped training data sets are generated by resampling the original model data set with replacement.
- Model Building. A separate model is built on each training data set based on the specified modeling approach (e.g., a tree model by CART).
- Outcome Averaging. As a final step, in case of quantitative outcome, the average of estimates or predictions is computed. In the case of classification trees, one may record the class estimated or predicted by each of the B trees and take a majority vote. The overall estimate or prediction is the most commonly occurring class among all B outcomes. For binary outcomes, if the focus is on the class probability instead of classification, averaging should be used instead of voting.

In the bagging procedure, when generating the bootstrapped data set, a natural test data set is created, which includes the observations not sampled into the bootstrapped data set and is referred to as the out-of-bag (OOB) observations. So, to check the model stability during model training, instead of checking test error on the specified test data set, we can use the OOB error. For each observation in the training data set, we can predict the outcome by means of the models for which that observation was OOB. We can average these predicted outcomes (if regression is the goal) or can take a majority vote (if classification is the goal) in order to obtain the prediction for a specific observation. The resulting OOB error is a valid estimate of

[6]Efron, B. (1979). Bootstrap methods: another look at the jackknife, Annals of Statistics 7: 1–26.

the test error for the bagged model, since the outcome for each observation is predicted using only the trees that did not fit using that observation.

With CART, bagging can be seen as an improved accuracy estimate over a single tree. However, model interpretation may become difficult. One can obtain a summary of the importance of each predictor, using the residual sum of squared (RSS) in the case of regression trees or the Gini index for bagging classification trees. In the first case, all RSS deductions due to the use of the variable as a splitter are summed over all B trees as the variable importance measure. Similarly, in the case of classification tree, Gini index decreases by splits over a given predictor are summed over all B trees.

Boosting works similarly to bagging except that the models are built sequentially. It is extended from the gradient optimization procedure, in which the optimal solution is approximated recursively by a small additional shrinkage step in the direction of the gradient. Each model is fitted using information from previously fitted model. Boosting does not involve bootstrap sampling. Instead, each model is fitted on a modified version of the original data set. Given the current model, a new model is fitted to the residuals from the previous model. That is, the response variable is not the original one, but residuals from a previous fitted model. The procedure is slightly more complex in the case of classification trees, for which the subsequent model focuses on the misclassified observations from the previous fitted model. The process is deeply impacted by the step shrinkage parameter, which controls the rate at which boosting learns. Boosting is trained to combine many weak learners to form a strong model instead of finding the best learner. So, when boosting, one should avoid overfitting by maintaining a relatively simple model structure.

Random forest (Breiman, 2001)[7] is a modified version of bagging. In addition to improving the single tree stability, it substantially de-correlates a large collection of trees built from bagging and then averages them. On many problems, the performance of random forests is very similar to boosting, and they are simpler to train and tune. As a consequence, random forests are popular and are implemented in a variety of packages.

The essential idea in bagging is to average many noisy but approximately unbiased models and hence reduce the variance. Trees are ideal candidates for bagging, since they can capture complex interaction structures in the data, and, if grown sufficiently deep, have relatively low bias. Since trees are notoriously noisy, they benefit greatly from the averaging. Moreover, since each tree generated in bagging is identically distributed (i.d.), the expectation of an average of B such trees is the same as the expectation of any one of them. This means the bias of bagged trees is the same as that of the individual bootstrap trees, and the only hope of improvement is through variance reduction. This is in contrast to boosting, where the trees are grown in an adaptive way to remove bias and hence are not i.d.

[7] Breiman, L. (2001). Random forests, Machine Learning 45: 5–32.

An average of B independent and identically distributed (i.i.d) random variables, each with variance σ^2, has variance $\frac{1}{B}\sigma^2$. If the variables are simply i.d. (identically distributed, but not necessarily independent) with positive pairwise correlation ρ, the variance of the average is:

$$\rho\sigma^2 + \frac{1-\rho}{B}\sigma^2 \tag{3.46}$$

As B increases, the second term disappears, but the first remains, and hence the size of the correlation of pairs of bagged trees limits the benefits of averaging. The idea in random forests is to improve the variance reduction of bagging by reducing the correlation between the trees, without increasing the variance too much. This is achieved in the tree-growing process through random selection of the input variables.

Algorithm *Random Forest for Regression or Classification* (Hastie et al., 2009)[8]

1. *For $b = 1$ to B:*
   (a) *Draw a bootstrap sample of size N from the training data.*
   (b) *Grow a random forest tree T_b using the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size $n_{\min}$ is reached:*
      I. *Select m variables at random from the K variables.*
      II. *Pick the best variable/split point among the m.*
      III. *Split the node into two child nodes.*
2. *Output the ensemble of trees $\{T_b\}_1^B$. To make a prediction at a new point x:*

   Regression: $\widehat{f}_{rf}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$

   Classification: Let $\widehat{C}_b(x)$ be the class prediction of the bth random forest tree.

   Then, $\widehat{C}_{rf}(x) = \textit{majority vote}\left\{\widehat{C}_b(x)\right\}_1^B$.

Typically values for m are $\sqrt{K}$ or even as low as 1. Not all estimators can be improved by shaking up the data like this. It seems that highly nonlinear estimators, such as trees, benefit the most. For bootstrapped trees, ρ is typically small (0.05 or lower is typical), while σ^2 is not much larger than the variance for the original tree. On the other hand, bagging does not change linear estimates, such as the sample mean (hence its variance either); the pairwise correlation between bootstrapped means is $N/(2N - 1)$, about 50%.

Random forests also use the OOB samples to construct a different variable importance measure, apparently to measure the prediction strength of each variable. When the bth tree is grown, the OOB samples are passed down the tree, and the

---

[8] Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer.

prediction accuracy is recorded. Then the values for the jth variable are randomly permuted in the OOB samples, and the accuracy is again computed. The decrease in accuracy as a result of this permuting is averaged over all trees and is used as a measure of the importance of variable j in the random forest. Compared to the addition/removal-based tree importance measure, although the rankings of the two methods are similar, the OOB importance is usually more uniform over the variables. The randomization effectively voids the effect of a variable, much like setting a coefficient to zero in a linear model. This does not measure the effect on prediction where this variable is not available, because if the model was refitted without the variable, other variables could be used as surrogates.

3.1.2 Competing Risk Framework

The competing risk framework was first developed in survival analysis, where termination could be caused by competitive diseases. It was used as a more accurate description of the risk decision process. For credit risk, the competing risk framework is designed to simultaneously model different target event risks, which don't have sequential order when making decision to take these risks. The target event risks are the main risks resulting in credit risks, for example, in both retail and wholesale lending businesses, the risks of default and payoff of a loan. The competing risk framework is considered as a more accurate description of the credit risk decision process when either default or payoff decision could terminate the loan and neither decision dominates the other.

The most popular model used for competing risks is the multinomial logit model (MNL), which is a special case of the conditional logit model derived from the random utility maximization (RUM) paradigm by the Nobel laureate Daniel McFadden (1974)[9] for multiple choices. Though MNL is theoretically sound for modeling competing risks, the estimation for MNL is notoriously expensive in computing, and a binary logit approximation can be used. We explain some cautions with such binary approximation beyond estimation. Finally, some machine learning models are explored for competing risks.

3.1.2.1 Multinomial Logit Model

A good interpretation of the multinomial logit model is the random utility maximization. Consider a decision-maker facing a finite set of mutually exclusive alterative: A: $\{1, 2,\ldots, m\}$. The utility for choosing different alternative is:

$$U_{ij} = x_i'\beta_j + z_{ij}'\alpha + \varepsilon_{ij}, j \in \{1, \ldots, m\} \tag{3.47}$$

[9] McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, (ed.), Frontiers in Econometrics, New York: Academic Press, 105–142.

where the subscript i denotes the individual and the subscript j denotes the choice. Notice that in (3.47), we distinguish two sets of factors, x_i (individual specific) and z_{ij} (individual and choice specific). For example, x_i might be age, gender, or income of the individual. z_{ij} can be the momentary costs to commute that depend on both individual spending and the choice of transportation modes (bus, car, or train) as in McFadden's original example. The utility maximizing rule is individual i chooses choice k: $y_i = k$ if and only if:

$$U_{ik} \geq U_{ij} \text{ for all } j \neq k \tag{3.48}$$

Lemma 3.2 If $\varepsilon_{i1}, \ldots, \varepsilon_{im}$ are i.i.d extreme value distributed with density $g(u) = \exp(-u - \exp(-u))$, then the conditional probability of choice k chosen is:

$$p_{ik} = P[y_i = k | x_i, z_{i1}, \ldots, z_{im}] = \frac{\exp\{V_{ik}\}}{\sum_{j=1}^{m} \exp\{V_{ij}\}} \tag{3.49}$$

where $V_{ij} = x_i'\beta_j + z_{ij}'\alpha$.

The model specification in (3.49) is called multinomial logit model when there are no factors that are choice specific. In this sense, the multinomial logit model is a special case of the conditional logit model. For credit risk modeling, we will focus on the multinomial logit model.

Let x_i be the vector of explanatory variables for individual i and y_i be the choice made by individual i. The multinomial logit model postulates that:

$$p_{ik} = P[y_i = k | x_i] = \frac{\exp\{x_i'\beta_k\}}{\sum_{j=1}^{m} \exp\{x_i'\beta_j\}}, k = 1, \ldots, m \tag{3.50}$$

It should be pointed out that MNL models for alternative choice with different set of factors are included by expanding the parameter vector and setting its components corresponding to alternative choices to 0's, though such setting will lead to restricted parameter estimation.

In the extreme case, each choice has its unique set of factors. Denote the choice-specific factor sets to be $x_{i1}, \ldots, x_{im}$. If we write $x_i = \left(x_{i1}', \ldots, x_{im}'\right)'$ and $\beta_1 = \left(\delta_1', 0', \ldots, 0'\right)'$, $\beta_2 = \left(0', \delta_2', \ldots, 0'\right)', \ldots$, $\beta_m = \left(0', \ldots, 0', \delta_m'\right)'$, then the model takes the form of (3.50). Again, the parameter estimation with $\beta_1, \ldots, \beta_m$ becomes restricted.

Independence of Irrelevant Alternatives (IIA)

MNL has a strong assumption on the probabilities of alternative choices. From (3.50), we obtain:

$$p_{ij}/p_{ik} = \exp\left(x_i'\left(\beta_j - \beta_k\right)\right), \text{ for all } j \neq k \tag{3.51}$$

which indicate that the ratio of probabilities of two alternatives is not impacted by the presence of any other alternatives – both a property and an assumption of MNL. This assumption can be relaxed with some other models (multinomial probit, nested logit).

The IIA property leads to linear log odds ratio (the proportional odds ratio as in the binary case) – the log odds ratio between any two alternatives is linear in the factors. This gives a convenient interpretation for the coefficients.

Model Estimation

As for binary logit model, maximum likelihood estimation (MLE) is the most popular method for MNL parameter estimation. Since $\sum_{j=1}^{m} p_{ij} = 1$, only $m - 1$ sets of parameters are identifiable. By parameter normalization convention, the last set $\beta_m = 0$. So:

$$\begin{aligned} p_{ij} = P[y_i = j | x_i] &= \frac{\exp\left(x_i'\beta_j\right)}{1 + \sum_{k=1}^{m-1} \exp(x_i'\beta_k)}, j = 1, \ldots, m-1 \\ p_{im} = P[y_i = m | x_i] &= \frac{1}{1 + \sum_{k=1}^{m-1} \exp(x_i'\beta_k)} \end{aligned} \tag{3.52}$$

Given the observations $\{x_i, y_i\}$, $i = 1, \ldots, N$, let $D_{ij} = 1$ if $y_i = j$ and $D_{ij} = 0$ if $y_i \neq j$ for $i = 1, \ldots, N$, $j = 1, \ldots, m$, and then the loglikelihood function is:

$$\log(L(\beta)) = \sum_{i=1}^{N} \sum_{j=1}^{m} D_{ij} \log\left(p_{ij}\right) = \sum_{i=1}^{N} \sum_{j=1}^{m} D_{ij} x_i'\beta_j - \sum_{i=1}^{N} \log\left(1 + \sum_{k=1}^{m-1} \exp\left(x_i'\beta_k\right)\right) \tag{3.53}$$

So, MLE $\widehat{\beta}$ solves the following $m - 1$ sets of first-order equations:

$$\begin{cases} \sum_{i=1}^{N} [D_{i1} - \widehat{p}_{i1}] x_i = 0 \\ \sum_{i=1}^{N} [D_{i2} - \widehat{p}_{i2}] x_i = 0 \\ \cdots\cdots \\ \sum_{i=1}^{N} \left[D_{i,m-1} - \widehat{p}_{i,m-1}\right] x_i = 0 \end{cases} \tag{3.54}$$

where $\widehat{p}_{ij} = \exp\left(x_i'\widehat{\beta}_j\right) / \left\{1 + \sum_{k=1}^{m-1} \exp\left(x_i'\widehat{\beta}_k\right)\right\}$. There is no analytics solution for $\widehat{\beta}$. However, as for the binary logit model MLE estimation in Sect. 3.1.1.1, under the following general conditions, $\widehat{\beta}_{\mathrm{ML}}$ is consistent, asymptotically normal, and efficient.

Sufficient Conditions

- y_i are independent condition on $X_i = x_i$.
- $X = [x_1, \ldots, x_N]'$ is full rank and bounded.
- $\frac{1}{N} X' X$ converges to a positive definite matrix.

In summary, the MLE of the MNL logit model has the following properties:

- $L(\beta)$ is globally concave and MLE $\widehat{\beta}_{\mathrm{ML}}$ is unique (and finite assuming no data separation).
- If the MNL model has an intercept, the normal equations show that the sample mean of response equals to the mean of predicted response:

$$N^{-1} \sum_i D_{ij} = N^{-1} \sum_i \widehat{p}_{ij}, j = 1, \ldots m$$

- If the MNL model includes categorical covariates, the same normal equation implies that the mean response equals to the mean of predicted response for each bucket of the covariate.
- Under very *general conditions* shown below, $\widehat{\beta}_{\mathrm{ML}}$ is consistent, asymptotically normal, and the most efficient.

Goodness-of-Fit

For MNL, all likelihood-based goodness-of-fit measures, R^2 and adjusted $\overline{R}^2$, AIC, and SIC, are the same as that in the binary logit model.

For the rank correlation measures, the predicted mean score of an observation is the sum of the ordered values of the responses minus 1, weighted by the corresponding predicted probabilities for that observation. So, the predicted mean score for the ith observations is:

$$\text{Preditcted Mean Score} = \sum\nolimits_{j=1}^{m} (j-1) \widehat{p}_{ij} \tag{3.55}$$

The predicted mean score depends on the order of the responses. If the responses have an order, that can be used naturally. If not, one needs to specify the order. If aligning with the parameter normalization convention and specifying rank 1 for non-event, the predicted mean score is just the predicted event probability in the binary case, and the rank correlation measures defined in Sect. 3.1.1.1 for the binary case can be extended to MNL.

- A pair of observations with different observed responses is said to be concordant if the observation with the lower response value has a lower predicted mean score than the observation with the higher response value. If the observation with the lower response value has a higher predicted score than the observation with the higher response value, then the pair is discordant. If the pair is neither concordant nor discordant, it is a tie (which means the same predicted scores for the pair). The following rank corrections are popularly used:

$$N - \#\text{of obs.} t - \#\text{of pairs}, n_c - \#\text{of concordant}, n_d - \#\text{of discordant}$$

$$c\ (\text{Concordance Index}) = \frac{n_c + 0.5(t - n_c - n_d)}{t}$$

$$\text{Somers}'\ \text{D}\ (\text{Gini}) = \frac{n_c - n_d}{t}$$

$$\text{Goodman} - \text{Kruskal Gamma} = \frac{n_c - n_d}{n_c + n_d}$$

$$\text{Kendall's Tau} - \text{a} = \frac{n_c - n_d}{0.5N(N-1)}$$

Note that without ties, $t = n_c + n_d$ and Somers' D (Gini) $= 2c - 1$. The higher these values, the better the fit.

For the ROC curve, the multinomial case needs to be binarized – taking one outcome as the risk event and the rest outcomes as non-event. Such binarized goodness-of-fit measures will show how well the MNL model fits the individual outcome.

Variable Specification and Selection
As for the binary generalized linear models, the classical variable selection methods such as forward, backward, and stepwise can be used for MNL. However, for large data, MNL variable selection is computational expensive due to a large number of parameters needed to be estimated. A common practice is that variable selection and specification has already been well studied by other approaches (e.g., CART) before fitting MNL and only some minor adjustments are carried within MNL modeling process.

3.1.2.2 Binary Logit Approximation

Simultaneously estimating $m - 1$ set of parameters as in MNL is costly and time-consuming even with today's most powerful computing techniques (especially with different sets of variables for alternative risk events as in some credit risk models). Statistics literatures have suggested estimating parameters separately using a sequence of $m - 1$ binary logit regression[10] as in the following process:

[10] R. L. Prentice, J. D. Kalbfleisch, A. V. Peterson, Jr., N. Flournoy, V. T. Farewell, N. E. Breslow. (1978). "The Analysis of Failure Times in the Presence of Competing Risks." Biometrics, Vol. 34, No. 4, pp. 541–554.

Judith E. Singer and John B. Willett. (1993). "It's About Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events." Journal of Educational Statistics, Vol. 18, No. 2, pp. 155–195.

The BLA Approach

- **Step 0.** From the choice set $A = \{1, \ldots, m\}$, choose a benchmark alternative, say $j = $ m.
- **Step 1.** To estimate β_1, take a subsample that consists of all observations that either $D_{i1} = 1$ or $D_{im} = 1$. That is, temporarily "delete" all the observations with "irrelevant" choices. With this subsample, estimate the binary logit model using D_{i1} as the dependent variable.
- **Step 2.** To estimate β_2, take a subsample that consists of all observations that either $D_{i2} = 1$ or $D_{im} = 1$. With this subsample, estimate the binary logit model using D_{i2} as the dependent variable.
- **Step $m - 1$.** To estimate β_{m-1}, take a subsample that consists of all observations that either $D_{im-1} = 1$ or $D_{im} = 1$. Estimate the binary logit model using D_{im-1} as the dependent variable on this subsample.

The BLA process generates unbiased parameter estimates for MNL due to the following facts:

Upon deletion from the data of all the observations for which $y_i \notin \{j, m\}$, the dependent variable y_i takes only two values $\{j, m\}$. By the conditional probability principle and the normalized model formula (3.52):

$$P(y_i = j|x_i, \{j, m\}) = \frac{P(y_i = j|x_i, A)}{P(y_i = j|x_i, A) + P(y_i = m|x_i, A)} = \frac{\exp\left(x_i'\beta_j\right)}{1 + \exp\left(x_i'\beta_j\right)} \tag{3.56}$$

which is the binomial logit model with β_{j} as its parameter. That means we can use the binomial logit model on the subsample to estimate the parameter of the original MNL model piece by piece. For all these binomial logit estimates, a common alternative choice m is proposed as the reference.

In practice, we not only need point estimates for $(\beta_1, \ldots .\beta_{m-1})$ but we also need to know how precisely those point estimates are. For efficiency considerations, the following condition is needed for the BLA approach to be reasonably accurate:

Condition A: *In the chosen set A, there exists an alternative that is the predominant choice in the data.*

This predominant choice will be used as the benchmark choice, taking the role of choice m in the above discussion. Using the predominant alternative as the benchmark assures each binary logit model fit has the maximum number of observations, therefore minimizing the sampling errors of the estimate from those model fits. In credit risk modeling, the non-event or "current" choice is commonly used as the predominant benchmark choice.

In summary, compared to the MNL, the BLA approach has the following advantages and disadvantages:

Advantages of the BLA Approach

- The main advantage of BLA is its saving in computing time (make feasible for some large data runs).

- More computer packages have binary logit regression than MNL.
- BLA is asymptotically unbiased due to the IIA property of MNL as shown by (3.56).
- Under certain conditions (existence of predominant benchmark choice), the BLA is highly efficient as compared with MNL.

Disadvantages of the BLA Approach

- Because $m - 1$ sets of parameters are estimated separately, the correlation structure between any two sets, β_j and β_k, is not estimated and precludes any attempts to test restriction involving parameters across sets.
- Event probability p_{ij} depends on all parameters by MNL construction, and confidence interval and other inference will be biased without proper correlation estimates across sets.
- BLA depends on IIA, which is a strong assumption.

3.1.2.3 Machine Learning Models

For multinomial data, one popular use of the machine learning models is the variable importance analysis using multiclass tree. Among the three impurity criteria listed in (3.45), since the Gini and entropy criteria more likely select purified nodes than the misclassification criterion, they are preferred in the variable importance analysis, which depends more on tree growing than tree pruning.

3.1.3 Multistate Transition Framework

Multistate transition model, also called transition matrix model (TMM), is popularly used for short-term credit loss forecasts. Instead of only focusing on terminal events as in the competing risk framework, the TMM projects the full trace of the account performance based on the current status and projected transition probabilities over a specified period of time. It is preferred for ACL (pre-CECL) and balance management due to its transparency with account or portfolio transition over the forecast period.

A typical TMM can be summarized by a transition matrix as shown in Table 3.2. Besides the terminal states charge-off and payoff, a full set of delinquency states

Table 3.2 A transition matrix

	D0	D30	D60	D90	D120	D150	D180	Chargeoff	Payoff
D0	X	X	X	X	X	X	X	X	X
D30	X	X	X	X	X	X	X	X	X
D60	X	X	X	X	X	X	X	X	X
D90	X	X	X	X	X	X	X	X	X
D120	X	X	X	X	X	X	X	X	X
D150	X	X	X	X	X	X	X	X	X
D180	X	X	X	X	X	X	X	X	X

from current to default depending on the periodic observations are tracked as events. Here D0 is defined as current or less than 30 days delinquent, D30 is defined as delinquent with 30 or more days but less than 60 days, and so on till D180, which is defined as delinquent with 180 or more days or default according to the SIFMA/BMA standards for certain credit products (e.g., mortgage and credit cards). The transition matrix defines the transition probabilities from one state to the next state, which could be a terminal state (Table 3.2 defines default as a non-terminal state).

Once the transition matrix is given, the full account trace over a specified period can be tracked, and the probability-weighted balance for each state can be computed over the full period. By summing up these weighted balances for each account, one can obtain the state-specific balances over the period for the entire portfolio. If account-level severity can be estimated for each state with potential gross loss, one can project the gross losses on the full trace of the account and sum up all account losses to get the portfolio losses.

So, for TMM, the key is to obtain the transition matrix. In the following sections, we discuss some common method to estimate the transition matrix and carry out loss forecasts.

3.1.3.1 Static Transition Matrix

A static transition matrix assumes that the transition probabilities from state to state do not change with time. This is a strong assumption, and very often it is used for some average transition analysis based on a pool of homogeneous accounts. The transition matrix can be estimated based on recent historical transition rates of a homogeneous pool.

3.1.3.2 Dynamic Transition Matrix

Static transition matrix can only be used for some summary or average transition analysis. In practice, the transition probabilities from one state to the next state will depend on both account characteristics and economic conditions. The dynamic transition matrix fits the account transition probabilities from state to state as a function of risk factors derived from the account characteristics and MEVs using historical transition rates of homogeneous account groups. Linear regression models are commonly used. However, such models can't guarantee the fitted probabilities are in the range between 0 and 1, and some control should be applied. Alternatively, one can use fractional response regression (FRR) technique to fit these transition probabilities to make sure the fitted values are within the expected range.

Due to a large number of transition probabilities in the transition matrix, finding proper models for all of them may not be feasible. One compromising method is combining the static and dynamic transition matrix approaches as a mixture:

- **Step 0**: Estimate a static transition matrix based on recent historical transition rates from a pool of homogeneous accounts in which the account falls in.
- **Step 1**: Estimate some key transition probabilities (e.g., those highlighted in yellow in Table 3.2) using regression techniques from risk factors.

- **Step 2**: Replace those key transition probabilities in the estimated transition matrix in Step 0 by the corresponding estimates obtained in Step 1.
- **Step 3**: Rescale the transition matrix obtained in Step 2 to make sure the probabilities in each row sum to 1.

There are different rescaling methods for Step 3. One can scale the changes caused by the replacement of these key transition probabilities across the entire row. Alternatively, one can just change some neighbor transition probabilities near the key transition probabilities. A justification should be given under either way.

3.1.4 Loss Given Default Models (LGD)

As a next step of credit risk measurement after the probability of default, loss given default measures the loss impact of a default event as a percentage of the exposure. It counts all post-default recoveries and costs. Based on the applications, for accounting and stress testing purposes, LGD (more often called loss severity) counts total net accounting loss including costs as a percentage of the exposure at default. For capital purpose, LGD counts total economic loss as the difference between the exposure at default and the total discounted (with an effective rate) post-default net recovery excluding costs and presents this difference as a percentage of the exposure at default. The LGD ratio should fall in the range of $[0, 1]$ with 0 and 1 as two extremes indicating full recovery and total loss, respectively. Though, conceptually, there could be cases that the total net loss may exceed the total exposure as well as the total net recovery may exceed the total exposure, such cases are rare (especially, when discounted with an effective rate). So, in practice, when the LGD ratio falls outside $[0, 1]$, some careful investigation should be carried out for data quality issues. This indicates that a proper reference default data set, which includes post-default recoveries and costs, account and loan characteristics, and macroeconomic and other risk factors at the default time, should be prepared for the targeted portfolio LGD modeling with high quality. LGD model complexity should not be motivated by low-quality reference default data.

The LGD model predicts the LGD as a percentage of the exposure at default. There are different approaches to build an LGD model. When post-default transaction data are available, which include the resolution time marks, gross losses, and exposures, the resolution (gross loss or payoff) probabilities and severities can be modeled along the post-default transaction path and used for the cash flow calculation. For capital purpose, the LGD is calculated as one minus the ratio of present value of the cash flow with an effective discount rate and the exposure at default. For accounting and stress testing, the LGD is just the resolution probability-weighted severity. The LGD model based on the resolution probabilities and severities is called the microstructure model and requires detailed post-default transaction data. It is suitable for collateral-based credit products for which a resolution is the main recovery through the disposal of the collateral, e.g., mortgage or auto loans. It is not suitable for credit products for which the post-default recovery may include many collections, e.g., credit cards or line of credits.

An alternative approach is to directly predict LGD as a function of the risk factors measured at the default time or simply use historical averages of the actual LGD when such relationship with risk factors could not be established. We call this approach the LGD regression method. The actual LGD is calculated from historical post-default recoveries and costs in the reference default data set. As we mentioned early, the actual LGD calculation varies according to applications. By tradition, for accounting and stress testing purposes, the actual LGD (or loss severity) is the net accounting loss including costs based on collateral deficiency and expenses/costs as a percentage of the exposure at default. For capital purpose, the economic loss is first calculated as the difference between the exposure at default and the net present value of post-default cash flow, and then the actual LGD is the economic loss as a percentage of the exposure at default. Regression models are preferred to fit the actual LGD as a function of the risk factors at default, which include Tobit regression, beta regression, and more general fractional response regression.

A finer version of the regression method is classifying the reference data set into homogeneous groups based on different post-default resolution methods, e.g., REO, short sale, third-party sale, charge-off, payoff, and modification with mortgage. A multinomial or tree-based model is used to predict the probabilities for these resolution methods, and different regression models fit the LGD for each of the resolution methods as a function of the risk factors at default. A probability-weighted average among all these resolution methods forms the final predicted LGD.

3.1.4.1 LGD Formulation and Components

As pointed out early, actual LGD is calculated differently for accounting and stress testing purposes (loss severity) from capital purpose. In the following, we describe the actual LGD formulation and components under different applications.

For capital purpose, the LGD for ith account is commonly calculated with the following formulas:

$$\mathrm{LGD}_{iT} = 1 - \frac{\mathrm{NPV}(\mathrm{RE}_{i,t>T}) - \mathrm{NPV}(\mathrm{EXP}_{i,t>T}) - \mathrm{NPV}(\mathrm{INT}_{i,t>T}) - \mathrm{ACC}_{i,T}}{\mathrm{EAD}_{iT}} + \mathrm{INCR} \tag{3.57}$$

where $\mathrm{RE}_{i,\ t\ >\ T}$ denotes all post-default recoveries from the account, $\mathrm{EXP}_{i,\ t\ >\ T}$ denotes all post-default expenses related to the account, $\mathrm{INT}_{i,\ t\ >\ T}$ denotes the missed interest payment from default till resolution, and $\mathrm{ACC}_{i,\ T}$ denotes the missed interest payment accrued to the default time from the last payment, and INCR denotes the incremental LGD adjustment for unresolved (open) loans in the reference default data set, NPV is the net present value, and EAD_{iT} is the exposure at default of the account.

For accounting and stress testing purposes:

$$\mathrm{LGD}_{iT} = \frac{\mathrm{CD}_{i,t>T} + \mathrm{EXP}_{i,t>T} + \mathrm{INT}_{i,t>T} + \mathrm{ACC}_{i,T}}{\mathrm{EAD}_{iT}} + \mathrm{INCR} \tag{3.58}$$

where $\mathrm{CD}_{i,\,t>T} = \mathrm{EAD}_{iT} - \mathrm{RE}_{i,\,t>T}$ is the collateral deficiency, and other items are the same as in (3.57).

While there might be some minor difference for the LGD components among different credit products, the components in (3.57) and (3.58) should cover most of the credit products. In the following, we illustrate each of these components and their subcomponents if more granular layers exist.

LGD Components

(a) ***Recoveries*** are comprised of the amounts from the following sources: (i) collateral sale proceeds[11]; (ii) collateral sale recovery[12]; (ii) customer deficiency, representing amounts collected from the borrower after charge-off[13]; (iii) loan sales; and (iv) other recoveries, such as dealer rebates, sales tax credits, insurance and credit enhancements, etc.
(b) ***Collateral deficiency*** is defined as the difference between EAD and *recoveries* here. It may also be defined as the difference between EAD and collateral-related recoveries, which exclude those defined as other recoveries in (**a**) and have those other recoveries treated separately.
(c) ***Expenses*** are comprised of all amounts from direct and indirect costs. The former refers to expenses specifically encountered for the account under investigation. On the other hand, indirect costs refer to general expenditure pro-quota allocated to defaulted account. Expenses consist of selling expenses and other expenses such as escrow expenses, attorney and broker expenses, asset maintenance, and other expenses.
(d) ***Forgone interest*** includes all interest payments lost from the time of default through the end of the loan resolution process (for resolved loans) or the last observation date (for unresolved loans). Forgone interest is calculated on the unpaid balance at an annualized rate equal to the discount factor.
(e) ***Accrued interest*** consists of the accrued interest in the period after the last payment and default. It is calculated using the annualized loan rate applied to the unpaid balance.
(f) ***Discount rate*** is used to calculate the net present value (NPV) in the capital application. By convention, it is defined as an index rate corresponding to the credit product class plus a premium decided by the management (e.g., the

[11] Recoveries from sale of collateral to third party, net of escrow expenses.

[12] A credit to the account post charge-off.

[13] Net of sales commissions due to the outside collection agencies.

30-year conventional mortgage rate plus 500 bps for regular time and 700 bps for downturn period).

(g) ***Exposure at default*** is the unpaid principal balance (UPB) for some credit products like mortgage, while it needs to be projected with a model for credit line-related products; see Sect. 3.1.5 for more details.

(h) ***Incremental LGD adjustment for unresolved loans*** is for unresolved loans, for which charge-offs and forgone interest need to be adjusted upward based on historical empirical evidence of resolved loans.

3.1.4.2 Default Data Processing

For LGD modeling, the reference default data set is critical. First, the selection of the reference defaults must be comprehensive and be consistent with the portfolio covered by the corresponding PD model such that the PD and LGD data are consistent. When the LGD reference default data are not sufficient and proxy reference default data are used, one must confirm the loss generating process for the target portfolio and the proxy default population are sufficiently similar. According to the different modeling approaches, the LGD reference default data are processed differently.

Post-default Transaction Data

Micro-structure model requires post-default transaction data. Given the resolution time marks, the post-default transactions should be expanded from the initial default date to the resolution date. For example, Table 3.3 illustrates one account's post-default transaction.

The account with an identification number ID1 became default at the second recording period (e.g., quarterly). After two periods of continuing being default, it was cured, which was marked as a payoff event for post-default resolution. After two periods of being current, it became default again and finally charged off after another five periods.

For this account, two post-default transaction series can be developed as shown in Table 3.4. Though they have the same account ID, they have different transaction attributes – time since default (Tdef), status, gross loss event indicator (GL), payoff event indicator (PO), and calendar time for the transaction (time). Such transaction data can be merged with account and loan characteristics data through the ID and macroeconomic variables through the calendar time to form the complete post-default dynamic modeling data.

Like the default/prepayment setup, either a sequential or competing risk framework can be used to model the conditional gross loss and payoff probabilities as a

Table 3.3 A single account post-default transaction

		ID1								ID1						
ID1	Dlq	Def	Def	Def	**Cured**	Current	Current	Dlq	Dlq	Def	Def	Def	Def	Def	Def	**CO**

Table 3.4 Post-default transaction data

ID	Tdef	Status	GL	PO	Time	Severity
ID1	t0	Def	0	0	mm/yy	
ID1	t1	Def	0	0	mm/yy	
ID1	t2	Def	0	0	mm/yy	
ID1	t3	Cured	0	1	mm/yy	XXX
ID1	t0	Def	0	0	mm/yy	
ID1	t1	Def	0	0	mm/yy	
ID1	t2	Def	0	0	mm/yy	
ID1	t3	Def	0	0	mm/yy	
ID1	t4	Def	0	0	mm/yy	
ID1	t5	Def	0	0	mm/yy	
ID1	t6	CO	1	0	mm/yy	XXX

function of account/loan characteristics, macroeconomic factors, and other default-related risk factors.

Severity for each of the resolution events needs to be aggregated at the resolution time, which includes collateral deficiency, expenses, forgone interest, and accrued interest and other recoveries and costs if available. Regression models are used to fit the severity as a function of account/loan characteristics, macroeconomic factors, and other default-related risk factors at this resolution time.

It should be pointed out that the definition of a resolution event may depend on the credit product loss mitigation process. For some products, e.g., mortgage or auto loan, one major asset sale or written-off could mark as the resolution if the remaining recoveries are minor. For other products, when the cumulative recoveries or written-offs exceed some limit, e.g., 95% of the exposure at default, then a resolution event is marked. Such open accounts can be included in the reference data by adjusting the severity up based on historical experience or a conservative judgment.

Account-Level Default Data with Actual LGD/Severity

Unlike the post-default transaction data that track the periodic evolving of the default account, the account-level default data collect all loss information at the default time. Table 3.5 shows a part of the account-level default data for mortgage products.

For each account, the main LGD components – collateral deficiency, expenses, forgone interest, and accrued interest – are either simply aggregated as the accounting loss or aggregated from the NPV of each component as the economic loss based on the applications. Then the actual loss severity or LGD can be obtained as the ratio of this aggregated sum and the EAD.

The account-level default data shown in Table 3.5 need to be combined with account/loan characteristics data and macroeconomic factors by merging through the loan ID and default time. Loss severity and LGD regression models can be fit based on the combined data.

Table 3.5 LGD reference data

ID	EAD	Resolution method	Net sale price	Collateral deficiency	Other expenses	Forgone interest	Accrued interest	Actual LGD/severity	Default time
ID1	XXX	REO	XXX	XXX	XXX	XXX	XXX	XXX	mm/yy
ID2	XXX	TPS	XXX	XXX	XXX	XXX	XXX	XXX	mm/yy
ID3	XXX	SS	XXX	XXX	XXX	XXX	XXX	XXX	mm/yy
ID4	XXX	PO	XXX	XXX	XXX	XXX	XXX	XXX	mm/yy
ID5	XXX	MOD	XXX	XXX	XXX	XXX	XXX	XXX	mm/yy
ID6	XXX	REO	XXX	XXX	XXX	XXX	XXX	XXX	mm/yy
ID7	XXX	MOD	XXX	XXX	XXX	XXX	XXX	XXX	mm/yy
.	.	.	.		.	.	.	.	.
.	.	.	.		.	.	.	.	.
.	.	.	.		.	.	.	.	.
IDn	XXX	CO	XXX	XXX	XXX	XXX	XXX	XXX	mm/yy

Table 3.5 also includes the resolution method for each account. The resolution method can be used to fit the selection model for resolution and predict the resolution selection probabilities.

3.1.4.3 Micro-structure Dynamic Model

The micro-structure dynamic model includes two sets of models – the resolution event probability models and resolution severity models.

The resolution event probability models are built on the post-default transaction data as we described early. The two resolution events are gross loss and payoff. While gross loss is the typical resolution event, payoff is particularly treated as its severity is close to 0. Similar to the default/payment setup before default, the post-default gross loss and payoff modeling can use either a sequential risk framework or a competing risk framework. For the sequential risk framework, payoff is first modeled using a binary logit or PH model by treating all alternative status default or gross loss as non-event without or with censoring, respectively. Then, by removing all payoff accounts from the reference default data, a second binary logit or PH model can fit for the gross loss event probability conditional on that the account is not payoff. The post-default sequential binary models are based on the logic that a payoff is always considered as a privileged resolution method before the gross loss. Given that the account is already in default, the argument of strategic default as in the pre-default default/prepayment competing risk situation may not be applaudable anymore. However, to avoid such choice assumption, one can still use the competing risk framework with binary approximation to fit these event probabilities additively. In practice, the two methods have minor impact on LGD prediction due to the dominance of gross loss events.

The resolution event probability models can incorporate open accounts when a major event of the resolution process has been executed, e.g., an asset sale or full principal written-off. As described in the processing of the post-default transaction data, some upward severity adjustments based on historical experiences or conservative judgment are needed for these open accounts.

The micro-structure dynamic model fits the severity at the resolution instead of default. This is more accurate since the sales of collateral depend more closely to the market and economic environments at the resolution time than the default time, especially when the resolution has been prolonged since the default. While a regression model is commonly used for the gross loss severity, the severity for payoff commonly takes a lower limit derived from either historical experiences or a management forward-looking.

The micro-structure dynamic model makes prediction for both post-default resolution event probabilities and severities based on information at future potential resolution time, which needs to be forecasted. So, the micro-structure dynamic model is forward-looking and is different from other LGD models based on information at the default time. The burden of more accurate severity models based on information at resolution time is the need to set up a forecast time limit for the micro-structure dynamic model. A conservative setup is using the maximum resolution time observed in the post-default loss mitigation process. An alternative is to set the cumulative projected resolution event probability to a target, e.g., 95%, and make upward adjustment for the final projected LGD. This is consistent on how we define the timing of a resolution event in the reference default transaction data.

The micro-structure dynamic model is more popularly used in the accounting and CCAR applications, which requires more accurate and forward-looking loss projections.

3.1.4.4 Regression Models

Different from the micro-structure dynamic model, LGD regression models take a one-to-go strategy. The models assume that the LGD of a default account either based on the accounting loss or economic loss has a direct relationship with the account/loan characteristics, macroeconomic factors, and other default-related risk factors measured at the default time and can be summarized in a regression model regardless of the post-default market or economic environments and details of the resolution process. Clearly, this is a strong assumption and needs assessment for model risk purpose. Nevertheless, for some credit products, especially for those with complex post-default recovery and collection processes and not so prolonged resolution process, the regression models have the advantage of being a simple summary of the post-default loss severity.

There are various regression models that can be used for LGD modeling. Here, we describe the most popularly used ones, Tobit, beta, and more general fractional response regression models.

Tobit Regression

Tobit[14] regression is popularly used for left-censored observations. Using Tobit regression in LGD modeling is due to the common practice that negative LGD observations are left-censored at 0.

The model assumes that the underlying non-censored LGD_i^* has a linear relationship with risk factors x_i for all accounts $i = 1, \ldots, n$:

$$\text{LGD}_i^* = x_i'\beta + \epsilon_i \tag{3.59}$$

and the observed LGD is:

$$\text{LGD}_i = \begin{cases} 0, & \text{LGD}_i^* \leq 0 \\ \text{LGD}_i^*, & \text{LGD}_i^* > 0 \end{cases} \tag{3.60}$$

The Tobit model can be extended to both left-censored and right-censored:

[14] Tobin, J., (1958). Estimation of relationship for limited dependent variables. Econometrica 26, 24–36.

$$\mathrm{LGD}_i = \begin{cases} 0, & \mathrm{LGD}_i^* \le 0 \\ \mathrm{LGD}_i^*, & 0 < \mathrm{LGD}_i^* < 1 \\ 1, & \mathrm{LGD}_i^* \ge 1 \end{cases} \tag{3.61}$$

Assume ϵ_i are i.i.d normal distributions with mean 0 and standard error σ and then LGD_i^* are i.i.d normal with mean $\mu_i = x_i'\beta$ and variance σ^2, and then the likelihood for the observed LGD_i can be computed as the following:

$$L(\mathrm{LGD}_i) = \begin{cases} \Phi\left(\frac{-\mu_i}{\sigma}\right), & \mathrm{LGD}_i \le 0 \\ \psi\left(\frac{\mathrm{LGD}_i - \mu_i}{\sigma}\right), & 0 < \mathrm{LGD}_i < 1 \\ \Phi\left(\frac{\mu_i - 1}{\sigma}\right), & \mathrm{LGD}_i \ge 1 \end{cases} \tag{3.62}$$

where Φ and ψ are the cumulative distribution function (CDF) and probability density function (PDF) of the standard normal distribution, respectively.

Parameter estimates for (β, σ) can be obtained through maximum likelihood estimation, and the predicted LGD for a given set of risk factors x_k can be obtained through $\widehat{\mu}_k = x_k'\widehat{\beta}_{\mathrm{ML}}$, which is the mean of the original uncensored LGD variable. To be comparable to the censored LGD observations, one can compute the mean of the censored LGD as pointed out by Tobin (1958).

Generally, the Tobit regression uses a linear model with censoring to fit the LGD. As a simple model, it has a strong assumption that there exists a linear relationship between LGD and risk factors with some white noises. Such assumption may only hold for defaults with certain homogeneous resolution processes. Nevertheless, as an initial step to explore the relationship between LGD and risk factors, the Tobit regression could be helpful in LGD modeling.

When there is no censoring for the observed LGD, the Tobit regression falls back to the regular linear regression. This is commonly seen in LGD modeling for capital purpose, for which the default loans in the reference data are classified into LGD grids based on LGD profiling derived from such linear regression models with neutral macroeconomic inputs. Once the LGD grids are built, downturn LGD (DLGD) based on actual LGD of identified downturn default accounts is estimated for the grids and used in the capital calculation.

Beta Regression

Instead of a censored normal distribution assumed in the Tobit regression, with beta regression, LGD is assumed to follow a beta distribution with PDF:

$$\mathrm{Beta}(\mathrm{LGD} = y|p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\, y^{p-1}(1-y)^{q-1}, 0 < y < 1 \tag{3.63}$$

where $p > 0$, $q > 0$ are parameters and Γ is the gamma function. The beta distribution has mean $E(y) = \mu = \frac{p}{p+q}$ and variance $\mathrm{Var}(y) = \frac{pq}{(p+q+1)(p+q)^2}$. Using $\mu = \frac{p}{p+q}$ and $\phi = p + q$, the beta distribution PDF can be reparametrized[15] as:

$$\mathrm{Beta}(\mathrm{LGD} = y|\mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)}\, y^{\mu\phi - 1}(1-y)^{(1-\mu)\phi - 1}, 0 < y < 1 \tag{3.64}$$

with $0 < \mu < 1$, $\phi > 0$, and $E(y) = \mu$ and $\mathrm{Var}(y) = \mu\,(1-\mu)/(1 + \phi)$. The parameter ϕ is known as the precision parameter since, for fixed μ, the larger the ϕ, the smaller the variance of y; ϕ^{-1} is a dispersion parameter.

Let $y_1, \ldots, y_n$ be a random sample such that $y_i \sim \mathrm{Beta}(\mu_i, \phi)$, $i = 1, \ldots, n$. The beta regression as a generalized linear model is defined as:

$$g(\mu_i) = x_i'\beta = \eta_i \tag{3.65}$$

with risk factors x_i, unknown regression parameters β, the linear predictor η_i, and the link function $g : (0, 1) \rightarrow R$, which links the linear predictor and LGD mean. Some commonly used link functions are logit $g(\mu) = \log(\mu/(1-\mu))$; probit $g(\mu) = \Phi^{-1}(\mu)$, where Φ is the standard normal distribution function; complementary log-log $g(\mu) = \log\{-\log(1-\mu)\}$; log-log $g(\mu) = -\log\{-\log(\mu)\}$; and Cauchy $g(\mu) = \tan\{\pi(\mu - 0.5)\}$. If LGD has a floor or cap, it can be rescaled back to $(0, 1)$ through a location/scale transformation $y = \frac{y^* - \theta}{\sigma}$.

The loglikelihood for an observation (x_i, y_i) is

$$\begin{aligned}\log(\mathrm{Beta}(y_i|\mu_i, \phi)) = {} & \log(\Gamma(\phi)) - \log(\Gamma(\mu_i\phi)) - \log(\Gamma((1-\mu_i)\phi)) \\ & + (\mu_i\phi - 1)\log(y_i) + ((1-\mu_i)\phi - 1)\log(1-y_i)\end{aligned} \tag{3.66}$$

where $\mu_i = g^{-1}\left(x_i'\beta\right)$ is a function of β and the parameters (β, ϕ) can be estimated through maximum likelihood estimation.

Beta regression fits a generalized linear model instead of linear model as used in Tobit regression. The popularity of using beta regression to fit LGD is due to its flexibility to catch the LGD shape as a function of risk factors.

When there are LGD observations polarized on extremes (i.e., zero loss or 100% loss), inflated beta (IBeta) may help the fitting process by means of the following distribution:

$$\mathrm{IBeta}(\mathrm{LGD} = y|\mu, \phi, \theta_0, \theta_1) = \begin{cases} \theta_0, & y = 0 \\ (1 - \theta_0 - \theta_1)\mathrm{Beta}(y|\mu, \phi), & 0 < y < 1 \\ \theta_1, & y = 1 \end{cases} \tag{3.67}$$

[15] Ferrari SLP, Cribari-Neto F. (2004). "Beta Regression for Modelling Rates and Proportions." Journal of Applied Statistics, 31(7), 799–815.

with $0 < \theta_0 < 1$ and $0 < \theta_1 < 1 - \theta_0$. Parameters $(\beta, \phi, \theta_0, \theta_1)$ can also be estimated through maximum likelihood estimation. The IBeta regression model just uses empirical frequency rates as estimates for the extremes without exploring the dependence of these extremes on risk factors. Very often the extremes can be classified into specific types of post-default resolutions and be treated independently with their own characteristics. In Sect. 3.1.4.5, we will discuss how a multiple resolutions model can be a better alternative for LGD modeling in such cases.

For both Tobit and beta regressions, the likelihood function can't be guaranteed concave, especially when sample size is small. To achieve certain stability of the MLE, good global optimization techniques with minimum sample size are required.

Fractional Response Regression

Like beta regression, fractional response regression (FRR)[16] assumes a generalized linear relationship between LGD and risk factors:

$$E(\mathrm{LGD} = y|x) = G(x'\beta) \tag{3.68}$$

where $G(z)$ is a known function satisfying $0 < G(z) < 1$ for all $z \in R$ and x is the vector of risk factors and β the parameter forming the linear predictor $\eta = x'\beta$. Instead of specifying the relationship between linear predictor, mean, and variance function as in the generalized linear models, the FRR directly specifies a Bernoulli loglikelihood function for an observation (x_i, y_i) as:

$$\mathrm{LL}_i(\beta) = y_i \log(G(x'\beta)) + (1 - y_i) \log(1 - G(x'\beta)) \tag{3.69}$$

Note that $y_i \in [0, 1]$, $i = 1. \ldots, n$ as the observed LGD can take extremes 0 or 1 and any values between.

Corresponding to the link functions suggested in the beta regression, the condition mean functions commonly suggested are the following cumulative distribution functions: logistic $G(z) = \exp(z)/(1 + \exp(z))$; normal $G(z) = \Phi(z)$, where Φ is the standard normal distribution function; extreme minimum $G(z) = 1 - \exp(-\exp(z))$; extreme maximum $G(z) = \exp(-\exp(-z))$; and Cauchy $G(z) = 0.5 + \frac{1}{\pi}\arctan(z)$. While the logistic, standard normal, and Cauchy specifications for $G(z)$ are symmetric about the point 0.5 and, consequently, approach 0 and 1 at the same rate, the extreme maximum and extreme minimum models are not symmetric: the former increases sharply at small values of $G(z)$ and slowly when $G(z)$ is near 1, while the latter exhibits the opposite behaviors. The Cauchy distribution presents the heaviest tails, which implies that this specification is more robust to outliers than the logistic and standard normal formulations.

---

[16] Papke, L.E. and Wooldridge, J.M. (1996), "Econometric methods for fractional response variables with an application to 401(k) plan participation rates," Journal of Applied Econometrics, 11(6), 619–632.

The FRR likelihood function in (3.69) is a quasi-likelihood function. However, since the Bernoulli distribution is a member of the linear exponential family (LEF), the quasi-maximum likelihood (QML) estimator that is defined by:

$$\widehat{\beta}_{\mathrm{QML}} = \mathrm{argmax}_{\beta} \sum\nolimits_{i=1}^{n} \mathrm{LL}_i(\beta) \tag{3.70}$$

is consistent and asymptotically normal, regardless of the true distribution of LGD conditional on x, provided that $E(\mathrm{LGD} = y|x)$ in (3.68) is indeed correctly specified (Gouriéroux et al., 1984).[17] Moreover, Papke and Wooldridge (1996) show there are some cases where this QML estimator is efficient in a class of estimators containing all LEF-based QML and weighted nonlinear least squares (NLS) estimators.

Though the FRR can handle LGD observations with extremes, it may not be able to interpret the LGD boundary phenome well due to the restriction of $0 < G(z) < 1$. As we pointed early, when there are a large number of extremes (here cured accounts with LGD ≤ 0), the extremes can be classified into specific types of post-default resolutions and be treated independently with their own characteristics. In the following section, we will discuss how a multiple resolutions model can be a better alternative for LGD modeling in such cases.

3.1.4.5 Multiple Resolutions Model

Multiple resolutions (MR) models are particularly useful for LGD modeling. First, while there is the issue of lack of good techniques to handle extremes as we pointed out with various regression models, LGD observed in extremes could be related to specific resolutions. Second, when the post-default transaction data are not sufficient, but information related to resolution processes is available, multiple resolutions model with simpler structure is a natural choice.

There are two types of multiple resolutions models – sequential multiple resolutions (SMR) models and competing multiple resolutions (CMR) models.

For the sequential multiple resolutions, a tree-based structure can be used to describe the modeling process as shown in Fig. 3.4. Sequential binary logit regression models are used to fit the resolution probabilities.

For competing multiple resolutions, a competing risk-based structure can be used to describe the modeling process as shown in Fig. 3.5. A multinomial logit model can be used to fit the resolution probabilities.

Once the resolution probabilities are obtained, the next step is to fit severity models for each of the resolutions using either regression models discussed in Sect. 3.1.4.4 or just historical averages if a regression relationship could not be established.

[17] Gouriéroux, C., Monfort, A. and Trognon, A. (1984), "Pseudo maximum likelihood methods: applications to Poisson models," Econometrica, 52(3), 701–720.

Fig. 3.4 SMR

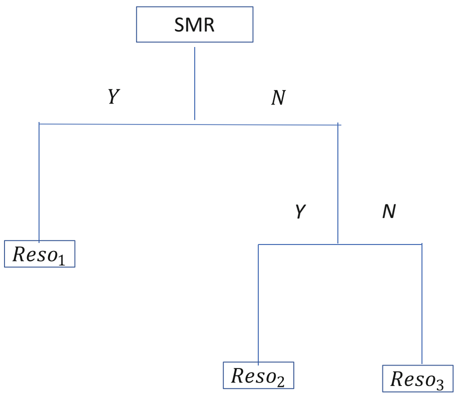
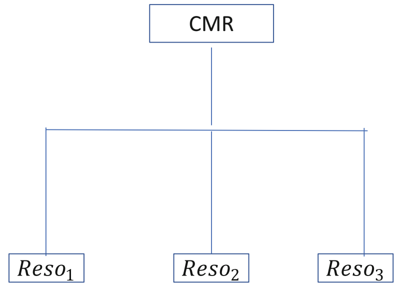


Fig. 3.5 CMR



For SMR in Fig. 3.4, the final LGD can be estimated as:

$$\mathrm{LGD} = P_{\mathrm{Reso}_1}\mathrm{Sev}_{\mathrm{Reso}_1} + (1 - P_{\mathrm{Reso}_1})P_{\mathrm{Reso}_2}\mathrm{Sev}_{\mathrm{Reso}_2} + (1 - P_{\mathrm{Reso}_1})(1 - P_{\mathrm{Reso}_2})\mathrm{Sev}_{\mathrm{Reso}_3} \tag{3.71}$$

where P_{Reso_i} for $i = 1, 2, 3$ are estimated event probabilities with the sequential binary logit models and $\mathrm{Sev}_{\mathrm{Reso}_i}$ for $i = 1, 2, 3$ are estimated severities for each of the three types of resolutions.

For CMR in Fig. 3.5, the final LGD can be estimated as:

$$\mathrm{LGD} = P_{\mathrm{Reso}_1}\mathrm{Sev}_{\mathrm{Reso}_1} + P_{\mathrm{Reso}_2}\mathrm{Sev}_{\mathrm{Reso}_2} + (1 - P_{\mathrm{Reso}_1} - P_{\mathrm{Reso}_2})\mathrm{Sev}_{\mathrm{Reso}_3} \tag{3.72}$$

where P_{Reso_i} for $i = 1, 2, 3$ are estimated event probabilities with the multinomial logit model and $\mathrm{Sev}_{\mathrm{Reso}_i}$ for $i = 1, 2, 3$ are estimated severities for each of the three types of resolutions.

SMR requires that the types of resolutions have a sequential order. This is largely true for collateral-based credit products. For example, for mortgage defaults, when loan to value (LTV) is low, the loss mitigation strategy may seek to get the mortgage default cured before considering alternative resolutions. If not cured, short sale by the owner is preferred to third-party sale or REO, among which REO may be the least preferred from the LGD consideration. Loan modification could be used as an alternative approach to cure the default when there is no write-off.

SMR has the advantage to use different set of risk drivers in the sequential binary logit models. Some disadvantages with SMR could be that there is not a clear order among different types of resolutions and there might not be sufficient data for certain types of resolutions to establish a stable regression relationship with risk factors. Model assumptions are needed to overcome these disadvantages, and model risk management for such assumptions will be discussed in later chapters with applications on specific risk management frameworks.

CMR does not discriminate different types of resolutions for a default. It is more suitable for credit products without collateral or other information at default and for which the choice of resolutions seems random. CMR also is conceptually sound when looking from a long-term through-the-cycle (TTC) view when loss mitigation strategies vary under different economic environments.

CMR has the advantage of model simplicity – with only one set of risk drivers for all types of resolutions – which could also be a disadvantage for model accuracy. Management judgments may often be required as overlay to adjust the model outputs to catch resolution dynamics.

Multiple resolutions model with either SMR or CMR uses a more granular approach than the one-to-go regression methods, which may not be a good summary for the practical loss mitigation processes that consist of different resolution strategies and deemed heterogeneous. In addition, LGD tends to be sensitive to account/loan characteristics and local market and economic environments when there is collateral involved. Initial comprehensive investigation can help to show more insights for LGD modeling. In the next section, we discuss how to use machine learning models for such purposes.

3.1.4.6 Machine Learning Models

Machine learning models have been popularly used in LGD modeling due to the heterogeneity embedded in LGD. In the micro-structure dynamic model with transaction data, classification and regression trees (CART) can be used to explore the gross loss and payoff probability models just as used in the pre-default default and prepayment probability models. CART are useful tools to explore data quality issues and outliers, segmentation, variable importance and selection, local nonlinear structures, and proper model interpretation. Most often, enhanced CART models using bagging, random forest, and boosting discussed in Sect. 3.1.1.3 are used as challenger or benchmark models to make sure the champion model performs as required.

For severity modeling in the micro-structure dynamic model or LGD modeling in the one-to-go LGD regression models, the dependent variable is numeric, and regression trees are preferred to explore the complex heterogeneity of severity and LGD with risk factors. High-level summary from these regression trees could lead to better segmentation, variable selection, and local structure capture than the classical bivariate analysis. Such regression tree-based analysis may also suggest when a proper type of regression method discussed in Sect. 3.1.4.4 should be adopted on a specific segment. Again, enhanced regression tree models using bagging, random forest, and boosting are used as challenger or benchmark severity/LGD models.

For the multiple resolutions model, either with SMR or CMR, classification trees can be used to explore the key risk factors by variable importance for either a binary or multinomial classification with the multiple resolutions. For the severity modeling within each resolution, similarly to the LGD regression modeling, regression trees can be used to explore when a proper regression model should be used or the sample mean is the final acceptable choice.

LGD modeling is heavily based on the reference default data and ML models as strong data exploratory tools have advantages over classical exploratory tools like frequency and bivariate analysis. Its strong ability to catch complex nonlinear relationships between the target and risk factors helps to identify better segmentation, key risk factors, important local structures, and intuitive interpretations.

As we pointed out early, ML models could easily overfit the data, and a balance between model accuracy and complexity must be achieved to avoid trivial or overfitted models. Due to interpretability and stability, even with the enhanced ML models, ML models are seldom used as the primary model, especially in the regulatory areas. They are more likely used as challenger or benchmark models. This is also the case with LGD modeling. In Chaps. 4, 5, and 6, we will present more details on how ML models are used for LGD modeling in different risk management frameworks and what and how model risks related to ML models should be managed.

3.1.5 Exposure at Default Models (EAD)

Exposure at default (EAD) measures the exposure of credit risk. As the loss rate measured by the combination of PD and LGD is based on the time of default, the exposure at default is required to compute the expected loss. Strictly speaking, PD is a periodic measure, while LGD is a pointwise measure, so the loss rate from the combination of PD and LGD can be considered as an average over the period defined by PD, and EAD is commonly measured at the beginning of this period.

For credit risk, EAD operates differently for committed and uncommitted credit products. For committed credit products like loans, at any time of the loan term, the unpaid principal balance (UPB) is the EAD. So, for committed credit products, actual UPB is actual EAD, and predicted UPB is future EAD. While actual UPB for a loan is recorded in the reference default data and used for the actual LGD calculation as discussed in Sect. 3.1.4.2, the predicted UPB needs combining default, prepayment, and scheduled amortization. For some products, besides prepayment (paid in full), there is also the overpayment (extra partial payment). However, due to its scarcity, overpayment is seldom considered in account level.

For uncommitted credit products like line of credit products, the balance at reporting time could be increased with extra credit draw before default if the customer seeks additional funds under an open line of credit. In such case, EAD is the combination of the current balance and the extra credit draw. Balance drawn between observation and default can be computed as a percentage of the undrawn credit line. This percentage is the loan equivalency (LEQ) factor. So, EAD under an open line of credit can be predicted through LEQ from the current balance and credit limit. Based on the relative size of the current balance to the credit limit (also called utilization), accounts are defined as high drawn and low drawn. LEQ for high-drawn and low-drawn accounts are modeled differently. An account also could be blocked, which indicates the customer can't draw additional funds under an open line of credit. EAD for blocked accounts is handled separately.

In the following sections, we focus on the EAD projection models for both committed and uncommitted credit products.

3.1.5.1 UPB for Committed Credit Products

For committed credit products, loss forecast from a reporting date requires the projection of credit exposure at any time starting from the reporting date. This is essentially the projection of the UPB starting from the reporting date. Assume, for ith account, the UPB at the reporting date is $\mathrm{UPB}_i(0)$, and then UPB at the beginning of period t can be projected as:

$$\mathrm{UPB}_i(t) = \mathrm{UPB}_i(t-1)(1 - \mathrm{PD}_i(t) - \mathrm{PP}_i(t) - \mathrm{AmFactor}_i(t)) \tag{3.73}$$

where $\mathrm{PD}_i(t)$ is the conditional default probability of the ith account over period t, $\mathrm{PP}_i(t)$ is the conditional prepayment probability of the ith account over period t, and $\mathrm{AmFactor}_i(t)$ is the amortization factor over period t and is defined as:

$$\mathrm{AmFactor}_i(t) = 1 - \frac{\text{Contract balance}_i(t)}{\text{Contract balance}_i(t-1)} \tag{3.74}$$

In (3.73), we assume all exposure reduction components – default, prepayment, and amortization – are competitive. This is a simplified case. A more practical formula as used in the general bond cash flow projection in the SIFMA/BMA manual[18] is:

$$\mathrm{UPB}_i(t) = \mathrm{UPB}_i(t-1) - \mathrm{Default}_i(t) - \mathrm{Prepayment}_i(t) - \mathrm{Amort}_i(t) \tag{3.75}$$

where $\mathrm{Default}_i(t) = \mathrm{UPB}_i(t-1) * \mathrm{PD}_i(t)$, $\mathrm{Amort}_i(t) = \mathrm{UPB}_i(t-1) * (1 - \mathrm{PD}_i(t)) * \mathrm{AmFactor}_i(t)$, and $\mathrm{Prepayment}_i(t) = \mathrm{UPB}_i(t-1) * (1 - \mathrm{AmFactor}_i(t)) * \mathrm{PP}_i(t)$.

The SIFMA/BMA formula (3.75) takes a conservative way and considers default prior the amortization and amortization before prepayment. However, both (3.73) and (3.75) consider default and prepayment as competitive. So, the inputs for both these formulas should come from the competitive risk framework. The difference between (3.73) and (3.75) is the amount of:

[18] SIFMA/BMA, Standard Formulas for the Analysis of Mortgage-Backed Securities and Other Related Securities. Uniform Practices/Standard Formulas. Feb 1, 1999.

$$\text{UPB}_i(t-1) * (\text{PD}_i(t) + \text{PP}_i(t)) * \text{AmFactor}_i(t) \tag{3.76}$$

So, (3.75) can be reformulated as:

$$\text{UPB}_i(t) = \text{UPB}_i(t-1)(1 - \text{PD}_i(t) - \text{PP}_i(t))(1 - \text{AmFactor}_i(t)) \tag{3.77}$$

Let $S_i^C(t) = (1 - \text{PD}_i(t) - \text{PP}_i(t))$ be the period t conditional survival probability from credit risk events and $S_i^A(t) = 1 - \text{AmFactor}_i(t)$ be the period t survival rate from amortization, and then:

$$\text{UPB}_i(t) = \text{UPB}_i(t-1) * S_i^C(t) * S_i^A(t) \tag{3.78}$$

Formula (3.78) is preferred since it can be extended to sequential risk framework, where the period t survival probability from credit risk events can be easily calculated from $S_i^C(t) = S_i^D(t) * S_i^{P|!D}(t)$ or $S_i^C(t) = S_i^P(t) * S_i^{D|!P}(t)$ by either considering default prior prepayment or vice versa, where $S_i^D(t)$ and $S_i^P(t)$ are conditional survival probability from default or prepayment, respectively, and $S_i^{P|!D}(t)$ and $S_i^{D|!P}(t)$ are sequential conditional survival probability from prepayment or default given no another risk event, respectively.

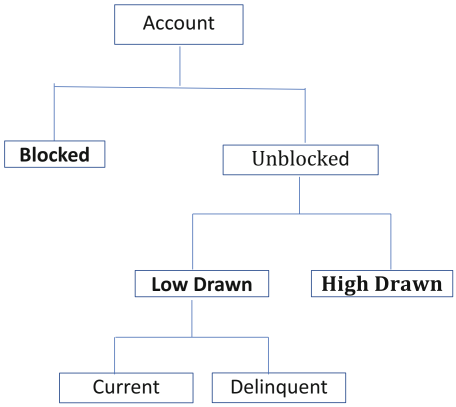
3.1.5.2 EAD Segmentation for Uncommitted Credit Products

For uncommitted credit products, the common loss mitigation procedure classifies accounts into three categories. For delinquent accounts, it is very often the customer is not allowed to draw additional funds when the delinquency is over some defined threshold (e.g., 65 days for mortgage-related line of credits). Such accounts are called in a “blocked” status. So, blocked accounts have a much higher concentration in delinquent segments than in current segments. For accounts that the customer can still draw additional funds under an open line of credit, they are classified as high drawn and low drawn. High drawn is defined as the observed utilization that is over a threshold percentage (e.g., 95%), while low drawn is below that threshold. EAD for these three categories behave differently. For blocked and high-drawn accounts, there is not much room to increase the exposure at default and thus less dynamics. For low-drawn accounts, the exposure at default could depend on the account and economic characteristics. Figure 3.6 shows the segmentation.

3.1.5.3 Default Data Processing for Uncommitted Credit Products

Similar to the LGD reference default data processing, EAD reference data also focus on default accounts. However, in addition to the post-default recovery and cost information used to calculate actual exposure at default and severity at default, EAD reference default data for uncommitted credit products also require pre-default credit utilization and credit limit information. So, for all default accounts in a portfolio book, snapshots are taken as observations to track the pre-default credit balance, utilization, and credit limit. Such snapshots can be directly extracted from the snapshot sampling data prepared for the credit event models as described in Sect. 2.4.2. Table 3.6 presents an example.

Fig. 3.6 EAD segmentation



Multiple snapshots can be taken from the same default account, and each snapshot across all accounts can be considered as a cohort. Balance at the snapshot/observation date is recorded as the credit drawn, and account utilization is defined as the ratio of the credit drawn over the credit limit at observation. Based on the utilization, the account is segmented into low drawn and high drawn. There is also the blocked segment for which the credit draw is frozen by the creditor.

The snapshot EAD data processing for uncommitted credit products corresponds to the cohort method of Moral (2011),[19] which can also be naturally extended to the variable time horizontal method if the snapshots are taken as the most granular (e.g., monthly).

Utilization is a rate that measures the percentage of the credit line that has been drawn. Empirical research shows that customers tend to tap available lines of credit when approaching default, which would increase utilization rate. But utilization can also change when the creditor increases or decreases account's credit line. During a recession, credit line decrease became an integral part of risk mitigation strategies and together with the closing of inactive accounts is the explanation of the shift in the cohort utilization distribution toward higher percentiles. So, account-level utilization, in some segments interacted with cohort-level utilization, contains great explanatory power for both credit event and exposure models.

EAD is the balance at default that can be decomposed into two parts, the balance observed at a given pre-default time and additional credit draw approaching default. Since the first part is observed, the prediction of the dynamic of EAD falls on the second part, which can be formulated and projected through loan equivalency (LEQ) factor. EAD can also be formulated and projected through some simpler measures, credit conversion factor (CCF) or EAD factor. These factors are considered EAD

[19] Moral, G. (2011), EAD Estimates for Facilities with Explicit Limits. Chapter 11 in *The Basel II Risk Parameters: Estimation, Validation, Stress Testing – with Applications to Loan Risk Management.* (eds.) Evelyn Hayden, Daniel Porath, Bernd Engelmann, Robert Rauhmeier. Springer-Verlag Berlin Heidelberg.

Table 3.6 EAD reference data

EAD reference default data

ID	Observation date	Balance	Credit limit	Utilization	Segment	EAD	EAD factor	LEQ	CCF	Actual LGD/severity	Default date
ID1	mm/yy	XXX	XXX	XXX	Low drawn	XXX	XXX	XXX	XXX	XXX	mm/yy
ID1	mm/yy	XXX	XXX	XXX	Low drawn	XXX	XXX	XXX	XXX	XXX	mm/yy
ID1	mm/yy	XXX	XXX	XXX	Low drawn	XXX	XXX	XXX	XXX	XXX	mm/yy
ID1	mm/yy	XXX	XXX	XXX	High drawn	XXX	XXX	XXX	XXX	XXX	mm/yy
ID2	mm/yy	XXX	XXX	XXX	High drawn	XXX	XXX	XXX	XXX	XXX	mm/yy
ID2	mm/yy	XXX	XXX	XXX	High drawn	XXX	XXX	XXX	XXX	XXX	mm/yy
ID2	mm/yy	XXX	XXX	XXX	High drawn	XXX	XXX	XXX	XXX	XXX	mm/yy
.	.	.	.	.		.	.	.	.	.	.
.	.	.	.	.		.	.	.	.	.	.
.	.	.	.	.		.	.	.	.	.	.
IDn	mm/yy	XXX	XXX	XXX	Blocked	XXX	XXX	XXX	XXX	XXX	mm/yy

risk measures and possibly driven by account characteristics and macroeconomic variables. So, data in Table 3.6 are merged with account information and macroeconomic variables to form the final EAD modeling data.

3.1.5.4 LEQ, CCF, and EAD Factor Models

For uncommitted credit products, an EAD model uses the information available at a pre-default observation or reporting date to project the exposure at default in order to calculate the projected loss through the PD, LGD, and EAD model framework. As described in the previous section, the available information includes the observed balance and credit limit.

Loan Equivalency (LEQ) Factor

For an account observed at time t and expected to default sometime in the future ($T > t$), exposure at default can be expressed as the sum of drawn balance at observation (t) and balance drawn in the time leading to default. Balance drawn between observation and default can be computed as a percentage of the undrawn line. This percentage is the loan equivalency factor (LEQ factor). EAD for an account is therefore defined as:

$$\mathrm{EAD}_{iT} = \mathrm{Draw}_{it} + \mathrm{LEQ}_{it}(\mathrm{Line}_{it} - \mathrm{Draw}_{it}) \tag{3.79}$$

where Draw_{it} is the drawn balance at observation (t) and Line_{it} is the credit line known at observation (t). So LEQ in the EAD reference data can be calculated through:

$$\mathrm{LEQ}_{it} = \frac{\mathrm{EAD}_{iT} - \mathrm{Draw}_{it}}{\mathrm{Line}_{it} - \mathrm{Draw}_{it}} \tag{3.80}$$

where $T > t$. Thus, LEQ represents the portion of the unused line that is drawn upon by the borrower in the time leading to default. LEQ ranges between 0 and 1, with an LEQ of 0 indicating EAD equals current exposure (no further transactions on the account) and an LEQ of 1 indicating that the entire unused line has been utilized by the borrower. Further, an LEQ of less than 0 indicates the account has paid down part or entire balance owed at observation. In some cases, the LEQ is greater than 1, where the customer has overdrawn on the account. To overcome such outliers, a cap (e.g., 150%) and a floor (e.g., 0%) are commonly applied for the observed LEQ in the EAD data preparation.

For blocked or high-drawn account, the undrawn part ($\mathrm{Line}_{it} - \mathrm{Draw}_{it}$) is likely near 0, so a more stable definition of LEQ is:

$$\mathrm{LEQ}_{it} = \frac{\mathrm{EAD}_{iT} - \mathrm{Draw}_{it}}{\mathrm{Line}_{it}} \tag{3.81}$$

For the low-drawn category, LEQ is usually modeled as a linear function of risk factors using least squares regression on the truncated data (truncation done with 5%

and 95% percentile of the observed LEQ within the reference data for each sub-portfolio):

$$\mathrm{LEQ}_{it} = \alpha + \sum_{j=1}^{p} \beta_j x_{jit} + \sum_{k=1}^{q} \gamma_k z_{kt} + \epsilon_{it} \tag{3.82}$$

In the above equation, LEQ for a particular account i that defaulted in the future since time t can be expressed as a linear combination of account-specific attributes (i.e., x_{jit} is the jth attribute for account i observed as of cohort date t) as well as other exogenous risk drivers (e.g., z_{kt} is the set of macroeconomic and/or portfolio characteristics observed as of cohort date t). A separate model should be developed for each product type, and coefficients should be evaluated to ensure the model is reasonable based on business knowledge and economic considerations. The predicted LEQ should be floored and capped (e.g., between 0% and 150%).

The truncated least squares regression (truncation done both on the input data and model outputs) is used to handle the extremes of LEQ data commonly seen in practice. Though other regression models may better handle the extreme LEQ observations, it is not recommended in the LEQ modeling because (i) a simpler LEQ model like linear regression is preferred to other complex models and (ii) extreme LEQ is either due to data quality issues or low undrawn balance for which the final EAD projection is not so sensitive to LEQ as shown by the following projection formula:

$$\widehat{\mathrm{EAD}}_{iT} = \mathrm{Draw}_{it} + \widehat{\mathrm{LEQ}}_{it}(\mathrm{Line}_{it} - \mathrm{Draw}_{it}) \tag{3.83}$$

where $\widehat{\mathrm{LEQ}}_{it}$ and $\widehat{\mathrm{EAD}}_{iT}$ are the projected LEQ and EAD at reporting time t, respectively.

For blocked and high-drawn accounts, the linear regression relationship in (3.82) may not be established, and the averages of LEQ observed in the EAD reference data are commonly used for the projection through the following formula corresponding to (3.81)

$$\widehat{\mathrm{EAD}}_{iT} = \mathrm{Draw}_{it} + \widehat{\mathrm{LEQ}}_{it} * \mathrm{Line}_{it} \tag{3.84}$$

For regulatory capital, a downturn LEQ similar to the downturn LGD discussed in Sect. 3.1.4.4 is used – accounts in the EAD reference data are classified into LEQ grids based on LEQ profiling derived from the linear regression model with neutral macroeconomic inputs. Blocked and high-drawn accounts may form independent grids. Once the LEQ grids are built, downturn LEQ (DLEQ) based on actual LEQ of identified downturn default accounts is estimated for the grids and used in the capital calculation.

Credit Conversion Factor (CCF) and EAD Factor
Besides LEQ, there are two more measures commonly used in EAD modeling. The credit conversion factor or CCF is defined as:

$$\mathrm{CCF}_{it} = \frac{\mathrm{EAD}_{iT}}{\mathrm{Draw}_{it}} \tag{3.85}$$

It measures the ratio of the drawn balance at default over the observed drawn balance at a pre-default time t. CCF has a range of $(0, \infty)$ and a linear regression model as in (3.82) can be used. The EAD projection based on the observed balance drawn and projected CCF is:

$$\widehat{\mathrm{EAD}}_{iT} = \mathrm{Draw}_{it} * \widehat{\mathrm{CCF}}_{it} \tag{3.86}$$

CCF does not use the credit limit information, so it suits for products without credit limits and having already had significant observed credit drawn balance since on the book.

The EAD factor or EADF is defined as:

$$\mathrm{EADF}_{it} = \frac{\mathrm{EAD}_{iT}}{\mathrm{Line}_{it}} \tag{3.87}$$

It measures the ratio of drawn balance at default over the given credit limit at a pre-default time t. EADF has a range of $[0, 1]$, and a Tobit, beta, or FRR model described in Sect. 3.1.4.4 for LGD modeling can be used. The EAD projection based on the observed balance drawn and projected CCF is:

$$\widehat{\mathrm{EAD}}_{iT} = \mathrm{Line}_{it} * \widehat{\mathrm{EADF}}_{it} \tag{3.88}$$

EADF does not use the observed credit drawn information, so it suits for new products without much credit drawn information yet but with known credit drawn limits, which are subject to dynamics of account credit risk profile and macroeconomic factors.

3.1.5.5 Machine Learning Models

As with LGD modeling, machine learning models have been popularly used in EAD modeling due to the heterogeneity embedded in EAD. For LEQ, CCF, and EADF, regression trees (CART) can be used to explore their complexity and relationship with risk factors. As we pointed out early, CART are useful tools to explore data quality issues and outliers, segmentation, variable importance and selection, local nonlinear structures, and proper model interpretation. Most often, enhanced CART models using bagging, random forest, and boosting discussed in Sect. 3.1.1.3 are used as challenger or benchmark models to make sure the champion model performs as required.

EAD modeling is heavily based on the reference data and ML models as strong data exploratory tools have advantages over classical exploratory tools like frequency and bivariate analysis. Its strong ability to catch complex nonlinear relationships between the target and risk factors helps to identify better segmentation, key risk factors, important local structures, and intuitive interpretations.

ML models could easily overfit the data, and a balance between model accuracy and complexity must be achieved to avoid trivial or overfitted models. Due to interpretability and stability, even with the enhanced ML models, ML models are seldom used as the primary model, especially in the regulatory areas. They are more likely used as challenger or benchmark models. This is also the case with EAD modeling. In Chaps. 4, 5, and 6, we will present more details on how ML models are used for EAD modeling in different risk management frameworks and what and how model risks related to ML models should be managed.

3.2 Cohort-Level Models

As we discussed in Sect. 2.2 , data could be available only in cohort format due to that either the account-level data are not available or too complex to handle. There are also situations that certain measures are difficult for individual account but easy for aggregation. One such example is the mortgage prepayment measure, which includes both full payoff and partial overpayment. In account level, handling partial overpayment is extremely tedious; while in a portfolio or cohort level, aggregating all prepayments either full or partial is natural in the mortgage servicing process, and the actual monthly prepayment rate for the entire cohort is just the ratio of this aggregated prepayment amount and the total aggregated unpaid principal balance at the beginning of the month.

Aggregating data into cohort data is also a way to simplify the data structure and reduce size for enormous large data sets. For credit products with long contract terms (e.g., 30-year mortgage), full transaction data for a large portfolio could be big. For some applications, handling such large data set is either not feasible even with the current computing power or may not be necessary. In Sect. 3.2.2, we will show when a pool/cohort-level prepayment model is preferred in such situation. There are also situations under which applications may prefer models based on cohort data for simplified interpretation without losing much accuracy. For example, for some applications in asset and liability management, vintage and other account characteristics-based cohort analysis is preferred to account-level analysis.

Cohort is commonly defined based on some account characteristics. The granularity of cohort could vary significantly. It could be defined as granular as by multiple account characteristics (e.g., loan to value or LTV and FICO combination for mortgage) or by the entire portfolio for sub-portfolio based on the application requirements and data availability (e.g., a sub-portfolio loss forecasting in CCAR based on its loss rate and a volume scalar).

The target of cohort-level models is usually a rate-based measure calculated from the corresponding balances. So, cohort-level models commonly fit a rate using either loglinear or nonlinear models. For forecasting, besides the forecasted rates based on risk factors, the cohort balance also needs to be forecasted. The cohort-level exposure at default similar to the account-level EAD discussed in Sect. 3.1.5 can be projected for committed and uncommitted products.

For cohort-level modeling, we lose the individual account information, which could impact our modeling approach. For cohort severity, instead of the post-default resolution and transaction information for individual account, we only have the cohort-level post-default loss information. So, the one-to-go regression methods are commonly used.

3.2.1 Default and Prepayment Rate Models

At the cohort level, default and prepayment are additive periodic measures. In the same period, default and prepayment can be modeled either independently or jointly. First, for a cohort h of credit product with active balance UPB_{ht} at the beginning of period t, the default and prepayment rate are defined as:

$$\mathrm{DR}_{ht} = \frac{\mathrm{DB}_{ht}}{\mathrm{UPB}_{ht}} \tag{3.89}$$

$$\mathrm{PR}_{ht} = \frac{\mathrm{PB}_{ht}}{\mathrm{UPB}_{ht}} \tag{3.90}$$

where DB_{ht} and PB_{ht} are default balance and prepaid balance during period t for cohort h. One can sum up these two if the total termination rate $\mathrm{TR}_{ht} = \frac{\mathrm{DB}_{ht}+\mathrm{PB}_{ht}}{\mathrm{UPB}_{ht}}$ is the concern.

Default rate and prepayment rate are commonly fitted with loglinear models:

$$\log(\mathrm{DR}_{ht}) = \alpha^{\mathrm{DR}} + \sum_{i=1}^{p^{\mathrm{DR}}} \beta_i^{\mathrm{DR}} x_{iht} + \sum_{k=1}^{q^{\mathrm{DR}}} \gamma_k^{\mathrm{DR}} z_{kt} + \epsilon_{ht} \tag{3.91}$$

$$\log(\mathrm{PR}_{ht}) = \alpha^{\mathrm{PR}} + \sum_{i=1}^{p^{\mathrm{PR}}} \beta_i^{\mathrm{PR}} x_{iht} + \sum_{k=1}^{q^{\mathrm{PR}}} \gamma_k^{\mathrm{PR}} z_{kt} + \epsilon_{ht} \tag{3.92}$$

where x_{iht} are cohort characteristics and z_{kt} are macroeconomic or other risk factors not specific on the cohort. These loglinear models can be fitted independently with least squares:

$$\min_{\alpha,\beta,\gamma} \sum_{h=1}^{N} \sum_{t=1}^{T} \left(\log(\mathrm{DR}_{ht}) - \log\left(r_{ht}^{\mathrm{DR}}\right)\right)^2 \tag{3.93}$$

$$\min_{\alpha,\beta,\gamma} \sum_{h=1}^{N}\sum_{t=1}^{T}\left(\log(\mathrm{PR}_{ht}) - \log\left(r_{ht}^{\mathrm{PR}}\right)\right)^2 \tag{3.94}$$

where r_{ht}^{DR} and r_{ht}^{PR} are observed cohort default and prepayment rates in period t, respectively. One can also fit the total termination rate TR_{ht} using a loglinear model, which is called the total termination model.

The loglinear model projects the rate in log scale, for example, the predicted cohort default rate:

$$\widehat{\mathrm{DR}}_{ht} = \exp\left(\hat{\alpha}^{\mathrm{DR}} + \sum_{i=1}^{p^{\mathrm{DR}}}\widehat{\beta}_i^{\mathrm{DR}} x_{iht} + \sum_{k=1}^{q^{\mathrm{DR}}}\widehat{\gamma}_k^{\mathrm{DR}} z_{kt}\right) \tag{3.95}$$

for cohort h with risk factors (x_{iht}, z_{kt}) at period t. Its standard error requires a log scale adjustment of the standard error of ϵ_{ht}.

Loglinear models are simple, but may not be sufficient for complex effects, especially for prepayment cohort models, for which complex nonlinear effects with interaction between cohort characteristics and macroeconomic factors are required. So, the following models are used:

$$\mathrm{DR}_{ht} = \exp\left(\alpha^{\mathrm{DR}} + \sum_{i=1}^{p^{\mathrm{DR}}}\beta_i^{\mathrm{DR}} f_i^{\mathrm{DR}}(x_{iht}) + \sum_{k=1}^{q^{\mathrm{DR}}}\gamma_k^{\mathrm{DR}} g_k^{\mathrm{DR}}(z_{kt}) + \sum_{j=1}^{s^{\mathrm{DR}}}\delta_j^{\mathrm{DR}} u_j^{\mathrm{DR}}\left(x_{jht}, z_{jt}\right) + \right) + \epsilon_{ht} \tag{3.96}$$

$$\mathrm{PR}_{ht} = \exp\left(\alpha^{\mathrm{PR}} + \sum_{i=1}^{p^{\mathrm{PR}}}\beta_i^{\mathrm{PR}} f_i^{\mathrm{PR}}(x_{iht}) + \sum_{k=1}^{q^{\mathrm{PR}}}\gamma_k^{\mathrm{PR}} g_k^{\mathrm{PR}}(z_{kt}) + \sum_{j=1}^{s^{\mathrm{PR}}}\delta_j^{\mathrm{PR}} u_j^{\mathrm{PR}}\left(x_{jht}, z_{jt}\right) + \right) + \epsilon_{ht} \tag{3.97}$$

where f_i^{DR}, f_i^{PR}, g_k^{DR}, g_k^{PR}, u_j^{DR}, and u_j^{PR} are nonlinear functions (e.g., splines). u_j^{DR} and u_j^{PR} are for the interactions between cohort characteristics and macroeconomic factors. A well-known example for these functions is the refinance incentive multiplier in mortgage prepayment models. Figure 3.7 shows a nonlinear effect of the refinance incentive measured as the difference between the cohort weighted average coupon WAC_h and market refinance rate r_{market}. Prepayment due to refinance quickly increases when the incentive becomes positive and becomes flat when the incentive exceeds certain level. This incentive effect has been observed in many refinance booms when market interest rate decreases deeply.

Nonlinear models are estimated from least squares on the scale of rates:

$$\min_{\alpha,\beta,\gamma,\delta} \sum_{h=1}^{N}\sum_{t=1}^{T}\left(\mathrm{DR}_{ht} - r_{ht}^{\mathrm{DR}}\right)^2 \tag{3.98}$$

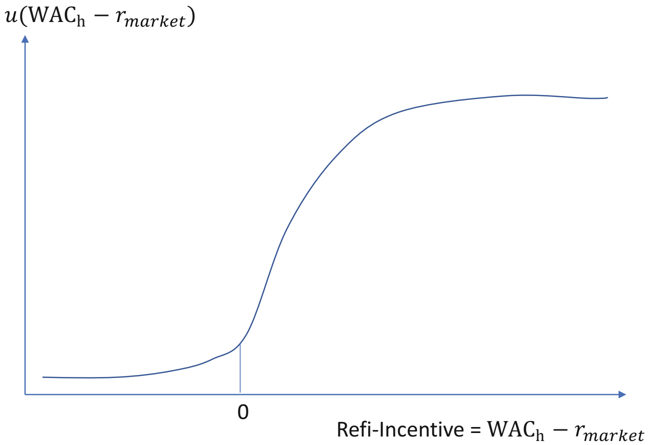


Fig. 3.7 S-curve

$$\min_{\alpha,\beta,\gamma,\delta} \sum_{h=1}^{N} \sum_{t=1}^{T} \left(\mathrm{PR}_{ht} - r_{ht}^{\mathrm{PR}}\right)^2 \tag{3.99}$$

Cohort-level risk event rate (or frequency) modeling has the power to fit complex model structures and local variations. However, overfitted models often perform poorly in projection of future trends. Balance between the model fitting and projection power has to be considered when these models are used for future cash flow projections.

3.2.2 Cohort Severity and Exposure Balance Models

At cohort level, we don't have the post-default transaction or loss resolution information any more alike at the account level. We only have the aggregated post-default cohort net loss NL_{ht} corresponding to default balance GL_{ht} measured at a snapshot time t. Based on whether applying discounting or not, the net loss NL_{ht} can be economic loss or accounting loss, respectively. So, the observed severity for cohort h in period t is:

$$\mathrm{Severity}_{ht} = \frac{\mathrm{NL}_{ht}}{\mathrm{GL}_{ht}} \tag{3.100}$$

The same as the account-level reference default data for LGD modeling, cohort-level reference default data include snapshots of reference cohorts with calculated severity measure and cohort characteristics measure at each snapshot. Macroeconomic factors are merged with the cohort-level reference default data at the snapshot date. At cohort level, severity is within (0, 1). So, the cohort-level severity is fitted with a linear regression model:

$$\text{Severity}_{ht} = \alpha + \sum_{j=1}^{p} \beta_j x_{jht} + \sum_{k=1}^{q} \gamma_k z_{kt} + \epsilon_{ht} \tag{3.101}$$

where x_{jht} are cohort-specific risk factors and z_{kt} are macroeconomic factors measure at snapshot time t.

At cohort level, the exposure at default also operates differently with committed and uncommitted credit products. For committed credit products, cohort-level unpaid principal balance (UPB) is the exposure. UPB can be projected based on projected default and prepayment rates as well as the cohort amortization factor from weighted average coupon (WAC) similar as at the account level discussed in Sect. 3.1.5.1.

For uncommitted credit products, cohort-level utilization and credit limit act the same as those at the account level. However, the exposure at default definition at account level doesn't apply at cohort level since we lose the individual account default information. Instead, we have the default balance measured at each snapshot, which has taken into account the extra credit draw before default. So, the DR defined by (3.89) in Sect. 3.2.1 for uncommitted product includes extra credit draw before default. Taking all these practices into consideration, for the final loss forecast with uncommitted credit product cohort, we only need the active balance at the beginning of each report time.

The dynamics of active cohort balance for uncommitted credit products could become complex due to the mixture of multiple factors – default, prepayment, minimum payment, and extra credit draw. So, in practice, top-down methods are preferred. Time series models, like ARMA(d_a, d_m) models, are popularly used for the cohort-level active balance projection:

$$\text{AR}(L)\ \text{UPB}_{h,t} = a + \text{MA}(L)\epsilon_{h,t} \tag{3.102}$$

where $\text{AR}(L) = \left(1 - b_1 L - b_2 L^2 - \ldots - b_{d_a} L^{d_a}\right)$ and $\text{MA}(L) = (1 - c_1 L - c_2 L^2 - \ldots - c_{d_m} L^{d_m})$ are the autoregression and moving average lagging operators with order d_a and d_m, respectively; a is the intercept and $\epsilon_{h,\ t} \sim N(0, \sigma^2)$ are white noise with standard error σ^2.

For uncommitted credit products, there is an alternative to the DR/UPB modeling by directly building models to project the gross loss rate (GLR) based on the default balance and cohort credit limit, which is parallel to the EADF at account level:

$$\text{GLR}_{ht} = \frac{\text{DB}_{ht}}{\text{L}_{ht}} \tag{3.103}$$

where GLR_{ht} is the gross loss rate, L_{ht} is the credit limit, and DB_{ht} is the default balance for cohort h observed in period t.

GLR_{ht} can be fitted with a loglinear or nonlinear model as for the DR or PR described in Sect. 3.2.1. Since L_{ht} is relatively stable and depends more on business loss mitigation decisions, some management judgments are commonly used to adjust the projection for L_{ht}. Then the final loss for cohort h in period t is projected as GLR_{ht} * $\mathrm{Severity}_{ht}$ * L_{ht} instead of DR_{ht} * $\mathrm{Severity}_{ht}$ * UPB_{ht}.

3.2.3 Loss Rate Models

For certain credit products (e.g., credit products without collaterals), loss rate can be obtained directly from periodic net loss (e.g., write-offs) and the corresponding exposure balance at the cohort level.

$$\mathrm{LR}_{ht} = \frac{\mathrm{NL}_{ht}}{\mathrm{UPB}_{ht}} \tag{3.104}$$

where UPB_{ht} is the unpaid principal balance for cohort h at the beginning of period t and NL_{ht} is the net loss corresponding to UPB_{ht}. UPB_{ht} only counts the active credit balance (excluding default balance). NL_{ht} records the cohort-level net loss corresponding to the risk events from the exposure of UPB_{ht}. So, there could be a delay in recording NL_{ht}. However, the recording of the write-offs must be consistent with the flow of "active balance – grow loss – net loss" within the same period t for the cohort, such that LR_{ht} is properly defined.

The loss rate LR_{ht} is commonly modeled with a loglinear model with a time series error model:

$$\log(\mathrm{LR}_{ht}) = \alpha^{\mathrm{LR}} + \sum_{i=1}^{p^{\mathrm{LR}}} \beta_i^{\mathrm{LR}} x_{iht} + \sum_{k=1}^{q^{\mathrm{LR}}} \gamma_k^{\mathrm{LR}} z_{kt} + \epsilon_{ht} \tag{3.105}$$

$$\epsilon_{\mathrm{ht}} \sim \mathrm{ARMA}(\mathrm{d_a}, \mathrm{d_m}) \tag{3.106}$$

where x_{iht} are cohort characteristics and z_{kt} are macroeconomic or other risk factors not specific on the cohort and ARMA(d_a, d_m) is the ARMA model defined in (3.102) with order d_a and d_m for the autoregression and moving average components, respectively.

The logic for using the ARMA model for the residuals of the loglinear model with the loss rate is that very often these residuals are correlated as time series, alike the residuals of default rate or prepayment rate fitted with loglinear models as described in (3.91) and (3.92).

The loss rate can be projected as:

$$\widehat{\mathrm{LR}}_{ht} = \exp\left(\widehat{\alpha}^{\mathrm{LR}} + \sum_{i=1}^{p^{\mathrm{LR}}} \widehat{\beta}_i^{\mathrm{LR}} x_{iht} + \sum_{k=1}^{q^{\mathrm{LR}}} \widehat{\gamma}_k^{\mathrm{LR}} z_{kt} + \widehat{\epsilon}_{ht} \right) \tag{3.107}$$

where α, β, γ are estimated through least squares with observed loss rates and $\widehat{\epsilon}_{ht}$ is projected from the fitted ARMA(d_a, d_m) model.

To project the loss, the future UPB as a scalar needs to be projected by ARMA model defined in (3.102) as in Sect. 3.2.1.

3.3 Model Selection

For all models we discussed in this chapter, there is a critical question in practical modeling applications – what models should be used? While this question needs to be answered in the initial model design stage and confirmed in the model validation process from the view of concept soundness of the model application, there are basic criteria and procedures that model selection including variable selection should follow to guarantee that the selected models have solid performance and certain required properties. We focus on the regression types of models that include linear, loglinear, and generalized linear models. Other types of models may need separate treatments and may be discussed individually in the book.

In this section, we first present some basic statistical criteria to follow for model and variable selection such that the selected models have some desired properties. Then, we explore how to incorporate business inputs into the modeling process. Once we have all proper inputs, we present two main model selection procedures: one is the commonly used stepwise variable selection procedure and the other is a more comprehensive model selection procedure called adaptive and exhaustive variable selection (AEVS).

3.3.1 Statistical Criteria and Tests

We describe the statistical criteria based on linear regression models. Most often, these criteria can be extended naturally to loglinear and generalized linear models. Assume the dependent variable $\boldsymbol{y}$ and the covariates $\boldsymbol{X} = (\boldsymbol{x_1}, \ldots, \boldsymbol{x_p})'$ have a linear relationship:

$$\boldsymbol{y} = \sum_{j=1}^{p} \beta_j \boldsymbol{x_j} + \epsilon \tag{3.108}$$

where $\beta = (\beta_1, \ldots, \beta_p)'$ is the coefficients vector (including the intercept) and $\boldsymbol{\epsilon}$ is a random variable with 0 mean and variance σ^2, i.e., $E(\epsilon) = 0$, $\text{var}(\epsilon) = \sigma^2$, and independent from the covariates $\boldsymbol{X}$.

Given a sample (y_i, X_i), $i = 1, \ldots, n$ of the dependent variable and covariates, the ordinary least squares (OLS) estimator of the coefficient is:

$$\widehat{\beta} = (X'X)^{-1} X'Y \tag{3.109}$$

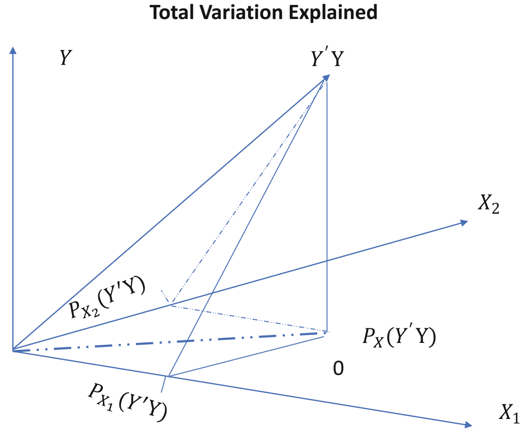


Fig. 3.8 Total variation

where $X = (X_1, \ldots, X_n)'$ and $Y = (y_1, \ldots, y_n)'$. Under model (3.108), the OLS estimator is the best linear unbiased estimator (or BLUE) that minimizes the squared errors:

$$\text{argmin}_\beta (Y - X\beta)'(Y - X\beta) \tag{3.110}$$

The OLS prediction of the dependent $\widehat{Y} = X\widehat{\beta} = X(X'X)^{-1}X'Y$ is the projection of Y on the p-dimension X space, and $\widehat{Y}'\widehat{Y} = Y'X(X'X)^{-1}X'Y = P_X(Y'Y)$ is the variation explained by the X space for the total variation $Y'Y$ in the Y direction. Clearly, the higher the dimension of the X space, the more variation $Y'X(X'X)^{-1}X'Y$ explained by the X space. See Fig. 3.8 for an illustration with two dimensions. The OLS estimator minimizes the residual variation since $\left(Y - X\widehat{\beta}\right)'\left(Y - X\widehat{\beta}\right) = Y'Y - \widehat{Y}'\widehat{Y}$.

F-statistics

The *F-statistics* of any dimension $\boldsymbol{x_i}$ is defined as the difference between the explained variations when $\boldsymbol{x_i}$ is included and when $\boldsymbol{x_i}$ is excluded:

$$F_{\boldsymbol{x_i}} = Y'X(X'X)^{-1}X'Y - Y'X_{-i}\left(X'_{-i}X_{-i}\right)^{-1}X'_{-i}Y \tag{3.111}$$

where X_{-i} is the n by $(p - 1)$ dimension matrix by removing the ith column from X as a sample of size n from $\boldsymbol{X} = (\boldsymbol{x_1}, \ldots, \boldsymbol{x_{i-1}}, \boldsymbol{x_{i+1}} \ldots, \boldsymbol{x_p})'$.

So, the F-statistics $F_{\boldsymbol{x_i}}$ measures marginal contribution in variation explained by adding the dimension $\boldsymbol{x_i}$. It is a valid measure of a variable contribution when it is added or removed from a model and is popularly used in the stepwise model selection procedure, which we will discuss later.

***R^2* and Adjusted $\overline{R}^2$**

The R^2 measures the model explained variation as a percentage of the total variation:

$$R^2 = \frac{\widehat{Y}'\widehat{Y}}{Y'Y} = 1 - \frac{\left(Y - X\widehat{\beta}\right)'\left(Y - X\widehat{\beta}\right)}{Y'Y} \tag{3.112}$$

If the effect of the intercept is not counted, the total variation $Y'Y$ should be replaced by the correct total variation $\left(Y - \overline{Y}\right)'\left(Y - \overline{Y}\right)$, where $\overline{Y}$ is the response mean vector. When the dimension of covariates is considered, R^2 can be adjusted as $\overline{R}^2 = \left(R^2 - \frac{p}{n}\right)/\left(1 - \frac{p}{n}\right)$.

The above definition of R^2 applies to linear and loglinear models. For generalized linear models, the generalization of coefficient of determination or R^2 can be defined based on the likelihood function as we discussed in Sect. 3.1.1.1 as $R^2 = 1 - \left\{\frac{L(0)}{L\left(\widehat{\beta}\right)}\right\}^{\frac{2}{N}}, R^2_{\max} = 1 - \{L(0)\}^{\frac{2}{N}}, \overline{R}^2 = \frac{R^2}{R^2_{\max}}$.

−2LogLikelihood

If $\boldsymbol{\epsilon}$ in the linear model (3.108) follows a normal distribution with mean 0 and variance σ^2, the conditional likelihood of a random sample y_i given X_i, $i = 1, \ldots, n$ is:

$$\left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)\right\} \tag{3.113}$$

and

$$-2\text{LogLikelihood} = n\left(\log\left(\sigma^2\right) + \log(2\pi)\right) + \frac{1}{\sigma^2}(Y - X\beta)'(Y - X\beta) \tag{3.114}$$

The maximum likelihood estimator (MLE) for β is identical to the OLS estimator $\widehat{\beta}$, and the MLE for σ^2 is the *mean squared error* (MSE) $\widehat{\sigma}^2 = \frac{1}{n}\left(Y - X\widehat{\beta}\right)'\left(Y - X\widehat{\beta}\right)$. Replace these parameters with their MLEs and ignore the constant terms, and we have:

$$-2\text{LogLikelihood} \sim n\log\left(\widehat{\sigma}^2\right) = n\log\left(\frac{1}{n}\left(Y - X\widehat{\beta}\right)'\left(Y - X\widehat{\beta}\right)\right) \tag{3.115}$$

which is the only quantity in the loglikelihood function depending on the p covariates and measures how good the covariates fit the dependent variable in the log scale of the residual variation.

If the variance σ^2 is assumed constant, then only replace β with the OLS estimator $\widehat{\beta}$ and ignore the constant terms, and we have:

$$-2\text{LogLikelihood} \sim \frac{1}{\sigma^2}\left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right) \tag{3.116}$$

which is equivalent to the MSE as a measure for the goodness-of-fit of the dependent variable using the p covariates.

In generalized linear models, the variance σ^2 corresponds to the dispersion parameter. So, whether this parameter is assumed constant or not, with the easy formulation of likelihood functions for generalized linear models, $-2LogLikelihood$ can be naturally extended as a measure of the goodness-of-fit of the dependent variable using the p covariates in generalized linear models.

AIC and SIC

$-2LogLikelihood$ is based on the residual variation and decreasing with the increase of the model dimension p, which means more complex models. In order to achieve a balance between model complexity and the goodness-of-fit, a penalty of model complexity is added to $-2LogLikelihood$:

$$\text{AIC} = n\log\left(\hat{\sigma}^2\right) + 2p = -2\text{LogLikelihood} + 2p \tag{3.117}$$

$$\text{SIC} = n\log\left(\hat{\sigma}^2\right) + p\log(n) = -2\text{LogLikelihood} + p\log(n) \tag{3.118}$$

AIC by Akaike (1974)[20] uses $2p$ as the model complexity penalty, while SIC by Schwarz (1978)[21] uses a more severe penalty p log (n), which increases with sample size. So, with large sample size, SIC prefers much simpler models. AIC can be explained using entropy theory (from which the name information criteria or IC was abstracted) under normal distribution, while SIC can be considered as a Bayes approximation under the exponential distribution family. Both AIC and SIC can be explained as approximations of the conditional expected prediction error under entropy of the exponential family with different priors shown by Chen (1993).[22] While the theoretic authentication of these criteria as model selection rules may not be rigor, in practice, they have been proved highly efficient and popularly used in likelihood-based model analysis and selection.

There are some other versions of information criteria using different penalty terms, but largely those penalties fall between that of AIC and SIC.

Cross-Validation

Cross-validation (CV) is the approach of analyzing and modeling on a part of data given and then validating on the left part. The approach was used as early as in the

[20] Akaike, H. (1974), "A New Look at the Statistical Model Identification," IEEE Transactions on Automatic Control, 19, 716–723.

[21] Schwarz, G. (1978), "Estimating the Dimension of a Model," Annals of Statistics, 6, 461–464.

[22] Chen, C. (1993). "Mean Loss of Prediction and Its Asymptotic Unbiased Estimators." MS Thesis. Institute of Systems Science, Academia, China.

1930s and 1940s.[23] After about 30 years, based on previous research, a series of important papers both on theory and application of CV were presented by Stone, M. (1974a, b, 1977a, b).[24] But the theory of CV itself is still incomplete, although its nonparametric characteristics have attracted much more attentions in the machine learning and AI applications in recent years.

Actually, CV was generated from the sense of prediction (Geisser, 1974, 1975).[25] Based on the prediction error, Chen and Zhang (2000)[26] proposed the mean loss of prediction (MLP) concept and proved that as AIC, CV statistics are also asymptotic unbiased estimators of MLP under moment estimation. For linear regression (3.108), the leave-one-out (LOO) CV is defined as:

$$\mathrm{CV}_{nn} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - X_i'\widehat{\beta}_{-i}\right)^2 \tag{3.119}$$

where $\widehat{\beta}_{-i}$ is the OLS with the ith observation (y_i, X_i) removed from the model estimation. And the conditional expected prediction error (or MLP) defined in Chen and Zhang (2000) is:

$$\mathrm{MLP} = \int \left(\boldsymbol{y} - \boldsymbol{X}'\widehat{\beta}\right)^2 dF(\boldsymbol{X}, \boldsymbol{y}) \tag{3.120}$$

Under some regular conditions, one can show that $E_F(\mathrm{CV}_{nn} - \mathrm{MLP}) = \frac{p\sigma^2}{n(n-1)} + O(n^{-3})$. This can be extended to the entropy loss, under which the asymptotic equivalence of CV and AIC can be established.

CV can be applied to *v-fold* by splitting the data into v equal-sized groups. The LOO CV is usually called *n-fold* CV. Depending on different applications, for model performance measure, LOO CV is commonly used.

[23] Larson, S. C. (1931). "The Shrinkage of the Coefficient of Multiple Correlation." J. Educ. Psychol. 22, 25–55.

Horst, P. (1941). "Prediction of Personal Adjustment." New York: Social Science Research Council (Bulletin 48).

[24] Stone, M. (1974a). "Cross-Validatory Choice and Assessment of Statistical Predictions." (With discussion) J. R. Statist. Soc. B 36,111–147.

Stone, M. (1974b). "Cross-Validation and Multinomial prediction." Biometrika 61, 509–515.

Stone, M. (1977a). "Asymptotics for and against Cross-Validation." Biometrika 64, 29–35.

Stone, M. (1977b). "An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's criterion." J. R. Statist. Soc. B 39, 44–47.

[25] Geisser, S. (1974). "A Predictive Approach to the Random Effect Model." Biometrika 61, 101–107.

Geisser, S. (1975). "The Predictive Sample Reuse Method with Applications." JASA. 70, 320–328.

[26] Chen, C. and Zhang, YG. (2000). "Variable Selection of Structural Models." Proceedings of World Multiconference on Systemics, Cybernetics, and Informatics, Vol. 8, 769–773.

In addition to the above statistical criteria, when time series models are required, the following statistical tests and criteria are also popularly used in model selection.

Stationarity Test

The independence constrain of $\boldsymbol{\epsilon}$ in (3.108) for a series of observations (y_t, X_t), $i = 1, \ldots, T$ can be relaxed as a time series with correlations. The time series models require stationarity, which means the variance of these ϵ_t is constant. The KPSS[27] test is the popular one used for this purpose. The KPSS test assumes that the residuals ϵ_t follow either a level or trend random process:

$$\epsilon_t = \xi t + r_t + w_t \tag{3.121}$$

where $\xi = 0$ means a level process, while $\xi \neq 0$ means a trend process; w_t is white noise $N\left(0, \sigma_w^2\right)$ and r_t is random walk:

$$r_t = r_{t-1} + u_t \tag{3.122}$$

where u_t are white noise $N\left(0, \sigma_u^2\right)$. The null hypothesis of the KPSS stationarity test is $H_0 : \sigma_u^2 = 0$, and the test statistics is:

$$\mathrm{LM} = \sum_{t=1}^{T} S_t^2 / \widehat{\sigma}_w^2 \tag{3.123}$$

where $S_t = \sum_{i=1}^{t} \widehat{e}_i$, $\widehat{\sigma}_w^2 = \frac{1}{T} \sum_{i=1}^{T} \widehat{e}_i^2$, and $\widehat{e}_i, i = 1, \ldots, T$ are the residuals from the regression (3.121) under H_0. The asymptotic distribution of the test statistics LM can be obtained, and its finite sample critical values can also be obtained through simulation.

Series Correlation Test

The residuals after the ARMA model fitting should be checked for correlation. A commonly used correlation test is the Ljung-Box test.[28]

Heteroscedasticity Test

The residuals after the ARMA model fitting should be checked for heteroscedasticity or constant variance. A commonly used heteroscedasticity test is the White test.[29]

[27] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin (1992): "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root." Journal of Econometrics 54, 159–178.

[28] G. M. Ljung and G. E. P. Box (1978). "On a Measure of a Lack of Fit in Time Series Models." Biometrika. 65 (2): 297–303.

[29] White, H. (1980). "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." Econometrica. 48 (4): 817–838.

Normality Test

The residuals after the ARMA model fitting should be checked for normality. A commonly used normality test is the Shapiro-Wilk test.[30]

3.3.2 Stepwise Procedure

Variable selection is essential in modeling and a critical step in model selection. For credit risk modeling as we discussed in this chapter, there are various risk factors available, and a variable selection process is necessary. First, for credit risk models, especially models for regulatory purposes, interpretability is required, and thus parsimonious models are preferred. So, only risk factors that show significance and provide appropriate interpretability should be selected in the model. Second, very often among available risk factors, some of them are highly correlated, and models with severe confounding effects could lead to erroneous model outputs and conclusions.[31]

So, careful variable analysis and selection are required in both qualitative and quantitative aspects.

For a quantitative view, automated procedures are required to process all available risk factors to filter out unimportant ones and have a pool of potential useful risk factors. The forward and backward variable selection procedures are the two popularly used procedures for this purpose.

The forward selection procedure starts from the null model and then evaluates all available variables for their contribution to the model based on the F-statistics defined in Sect. 3.3.1 when it is added into the model. Then, the variable with the largest contribution to the model is first added into the model. Based on the new model, the rest available variables are examined, and the one with the largest contribution based on F-statistics is added to the new model, and this process keeps going until all available variables are added into the final model or stops in the middle according to some stopping rule. The backward selection procedure is the opposite, and it starts from the full model with all available variables. All variables in the model are checked to see which one is the least significant in F-statistics when it is removed from the model, and then it is removed from the model. Based on the new model, then the procedure repeats until all variables are removed from the model or stops in the middle according to some stopping rule.

Both the forward and backward selection procedures couldn't guarantee the selected subset of variables has the best performance according to the selection criterion like F-statistics due to the one-directional processing of these procedures.

[30] Shapiro, S. S. and Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)." Biometrika. 52 (3–4): 591–611.

[31] Chen C., Chock D. and Winkler S.L. (1999). "A Simulation Study of Confounding in Generalized Linear Models for Air Pollution Epidemiology." Environmental Health Perspectives, Vol. 107, 217–222.

For forward selection procedure, the latter added variable could make some of the current variables in the model insignificant, while for backward selection procedure, the removing of an insignificant variable could also lead to a previous removed one becoming significant. The stepwise procedure intends to overcome these issues by combining forward and backward procedures. It starts as a forward selection procedure, but with a backward elimination check after each forward selection. In a traditional implementation of stepwise selection procedure, the same entry and removal of F-statistics threshold for the forward selection and backward elimination are used to access contributions of factors as they are added into or removed from a model. If, at a step of the stepwise procedure, any factor is not significant at a specific level, then the least significant one is removed from the model, and the procedure proceeds to the next step. This ensures that no factor can be added into the model while some factor currently in the model is deemed not significant. Only after all necessary deletions have been accomplished can another factor be added into the model. In this case, the factor whose addition yields the most significant F-statistics is added into the model, and the procedure proceeds to the next step. The stepwise procedure stops when none of the factors outside the model has an F-statistics significant at the specified level and every factor in the model is significant at this level. In some cases, neither of these two conditions for stopping is met and the sequence of models cycles. In this case, the stepwise processing stops at the second cycle. Some further investigation is required to assess the cycling. Due to this cycling issue, some stepwise selection procedures modify the stopping rule by adding an early stopping rule like limiting the number of factors in the final selected model.

For stepwise selection, besides the F-statistics, there are other criteria that can be used, e.g., AIC and SIC. In such cases, the adding or removing is based on a smaller value of these criteria that can be achieved. Also, stepwise selection can be naturally extended to generalized linear models and other models with proper criteria for adding and removing factors.

3.3.3 Lasso

In addition to the traditional stepwise selection procedure, there are other alternative variable selection procedures. The most popular ones are Lasso (least absolute shrinkage and selection operator) and LAR (least angle regression). These methods add or remove factors based on a version of the ordinary least squares where the sum of the absolute regression coefficients is constrained or shrunken in the case of LAR. Due to the similarity between Lasso and LAR, in this section, we focus on Lasso. Lasso extends the ridge regression, which overcomes the degenerated case when risk factors are (or near) linearly correlated by adding a sum of squared coefficient penalty to the least square objective function as for OLS:

$$\text{argmin}_{\beta}\left\{(Y - X\beta)'(Y - X\beta) + \lambda\sum_{i=1}^{p}\beta_i^2\right\} \tag{3.124}$$

This leads to the weighted LS (or WLS) estimator of ridge regression:

$$\widehat{\beta}_{\text{Ridge}} = (X'X + \lambda I)^{-1}X'Y \tag{3.125}$$

Even with a small value $\lambda > 0$, $\widehat{\beta}_{\text{Ridge}}$ is much more stable than the OLS $\widehat{\beta}$ when $X'X$ is near degenerated, though it adds bias and is not BLUE anymore. Instead of sum of squared coefficient penalty, Lasso uses a sum of absolute coefficient penalty:

$$\text{argmin}_{\beta}\left\{(Y - X\beta)'(Y - X\beta) + \lambda\sum_{i=1}^{p}|\beta_i|\right\} \tag{3.126}$$

It turned out that the Lasso estimator $\widehat{\beta}_{\text{Lasso}}$ assigns 0 values to its components based on the magnitude of λ, and furthermore, based on a sequence of increasing or decreasing λ values, the full path of the factors removing from or adding into the model can be found efficiently through the LAR algorithm. So, Lasso can function as a backward or forward variable selection operator, from which it's named. Also, it can be naturally extended to generalized linear models and other areas.

3.3.4 Expert Inputs

All criteria and procedures for model selection we discussed so far are either based on theory or data. There is an additional important input for model selection, which is expert judgment. The expert judgment is based on long-term expert experiences that could include critical risk factors, characteristics of risk factors, and expert causality analysis. Quantitatively, these can be transferred into required or rejected risk factors in the model, known sign of certain risk factors, and ranking of the final models.

To incorporate expert inputs, we have several options. First, we can combine quantitative criteria or components derived from the expert inputs into the model selection procedure. As an example, the variable selection procedure described in the next section has an adaptive way to include these inputs. Second, expert reviews of the top models selected by the automated procedure can be another way to include expert inputs.

Model selection incorporating expert inputs is a Bayesian procedure, which combines both prior information and historical data to achieve a better decision in case that historical data may not be complete.

3.3.5 Adaptive and Exhaustive Variable Selection (AEVS)

Variable selection procedures simply based on statistical criteria and tests have limitation in credit risk modeling. In credit risk modeling, very often, besides statistical criteria, qualitative rules based on business intuitions are required to select the most appropriate variables among a large number of candidate variables and their transformations. Simply applying these criteria sequentially is not sufficient to meet these requirements. For example, first applying statistical criteria to select a subset of variable combination and then applying the qualitative rules from business may either result in no model selected (qualitative rules reject all survived variable combinations from those criteria) or too many variable combinations selected. The sequential use of statistical criteria and qualitative rules neither can guarantee the selected mode is optimal even locally. On the other hand, applying the qualitative rules to all variable combination (exhaustive search) is usually not computationally feasible due to a huge number of such combinations.

In this section, we describe our proposed variable selection procedure named adaptive and exhaustive variable selection or AEVS. In AEVS, we first apply stepwise selection on all variables in the variable inventory with different selection criteria, e.g., F-statistics, Lasso, AIC, or SIC. Based on a combination of different criteria, we reduce the general candidate variable set to an initial smaller set of variables (e.g., less than 20 variables) on which the exhaustive search can be applied. We also guarantee that variables from the business intuition are included in the initial candidate variable set. A set of rejection rules would be applied on the model results from each of the variable combination from the exhaustive search on the initial variable set. The rules consist of qualitative checks (e.g., signs of coefficients matching business intuitions), statistical tests, and residual analysis. Models that pass rejection criteria then will be ranked according to a set of criteria such as goodness-of-fit and cross-validation as well as qualitative aspect of the model. The set of top ranked models will then be assessed by the model developers and reviewed by the business, and the final model is determined at the end of the procedure by considering all inputs.

In the following, we describe an application of the AEVS procedure on a loss rate regression model selection process. The application can be extended naturally to other modeling areas.

Loss Rate Modeling Process with AEVS

1. *Data Input Preparation*
   (a) Loss rate $Y = \log\left(\frac{\text{Quarterly Total Loss}}{\text{Loss Scalar.}}\right)$
   (b) Macro data X_{macro}, the list of macros chosen for a line of business (LOB)
   Calculate $T(X_{\text{macro}})$, the list of chosen meaningful transformations for each macro.
   Calculate $\text{Lag}(X_{\text{macro}})$, $\text{Lag}(T(X_{\text{macro}}))$, moving average of X_{macro}, and $T(X_{\text{macro}})$ over $i + 1$ quarters, $i \leq 5$.

Calculate forecasts of $T(X_{\text{macro}})$, $\text{Lag}(X_{\text{macro}})$, and $\text{Lag}(T(X_{\text{macro}}))$.

Determine $\text{sign}(X_{\text{macro}})$, the meaningful relationship (positive or negative) between X_{macro} and Y.

Determine $\text{Cate}(X_{\text{macro}})$, the category of macros, for instance, labor market, equity market, etc.

(c) N, the total number of X variables including X_{macro}, $T(X_{\text{macro}})$, $\text{Lag}(X_{\text{macro}})$, and $\text{Lag}(T(X_{\text{macro}})$ called four variations of X_{macro};select top $D = 10$ candidate variables from these N variables using each of the following methods:
- Lasso top D (10)
- Standard stepwise top D (10)

Combine variables selected from Lasso and stepwise.

Add the variables by chosen business intuitions.

Determine TopN, the combined list for selected variables derived above.

2. *Regression: Model Selection*
   (a) Generate M_j regression models from possible subset of TopN variables using exhaustive search, $j = 1, \ldots, n \sim 2^{\text{TopN}}$.
   (b) Reject M_j if any regression coefficient doesn't satisfy $sign(X_{\text{macro}})$.
   (c) Reject M_j if more than two variations of one X_{macro} are included in the model.
   (d) Reject M_j if the model F-statistics p-value $>5\%$.
   (e) Reject M_j if any regression coefficient p-value $>5\%$.
   (f) Reject M_j if residual stationarity KPSS test p-value $<5\%$.
   (g) Flag M_j for ARMA model fitting if any regression residual autocorrelation Ljung-Box test p-value $<1\%$. Based on flag, fit ARMA model on regression residuals (order p and q decided by the lags of ACF/PACF lags).
   (h) Flag M_j for attention if regression model or ARMA model residual heteroskedasticity White test p-value $<5\%$.
   (i) Reject M_j if regression model or ARMA model residual normality Shapiro-Wilk test p-value $<1\%$.

   This determines the selected candidate regression models.
3. *Model Ranking*

   For the survived candidate regression models after Step 2:
   (a) Rank by LOO cross-validation CV_{nn} (the lower the CV_{nn}, the higher the rank M_j has).
   (b) Rank by LOO cross-validation R-square (the higher the LOO R-square, the higher the rank M_j has).
   (c) Rank by in-sample fitting AIC (the lower the AIC, the higher the rank M_j has).
   (d) Rank by in-sample fitting R-square (the higher the R-square, the higher the rank M_j has).
   (e) Calculate the average rank of the four performance rankings above, and select the top 5 models.

From the five models, choose the final model based on LOB review and model assessment.

In the above AEVS application for regression models, we classify the rejection rules into two types: one type is hard rejection rules and the other type is soft rejection rules. Model subject to hard rejections will not be reviewed in the following steps. Models flagged with a soft rejection by rule like the residual heteroscedasticity White test will continue to be reviewed and ranked lower if selected as one of the top models.

AEVS can be naturally applied to loglinear and generalized linear models, though some of the rejection rules may not be applicable anymore. The procedure can also be modified to meet special model selection needs since the two characteristics of the procedure, adaptability to incorporate business inputs and optimality through exhaustive search, are integrated in the procedure naturally.

In summary, AEVS has the following advantages:

- AEVS integrates both business qualitative rules and quantitative statistical criteria naturally into an optimal variable selection procedure.
- AEVS guarantees that the selected model has local optimality with both qualitative rule and quantitative criteria combined through an exhaustive model search.
- AEVS extends optimality of its selected model by combining subsets selected from different variable selection techniques to form its initial variable candidate sets, e.g., Lasso and stepwise.
- AEVS is fully automated till the final model review.

4 Allowance for Credit Loss and CECL

In this chapter, we focus on the allowance for credit loss or ACL, especially with CECL, by applying the theories and methodologies in the previous chapters on both retail and wholesale portfolios.

We first introduce the commonly implemented ACL processes. Different firms may have their legacy financial reporting systems prior to CECL implementation. Such systems will require enhancements or even redevelopment to incorporate CECL. We provide some practices on CECL implementation on top of some existing ACL processes. Then we introduce two synthetic data sets, which will be extensively used in this chapter as examples for various stages of ACL modeling process. Programming is first provided in the prototype format for the purpose that readers can practice using their own preferred languages. Then R or Python code is partially provided as examples of implementation of these prototypes.

Model validation is a key component of the ACL process, especially with the implementation of CECL models. We present a full model validation framework illustrated with our synthetic data, on which we can manipulate for clear insights.

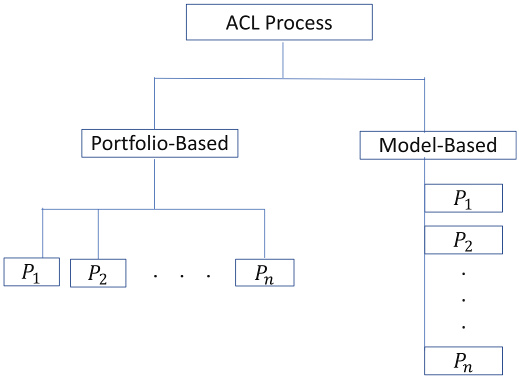
4.1 Introduction to ACL Processes

Over the years, financial institutes have built their ACL processes. These processes are usually first developed based on line of business, for example, retail or wholesale businesses. Retail businesses are related to consumer credit products, while wholesale businesses are more focused on the commercial credit products. Credit products from different line of businesses are managed differently as well as their performance is measured differently. These differences impact source information collection, data availability, and thus ACL modeling methodology.

Within the same type of credit products, commonly two types of ACL processes have been built. One is developing ACL models on each of the portfolio or called portfolio-based, and the other is developing a common ACL modeling process for

C. Chen, *Practical Credit Risk and Capital Modeling, and Validation*, Management for Professionals, https://doi.org/10.1007/978-3-031-52542-1_4

Fig. 4.1 ACL process



multiple portfolios or called model-based. Figure 4.1 shows the two types of ACL implementation process.

Under the incurred loss method, there is not much difference for these ACL processes due to their full dependence on the portfolio-specific historical data. However, some large institutions may have built their internal credit loss forecasts processes (e.g., ALLL) with long-term forecasts (e.g., 72 months for mortgage portfolios). For these large institutions, the adoption of CECL can be naturally executed based on their existing long-term forecast systems without rebuilding a full CECL system. For institutions without such long-term forecast systems, it will be hard to comply with CECL requirements by just realigning their current methods, especially for credit products with longer terms. In the following, we discuss how different ACL platforms can be built based on the two types of ACL processes.

4.1.1 Platforms

The portfolio-based and model-based ACL implementation processes result in two ACL platforms. The portfolio-based process builds the ACL platform based on individual portfolio. It starts from the line of business that manages the portfolio. The platform starts from resource allocation, which includes hardware, software, IT, and modeling groups. Then a project plan should be designed and reviewed by all parties and management. The project plan should describe the scope of the project, goal and periodic targets, responsibility assigning, testing, and production delivery processes.

The portfolio-based platform has the flexibility to select portfolio-specific methods for ACL modeling, which starts from data collection and extraction (or other ETL processing), model development and validation, testing, and monitoring. The disadvantage of the portfolio-based platform is higher cost due to resource request for each portfolio, and such resources may not be shared efficiently among portfolios. There is also the possibility that similar portfolios are handled differently and lack of the consistency.

The model-based ACL platform is an alternative to the portfolio-based. Instead of starting from the line of business, building such a platform is usually starting from the model development and governance groups per business requests. The production of the platform is a joint effort between line of business, finance, IT, and modeling groups. The platform is commonly designed for a group of similar portfolios, for example, retail products including mortgage, credit cards, auto loans, and small business lending products. Such portfolios share similar data processes and ACL modeling structures on the account level. The disadvantage of the model-based platform is its less flexibility when adding a new portfolio if the portfolio has differences in either data processing or modeling structures. So, when designing the model-based ACL platform, make sure its scope and the required extendibility.

4.1.2 CECL and CCAR

In Sect. 1.2.3, we have discussed the CECL impact on CCAR. There, we focused on the CECL impact on CCAR results and the integration of the two suites of modeling methodologies. Here we would explore the possibility of integration of the two risk management practices. Since from cost and resource perspectives, it is attractive to integrate the two risk management frameworks into one platform.

For most institutions, CCAR was implemented in a model-based process due to the fact that the CCAR components are strongly model-based, though within each component, income and loss projections are carried out separately across portfolios or line of businesses before aggregating the final results. Fortunately, as described in Sect. 1.4.2, ACL can be considered as a subcomponent of CCAR for the provision and loss forecasts, which leads to a natural integration of ACL into CCAR. So, the main platform integration overloading will fall on a consistency integration of data, portfolio segmentation, and, most importantly, the modeling methodologies.

For a successful CCAR/CECL data integration, some data reconciliation is needed. Very often, CECL models require account-level data, while CCAR models may just use the aggregated portfolio data. So, data reconciliation for the two risk management frameworks is the basic requirement for the integration of CCAR and CECL into one platform. On the other hand, macroeconomic variables for CCAR and CECL could be different due to their separate purposes as we discussed early. Still certain consistency checks are needed to make sure the two systems are not too extremely far from each other to be interpretable.

Portfolio segmentation will have an impact on the data reconciliation such that more granular segmentation may be required for the reconciliation. However, it is going to be an improvement in data and model transparency with such portfolio segmentation when combining CCAR and CECL into one platform.

The hardest part for a consistency integration of the two risk frameworks will be the modeling consistency as we discussed. Here we have two sets of loss forecast models – CCAR that focuses on the 9–13 quarters of loss forecast under different scenarios and CECL that focuses on the lifetime of loss forecast under a reasonable and supportable scenario. Some reconciliation methods we discussed in Sect. 1.2.3 require to be examined in practice though we think it is feasible to recalibrate the CECL model to adapt to both baseline and stressed scenarios.

4.2 Model Data for CECL

In this section, we present several synthetic data sets. These data sets will be used throughout the book.

4.2.1 A Sample of Mortgage Portfolio

We simulate two account-level transaction data sets. They are synthetic data sets based on a real portfolio of subprime mortgages. The subprime mortgage portfolio consists of 69,241 loans. We randomly select 10% from this portfolio and obtained 6924 loans, which mostly were originated from year 2000 to 2008, right before the housing crash and financial crisis. While the loan origination information is available from the original data, the loan transaction data, which records the monthly transactions of the portfolio, was simulated based on prespecified models. For each loan, the transaction path starts either from the booking date or its origination date if the loan was originated after the booking date and ends either due to a risk event (default or prepayment) or censored at the end of the observation window.

Table 4.1 shows a summary of these subprime loans with some origination information.

In Table 4.1, both the loan ID and origination date (OrigDate) indicate there are 6924 loans in the subprime loans sample. The acquisition index (aci) is a risk index based on the loan acquisition information (alike the credit score) that ranged from 300 to 850. The origination loan to value ratio (oltv) is the ratio of loan amount at origination (LoanSize in dollar shown in Table 4.1) to the appraisal house value at

Table 4.1 Summary information for the subprime mortgage portfolio sample

ID	OrigDate	aci	oltv	LoanSize	OHPI	OUER
Length: 6924	Length: 6924	Min.: 300	Min.: 1.528	Min.: 4568	Min.: 97.35	Min.: 1.800
Class: character	Class: character	1st Qu.: 437	1st Qu.: 71.429	1st Qu.: 167,223	1st Qu.:232.66	1st Qu.: 4.100
Mode: character	Mode: character	Median: 577	Median: 79.511	Median: 262,500	Median: 281.11	Median: 4.800
NA	NA	Mean: 574	Mean: 80.135	Mean: 287,784	Mean: 267.95	Mean: 5.464
NA	NA	3rd Qu.: 709	3rd Qu.: 80.000	3rd Qu.: 378,750	3rd Qu.: 309.82	3rd Qu.: 5.893
NA	NA	Max.: 850	Max.: 363.459	Max.: 2,812,500	Max.: 635.08	Max.: 27.742

origination. OHPI and OUER are the house price index (HPI) and unemployment rate (in percentage) at origination. The five quantiles of these factors and the means are presented in Table 4.1. These summary statistics do not show extremes as a typical subprime portfolio before the financial crisis.

Our first loan-level transaction data set was simulated based on static risk event probabilities. For each loan, the monthly default and prepayment rates are based on loan origination factors and kept constant over the transaction path. Over the transaction path, the loan could be terminated by either a default or a prepayment risk event or continue to be active. A binary variable is used to record each risk event. All loans have the 30-year term indicating a maximum number of 360 monthly transactions. Given that these subprime loans were high-risk loans with much higher default rates than conventional loans, the synthetic data assume faster termination speed than the real transaction history of these loans, which mostly are interest-only loans or pick-a-pay (PAP) loans.

Table 4.2 shows a summary of the synthetic transaction data of the sample subprime loans with some origination information.

There are 77,924 monthly transactions generated by the 6924 loans. This is indicated by the mean age of 11.27 (month). The two risk event indicator variables (DF-Event and PP-Event) show the average monthly default rate $\mathrm{DR} = 0.03819$ and prepayment rate $\mathrm{PR} = 0.05066$, corresponding to high 1-year cumulative default rate (CDR of 37.3%) and cumulative prepayment rate (CPR of 46.4%) as $\mathrm{CDR} = 1 - (1 - \mathrm{DR})^{12}$ and $\mathrm{CPR} = 1 - (1 - \mathrm{PR})^{12}$. Note that the summary statistics for those three risk factors at origination (aci, oltv, and LoanSize) with the transaction data are different from those with the origination data shown in Table 4.1 due to different transaction ages for different loans, while the summary statistics for the two macroeconomic variables measured at origination (OHPI and OUER) do not change due to not relating to individual loans.

Our second loan-level transaction data set was simulated based on dynamic risk event probabilities. For each loan, the monthly default and prepayment rates are based on loan origination factors and dynamic macroeconomic variables, which are changing over the transaction path. This is more realistic as we know that the loan performance changes with the economic environments. Similarly, over the transaction path, the loan could be terminated by a risk event (default or prepayment) or continue to be active; binary variables are used to record each risk event; and all loans have the 30-year term indicating a maximum number of 360 monthly transactions.

Table 4.3 shows a summary of the synthetic transaction data of the sample subprime loans with some origination information.

For the dynamic transaction data, there are 83,095 monthly transactions generated by the 6924 loans. This is indicated by the higher mean age of 12.42 (month) compared to the static transaction data. The two risk event indicator variables (DF-Event and PP-Event) show the average monthly default rate $\mathrm{DR} = 0.04032$ and prepayment rate $\mathrm{PR} = 0.04301$, corresponding to high 1-year cumulative default rate (CDR of 39.0%) and cumulative prepayment rate (CPR of 41.0%). Note that the summary statistics for those two risk factors at origination (aci, oltv) are similar to

Table 4.2 Summary information for the static transaction data

ID	ActDate	Age	aci	oltv	DF_Event	PP_Event	LoanSize	OHPI	OUER
Length: 77,924	Length: 77,924	Min.: 1.00	Min.: 300.0	Min.: 1.528	Min.: 0.00000	Min.: 0.00000	Min.: 4568	Min.: 97.35	Min.: 1.800
Class: character	Class: character	1st Qu.: 4.00	1st Qu.: 421.0	1st Qu.: 73.091	1st Qu.: 0.00000	1st Qu.: 0.00000	1st Qu.: 168,750	1st Qu.: 232.84	1st Qu.: 4.100
Mode: character	Mode: character	Median: 8.00	Median: 554.0	Median: 79.900	Median: 0.00000	Median: 0.00000	Median: 270,000	Median: 281.23	Median: 4.800
NA	NA	Mean: 11.27	Mean: 560.4	Mean: 80.441	Mean: 0.03819	Mean: 0.05066	Mean: 292,244	Mean: 267.41	Mean: 5.467
NA	NA	3rd Qu.: 15.00	3rd Qu.: 694.0	3rd Qu.: 80.000	3rd Qu.: 0.00000	3rd Qu.: 0.00000	3rd Qu.: 384,000	3rd Qu.: 310.00	3rd Qu.: 5.900
NA	NA	Max.: 94.00	Max.: 850.0	Max.: 363.459	Max.: 1.00000	Max.: 1.00000	Max.: 2,812,500	Max.: 635.08	Max.: 27.742

Table 4.3 Summary information for the dynamic transaction data

ID	ActDate	Age	aci	oltv	DHPI	DUER	DF_Event	PP_Event
Length: 83,095	Length: 83,095	Min.: 1.00	Min.: 300.0	Min.: 1.528	Min.: –212.3015	Min.: –13.7672	Min.: 0.00000	Min.: 0.00000
Class: character	Class: character	1st Qu.: 4.00	1st Qu.: 422.0	1st Qu.: 72.738	1st Qu.: –42.2391	1st Qu.: –2.9199	1st Qu.: 0.00000	1st Qu.: 0.00000
Mode: character	Mode: character	Median: 9.00	Median: 556.0	Median: 79.900	Median: –21.7133	Median: –1.7656	Median: 0.00000	Median: 0.00000
NA	NA	Mean: 12.42	Mean: 559.1	Mean: 80.546	Mean: –30.1295	Mean: –2.1018	Mean: 0.04032	Mean: 0.04301
NA	NA	3rd Qu.: 17.00	3rd Qu.: 690.0	3rd Qu.: 80.000	3rd Qu.: –9.5693	3rd Qu.: –0.8718	3rd Qu.: 0.00000	3rd Qu.: 0.00000
NA	NA	Max.: 114.00	Max.: 850.0	Max.: 363.459	Max.: –0.9735	Max.: –0.0900	Max.: 1.00000	Max.: 1.00000

that of the static transaction data. The two new macroeconomic variables are cumulative home price index change since origination (DHPI) and cumulative unemployment rate change since origination (DUER), which are changing over the transaction path.

Based on these two transaction data sets, we will show how to build generalized linear models for default and prepayment risks and carry out a simpler version of the AEVS model selection procedure illustrated in Sect. 3.3.5 to show how AEVS selects the best models.

4.2.2 MEV Projection

MEV forecasting is critical for CECL loss forecasting. In general, there are two approaches for MEV forecasting as described in Sect. 2.5.2 – the mean reversion approach and the full distribution approach – both supported by the regulatory guidance. With limited number of MEVs for our subprime mortgage portfolio, we select the input mean reversion method described in Sect. 2.5.2.

With the input mean reversion approach, we predict the MEVs using time series models for a relatively short period, in which we consider the forecasting is reasonable and supportable. After that short period, we reverse back to the historical trend. For example, the home price index (HPI) is a critical MEV used in our subprime mortgage CECL models. Figure 4.2 shows the Case-Shiller home price index since 1987 (using 100 at January 2000). Shaded bars indicate the US recessions.

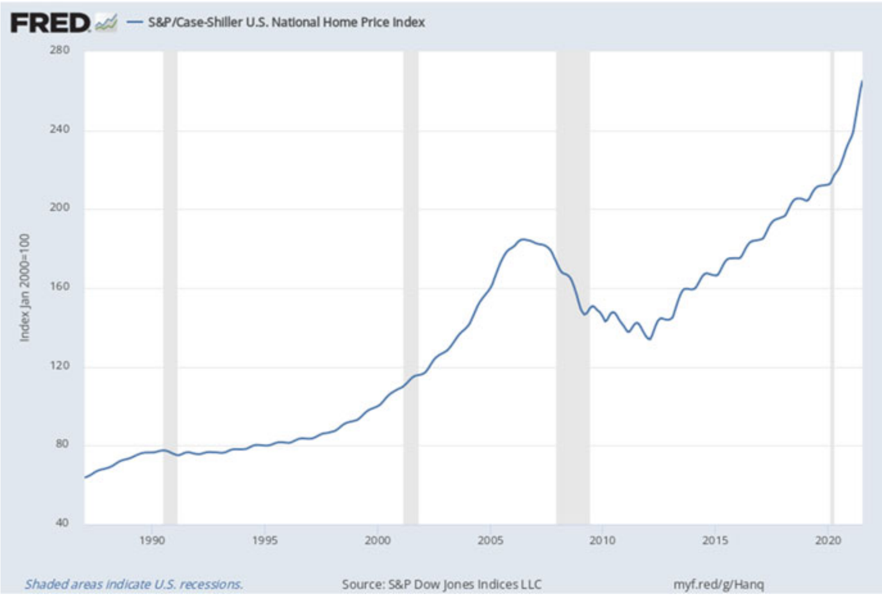


Fig. 4.2 Case-Shiller US national home price index

HPI downturn may not happen with every recession, like the 2000 recession due to dot-com bubble. A mild recession like the 1990 one led to about 3% 1-year HPI decline, while the 2008 great recession due to the housing bubble saw a doubled decline rate, about 6% annual HPI decline over the next 5 years starting from 2006. So, for an average housing downturn, we adopt an autoregression leading to 6% annual HPI decline over the next 3 years. After the 3 years' decline, we reverse back to the long-term housing trend – a 3% annual HPI appreciation for the rest of the life of the loan. Similarly, UER is assumed a 6% annual decline for the next 3 years before reversing back to the long-term average of 5%.

For our static data, MEVs do not impact the loan transactions (default and prepayment). One can consider such transactions as long-term average. For the dynamic transaction data, both default and prepayment are impacted by the projections of MEVs, and the model results are sensitive to the strategies applied to the projections of MEVs.

For both static and dynamic data, we have one suite of severity models for loss forecasting. The severity models for collateral-based credit products are likely including MEVs as drivers since the disposing of the collaterals highly depends on the post-default economic environments. With our subprime mortgage portfolio, we adopt the multi-resolution severity model, and both the projection of a resolution type and the severity for a specific resolution depend on the home price index. We will show how the HPI projection is incorporated into the severity projection, which in turn is integrated into the final loss forecasting in CECL.

For committed credit products, MEVs don't impact exposure at default (EAD) directly, but through the transactions. As described in Sect. 3.1.5.1, the unpaid balance (UPB) during transactions depends on the survival rate on both credit events and amortization. This highlights the complexity of the impacts from MEVs for the final CECL loss forecasting.

For uncommitted credit products, MEVs could have a direct impact on the final exposure at default, since the final draw of credit likely depends on the economic environments near default.

4.3 CECL Models

In this section, we build the CECL component models based on the two transaction data sets described in Sect. 4.2.1. For the default and prepayment risks, we adopt the competing risk modeling framework illustrated in Sect. 3.1.2 due to the evidence of strategic default especially seen in the subprime mortgage portfolios during the finance crisis. With the competing risk framework, we prefer using the binary approximation to estimate the default and prepayment probabilities due to its computing efficiency.

For loss forecasting models, e.g., CECL and CCAR models, model selection is important since good model performance in forecasting accuracy is commonly required. Under the competing risk framework, model selection for multinomial models is complex and there is a lack of a good procedure. So, our solution is carrying out model selection under the target risk framework as illustrated in Sect. 3.1.1 for individual risk event independently. Such selected models for both default and prepayment risks are recalibrated using the binary approximation to obtain the unbiased parameter estimates under the competing risk framework for loss forecasting. Although the recalibrated parameter estimates using the binary approximation are not much different from those under the target risk framework due to the sparsity of the risk events in the transaction data set, for the purpose of lost forecasting accuracy and theoretic soundness, the parameter recalibration is a necessary step.

We also build LGD models based on a reference default data set from this subprime portfolio. EAD is calculated according to the formulas given in Sect. 3.1.5.1 for committed credit products.

These component models are integrated with the CECL MEV projections described in Sect. 4.2.2 in the following section to produce the final CECL loss forecasts for our sample subprime mortgage portfolio.

4.3.1 PD and PP Models

As a reminder, under the target risk framework, a binary variable marks the target risk event (default or prepayment) as 1 if it occurs and as 0 else. In our transaction data sets, two binary variables PD-Event and PP-Event are used for marking default or prepayment, respectively. They act as the dependent variable in the binary logit model we will build. We focus on the default model, as the prepayment modeling fits in the same modeling process.

First, as the core of the modeling process, we present a simpler version of AEVS for generalized linear models:

GLM Modeling Process with AEVS

1. *Data Input Preparation*
   (a) Target risk event variable $Y = \begin{cases} 1, & \text{Event} \\ 0, & \text{No} - \text{Event.} \end{cases}$
   (b) Macro data X_{macro}, the list of macros chosen by line of business (LOB)
   Calculate $T(X_{\text{macro}})$, the list of chosen meaningful transformations for each macro.
   Calculate $\text{Lag}(X_{\text{macro}})$, $\text{Lag}(T(X_{\text{macro}}))$ meaningful over certain periods (e.g., quarters).
   Determine $\text{sign}(X_{\text{macro}})$, the meaningful relationship (positive or negative) between X_{macro} and Y.
   (c) N, the total number of X variables including X_{macro}, $T(X_{\text{macro}})$, $\text{Lag}(X_{\text{macro}})$, and $\text{Lag}(T(X_{\text{macro}})$ called four variations of X_{macro}; select top $\text{D} = 10$ candidate variables from these N variables using each of the following methods:

- Concordance Index (i.e., area under ROC) top D (10)
- Standard stepwise (based on -2loglikelihood) top D (10)
  Combine variables selected from Concordance Index and stepwise.
  Add the variable chosen by business intuitions.
  Determine TopN, the combined list for selected variables derived above.

2. *GLM Model Selection*
   (a) Generate M_j GLM models from possible subset of TopN variables using exhaustive search, $j = 1, \ldots, n \sim 2^{\text{TopN}}$.
   (b) Reject M_j if any regression coefficient doesn't satisfy sign(X_{macro}).
   (c) Reject M_j if more than two variations of one X_{macro} are included in the model.
   (d) Reject M_j if any GLM coefficient p-value $>5\%$.
   This determines the selected candidate GLM models.
3. *Model Ranking*
   For the survived candidate models after Step 2:
   (a) Rank by SIC (the lower the SIC, the higher the rank M_j has), optionally reject low-rank models.
   (b) Rank by AIC (the lower the AIC, the higher the rank M_j has), optionally reject low-rank models.
   (c) Rank by v-fold cross-validation (the lower the CV_v, the higher the rank M_j has) on survived models.
   (d) Calculate the average rank of the three performance rankings above, and select the top 5 models based on the average rank. Note that we use v-fold cross-validation CV_v as the goodness-of-fit measure instead of the leave-one-out (LOO or CV_{nn}) due to computational efficiency for GLM models.

From the five models, choose the final model based on LOB review and model assessment.

Example 1: PD/PP Modeling with Static Transaction Data
In the following, we execute the above GLM modeling process with AEVS on our static transaction data.

Data Preparation
Transaction data are extracted from the synthetic portfolio transaction data, which include the loan ID (ID), loan transaction date (ActDate), loan age at each transaction date (age), as well as the three loan characteristic measures at origination – acquisition index (aci), loan to value ratio at origination (oltv), and loan amount at origination (LoanSize). The performance of the loan in the transaction data is recorded by the default risk event variable DF-Event, which is a binary variable taking value 1 when there is a default event at the transaction date and 0 otherwise.

For the macroeconomic variables, mimicking the business inputs, we include two static macroeconomic variables (MEVs) – the house price index and unemployment rate at origination, OHPI and OUER. For simplicity, we won't consider the other three variants (transformation, lag, and their combination) of these MEVs as descried in the modeling process. As a consequence of the small number of risk factors, we skip the step to form the candidate risk factor pool by taking all these five factors –

aci, oltv, LoanSize, OHPI, and OUER – as the candidate risk factors. These MEVs are merged with the transaction data obtained earlier.

So, the model data for our static transaction example is simple: the binary default event variable DF-Event as the dependent variable and the five static measures – aci, oltv, LoanSize, OHPI, and OUER – as the risk factors. Due to the simple relationship between risk events and risk factors, no special sampling approach is required for the forecasting purpose, and we take the full loan-level transaction observations as our panel data. In the following, we will show how to carry out variable selection and modeling ranking with the AEVS procedure.

Model Selection

In credit risk modeling, model selection is no doubt the most expensive step, either manually or computationally. While automation should be adopted as much as possible in model selection, some decisions are still needed from business judgment. Our model selection with AEVS tries to achieve a balance between these two aspects.

For our simple example with the static transaction data, we only have five risk factors in our candidate variable pool. So, there are totally $2^5 = 32$ combinations of the risk factors (including the intercept-only model). It is still a lengthy job if we go over these models one by one to manually apply the rejection rules described in our modeling process, not even say a little larger candidate pool, for example, with 20 candidate variables, which will generate $2^{20} = 1{,}048{,}576$ combinations. An automation with an algorithm is the only solution. In the following, we present algorithms in pseudocode to implement the model selection step with AEVS.

Algorithm 1 (Block AEVS)

1. *Form the TopN binary vector V_Index, in which each component corresponding to a candidate variable and generate the* $n = 2^{TopN}$ *possible V_Index vectors. Mark feasibility according to variable combination rules,* e.g.*, no more than two variants of the same candidate variable are included.*
2. *Split the* $n = 2^{TopN}$ *possible V_Index vectors into m (*e.g.*,* $m = 16$ *for TopN* ≥ 20*) blocks.*
3. *Go over m blocks and fit only the feasible GLM based on selected variables by each of the binary vector V_Index in a block. Variable names are picked up from the TopN variable names vector V_Names.*
4. *Apply the model rejection rules for each fitted model (*e.g.*, the sign matching rule for each selected variable and its p-value rule for the fitted GLM model).*

The Block-AEVS splits the total variable combinations $n = 2^{\mathrm{TopN}}$ into m blocks and is suitable for parallel processing, since each block can be processed independently before merging the survival models. When the number of variables in the candidate pool is relatively large (e.g., 20 or more) and multiprocessing is available, the Block-AEVS is recommended.

Correspondingly, there is also a recursive version of AEVS based on model size. The pseudocode to implement the recursive AEVS is summarized in the following algorithm:

Algorithm 2 (Recursive AEVS)

1. *Loop over the number of candidate variables added into the model from 1 to TopN:*
   for (Msize = 1 to TopN)
      {Select Msize variables from the candidate variable pool;
      Check model feasibility (e.g.*, when Msize >2, no more than two variants of the same candidate variable are included.}*
2. *Fit GLM model for each of the feasible models of size = Msize*
3. *Apply the model rejection rules for each fitted model* (e.g.*, the sign matching rule for each selected variable and its p-value rule for the fitted GLM model).*

The recursive algorithm follows the steps of model expansion and does not need to maintain the variable index matrix. However, due to its sequential processing of model fitting, parallel computing is not feasible and may only be good for small-sized candidate pools. For our examples, we choose the recursive algorithm. Also, with our synthetic data, we relax the *p*-value rejection rule.

Model Ranking
Survived models are ranked by the three goodness-of-fit measures – SIC, AIC, and CV_v. For the *v*-fold cross-validation measure, due to computational efficiency for our relatively large data set, we choose $v = 11$. The negative 2loglikelihood (also called deviance) usually tends to select larger model and thus is not used as a goodness-of-fit measure in the model ranking. Note that we have a specific order for applying the three goodness-of-fit measures due to the optional further model rejection based on these measures. If these optional model rejections were not used, then the order of applying these measures on model ranking is not relevant.

The optional further model rejections based on these goodness-of-fit measures have several advantages than directly applying these measures to all survived models from the model selection step. Firstly, it is computationally more feasible for AEVS with large data since some of these goodness-of-fit measures (e.g., measures from the cross-validation family) are computationally intensive and computing such measures for a large number of survived models from the model selection step may not be feasible in some cases. Secondly, the order defines a preference on different goodness-of-fit measures in the model ranking step since the preferred measure can be used first to further reject low-performance models based on this measure and less preferred measures only have the chance to do so on the survived models after the preferred measure. This is especially useful when we have a strong preference to certain models, e.g., SIC tends to give higher rank to parsimonious models having strong predictive power and simpler structure. Lastly, by adjusting the rejection threshold applied to different measures, one can see how consistently difference measures behave in the model ranking step. In our static transaction example, we can see SIC and AIC are largely consistent from the relatively stable ranks these two measures present when adjusting the SIC rejection threshold; however, the ranks given by the deviance measures are strongly inconsistent.

For our static transaction data, we simply select the top 5 models by SIC and obtain ranks of these models by AIC and CV_v with $v = 11$. These top 5 models with the three goodness-of-fit measures, as well as their ranks based on these measures and average ranks, are shown in Table 4.4.

Though ranks by AIC and CV_v are different from that of SIC, the average ranks of all three measures are consistent with that of SIC. So, the top model selected by the average rank is the true model that was used to simulate the synthetic data. Note that the average is used for tied ranks with CV_v. Simple R code and the top model fit results are shown in the following figures (Fig. 4.3).

The "boot" package in R was used to calculate the cross-validation measures (cv. glm function call). "SIC _ rank" holds the ranks by SIC for the survived models with their names defined by the model formula in the "modelnames" vector. The code loops over all the top 5 selected models by SIC and obtains the model fitting results including deviance, AIC, and CV_v using the transaction data set "PDate." AIC and CV_v are ranked, and full average ranks for the three goodness-of-fit measures are calculated and presented in Table 4.4. For a simple display, only the top model fitting results are presented in Fig. 4.4.

As a separate target event, the prepayment event is independently simulated from the underlying prepayment model and recorded in the same transaction data. Model selection is carried out similarly as that for the default model. Figure 4.5 displays the final top model selected.

Example 2: PD Modeling with Dynamic Transaction Data

In the following, we execute the GLM modeling process with AEVS on our dynamic transaction data.

Data Preparation

Similar to the static transaction data, dynamic transaction data are also extracted from a synthetic portfolio transaction data with additional dynamic macroeconomic variables. Similarly, no special sampling is needed, and we take the full loan-level transaction observations as our panel data.

The two new macroeconomic variables are cumulative home price index change since origination (DHPI) and cumulative unemployment rate change since origination (DUER), which are changing over the transaction path. In addition to the loan origination characteristics (aci and oltv), the synthetic risk events are simulated based on these dynamic macroeconomic variables.

Model Selection

Similar to the static transaction data, for simplicity, we skip the step to generate other three variants of the initial selected variables. So, our pool of candidate variables includes these seven variables – aci, oltv, DHPI, DUER, OHPI, OUER, and

Table 4.4 Top 5 default models selected (static transaction data)

Model	SIC	AIC	CV11	Rank SIC	Rank AIC	Rank CV	AveRank
DF_Event ~ aci + oltv	25,038.2	25,010.41	0.038191	1	2	2	1.67
DF_Event ~ aci + oltv + OUER	25,046.86	25,009.81	0.038204	2	1	4	2.33
DF_Event ~ aci + oltv + LoanSize	25,048.26	25,011.2	0.038191	3	4	2	3.00
DF_Event ~ aci + oltv + OHPI	25,048.83	25,011.77	0.038191	4	5	2	3.67
DF_Event ~ aci + oltv + LoanSize + OUER	25,056.87	25,010.55	0.038204	5	3	5	4.33

```
>
> cv11<-NULL
> cost <- function(r, pi = 0) mean(abs(r-pi) > 0.5)
> i<-0
> for(mid in top5_SIC){
+   i<-i+1
+   cat("Top", i, "Model:", mid, "\n")
+
+   frm_df <- as.formula(mid)
+   glmmodel_df<-glm(formula=frm_df, family=binomial(link='logit'), data=PData)
+
+   cv_delta<-cv.glm(PData, glmmodel_df, cost, K = 11)$delta
+   cv11<-rbind(cv11, cv_delta)
+
+   print(summary(glmmodel_df))
+   print("11 Fold Cross Validation")
+   print(cv_delta)
+ }
```

Fig. 4.3 Partial R-code

```
Top 1 Model: DF_Event ~ aci + oltv

Call:
glm(formula = frm_df, family = binomial(link = "logit"), data = PData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0958  -0.2938  -0.2705  -0.2483   2.8356

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.4936916  0.0902483 -38.712   <2e-16 ***
aci         -0.0011492  0.0001205  -9.537   <2e-16 ***
oltv         0.0108875  0.0007261  14.994   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 25271  on 77923  degrees of freedom
Residual deviance: 25004  on 77921  degrees of freedom
AIC: 25010

Number of Fisher Scoring iterations: 6

[1] "11 Fold Cross Validation"
[1] 0.03819106 0.03819106
```

Fig. 4.4 Top default model selected (static transaction data)

```
Top 1 Model: PP_Event ~ aci + oltv

Call:
glm(formula = frm_pp, family = binomial(link = "logit"), data = PData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8117  -0.3565  -0.3081  -0.2643   3.3376

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.1984261  0.1237922  -17.76   <2e-16 ***
aci          0.0019344  0.0001049   18.43   <2e-16 ***
oltv        -0.0237629  0.0014192  -16.74   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 31243  on 77923  degrees of freedom
Residual deviance: 30577  on 77921  degrees of freedom
AIC: 30583

Number of Fisher Scoring iterations: 6

[1] "11 Fold Cross Validation"
[1] 0.05066475 0.05066475
```

Fig. 4.5 Top prepayment model selected (static transaction data)

LoanSize. There are $2^7 = 128$ variable combinations, which we consider relatively small and still choose the recursive AEVS algorithm. Again, due to our synthetic data, we relax the p-value rejection rule.

Model Ranking

Survived models are ranked by the three goodness-of-fit measures – SIC, AIC, and CV_v. Similar to the static transaction data, we choose $v = 11$ for the v-fold cross-validation measure due to computational efficiency. For the three goodness-of-fit measures, we prefer SIC and use it as the first ranking criteria. Top 5 models based on SIC are selected, and their ranks based on AIC and CV_v are used to compute the full average ranks.

The goodness-of-fit measures and their ranks for the top 5 selected models based on SIC are presented in Table 4.5. Note that the top 5 default models have the same CV_v measure (ties for all five models), which means these models have the same classification error defined by the cost function (see Fig. 4.3 R-code) for the binary outcome. Figure 4.6 presents the model fitting results for the top model, which we used to simulate the synthetic dynamic transaction data. This example demonstrates

Table 4.5 Top 5 default models selected (dynamic transaction data)

Model	SIC	AIC	CV11	Rank SIC	Rank AIC	Rank CV	AveRank
DF_Event ~ aci + oltv + DHPI + DUER	27,837.51	27,790.88	0.040315	1	1	3	1.67
DF_Event ~ aci + oltv + DHPI + DUER + OUER	27,847.38	27,791.42	0.040315	2	2	3	2.33
DF_Event ~ aci + oltv + DHPI + DUER + OHPI	27,847.55	27,791.59	0.040315	3	3	3	3.00
DF_Event ~ aci + oltv + DHPI + DUER + LoanSize	27,848.72	27,792.75	0.040315	4	5	3	4.00
DF_Event ~ aci + oltv + DHPI + DUER + OHPI + OUER	27,857.17	27,791.88	0.040315	5	4	3	4.00

```
Top 1 Model: DF_Event ~ aci + oltv + DHPI + DUER

Call:
glm(formula = frm_df, family = binomial(link = "logit"), data = PData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8719  -0.3017  -0.2762  -0.2525   2.9592

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.5452094  0.0914864 -38.751  < 2e-16 ***
aci         -0.0010324  0.0001144  -9.027  < 2e-16 ***
oltv         0.0100545  0.0008442  11.910  < 2e-16 ***
DHPI        -0.0098546  0.0008217 -11.992  < 2e-16 ***
DUER         0.0917554  0.0157832   5.813 6.12e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 28077  on 83094  degrees of freedom
Residual deviance: 27781  on 83090  degrees of freedom
AIC: 27791

Number of Fisher Scoring iterations: 6

[1] "11 Fold Cross Validation"
[1] 0.0403153 0.0403153
```

Fig. 4.6 Top default model selected (dynamic transaction data)

that AEVS performs well on dynamic transaction data as on static transaction data. Using the same model selection process, the top prepayment model selected is also the one we used in simulation, and the model fitting results are shown in Fig. 4.7.

4.3.2 LGD Models

For CECL, LGD is more often called loss severity, which counts total net accounting loss including costs as a percentage of the exposure at default. The actual LGD for default loans in CECL can be calculated according to the formula (3.58) and the components described in Sect. 3.1.4.1.

For CECL modeling data processing, the actual LGD calculation is a major step and requires high-level automation and maintenance. So, very often, financial institutions build up database for specific portfolio loss information, which include some account characteristics at origination, all information about the account default

```
Top 1 Model: PP_Event ~ aci + oltv + DHPI + DUER

Call:
glm(formula = frm_pp, family = binomial(link = "logit"), data = PData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8873  -0.3323  -0.2785  -0.2286   3.1067

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.478814   0.117698  -21.06   <2e-16 ***
aci          0.002076   0.000111   18.71   <2e-16 ***
oltv        -0.021371   0.001297  -16.48   <2e-16 ***
DHPI         0.021900   0.001228   17.83   <2e-16 ***
DUER        -0.217814   0.017293  -12.60   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 29482  on 83094  degrees of freedom
Residual deviance: 28563  on 83090  degrees of freedom
AIC: 28573

Number of Fisher Scoring iterations: 6

[1] "11 Fold Cross Validation"
[1] 0.04301101 0.04301101
```

Fig. 4.7 Top prepayment model selected (dynamic transaction data)

(default date, exposure – EAD, default reason, etc.), and loss and recovery transactions (accrued interests, costs and expenses, write-downs, recoveries, collateral disposals, collections, etc.). As defined in Sect. 3.1.4.1, the accounting loss is the summation of all these losses net of recoveries and collections. All these components need to be recorded correctly and updated promptly in the loss database. Ideally, the actual LGD is automatically calculated and updated in this loss database with periodic validation and reconciliation. Due to the complexity of actual LGD calculation and traditional low quality of LGD reference default data, a strong data review and validation process should be adopted. Abnormal data without proper explanation should be identified as outliers and treated appropriately, and they should not be the motivation to develop complex LGD models.

For LGD modeling, as described in Sect. 3.1.4, there are three types of LGD models popularly used in credit risk modeling – the micro-structure dynamic models, regression models, and multiple resolutions models. There are different preferences with each type of models for different credit products. For mortgage and other collateral credit products, multiple resolutions models are preferred when post-default transactions are not well defined and maintained. For our subprime mortgage portfolio, we will show how to build multiple resolutions models.

Data Preparation

For our subprime mortgage loan portfolio, we have the full historical loss information and are able to build up the reference default data. Table 4.6 presents summary statistics for some of the attributes of this sample reference default data.

The reference default data include the mortgage loan ID, origination date (OrigDate) and default date (DefDate), and some loan characteristics – loan to value at origination (oltv), occupation types (OccType) and property types (not shown), and most importantly the loan resolution types (Reso_Type) and actual LGD in accounting (LGD Acc).

Macroeconomic variables are created separately as described in Sect. 4.2 and then merged with the reference default data. In Table 4.6, we have the cumulative home price index (chpi), which is defined as the ratio of the current hpi and the hpi at origination at MSA (metropolitan statistical area) level. The chpi in the reference default data is calculated at the default time for each loan. For our subprime portfolio, the chpi turned out to be the most important macroeconomic driver in the LGD models (and the only macroeconomic variable used in our final model).

Model Selection

For LGD model selection, there are two layers. The first layer lies on the choice of the three types of popular LGD models used in credit modeling as we mentioned early. For collateral-based credit products, especially for mortgages, for which the loss severity is heavily impacted by the collateral disposal processes and loss mitigation strategies, the multiple resolutions models are preferred. The second layer of model selection is within the multiple resolutions (MR) models. As we described in Sect. 3.1.4, there are both sequential multiple resolutions (SMR) and competing multiple resolutions (CMR) approaches, which lead to different LGD modeling processes.

For our subprime mortgage portfolio, the mortgage servicer applied various post-default loss mitigation strategies. These strategies lead to different resolution approaches for default properties. In summary, these resolution approaches can be classified into three categories – short sale (SS), third-party sale (TPS), and real estate owned (REO). Short sale is a loss mitigation strategy, which allows the property owner to sell the property with some discount to partially cover the mortgage loan exposure at default, since, in most cases, the property owner still lives in the property and the property is maintained in basic conditions. Not often, but there are cases that the short sale proceeds can cover the exposure at default in full, which we call payoff. So, this category has the best performance in loss recovery. Without the cost of maintaining the property, the lender can sell these default loans to third parties, who may have specific expertise in handling default loans. The sale of default loans usually involves large discount, and so the loss severity for third-party sale is higher than short sale in average. The last category,

Table 4.6 Summary information for a sample reference default data

ID	OrigDate	DefDate	oltv	chpi	OccType	Reso_Type	LGD_Acc
Length: 4847	Length: 4847	Length: 4847	Min.: 1.528	Min.: 0.3545	Length: 4847	Length: 4847	Min.: 0.0000
Class: character	Class: character	Class: character	1st Qu.: 71.339	1st Qu.: 0.6059	Class: character	Class: character	1st Qu.: 0.3847
Mode: character	Mode: character	Mode: character	Median: 79.300	Median: 0.7217	Mode: character	Mode: character	Median: 0.5892
NA	NA	NA	Mean: 80.155	Mean: 0.7531	NA	NA	Mean: 0.5991
NA	NA	NA	3rd Qu.: 80.000	3rd Qu.: 0.8675	NA	NA	3rd Qu.: 0.7849
NA	NA	NA	Max.: 328.828	Max.: 2.8278	NA	NA	Max.: 1.5000

real estate owned, means that the lender takes the property back and takes care of the maintenance costs and sales. This category also includes loans due to owner bankruptcy. Due to various costs and high recovery deficiency, this category usually has the highest loss severity in average. Since both the agreement with the property owner for a short sale and the selection of loans to sell to third party are random for our subprime portfolio, we select the competing multiple resolutions models instead of the sequential multiple resolutions models. A multinomial logistic regression is used to model the probabilities of a default loan falling into one of these three categories.

Once a resolution type is selected, the next step is to decide the loss severity within that resolution type. Since the loss mitigation strategies within a resolution type are considered homogeneous, the simplest way for the loss severity estimate within a resolution type is using the historical loss severity average observed within that type. While such simple method was popularly accepted in the ACL practices with the incurred loss approach, the forward-looking characteristics for lifetime loss estimate in CECL require more dynamics in future loss severity. So, a regression method within a resolution type based on drivers including macroeconomic variables is preferred for CECL LGD modeling. For our subprime portfolio, we fit simple linear regression models for loss severity within each resolution type with both loan characteristics at origination and macroeconomic variables observed at default time. We prefer parsimonious loss severity model within each resolution type for stability without sacrificing too much of accuracy.

Model Results and Performance

For the multiple resolutions model, we use multinomial logistic regression with the same risk factors for all resolutions types. For the variable selection with this multinomial logistic regression, we adopt the backward stepwise selection procedure – we start with the full model having all variables in the candidate pool, then removing variables with less contribution to the F-statistics. Since we have a small pool of candidate variables, the backward stepwise selection procedure is reasonable in modeling complexity. The final model includes two variables, loan to value at origination (oltv) and cumulative home price index (chpi). We fit multinomial logistic regression using the R package *nnet*. Figure 4.8 presents the final multinomial logistic regression model results.

The model summary in Fig. 4.8 shows that the resolution type REO is taken as the reference level for the multinomial logistic regression. The coefficients for both oltv and chpi are representing the impacts of these drivers to the probabilities of these resolution types. For example, the coefficient 10.533602 for chpi with SS means that a 10% increase in HPI will lead to 2.87 ($= e^{10.533602 * 0.1}$; see (3.51)) times chance of the resolution being SS over REO. Similarly, the coefficient -0.08823604 for oltv with SS means that a 10-point oltv increase will lead to 0.41 times chance of the resolution being SS over REO. These are all reasonable interpretation with our subprime portfolio. The same signs of the coefficients for oltv and chpi with TPS as SS indicate that these drivers have the same directional impact to the chance of the resolution being TPS over REO, but with less magnitude compared to that of SS as

```
> library(nnet)
> frm_df <- as.formula(paste("Reso_Type ~", "oltv", "+", "chpi"))
> multimodel<-multinom(formula=frm_df, data=LGD_Data)
# weights:  12 (6 variable)
initial  value 5324.973763
iter  10 value 4189.945889
final  value 4185.310162
converged
>
> model_summary <- summary(multimodel)
>
> model_summary
Call:
multinom(formula = frm_df, data = LGD_Data_Show)

Coefficients:
    (Intercept)        oltv      chpi
SS    -2.222558 -0.08823604 10.533602
TPS   -1.383740 -0.01790398  3.960228

Std. Errors:
    (Intercept)        oltv      chpi
SS    0.3533837 0.004701374 0.3408198
TPS   0.1827943 0.001880432 0.2386411

Residual Deviance: 8370.62
AIC: 8382.62
>
> z <- model_summary$coefficients/model_summary$standard.errors
> z
    (Intercept)       oltv     chpi
SS    -6.289362 -18.768140 30.90666
TPS   -7.569933  -9.521203 16.59491
>
> p <- (1 - pnorm(abs(z), 0, 1)) * 2
> p
     (Intercept) oltv chpi
SS   3.187743e-10    0    0
TPS  3.730349e-14    0    0
```

Fig. 4.8 Final multinomial model for resolution types

these coefficients are smaller in magnitude. Such impacts suggest that TPS is a middle type of resolution between SS and REO. The multinomial function in *nnet* does not provide *p*-values for the independent variables, and the *z*-values and *p*-values need to be computed separately. Both these values suggest highly significance of these variables.

```
> frm_df0 <- as.formula(paste("Reso_Type ~", "1"))
> multimodel0<-multinom(formula=frm_df0, data=LGD_Data)
# weights:  6 (2 variable)
initial  value 5324.973763
final  value 5072.148860
converged
> summary(multimodel0)
Call:
multinom(formula = frm_df0, data = LGD_Data_Show)

Coefficients:
    (Intercept)
SS  -0.78402137
TPS -0.03619045

Std. Errors:
    (Intercept)
SS   0.03991868
TPS  0.03189648

Residual Deviance: 10144.3
AIC: 10148.

> anova(multimodel0, multimodel)
Likelihood ratio tests of Multinomial Models

Response: Reso_Type
        Model Resid. df Resid. Dev   Test    Df LR stat. Pr(Chi)
1           1      9692   10144.30
2 oltv + chpi      9688    8370.62 1 vs 2     4 1773.677       0
```

Fig. 4.9 ANOVA for the final multinomial model

To further see the performance of the final model, we also fit the intercept-only (or null) model and carry out the chi-square goodness-of-fit test for the final model as shown in Figs. 4.8 and 4.9.

Both variables selected in the final model, oltv and chpi, have high importance. Models without any one of them are showing lack of fit. These two drivers represent two important aspects of a default loan in loss severity. First, the loan to value ratio at origination represents how much deposit the owner put in the mortgage at origination. The lower of this ratio means the closer bond between the owner and the mortgage and the incentive to recover the loss and maintain a good condition of the property. On the other aspect, chpi as a macroeconomic driver represents the cumulative home price appreciation (or depreciation) over the time, which has a close relation with the current and future housing market and gives the owner an indication of current and future house equity.

```
> multi_pred <- predict(multimodel)
> table(multi_pred, LGD_Data$Reso_Type)

multi_pred  REO   SS  TPS
       REO 1308   84  837
       SS    62  454  182
       TPS  632  376  912

> prop1 <- with(LGD_Data, table(multi_pred, Reso_Type)) %>% prop.table(margin=1)
> prop1
          Reso_Type
multi_pred        REO         SS        TPS
       REO 0.58681023 0.03768506 0.37550471
       SS  0.08882521 0.65042980 0.26074499
       TPS 0.32916667 0.19583333 0.47500000
```

Fig. 4.10 Confusion matrix for the final multinomial model

One caution for the multinomial model is using the prediction based on the fitted probabilities for different outcomes. Such prediction is usually based on the maximum (or majority) rule, which predicts the outcome based on the maximum fitted probabilities among all possible outcomes and acts as a classifier. Performance measures, such as the confusion matrix based on such predicted outcomes shown in Fig. 4.10, may heavily depend on the data and not necessarily represent the goodness-of-fit of the model, which focuses on the fitted probabilities of the different outcomes.

Figure 4.10 shows the confusing matrix and percentage of predicted outcomes among the observed outcomes. The multinomial logistic model does not perform well as a classifier for our LGD reference default data. For multiple outcomes, tree-based models usually perform better than multinomial models. We will continue this discussion in Sect. 4.5.

The severity within each resolution type is fitted with a linear regression model. Figures 4.11, 4.12, and 4.13 show the fitted models. For REO and TPS, both oltv and chpi are strong predictors, while for SS, oltv is not a significant driver anymore. One explanation is that for short sales, LGD is largely a factor of allowed sale discount of the exposure at default based on the current market. So, deposit at origination may be less relevant once the owner agrees to the short sale option. At least this seems the case with our subprime portfolio. With REO and TPS, oltv is either related to the recovery deficiency or negotiated loan sale price. All resolution types are highly depending on the market conditions represented by chpi.

As a well-known fact, severity models by linear regression have poor goodness-of-fit due to the irregular LGD observations.

```
> summary(lm(REO$LGD_Acc~ REO$chpi + REO$oltv ))

Call:
lm(formula = REO$LGD_Acc ~ REO$chpi + REO$oltv)

Residuals:
     Min       1Q   Median       3Q      Max
-0.85149 -0.16764 -0.02548  0.13640  0.99762

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.073346   0.027285  39.338   <2e-16 ***
REO$chpi    -0.720211   0.040623 -17.729   <2e-16 ***
REO$oltv     0.001959   0.000236   8.302   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2518 on 1999 degrees of freedom
Multiple R-squared:  0.1367,    Adjusted R-squared:  0.1358
F-statistic: 158.2 on 2 and 1999 DF,  p-value: < 2.2e-16
```

Fig. 4.11 REO severity model

```
> summary(lm(SS$LGD_Acc~ SS$chpi  ))

Call:
lm(formula = SS$LGD_Acc ~ SS$chpi)

Residuals:
     Min       1Q   Median       3Q      Max
-0.36524 -0.13474 -0.01747  0.09706  0.72128

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.53813    0.02613  20.597  < 2e-16 ***
SS$chpi     -0.22406    0.02703  -8.288 4.09e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.189 on 912 degrees of freedom
Multiple R-squared:  0.07004,   Adjusted R-squared:  0.06902
F-statistic: 68.69 on 1 and 912 DF,  p-value: 4.095e-16
```

Fig. 4.12 SS severity model

```
> summary(lm(TPS$LGD_Acc~ TPS$chpi + TPS$oltv ))

Call:
lm(formula = TPS$LGD_Acc ~ TPS$chpi + TPS$oltv)

Residuals:
     Min       1Q   Median       3Q      Max
-0.54410 -0.15377 -0.02393  0.12664  0.89763

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.875067   0.029216  29.951  < 2e-16 ***
TPS$chpi    -0.610047   0.032968 -18.504  < 2e-16 ***
TPS$oltv     0.001881   0.000314   5.989 2.51e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2217 on 1928 degrees of freedom
Multiple R-squared:  0.1508,    Adjusted R-squared:  0.1499
F-statistic: 171.2 on 2 and 1928 DF,  p-value: < 2.2e-16
```

Fig. 4.13 TPS severity model

4.4 Model Integration and Loss Forecasting

Loss forecasting is a process consisting of several critical components, including portfolio identification, portfolio forecasting data processing, model integration and implementation, and forecasting result analysis and reporting. We will briefly go over these components, especially the model integration, for which proper modeling frameworks are required to meet the concept soundness of the loss forecasting process as a part of the requirement by model validation in the next section.

For CECL, the forecasting implementation also needs to take care of the lifetime terms by providing proper projections for macroeconomic and other dynamic drivers needed. Although most of these executions are carried out during the model implementation process and consolidated in the production system, each component needs to be architected correctly to fit into the forecasting process. In the following, using the subprime mortgage portfolio, we demonstrate a concise flow of forecasting process with an intention of automation.

Table 4.7 Summary information for the static portfolio snapshot

ID	OrigDate	ActDate	Age	aci	oltv	UPB
Length: 1328	Length: 1328	Length: 1328	Min.: 1.000	Min.: 300.0	Min.: 36.11	Min.: 9700
Class: character	Class: character	Class: character	1st Qu.: 2.000	1st Qu.: 422.0	1st Qu.: 73.00	1st Qu.: 183,330
Mode: character	Mode: character	Mode: character	Median: 4.000	Median: 560.0	Median: 79.95	Median: 273,176
NA	NA	NA	Mean: 5.923	Mean: 561.5	Mean: 80.17	Mean: 283,733
NA	NA	NA	3rd Qu.: 8.000	3rd Qu.: 695.0	3rd Qu.: 80.00	3rd Qu.: 367,691
NA	NA	NA	Max.: 45.000	Max.: 849.0	Max.: 260.67	Max.: 873,000

4.4.1 Portfolio Identification and Forecasting Data Processing

An important issue on model governance is the proper use of models. Models could be developed based on data from multiple portfolios or even third-party data. Model use coverage is a top issue to be addressed at the beginning of model development. Once the model coverage for the specified portfolio is certified, the forecasting implementation can start from processing the portfolio data.

For our subprime mortgages, we generated two portfolios. One portfolio has static transaction behavior, while the other has dynamic transaction behavior. So, we will show how CECL forecasting can be done on snapshots of these portfolios with different models we have developed through proper model integration.

Table 4.7 shows summary statistics for the static portfolio at a specific snapshot of January 1, 2006 (defined by the ActDate). There are 1328 loans active at January 1, 2006, for our static portfolio. The current balance of the loan is recorded by the UPB variable.

Table 4.8 shows summary statistics for the dynamic portfolio also at the snapshot of January 1, 2006. There are 1368 loans active at January 1, 2006, for our dynamic portfolio. In addition to the two static drivers (aci and oltv) included in the static portfolio, we have two dynamic drivers (HPI and UER) used in the model to predict risk event probabilities.

4.4.2 Model Implementation and Integration

The CECL loss forecast is the aggregation of predicted periodic (monthly) losses over the remaining life of each loan in a portfolio:

Table 4.8 Summary information for the dynamic portfolio snapshot

ID	OrigDate	ActDate	Age	aci	oltv	HPI	UER	UPB
Length: 1368	Length: 1368	Length: 1368	Min.: 1.000	Min.: 300.0	Min.: 29.96	Min.: 91.39	Min.: 0.5548	Min.: 9700
Class: character	Class: character	Class: character	1st Qu.: 2.000	1st Qu.: 426.8	1st Qu.: 73.05	1st Qu.: 219.47	1st Qu.: 2.9937	1st Qu.: 179,632
Mode: character	Mode: character	Mode: character	Median: 5.000	Median: 556.0	Median: 79.89	Median: 258.46	Median: 3.7725	Median: 266,847
NA	NA	NA	Mean: 6.093	Mean: 561.9	Mean: 80.22	Mean: 249.41	Mean: 4.1471	Mean: 278,652
NA	NA	NA	3rd Qu.: 8.000	3rd Qu.: 697.2	3rd Qu.: 80.00	3rd Qu.: 287.67	3rd Qu.: 4.5612	3rd Qu.: 361,070
NA	NA	NA	Max.: 44.000	Max.: 850.0	Max.: 236.06	Max.: 608.69	Max.: 16.1375	Max.: 800,250

```
Call:
glm(formula = frm_df, family = binomial(link = "logit"), data = PData_df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0583  -0.3005  -0.2784  -0.2571   2.8019

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.4499357  0.0904256 -38.152   <2e-16 ***
aci         -0.0010576  0.0001206  -8.766   <2e-16 ***
oltv         0.0103704  0.0007303  14.201   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 24955  on 73975  degrees of freedom
Residual deviance: 24720  on 73973  degrees of freedom
AIC: 24726

Number of Fisher Scoring iterations: 6
```

Fig. 4.14 Binary approximation default model for static data

$$\mathrm{Loss}_{\mathrm{CECL}}(p) = \sum_{i=1}^{N_p} \sum_{t=1}^{T_i} \mathrm{UPB}_i(t) * \mathrm{PD}_i(t) * \mathrm{LGD}_i(t) \tag{4.1}$$

where p indicates the target portfolio with total N_p loans. $\mathrm{UPB}_i(t)$, $\mathrm{PD}_i(t)$, and $\mathrm{LGD}_i(t)$ are unpaid balance, default probability, and severity of the ith loan at period t of the remaining lifetime of T_i.

Given the conditional periodic default and prepayment probabilities projected from the models, as well as the amortization rate, $\mathrm{UPB}_i(t)$ can be computed from (3.78). It should be pointed out that the models we fit in Sect. 4.3.1 are under the independent risk event framework. To apply (3.78), the default and prepayment probabilities under the competing framework should be projected, for which we use the binary approximation as described in Sect. 3.1.2.2. Figures 4.14 and 4.15 show the binary approximation model results for the static data, and Figs. 4.16 and 4.17 show the binary approximation model results for the dynamic data. Compared to the model results from Figs. 4.4, 4.5, 4.6, and 4.7, these results are very close.

```
Call:
glm(formula = frm_pp, family = binomial(link = "logit"), data = PData_pp)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8123  -0.3631  -0.3145  -0.2706   3.3120

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.167598   0.124071  -17.47   <2e-16 ***
aci          0.001892   0.000105   18.02   <2e-16 ***
oltv        -0.023348   0.001423  -16.40   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 30927  on 74947  degrees of freedom
Residual deviance: 30294  on 74945  degrees of freedom
AIC: 30300

Number of Fisher Scoring iterations: 6
```

Fig. 4.15 Binary approximation prepayment model for static data

```
Call:
glm(formula = frm_df, family = binomial(link = "logit"), data = PData_df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8374  -0.3076  -0.2833  -0.2607   2.9201

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.4959375  0.0916368 -38.150  < 2e-16 ***
aci         -0.0009478  0.0001145  -8.279  < 2e-16 ***
oltv         0.0094932  0.0008490  11.182  < 2e-16 ***
DHPI        -0.0092442  0.0008245 -11.212  < 2e-16 ***
DUER         0.0858430  0.0158338   5.421 5.91e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 27776  on 79520  degrees of freedom
Residual deviance: 27519  on 79516  degrees of freedom
AIC: 27529

Number of Fisher Scoring iterations: 6
```

Fig. 4.16 Binary approximation default model for dynamic data

```
Call:
glm(formula = frm_pp, family = binomial(link = "logit"), data = PData_pp)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8853  -0.3388  -0.2849  -0.2347   3.0854

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.4452986  0.1179463  -20.73   <2e-16 ***
aci          0.0020431  0.0001111   18.39   <2e-16 ***
oltv        -0.0211060  0.0013016  -16.22   <2e-16 ***
DHPI         0.0215910  0.0012306   17.55   <2e-16 ***
DUER        -0.2162531  0.0173258  -12.48   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 29181  on 79744  degrees of freedom
Residual deviance: 28298  on 79740  degrees of freedom
AIC: 28308

Number of Fisher Scoring iterations: 6
```

Fig. 4.17 Binary approximation prepayment model for dynamic data

Example 1 *Binary Approximation*

Example 2 *Binary Approximation*

For the two MEVs, HPI and UER, as described in Sect. 4.2.2, we use mean reversion:

$$\mathrm{HPI}(t) = \begin{cases} \mathrm{HPI}(t-1) * 0.995, & t \leq 36 \\ \mathrm{HPI}(t-1) * 1.0025, & t > 36 \end{cases} \tag{4.2}$$

$$\mathrm{UER}(t) = \begin{cases} \mathrm{UER}(t-1) * 0.995, & t \leq 36 \\ 0.05, & t > 36 \end{cases} \tag{4.3}$$

The amortization factor is calculated from the monthly interest rate $I_i = r_i/12$ of the loan as:

$$\mathrm{Amor}_i = \frac{I_i}{1 - (1 + I_i)^{-360}} \tag{4.4}$$

across all periods for 30-year mortgages.

$\mathrm{LGD}_i(t)$ is predicted according to (3.72) for the multiple resolutions model we choose.

Table 4.9 Summary information of CECL loss forecast for the static portfolio

ID	Forecast Date	PD	PP	amor	Loss	UPB
Length: 1328	Length: 1328	Min.: 0.02000	Min.: 0.0006293	Min.: 0.004491	Min.: 919	Min.: 9700
Class: character	Class: character	1st Qu.: 0.03280	1st Qu.: 0.0380422	1st Qu.: 0.005141	1st Qu.: 49,334	1st Qu.: 183,330
Mode: character	Mode: character	Median: 0.03823	Median: 0.0499646	Median: 0.005828	Median: 89,133	Median: 273,176
NA	NA	Mean: 0.03993	Mean: 0.0530658	Mean: 0.005850	Mean: 104,709	Mean: 283,733
NA	NA	3rd Qu.: 0.04446	3rd Qu.: 0.0657090	3rd Qu.: 0.006544	3rd Qu.: 143,103	3rd Qu.: 367,691
NA	NA	Max.: 0.22434	Max.: 0.1705625	Max.: 0.007337	Max.: 501,736	Max.: 873,000

Table 4.10 Summary information of CECL loss forecast for the dynamic portfolio

ID	Forecast Date	PD	PP	amor	Loss	UPB
Length: 1368	Length: 1368	Min.: 0.01920	Min.: 0.002711	Min.: 0.004493	Min.: 943.4	Min.: 9700
Class: character	Class: character	1st Qu.: 0.03694	1st Qu.: 0.019748	1st Qu.: 0.005178	1st Qu.: 47,575.2	1st Qu.: 179,632
Mode: character	Mode: character	Median: 0.04445	Median: 0.031386	Median: 0.005817	Median: 85,530.6	Median: 266,847
NA	NA	Mean: 0.04611	Mean: 0.037456	Mean: 0.005860	Mean: 98,863.5	Mean: 278,652
NA	NA	3rd Qu.: 0.05366	3rd Qu.: 0.048695	3rd Qu.: 0.006553	3rd Qu.: 134,766.5	3rd Qu.: 361,070
NA	NA	Max.: 0.11830	Max.: 0.210086	Max.: 0.007337	Max.: 386,324.6	Max.: 800,250

4.4.3 Forecasting Result Analysis

Table 4.9 shows the summary of the CECL loss forecast for the static portfolio. For this subprime portfolio with a total of 1328 loans, the loss rate is about 37% (ratio of the means 104,709/283,733 of loss and UPB). This is reasonable given that the average monthly default rate is 0.03993 and average severity near 0.9 (not shown).

Table 4.10 shows the summary of the CECL loss forecast for the dynamic portfolio. For this subprime portfolio with a total of 1368 loans, the loss rate is about 35.5% (ratio of the means 98,863.5/278,652 of loss and UPB). This is largely due to the lower average default rate (PD) and higher average prepayment rate (PP) in the early periods (in 36 months), but offset by the later higher unemployment rate reversion assumption (5%). Note that the summary statistics for default and prepayment probabilities (PD and PP) in Table 4.10 are for the entire remaining loan lifetime.

4.5 Model Validation and Performance Monitoring

Model risk has been treated as an independent tier of risk for over a decade since the milestone regulatory guidance of OCC Bulletin 2011-12 and FRB Supervisory and Regulation Letter 11-7. Over the decade, financial institutes have been building up proper model risk management framework. Such framework generally inherits from risk management framework popularly used by operational risks where model risk once belonged to. The framework usually consists of risk identification, assessment, measurement, mitigation, monitoring, and reporting. These components are commonly classified into two functions, model governance and independent model validation, and are commonly summarized in the following process cycle (Fig. 4.18):

In this section, we will describe these risk management components, with a focus more on the model validation side, which covers model risk assessment, measurement, mitigation, and some methodologies used by model performance monitoring. Using the CECL models as a blueprint, we go over the major steps in a formal model validation procedure according to the different components of a model.

A *model* refers to a quantitative method, system, or approach that applies statistical, economic, financial, or mathematical theories, techniques, and assumptions to process observed, assumed input data into quantitative estimates. A model produces an estimate of an uncertain quantity of varying precision and

Fig. 4.18 Model risk management framework

accuracy, depending on objective, assumptions, input data quality, and model structure. A model consists of three components:

- An information input component, which delivers assumptions and data to the model
- A processing component, which transforms inputs into estimates
- A reporting or output component, which translates estimates into useful business information

On top of these model components are the model scope and usage, which define the limits of model risk management.

A formal model validation procedure should cover all these components, and we will discuss in detail of these components in the following subsections.

4.5.1 Scope and Usage

Model scope defines the model coverage on products, and model usage defines the applications in which the model can be used. Although these two factors in model risk management are more on the governance side, they are critical to both model development and model validation as the initial steps to be discussed in details. Very often, the model development failure and model review rejection are due to wrong recognitions on model scope and usage. Such errors are critical, especially in the later stage of the processes. More catastrophic outcome could be resulted from a long-term misuse of model. To avoid such risks, model scope and usage should be clearly stated in the initial proposal with justification and approval from higher management. Over the time, the model scope and usage should be closely tracked, and limitations should be strictly abided.

For CECL models, the usage is clearly defined as credit loss projection for the lifetime of the products with reasonable and supportable economic conditions. As for scope, portfolio identification as we described in Sect. 4.4.1 is a good starting point. Model scope extension after model has been approved requires both business judgment and model risk reassessment, or even model revalidation.

For our subprime portfolios, we have two separate sets of risk event models for the two portfolios, while we only have one set of severity models for both portfolios. The reason for our model scope design is that the two portfolios have different transactional behaviors toward risk events, while their post-default loss behaviors are considered similar. These could be further investigated from the view of concept soundness discussed in Sect. 4.5.3 from the aspects of theory and design.

4.5.2 Inputs and Assumptions

Model inputs and assumptions belong to the first components of a model. Input data can be observed or assumed. Assumed data are related to forecasting of future with subjective. Under CECL, the focus for such data risk is the assessment of being "reasonable and supportable" as we discussed in Sect. 4.2.2 with some MEVs. For observed input data, there could be both random and systematic errors during data collection and processing. For random input data errors, good data quality checks could identify them. Such checks could include outlier detection, hard limits, similarity tests, and machine learning-based data quality checks. Over the last decade, the use of machine learning techniques through data quality control has greatly improved the accuracy and efficiency in data quality checks, especially in automation of data quality control processes, and reduces false-positive rates. Input data systemic errors are hard to discover and could result in higher risks. A commonly used method is comparing with benchmark data. Large deviation from benchmark data will draw attention for investigation of potential systemic errors.

One type of systemic error for input data often happens in bluntly using proxy input data, when the target portfolio and the portfolio generating the proxy data have fundamental different loss generating mechanisms. This is commonly seen in modeling for relatively new wholesale portfolios with less or no historical data recording. Although proxy input data are not "fabricated or synthetic" data defined as a type of Dark Data by Hand (2020)[1], similarity in credit loss between portfolios needs to be assessed, and potential gap needs to be measured in a conservative view.

Assessing the impact of input data errors to model output is an essential part of model validation as sensitivity analysis. Under the reality that input data errors can't be avoided completely, sensitivity analysis on input data is a way to assess the range of model output deviation by assuming certain degree of input data errors. Model output confidence intervals are commonly used to measure the impact of random input data errors under certain distributions.

For our subprime portfolio, input data errors to the actual LGD calculation are common due to the complex post-default processes and manual recording. Over the years, mortgage servicers have been improving their loss recording systems. However, there are abnormal LGD values from the reference loss data, and the simplest way is putting a threshold (e.g., 3) to filter out the larger LGD values to investigate whether this is due to a small exposure at default or some errors. As shown in Table 4.6, the largest actual LGD is 1.5, which is within a reasonable limit.

Model assumption risk has been a topic not being fully addressed since the initial regulatory model risk management setup. A critical reason is the difficulty in identifying critical model assumptions and how to assess and measure the risks of these assumptions. Most often, a model is used in a specific area due to the modeling tradition without careful investigation of what assumptions have already been embedded in the model development process. Further, model assumptions could be convolved into the model concept soundness. So, a sound model validation practice should have a specific section on model assumption identification and assessment. Assumptions make the model conceptually sound; however, specific

[1] David J. Hand. Dark Data: Why What You Don't Know Matters, Princeton University Press, February 2020.

assumptions on model implementation and model use should all be assessed for potential risks if these assumptions break. Statistics tests are popular tools used for the assessment and measurement of the assumptions.

For the GLM models we used for risk event modeling, we assume the risk event probability meets the proportional odds ratio assumption with the logit link to the risk drivers. The goodness-of-fit statistics in Sect. 3.1.1.1 are the common measurements of how this assumption is satisfied based on the observed data. So, such statistics must be provided by default as a part of the model validation.

Specific assumptions related to model choice and implementation must be independently and continuously assessed. In our model selection process with AEVS, we used various criteria to measure the final selected model as the best model.

Robust controls and processes must exist to ensure the completeness, timeliness, and integrity of key data inputs. This may include data reconciliations, independent review of manually input data, or other data quality controls. Developers should be able to demonstrate that data inputs are suitable for the model and consistent with model methodology. Any data proxies used must be identified, justified, and documented. All key assumptions underlying the model should be supported by initial and ongoing analysis and documentation, so that users are aware of any model limitations.

4.5.3 Theory and Design

Model theory and design are strongly linked to the concept soundness. Model development should provide sufficient evidence to demonstrate that the proposed models are mathematically and statistically correct before building the model. Model validation should confirm the correctness of model theory and identify any theoretical deficiency. Very often, model failure and model rejection are due to insufficient theoretical support or obvious deficiency in the design of modeling processes. Such model risks are classified as lack of concept soundness and are often hard to be mitigated. The best mitigation approach for such risks is making sure there are no errors in model theory and design before the model development.

Identifying and assessing model errors in theory and design requires understanding in depth the specific modeling areas and practices and sufficient training in model theories. For credit modeling, in Chap. 3, we provide the most commonly used models with fundamental theories and sound practical uses. For model validation purposes, the modeling background provided in Chap. 3 is a foundation. On top of this foundation, simulation testing is a common practice to validate model theory and design, though it is considered conservative due to the strong conditions required in simulations.

A more appealing approach for validating model theory and design is comparison with alternative theories and approaches, since such comparison directly shows the advantages and disadvantages of proposed and alternative models. Further, the comparison may demonstrate if the proposed modes work as intended, are appropriate for the intended business purpose, and are conceptually sound and

```
> frm_tr <- as.formula(paste("Reso_Type ~", "oltv", "+", "chpi", "+", "LoanSize", "+", "OccType", "+", "PropType"))
> tree_fit <- rpart(frm_tr, method="class", data=LGD_Data)

> printcp(tree_fit) # display the results

Classification tree:
rpart(formula = frm_tr, data = LGD_Data, method = "class")

Variables actually used in tree construction:
[1] chpi oltv

Root node error: 2845/4847 = 0.58696

n= 4847

        CP nsplit rel error  xerror     xstd
1 0.106151      0   1.00000 1.00000 0.012049
2 0.096661      1   0.89385 0.91986 0.012197
3 0.015817      2   0.79719 0.81547 0.012224
4 0.010000      4   0.76555 0.79895 0.012212
```

Fig. 4.19 LGD resolution model by decision tree

mathematically and statistically correct. Using our subprime reference default data, we show the comparison of LGD resolution models by multinomial logit regression used in Sect. 4.3.2 and the alternative decision tree approach.

We use the R package “rpart” for decision tree modeling. Figure 4.19 presents the decision tree fit results by using all five risk drivers – oltv, chpi, LoanSize, OccType, and PropType – as we described in Sect. 4.3.2. As discussed in Sect. 3.1.1.3, a decision tree minimizes the cost complexity criteria (3.43), which is the sum of the tree impurity measure and its complexity penalty. Among the three impurity measures, the “rpart” package takes the classification error as the default.

Figure 4.19 shows that the optimal decision tree only uses oltv and chpi as split variables, which is consistent with the best multinomial logit model selected. The “rpart” package presents the optimal tree under a series of complexity parameters (CP) with the number of splits (nsplit), the relative classification error (rel error) scaled by the initial number of misclassification (2845), the scaled cross-validation error (xerror), and its standard error. The final optimal decision tree with 4 splits corresponds to the complexity parameter 0.01 decided by the cross-validation criterion as we described in Sect. 3.1.1.3. This can be shown by growing the tree with more splits through specifying a smaller complexity parameter ($cp = 0.001$). Figure 4.20 presents the optimal trees correspond to a series of smaller complexity parameters. In our case, the cross-validation criterion directly decides the final optimal decision tree has 4 splits and corresponds to the complexity parameter 0.0028822. Actually, this decision tree is also optimal for any complexity parameters that fall in (0.0158172, 0.0028822) due to the cost complexity function defined in (3.43). By chance, the default CP value used by the “rpart” package 0.01 falls in this range. In general, the 1-SE rule is used to decide the optimal decision tree – the simplest tree with the least number of cross-validation errors falling in its cross-validation error range with one standard error (xstd) is selected.

To further show the decision tree optimization process, Fig. 4.21 presents the splitting process for the first two nodes. Since the LGD data set has more REO resolution types than the other, the entire data set is initially classified as REO and the classification error $2845/4847 = 0.586961$. Then, all five drivers are tested as a primary splitter, and chpi has the largest improvement in reducing classification

```
> tree_fit0 <- rpart(frm_tr, method="class", data=LGD_Data, cp=0.001)

> printcp(tree_fit0) # display the results

Classification tree:
rpart(formula = frm_tr, data = LGD_Data, method = "class", cp = 0.001)

Variables actually used in tree construction:
[1] chpi     LoanSize OccType  oltv     PropType

Root node error: 2845/4847 = 0.58696

n= 4847

          CP nsplit rel error  xerror     xstd
1  0.1061511      0   1.00000 1.00000 0.012049
2  0.0966608      1   0.89385 0.87452 0.012231
3  0.0158172      2   0.79719 0.81019 0.012221
4  0.0028822      4   0.76555 0.78348 0.012196
5  0.0024605     13   0.73603 0.79262 0.012206
6  0.0022847     15   0.73111 0.79086 0.012204
7  0.0021090     17   0.72654 0.78946 0.012203
8  0.0017575     18   0.72443 0.78910 0.012202
9  0.0014938     19   0.72267 0.79473 0.012208
10 0.0014060     32   0.69877 0.79789 0.012211
11 0.0012888     41   0.68612 0.80562 0.012218
12 0.0012302     44   0.68225 0.80387 0.012216
13 0.0011716     48   0.67733 0.80281 0.012215
14 0.0010545     54   0.67030 0.81230 0.012222
```

Fig. 4.20 LGD resolution model decision tree selection

error. So, the first split is done by chpi. Recursively, the best driver for the next splitting is chosen, which is also chpi. It also provides surrogate splits if the primary split driver is missing.

The final optimal tree is presented in Figs. 4.22 and 4.23 by counts and percentages, respectively. Note that in each node of the tree shown by Fig. 4.23, the first line percentage numbers are for the three resolution types REO/SS/TPS within that node, while the second line percentage number is for the node as its percentage in the entire population. The confusing matrix for the optimal decision tree is shown in Fig. 4.24. Compared to the confusing matrix from the multinomial logit model shown in Fig. 4.10, they are very close. The difference is that tree classifies less REO correctly, but more TPS correctly.

The decision tree model shows close performance to that of the multinomial logit model. Given that the decision tree is a nonparametric method and should have high accuracy in model fitting, this validates that the multinomial model is performing well in model fitting accuracy with our LGD data. Based on such performance comparison and the fact that a parametric model is always more stable and easier for implementation, the multinomial logit model is selected as the primary model for the

```
> summary(tree_fit) # detailed summary of splits

Node number 1: 4847 observations,    complexity param=0.1061511
  predicted class=REO  expected loss=0.586961  P(node) =1
    class counts:  2002   914  1931
   probabilities: 0.413 0.189 0.398
  left son=2 (4319 obs) right son=3 (528 obs)
  Primary splits:
      chpi     < 0.9903897 to the left,  improve=209.552000, (0 missing)
      oltv     < 58.07     to the right, improve= 28.877950, (0 missing)
      LoanSize < 136650    to the right, improve= 21.821690, (0 missing)
      PropType splits as  RLLRLRL, improve= 14.472310, (0 missing)
      OccType  splits as  RLL, improve=  2.055641, (0 missing)

Node number 2: 4319 observations,    complexity param=0.09666081
  predicted class=REO  expected loss=0.547812  P(node) =0.8910666
    class counts:  1953   563  1803
   probabilities: 0.452 0.130 0.417
  left son=4 (1543 obs) right son=5 (2776 obs)
  Primary splits:
      chpi     < 0.6474895 to the left,  improve=97.991250, (0 missing)
      oltv     < 57.40533  to the right, improve=29.621220, (0 missing)
      PropType splits as  RLLRLRL, improve= 9.157807, (0 missing)
      LoanSize < 527600    to the left,  improve= 4.650024, (0 missing)
      OccType  splits as  RLL, improve= 1.953092, (0 missing)
  Surrogate splits:
      oltv     < 48.7165   to the left,  agree=0.644, adj=0.003, (0 split)
      PropType splits as  RRLRRRR, agree=0.643, adj=0.002, (0 split)
      OccType  splits as  RLR, agree=0.643, adj=0.001, (0 split)
```

Fig. 4.21 LGD decision tree splitting process

LGD resolution type modeling. In the meantime, the tree model is a valuable benchmark model to be used to monitor the performance of the primary model.

4.5.4 Implementation and Output Analysis

Conceptually, a model is independent of the system or platform in which it resides. However, in practice, the performance of a model critically depends on the correct configuration and implementation in the system or platform.

A full modeling cycle includes model development and redevelopment, user acceptance testing (UAT), and the production delivery. Usually, these processes are carried out in different computing platforms and environments; however, in recent years, the concepts of model continuous integration and continuous deployment or delivery (CI/CD) have become more and more popular. The CI/CD model implementation brings both efficiency in model production and new challenges for model implementation validations. In the following, we will cover different stages of modeling and implementation processes to address potential issues model validation could embrace.

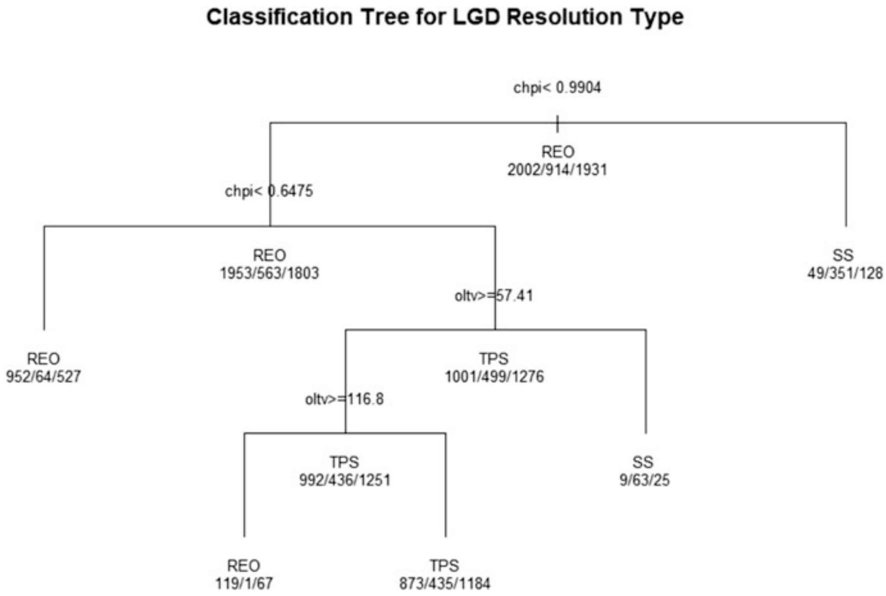


Fig. 4.22 Final decision tree by counts

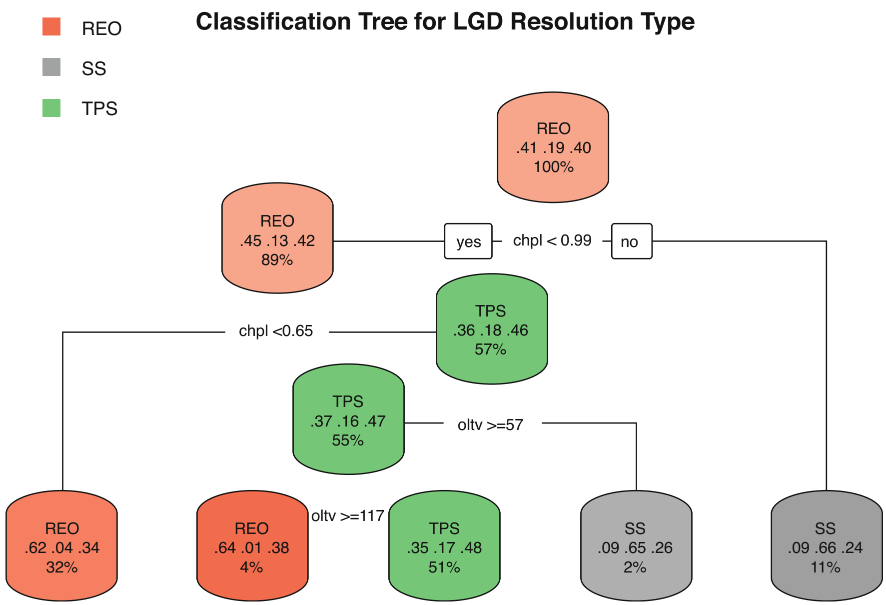


Fig. 4.23 Final decision tree by percentages

```
> tree_pred = predict(tree_fit, type="class")
> table(tree_pred, LGD_Data$Reso_Type)

tree_pred  REO   SS  TPS
      REO 1071   65  594
      SS    58  414  153
      TPS  873  435 1184


> prop3 <- with(LGD_Data, table(tree_pred, Reso_Type)) %>% prop.table(margin=1)
> prop3
         Reso_Type
tree_pred        REO         SS        TPS
      REO 0.61907514 0.03757225 0.34335260
      SS  0.09280000 0.66240000 0.24480000
      TPS 0.35032103 0.17455859 0.47512039
```

Fig. 4.24 Final decision tree by counts

Dev Environment

Model development is carried out in the model development environment, which is commonly called the Dev environment. The model Dev environment is the computing platform built for model developers to carry out all model development task, including data analysis, initial model design, and various testing related to model development. In recent years, the concept of integrated development environment (or IDE) has become popular. The IDE commonly defines a software platform including capabilities of data loading and visualization, user-friendly code editor that can provide smart programming, build and execution, debugging, and profiling for one or more programming languages. More powerful IDEs can even carry out unit testing, code integration, version control, deployment, and delivery. There are both open-source and commercial IDEs. In recent years, open-source IDEs have grown quickly, for example, in modeling, RStudio, Jupyter Notebook, Apache NetBeans, Apache Spark, and many others.

For independent model validation and auditing, model validators and auditors are often granted the same level of accesses as model developers with the Dev environment. Such requirements are not just for independently replicating model results; they are also critical for model validation and auditing to assess the model development environment for potential model risks.

UAT Environment

The user acceptance testing (UAT) environment is the next level of software development platform after the model development. For large and complex model, the UAT layer is necessary for model usage testing before model production. The UAT environment requires efficient ETL processes and friendly interfaces for user to test their own data with the deployed model. A data lake is usually a great help for this purpose.

UAT is a safe buffer between model development and production, which allows model users to test and learn the model implementation. Any issues discovered in the

UAT level can be sent back to model development for further investigation and testing. Since UAT is a less restricted environment compared to the production environment, this can speed up the model redevelopment and testing.

For high-quality model production, the UAT layer should not be skipped. Very often, model users could miss the UAT for some model releases and cause auditing failure. In recent years, automation of the UAT layer has become popular for model production efficiency.

PROD Environment

Model production environment requires higher restriction for access and data security, since very often, the production is delivered to clients (internal or external) and has exposure security requirements. Model production also has much less error tolerance, and any implementation issues could result in critical model risks.

Monitoring system is built on production environment for continued assessment of the model production, which is a critical part of model validation and will be discussed further later.

CI/CD Design

In recent years, to enhance the efficiency of model production process, the concept of continuous integration, deployment, and delivery (CI/CD) has become popular, and CI/CD implementation has become a trend in new software platforms, especially in the cloud computing platforms. From the model development side, this is an automation of the coding, building, packaging, and delivery process and greatly shortens the model production timelines. However, for the validation and auditing sides, there could be some transitions with less transparency.

To overcome the transparency issue with CI/CD, one option is using the layered model development processes as we described with Dev, UAT, and PROD for the initial model release and only using CI/CD for model redevelopment and production update. The other option is adding intermediate testing within CI/CD as a UAT layer. Either way, CI/CD should not become the reason for a less transparent modeling process, especially for large complex models.

Model Output Analysis

A basic requirement for output analysis is the ability to replicate the model outputs on appropriate platforms, especially in the production platform. For outputs with random number generation, a random seed has to be set for result replication. Output replication also presents validator a chance to review the model development and implementation logic, design, and coding. It also helps to check the model documentation consistency with model development and implementation. A full model output results replication should be done periodically to check gaps that could be created by ongoing model updates.

Variation analysis is an essential part of model output analysis, especially for loss forecasting models like CECL models. For model validation, variation analysis can be useful to access model behaviors and discover any model issues contradictory to business intuitions, because, very often, line of business requires to assess and explain the variation of loss projections from quarter to quarter as a way to understand projection dynamics and take proper business actions on top of these projections.

Sensitivity analysis is used to assess how sensitive a model's outputs are to the change of model inputs, which could be data and assumptions. Models should have proper sensitivity to corresponding inputs. In general, CECL models should not be too sensitive to input data, since it is assumed that CECL models are built on data from different economic cycles. Models show extreme sensitivity to data changes seen in different business and economic environments could indicate instabilities. Further investigation should be carried out to check if models have fundamental issues or the modeling data do not cover sufficient business and economic scenarios. Such issues could lead to model rejection.

Models could be sensitive to model assumptions. The assessment of model assumptions should be a part of the model concept soundness. Critical model assumptions should have been included in model limitations. Model outputs should be assessed for some minor deviations from critical model assumptions. For example, time series models are built under the assumptions that the target time series variables are stationary as discussed in Sect. 3.3.1 and need to be tested. Time series models rejected by a strict stationarity test with a lower test level may pass a less strict test with a higher test level, and these models could still be used for CECL forecasting if they pass all other tests.

4.5.5 Performance Monitoring

Model performance monitoring is critical, especially for loss forecasting models, which require continuous assessments. Ongoing performance assessment (OPA), back testing, and benchmarking are common practices for CECL model performance assessments.

Ongoing performance assessment monitors whether production models continue to perform with the time advancing. With the changes in markets, products, exposures, activities, clients, or business practices, production models not updated promptly could deviate from the business trend and results in poor performance as measured by both statistical and business criteria. There is also the possibility that model assumptions could be broken and new model limitations are needed. Models on which business decisions depend require close performance monitoring in a timely manner; otherwise, flawed and costly business decisions could be made before deterioration in model performance becomes apparent.

For forecasting models, there are too many factors that could be subject to change, and the loss forecast results requires timely checking with the actual observations. Large deviation of the forecasts from the actuals indicates the production model is not performing at least with the current observation and model enhancements or adjustments are required. There is also the requirement that when new data and information inputs are added, the production model should be competitive in the model selection process as required by the model selection criteria. So, both the performance measured by those statistical criteria in the model selection process and the model ranking in the model selection process should not deteriorate significantly.

For CECL models, there is the specialty that lifetime measurements are the targets, and those measurements are not observable currently. So, performance monitoring should be designed to reflect what we can observe since the assumed CECL expectation requires to be supportable. Under such situation, a confidence interval should be agreed upon all parties that the model performance based on actual observations in specified period should be within that interval. This is usually done through back testing. As for the CECL lifetime expectations, benchmarks are commonly used as the defending bars. We discuss these techniques next.

Back testing involves the comparison of actual outcomes with model forecasts during a sample time period not used in model development and at an observation frequency that matches the forecast horizon or performance window of the model. The comparison is generally done using expected ranges or statistical confidence intervals around the model forecasts. When outcomes fall outside those intervals, one should analyze the discrepancies and investigate the causes that are significant in terms of magnitude or frequency. The objective of the analysis is to determine whether differences stem from the omission of material factors from the model, whether they arise from errors with regard to other aspects of model specification such as interaction terms or assumptions of linearity, or whether they are purely random and thus consistent with acceptable model performance. Analysis of in-sample fit and of model performance in holdout samples (data set aside and not used to estimate the original model) are important parts of model development but are not substitutes for back testing.

One can extend the one-period back test to a more comprehensive *walk-through test*. In a walk-through test, starting from a specified historical timeline with a specified forecast period (e.g., quarterly or yearly), the target model being tested (usually the production model) is refitted on the data before the selected historical timeline, and its projection on the next period is compared to that period's actual observations, and then these actual observations are added to the historical data to refit the target model, and its projection for the next period is used to compare with the actual observation in the corresponding period. Continue this process to measure the target model's projection performance while walk through all these periods to the most recent period. The walk-through test continuously measures the target model's performance over several periods to see its pertinent strength in performance instead of only the most recent period. It is also often used to detect whether the model catches the trend of the underlying changes.

Benchmarking is the comparison of a given model's inputs and outputs to estimates from alternative internal or external data or models. It can be incorporated in model development as well as in ongoing monitoring. For credit risk models, examples of benchmarks include models from different methodologies, vendor firms, or industry consortia and data from retail credit bureaus. Whatever the source, benchmark models should be rigorous, and benchmark data should be accurate and complete to ensure a reasonable comparison.

Discrepancies between the model output and benchmarks should trigger investigation into the sources and degree of the differences and examination of whether they are within an expected or appropriate range given the nature of the comparison. The results of the analysis may suggest revisions to the model. However, differences do not necessarily indicate that the model is in error. The benchmark itself is an alternative prediction, and the differences may be due to the different data or methods used. If the model and the benchmark match well, that is evidence in favor of the model, but it should be interpreted with caution so we do not get a false degree of comfort.

The decision tree model we used in Sect. 4.5.3 for the LGD resolution types is a good example of a benchmark model. It is a type of highly nonlinear model and is fundamentally different from the multinomial logit model. The match of driver variables selected by both models is a confirmation of the backward variable selection process for the multinomial logit model. Even the confusion matrices from the two approaches are close given that decision tree should have an advantage in classification in theory. Decision tree models are also good benchmarks for default and prepayment binary logit models.

4.5.6 Model Governance

On top of the previous components of model risk management is the model governance, which sets an effective framework with defined roles and responsibilities for clear communication of model limitations and assumptions, as well as the authority to restrict model usage.

The model risk management framework as shown in Fig. 4.18 is set up by the model governance through policies and procedures. The common practice is that an institute has an overall model risk policy, which covers all aspects of model risk management, including model and model risk definitions; assessment of model risk; acceptable practices for model development, implementation, and use; appropriate model validation activities; and governance and controls over the model risk management process. Then, within different legal entities, line of businesses, or modeling areas, there may be different model risk management policies and procedures. These policies and procedures cover specific model risk management activities, for example, specific policies for certain legal entities, or specific procedures for model development, validation, and monitoring and reporting for a specified modeling area. Procedures usually provide more detailed guidance on the required activities.

CECL models should comply with these policies and procedures. Additional model risk policies may be added to CECL models due to its importance on accounting and capital and new models adopted to replace the incurred loss estimation, macroeconomic projections, and critical data risks.

4.6 ACL Based on Cohort-Level Data

For credit portfolios with only cohort-level data, the lifetime credit loss forecasting for CECL could face challenges in selecting a proper forecasting methodology. In this section, we present the popularly used methods and discuss their pros and cons in CECL implementation. Some of these methods may not qualify as a model under the formal model risk management framework described in Sect. 4.5; therefore, monitoring the risks on these methods may require higher standards, and their usage may be only limited on less important exposures. On the other hand, such methods may have the advantage of simplicity and be preferred by line of business. In summary, there should be a proper balance in the choice of these methods guided by the risk management governance to avoid excess use of these methods such that the regulatory requirements of CECL implementation on historical evidences, current conditions, and reasonable and supportable forward-looking are satisfied.

Aligning with the models discussed in Sect. 3.2 for cohort-level data, we cover the following methods – frequency and severity rate, loss rate, and vintage based.

4.6.1 Frequency and Severity Rate

For portfolios with risk event frequency (default and prepayment) and post-default loss data available at cohort level, loglinear models for default and prepayment rates and linear model for severity rate can be built based on historical data as descried in Sect. 3.2. These models are equivalent to the PD, PP, and LGD models for the loan-level data, respectively.

For CECL loss forecasting, similar to the loan-level model integration, the lifetime loss projection for a portfolio p with N_p cohorts can be computed accordingly:

$$\text{Loss}_{\text{CECL}}(p) = \sum_{h=1}^{N_p} \sum_{t=1}^{T_h} \text{UPB}_{ht} * \text{DR}_{ht} * \text{Severity}_{ht} \tag{4.5}$$

where UPB_{ht} is the active balance for cohort h at the beginning of period t, DR_{ht} (PR_{ht}) is the default (prepayment) rate for cohort h during period t, and Severity_{ht} is the loss severity for cohort h with default incurred during period t. T_h is the cohort weighted average maturity age (WAMA):

$$\text{T}_h = \frac{\sum_{i \in h} \text{UPB}_i * T_i}{\sum_{i \in h} \text{UPB}_i} \tag{4.6}$$

where UPB_i and T_i are unpaid principal balance and maturity age of ith loan in cohort h.

For committed credit products, the cohort-level active balance can be projected similarly to the loan-level UPB projection:

$$\mathrm{UPB}_{ht} = \mathrm{UPB}_{h(t-1)} * S_{ht}^{C} * S_{h}^{A} \tag{4.7}$$

where $S_{ht}^{C} = (1 - \mathrm{DR}_{ht} - \mathrm{PR}_{ht})$ is the conditional survival rate and $S_{h}^{A} = (1 - \mathrm{AmFactor}_{h})$ is the amortization survival rate.

The amortization factor for cohort h at period t, $\mathrm{AmFactor}_{ht}$, can be calculated from cohort weighted average coupon (WAC_{ht}) and cohort weighted average maturity age (T_h) at forecasting:

$$\mathrm{AmFactor}_{h} = \frac{\mathrm{WAC}_{h}/12}{1 - (1 + \mathrm{WAC}_{h}/12)^{-T_h}} \tag{4.8}$$

DR_{ht} and PR_{ht} can be projected using models (3.91) and (3.92) or more granular models (3.96) and (3.97). Severity can be projected using model (3.101).

For uncommitted credit products, UPB_{ht} is commonly projected through time series models, such as the ARMA model in (3.102), and the model integration is done through (4.5) with projected default rate and severity rate as for the committed credit products.

An alternative model integration for uncommitted credit products is through the projection of gross loss rate (GLR) as defined in (3.103):

$$\mathrm{Loss}_{\mathrm{CECL}}(p) = \sum_{h=1}^{N_p} \sum_{t=1}^{T_h} L_{ht} * \mathrm{GLR}_{ht} * \mathrm{Severity}_{ht} \tag{4.9}$$

where L_{ht} is the credit limit, GLR_{ht} is the projected gross loss rate, and $\mathrm{Severity}_{ht}$ is the projected severity for cohort h in period t. A loglinear or nonlinear model is commonly used for GLR.

As for the loan-level case, CECL requires reasonable and supportable forecast for macroeconomic variables used in all component models. Such forecasts for cohort-level CECL loss forecasts are the same as for loan level. Refer to (4.2) and (4.3) as examples we used in loan-level cases.

4.6.2 Loss Rate

For portfolios with only net loss rate available at cohort level, the loss rate and unpaid principal balance modes described in Sect. 3.2.3 can be aggregated to project the CECL lifetime loss:

$$\mathrm{Loss}_{\mathrm{CECL}}(p) = \sum_{h=1}^{N_p} \sum_{t=1}^{T_h} \mathrm{UPB}_{ht} * \mathrm{LR}_{ht} \tag{4.10}$$

where UPB_{ht} is the unpaid principal balance for cohort h at the beginning of period t. For uncommitted credit products, UPB_{ht} could also be the credit limit. It can be projected from a time series model. LR_{ht} is the net credit loss rate projected from the loglinear model with time series errors as described in (3.107).

The loss rate approach is a relatively simple approach for CECL loss forecasting. It does not track the loss generating process for credit products but instead only focuses on the net loss. The approach can be further simplified if the loss rate is projected based on historical averages as natural extension of the incurred loss approach. In the following, we present an example for this simple method.

Table 4.11 shows a one-million-dollar cohort with a 5-year term from 2017 to 2021 has a total incurred loss of $15,000. So, the total loss rate is 1.50%. To forecast the lifetime loss for a similar cohort with two-million-dollar balance from 2022 to 2026, we adjust the loss rate based both on the housing price factor (0.30%) and the unemployment rate factor (0.20%). The final projected loss rate is 2.00% and the lifetime loss is $40,000.

One can see that there are some key assumptions the simple loss rate method adopts. First, the historical loss rate is considered as the base loss rate for the cohort. The adjustments based on housing price and unemployment rate are assumed some fixed values. While the first assumption sounds reasonable, the second assumption is hard to justify and likely required to be more conservative. As we mentioned earlier, the method based on some extension of incurred loss rate is simple to implement, but hard to justify.

Table 4.11 CECL loss forecast based on loss rate for cohort

Time (year)	Amortized cost ($ thousands)	Actual net loss ($ thousands)	Loss rate (%)
2016	1000		
2017		5	
2018		4	
2019		3	
2020		2	
2021		1	
5-year total		15	1.50
Adjustment current condition			0
Adjustment housing price			0.30
Adjustment unemployment rate			0.20
Adjustment other factors			0
Total ECL %			2.00
Total ECL amount	2000	40	

Table 4.12 CECL loss forecast based on vintage loss rate for cohort

Orgi Year	Orig Balance ($Thousands)	Loss Rate					Incurred Loss ($Thousands)	Remaining Expected Loss %	ECL ($Thousands)
		Year1	Year2	Year3	Year4	Year5			
2013	100,000	0.20%	0.30%	0.40%	0.50%	0.10%	1,500	0.00%	0
2014	150,000	0.30%	0.25%	0.50%	0.20%	0.15%	2,100	0.00%	0
2015	200,000	0.20%	0.35%	0.40%	0.45%	0.10%	3,000	0.00%	0
2016	250,000	0.20%	0.30%	0.40%	0.50%	0.10%	3,750	0.00%	0
2017	300,000	0.25%	0.30%	0.40%	0.45%	0.15%	4,200	0.15%	450
2018	300,000	0.25%	0.30%	0.40%	0.50%	0.15%	2,850	0.65%	1,950
2019	300,000	0.30%	0.30%	0.45%	0.50%	0.15%	1,800	1.10%	3,300
2020	400,000	0.30%	0.35%	0.45%	0.50%	0.15%	1,200	1.45%	5,800
2021	500,000	0.30%	0.35%	0.45%	0.50%	0.15%	0	1.75%	8,750
Media Loss Rate		0.25%	0.30%	0.40%	0.45%	0.10%			
Adjustment		0.05%	0.05%	0.05%	0.05%	0.05%			
Final Loss Rate		0.30%	0.35%	0.45%	0.50%	0.15%			
Total									20,250

4.6.3 Vintage

The vintage method is a special case of the loss rate method. It is implemented only for closed portfolios (cohorts), whose lifetime is the remaining term of the running credit products.

Table 4.12 shows a cohort with a 5-year term and different vintages since 2013. Annual loss rates are observed based on the vintage of product origination. There are four vintages observed with complete loss rates, and the rest of the vintages have partial or no observed losses yet. To project the lifetime loss for this cohort, we estimated the annual loss rate based on the observed annual loss rates for all vintages by using the media and adjusting it upward with a fixed percentage (0.05%). The estimated annual loss rates are highlighted. The total CECL loss forecast is the sum of all the remaining expected losses, which is $20.25 million.

The vintage method is a granular loss rate method by further dividing the cohort at forecasting time into sub-cohorts based on maturity age. Similarly, the loss rate adjustment is hard to justify.

The change from incurred loss approach to lifetime loss projection by CECL imposes the difficulty for portfolios only having cohort data available. However, we present loss rate-based models that should be preferred as deeper insights and more reasonable and supportable assumptions can be applied to these models. Using models, one can at least understand where the risk could be instead of some subjective adjustments. It should be pointed out that any effort to make such adjustments on incurred losses as standard approaches could be misleading and embeds more risks into the approach and impairs transparency.

5 Capital Management and Risk Weighted Asset

This chapter is a continuation of Sect. 1.3 on capital management by focusing on the credit products. Following the framework introduced in Sect. 1.3, we cover both regulatory capital and economic capital. For both areas, we will go deeper into the modeling techniques introduced in Chap. 3 and show how those techniques are implemented in capital calculation.

For regulatory capital, we focus on the Risk Weighted Asset (RWA) calculation, which uses credit models for both loan segmentation based on credit risk profiles and unexpected credit loss calculations. We will introduce the concepts of through-the-cycle (TTC) and point-in-time (PIT) in credit models and show how these concepts are implemented in the Internal Rating Based (IRB) capital modeling. We will also cover loan segmentation techniques like classification tree based on probability of default (PD) estimates, final grid level PD estimates for RWA calculation, and grid migration monitoring on segmentation stability. For loss given default (LGD) modeling, regulation requires downturn LGD estimation based on identified portfolio downturn period since Basel II. We present a downturn identification method and its application on grid level LGD estimation. For regulatory capital, the exposure at default (EAD) is either observed balance at the reporting date for committed credit products or estimated exposure at the reporting date for uncommitted credit products. The final unexpected credit loss and RWA are aggregations of the PD, LGD, and EAD estimates as described in (1.9) and (1.10).

For economic capital, we focus on the simulation-based approach, which uses the developed PIT credit models (PD, LGD, EAD) in the loss generating process. First, we will build the Monte Carlo (MC) simulation framework as described in Sect. 1.3.2. We will show how to simulate the conditional credit loss distribution based on one set of macroeconomic factors as well as the full credit loss distribution based on the full joint distribution of all macroeconomic factors. Using the MC simulation framework, we demonstrate how stress testing and reverse stress testing can be carried out on economic capital.

For both regulatory and economic capital, the performance measure window is 1 year (12 months) as defined for credit risk. This requires aggregation if granular

C. Chen, *Practical Credit Risk and Capital Modeling, and Validation*, Management for Professionals, https://doi.org/10.1007/978-3-031-52542-1_5

observations are available in the data, for example, monthly or quarterly transaction observations. In this chapter, we will reuse the subprime mortgage data used in the previous chapter. The loan level transaction data are first transferred to 1-year performance data. Then for the probability of default (PD) model, we investigate various sampling methods discussed in Chap. 2, especially the Full Observation Stratified Sampling (FOSS) approach when delinquency information is added in the observations. The transferred loan level subprime mortgage data are used throughout this chapter to demonstrate how IRB approach is implemented for a popular credit product.

Again, as in the previous chapter, programming is first provided in the prototype format for the purpose that readers can practice using their own preferred languages. Then R code is partially provided as examples of implementation of these prototypes.

Model validation is a critical component for a successful capital modeling process. Although for both regulatory capital and economic capital modeling, best practices and concept soundness have been established over the years, there are changes and updates on both data and modeling from either industry or regulation. We present a full model validation framework illustrated with our synthetic data, on which we can manipulate for clear insights.

Although Basel Capital management has been criticized for its sufficiency to manage the unexpected losses shown in the 2008 financial crisis (and thus the introduction of CCAR and CECL), both regulatory and economic capital processes are still the basic requirements in capital management. In addition, modeling components in both regulatory capital and economic capital are foundations of other risk management processes. So, we will make an effort to cover the modeling details for these capital management processes.

5.1 Introduction to Capital Modeling Processes

Since Basel II, financial institutes have been building up and updating their capital management platforms, on which the capital modeling processes act as the core. For large banks, regulatory capital and economic capital modeling processes are essential to meet the regulatory requirements and internal capital adequacy assessment and management. In the following, we introduce the commonly used modeling processes for both regulatory capital and economic capital management.

5.1.1 Regulatory Capital

Basel II introduced the Internal Rating Based (IRB) regulatory capital calculations and recommended large banks to adopt IRB-based capital processes, while small banks are allowed to use standard approaches with simplified calculations. We will focus on the IRB approaches. Figure 5.1 describes a commonly implemented regulatory capital modeling process.

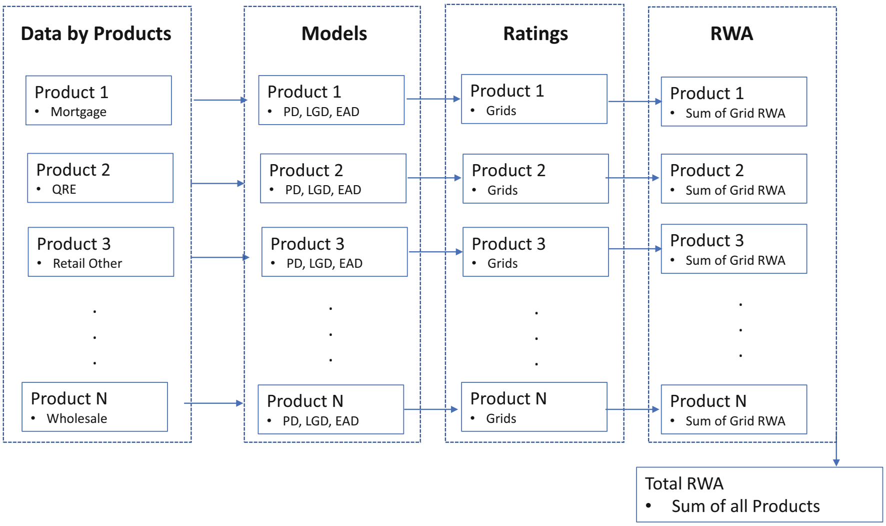


Fig. 5.1 Regulatory capital modeling process

The IRB approach starts from asset segmentation and data processing by segments. Different from CECL, CCAR, and other loss forecasting-based processes, for which segmentation requires granularity at least in the portfolio level, IRB can start from the full product (e.g., mortgage) and use the developed rating system for further segmenting into grids. Each grid is considered a portfolio of the product with similar risk profile.

Data are aggregated for the entire product; thus models (PD, LGD, and EAD) are built on the product. For capital modeling, performance is only required to be projected once for the next period (12 months), and cross-sectional historical performance data are needed for the PD model. Since exposures are not needed to be tracked for multiple periods, the prepayment (PP) model is not needed in capital modeling. For IRB, since the main use of the PD model is to create the internal rating system and classify the product into grids with similar risk profiles, the through-the-cycle (TTC) concept is introduced for this purpose. For LGD, regulation recommends the downturn concept.

The internal rating based on the PD model is the core of IRB approach. The grids created based on the internal rating system are considered homogenous in risk profiles. So, the migration of loan flows between these grids is closely monitored to making sure the internal rating system is stable.

Once a stable internal rating system is built, observed historical default averages within each grid are treated as the estimated PD for the grid and are used in the standard formula (1.9) for grid level unexpected loss projection. The RWA is calculated using the standard formula (1.10), and the final total RWA is the summation of all RWAs by products or even different risks as in (1.1).

The regulatory capital modeling process is relatively straightforward, and the main components are the capital data processing and model building. We will illustrate these components in the following sections using our subprime mortgage portfolio. Since the calculations are all formula based, the entire process is computationally inexpensive.

5.1.2 Economic Capital

The economic capital modeling process is based on simulation. We have presented a portfolio economic capital modeling process in Sect. 1.3.2 (see Fig. 1.7), which is a conditional economic capital framework based on a set of given macroeconomic factors and other risk factors. If we simulate these macroeconomic factors and risk factors from appropriate models, then we have the full marginal economic capital framework as described in Fig. 5.2.

For segmentation, EC framework has the flexibility to include both full product and a sub-portfolio into the framework. Historical data are collected in the appropriate level corresponding to the segments included in the EC frameworks. Risk parameters are estimated for the PD, LGD, and EAD models based on the historical data. Different from the regulatory capital process, these models are used to project the point-in-time (PIT) estimates of PD, LGD, and EAD. Market data, which include macroeconomic and other risk factors, can be given as conditions for conditional EC or simulated from the models built for marginal EC.

In the simulation engine, random events are simulated from the PD model to mimic the loss generating process by incorporating the simulated LGD and EAD. These simulated losses can be aggregated in various levels to approximate the loss

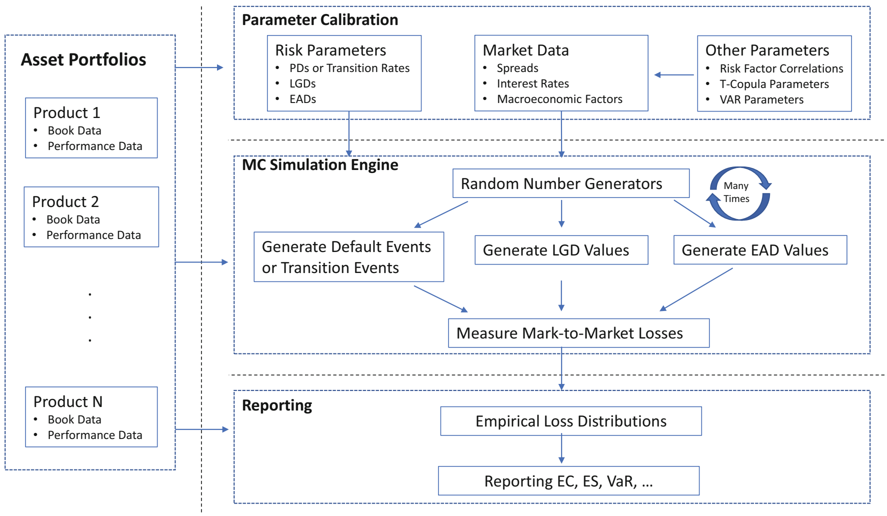


Fig. 5.2 Economic capital modeling process

distributions. Unexpected losses are estimated directly from these empirical loss distributions.

As we discussed in Sect. 1.3.2, the EC framework has the advantage to incorporate correlations among risk factors as well as products. Such correlations are hard to measure among products and portfolios, while EC can efficiently incorporate them into the framework based on their loss generating processes summarized from the models built for each product or portfolio. This provides a convenient way to investigate the diversification benefit and carry out capital allocation as we described in Sect. 1.3.2.

Since the EC process traces each loss events by simulating the loss generating process instead of using analytical approximations as in the RC process, we can obtain full loss distributions at any granularity level. This allows the insights for any broken of capital controls when imposed on the required level. Thus, EC framework can be used for reverse stress testing and other capital management programs.

An obvious disadvantage of the EC process is expensive computation. To obtain the full distributions of losses at all levels, a large number of simulations are required, especially for the better accuracy of high quantiles required by the capital estimates. However, with the great advance on computing power, parallel algorithms, and distributed systems, the computational bottleneck is not that critical anymore.

5.2 Capital Model Data

Regulatory capital models are usually built on data with less granularity compared to loss forecasting models. Part of the reason is that PD, LGD, and EAD models in regulatory capital modeling are used for risk profile segmentation. It is the actual observed historical averages of these parameters in each segment that are used in the final capital calculation. So, the historical data assembled for capital modeling, also called reference data, are critical for capital modeling and calculation. To make sure the consistency on reference data usage and the efficiency as a regulatory measure of capital sufficiency, it is not surprised that regulatory guidelines have been continuously issued to standardize the data requirements in capital modeling, at least among large banks in the use of advanced IRB approach.

We will not track the history of regulatory requirements on reference data in capital modeling here. However, we will focus on some major adoptions on reference data since Basel II through the best practice recommended by regulators. The first adoption on reference data is the coverage of a full economic cycle, which corresponds to the though-the-cycle (TTC) concept in regulatory capital. The second is the conservative adoption on LGD with a downturn period.

In addition to the regulatory requirements on reference data, finance institutes are facing continued data processing and platform upgrades in capital modeling. New data architectures and technologies with enhanced automation processes as we described in Sect. 2.6 challenge the existing capital modeling infrastructures, and an industry-scale upgrade can be foreseen.

5.2.1 Economic Cycles

Market economy is cyclical. At the end of an economic recession, economic activities reach the trough, and a new cycle of economic growth starts. This economic growth may slow down but it will continue till a peak. From the peak, the economy starts to shrink and quickly forms a new recession. Then a new economic cycle starts. This cyclical pattern can be represented by the national unemployment rate in Fig. 5.3, which shows over 70 years of national unemployment rate cycles till the recent huge jump due to the 2020 pandemic. Although the pandemic impact on economic cycles may be a one-time event, the economic cyclical pattern is long-term.

In consideration of the existence of economic cycles, capital estimates should include loans issued both at economic trough and peak in the reference data, since the actual performance of loans in the reference data decides the risk parameters for regulatory capital calculations. To avoid directional biases, reference data should include loans over at least one full economic cycle. For large financial institutes, regulatory guidance recommends reference data should cover loans with performances over multiple economic cycles including the recent 2008 financial crisis.

In addition, the IRB method maps individual loans into the grids which are constructed from the reference data based on risk profile. To cover the full range of risk profiles, the reference data should be comprehensive by including loans with different underwiring standards. During economic growth, financial institutes tend to

Fig. 5.3 US unemployment rate

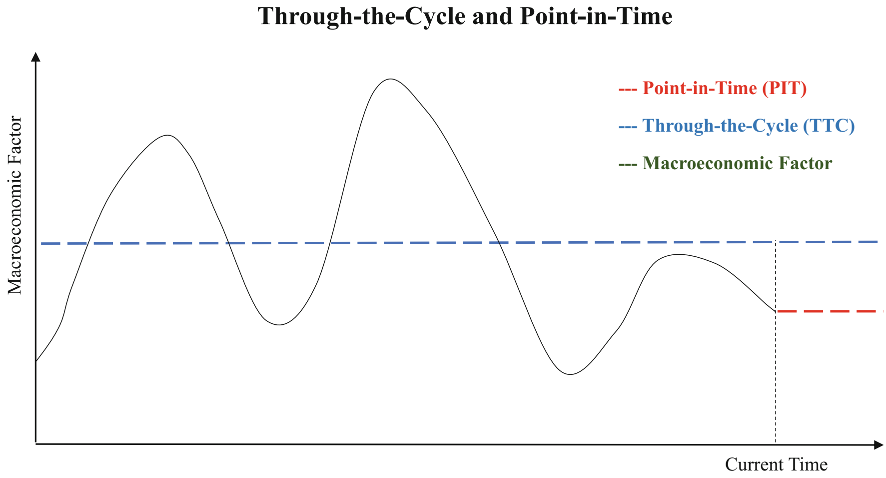


Fig. 5.4 TTC and PIT

loosen their underwriting standards to obtain more market shares, while in economic shrinking periods they tend to tighten the underwriting standards.

The through-the-cycle (TTC) concept was introduced to alleviate the pro-cyclical effect of capital estimation based on the current economic conditions when point-in-time (PIT) values are used for the economic factors. Since using PIT values of economic factors, the capital estimates at economic trough will be larger, while the estimates will be smaller at economic peak. This pro-cyclical effect will lead to banks tightening liquidity at economic trough while loosening liquidity at economic peaks. Besides manually adjusting such pro-cyclical effects in capital management, using the TTC values for economic factors in regulatory capital calculation is an alternative.

Figure 5.4 shows the difference between TTC value and PIT value for a macroeconomic factor which is used in the capital models. The TTC value is the average of the macroeconomic factor over a period, while the PIT value is the current observed value.

The TTC and PIT values for risk factors used in the credit models produce TTC and PIT estimates of the risk parameters (PD, LGD, and EAD). While TTC estimates are commonly used for risk profile segmentation, the PIT estimates are used for model performance validation and monitoring.

5.2.2 Downturn Periods

In the IRB approach, besides the full economic cycle requirements by risk parameter estimation, the Basel II final rule specifically requires the reference data include a downturn period. The downturn period is defined when defaults are “significantly higher than average." Adjustments are required as provision when downturn data are not available.

There are two purposes of incorporating downturn data in the reference data. The first is to incorporate a full economic cycle including downturn periods in estimating PD as we explained early. The second purpose is to estimate the downturn LGD and EAD as inputs in the Basel capital formula. Here, we focus on the downturn LGD (DLGD) quantification, while the downturn EAD quantification for uncommitted credit products is similar.

The downturn LGD quantification includes the following steps – downturn period identification, PD and LGD correlation confirmation, DLGD quantification, and DLGD adjustment due to incomplete data. We briefly introduce these steps, and details will be illustrated by an example in Sect. 5.3.2.

As required by the Basel Final Rule[1], the downturn period is identified by the probability of default:

> Economic downturn conditions mean, with respect to an exposure held by the [bank], those conditions in which the aggregate default rates for that exposure's wholesale or retail exposure subcategory (or subdivision of such subcategory selected by the [bank]) in the exposure's national jurisdiction (or subdivision of such jurisdiction selected by the [bank]) are significantly higher than average. (p. 69399)

For regulatory capital, the default rate is the 1-year (12 months) forward cumulative default rate from the reporting date. The length of the downturn period as well as the magnitude of downturn aggregate default rate from the average was not clearly defined in the Basel rule. Finance institutes are encouraged to develop some internal rules on these measurements.

The downturn period is defined according to PD, not LGD. So, it is necessary to confirm that the actual LGD values through the economic cycle and during the downturn period have a positive correlation with the default rates, or equivalently higher PD corresponds to higher LGD. As a commonly accept practice, a downturn LGD quantification is needed only when such a positive correlation is confirmed; otherwise a long-run average of default weighted LGD is sufficient for the downturn capital calculation.

Once the positive dependency between default rate and severity rate is confirmed, a quantification of the downturn LGD should be developed. The TTC LGD, projected by the LGD model with neutral macroeconomic inputs and other risk factors valued at default, was used to develop a LGD grid scheme. This is done by a tree-based clustering model. Then the LGD grid maps the projected TTC LGD to the actual LGD. The default weighted averages of actual LGD in these grids for the downturn defaulted loans are defined as the DLGD and are used for the Basel capital calculation.

[1] Final Rule: Risk-Based Capital Standards: Advanced Capital Adequacy Framework – Basel II – Federal Register/Vol. 72, No. 235, page 69288/Friday, December 7, 2007/Rules and Regulations.

Very often, due to censoring of resolution observation in the reference loss data and the observation that quickly resolved loans have lower LGD, the actual LGD as well as the final DLGD needs adjustment. The adjustment is based on the censoring time measurement and an empirical percentage of adjustment based on the censoring time.

5.2.3 Mortgage Reference Data

We use the subprime mortgage portfolio introduced in Sect. 4.2.1 for an example of the capital modeling process. Again, we use the synthetic data by simulating the performance of the loans from a given model, which is similar to the dynamic model we used in Sect. 4.2.1 to form the dynamic transaction data. As pointed out earlier, the reference data for capital modeling should be comprehensive, so we should include all loans from the portfolio in the reference data and not a sample as we have done for loss forecast modeling.

Table 5.1 presents partial information of the full subprime mortgage portfolio. There are 69,241 loans in the portfolio, mostly originated between 2000 and 2008. The data include five risk factors. The acquisition index (aci) as a risk index based on the loan acquisition information (alike the credit score) ranged from 300 to 850. The origination loan to value ratio (oltv) is ratio of loan amount at origination (LoanSize) to the appraisal house value at origination. OHPI and OUER are the house price index (HPI) and unemployment rate (in percentage) at origination.

Table 5.2 presents a summary of the synthetic transaction data, which is similar to that of the dynamic transaction data in Sect. 4.2.1.

The transaction data are used to form the PD reference data by snapshot sampling as we explained in Sect. 2.4.2. First, we aggregate the monthly observed default indicator (DF_Event) into a 1-year forward cumulative default indicator (CDF_1Y). Then, loans active at each snapshot (usually at quarters) from the transaction data are taken as independent loans with its 1-year forward performance measurement

Table 5.1 Summary information for the subprime mortgage portfolio

ID	OrigDate	aci	oltv	LoanSize	OHPI	OUER
Length: 69241	Length: 69241	Min.: 300.0	Min.: 1.528	Min.: 4191	Min.: 97.35	Min.: 1.300
Class: character	Class: character	1st Qu.: 437.0	1st Qu.: 71.053	1st Qu.: 165442	1st Qu.: 230.50	1st Qu.: 4.100
Mode: character	Mode: character	Median: 574.0	Median: 79.487	Median: 260000	Median: 280.27	Median: 4.800
NA	NA	Mean: 574.6	Mean: 79.913	Mean: 286529	Mean: 267.07	Mean: 5.491
NA	NA	3rd Qu.: 711.0	3rd Qu.: 80.000	3rd Qu.: 377000	3rd Qu.: 309.03	3rd Qu.: 5.900
NA	NA	Max.: 850.0	Max.: 460.309	Max.: 5000000	Max.: 714.1	Max.: 27.742

Table 5.2 Summary information for the subprime mortgage transaction data

ID	ActDate	age	aci	oltv	DHPI	DUER	DF_Event	PP_Event
Length: 806847	Length: 806847	Min.: 1.00	Min.: 300	Min.: 1.528	Min.: –217.1216	Min.: –20.0470	Min.: 0.00000	Min.: 0.00000
Class: character	Class: character	1st Qu.: 4.00	1st Qu.: 426	1st Qu.: 72.235	1st Qu.: –40.7527	1st Qu.: –2.8934	1st Qu.: 0.00000	1st Qu.: 0.00000
Mode: character	Mode: character	Median: 8.00	Median: 556	Median: 79.876	Median: –20.9244	Median: –1.7248	Median: 0.00000	Median: 0.00000
NA	NA	Mean: 12.06	Mean: 561	Mean: 80.384	Mean: –29.2323	Mean: –2.0954	Mean: 0.04054	Mean: 0.04528
NA	NA	3rd Qu.: 17.00	3rd Qu.: 693	3rd Qu.: 80.000	3rd Qu.: –9.3339	3rd Qu.: –0.8533	3rd Qu.: 0.00000	3rd Qu.: 0.00000
NA	NA	Max.: 117.00	Max.: 850	Max.: 460.309	Max.: –0.9735	Max.: –0.0650	Max.: 1.00000	Max.: 1.00000

Table 5.3 Summary information for the subprime mortgage PD reference data

ID	ActDate	aci	oltv	DHPI	DUER	CDF_1Y
Length: 245423	Length: 245423	Min.: 300.0	Min.: 1.528	Min.: −208.967	Min.: −19.21560	Min.: 0.0000
Class: character	Class: character	1st Qu.: 425.0	1st Qu.: 72.364	1st Qu.: −40.921	1st Qu.: −2.89337	1st Qu.: 0.0000
Mode: character	Mode: character	Median: 554.0	Median: 79.900	Median: −21.061	Median: −1.72543	Median: 0.0000
NA	NA	Mean: 559.8	Mean: 80.431	Mean: −29.303	Mean: −2.09675	Mean: 0.3153
NA	NA	3rd Qu.: 691.0	3rd Qu.: 80.000	3rd Qu.: −9.374	3rd Qu.: −0.85327	3rd Qu.: 1.0000
NA	NA	Max.: 850.0	Max.: 414.505	Max.: −1.127	Max.: −0.08724	Max.: 1.0000

conditional on the macroeconomic factors and other risk factors measured at the snapshot to form the PD reference data.

Table 5.3 presents a summary of the PD reference data created from the snapshot sampling. We take the quarterly snapshots starting from January 2000 and ending on October 2013. The selection of the snapshot quarters is a combination of the observation of transaction data distribution and a near two-economic-cycle period based on the unemployment rate in Fig. 5.3.

As we explained in Sect. 2.4.2, the PD reference data are treated as cross-sectional data, which means the individual loans taken from the snapshots are treated as independent loans, not as the panel data which count on the time continuation on each loan.

The LGD reference data is similar to the LGD reference data for CECL modeling and usually called reference default data. Since for capital modeling, the focus is on the economic loss, which is calculated based on the net present value of all the LGD components as described in (3.57) of Sect. 3.1.4.1. So, besides the loss severity calculated for all these default loans in the reference default data, economic losses discounted back to the default time are also calculated. More details including our subprime mortgage reference default data will be presented in Sect. 5.3.2.

For EAD, the observed balance at reporting time is simply used for a subprime mortgage loan as committed credit product. However, for uncommitted credit product, EAD models are built on a reference EAD data based on default loans as described in Sect. 3.1.5.

5.3 Capital Models

In this section, we build the capital component models based on the reference data sets described in Sect. 5.2.3. For the 1-year forward cumulative default risk, we adopt the target event risk modeling framework illustrated in Sect. 3.1.1, since capital modeling only considers one-time credit risk measured by the 1-year forward cumulative probability of default. As described in the previous section, the 1-year PD is measured on reference data, which is constructed by snapshots of the transaction data (or other sampling techniques, e.g., FOSS) and aggregating the monthly default indicator into a 1-year forward cumulative default indicator.

Similar to forecasting models, model selection is also important for capital models though the focus of capital models is more on capturing of critical risk factors for the risk profiling and risk rating. Under the target risk framework and without the need of considering prepayment, the capital PD modeling is relatively simpler compared to the loss forecasting modeling. However, a sound model selection procedure is required for the purposes of both rigorous model development and validation. A modified version of the adaptive and exhaustive variable selection (AEVS) procedure for generalized linear models (GLM) used for loss forecasting modeling in Sect. 4.3.1 is adopted for the capital PD modeling.

Although the capital PD modeling processes are similar for both Regulator Capital (RC) and Economic Capital (EC), there are different focuses. As we pointed out at the beginning of this chapter, RC is more focusing on the TTC PD while EC on the PIT PD. Traditionally, RC and EC PD models are built together, but segmentation, reference data processing and model selection processes could be different when different portfolio or product granularity is required.

We build LGD models based on a reference default data set from the subprime portfolio. For capital modeling, economic loss given default or ELGD, based on the net present value (NPV) at default, is the loss measurement given default. Commonly, actual ELGD has fewer extreme values, and a linear regression model is sufficient and preferred due to its simplicity. For RC, as the PD model, the LGD model is used to create the internal risk ratings under the TTC implementation, and DLGD is the final estimate. For EC, PIT LGD is the final estimate.

These component models are integrated into either the IRB framework for RC or the MCMC simulation framework for EC to calculate the regulatory capital (or RWA) or the loss distributions for EC. We will cover the model integration and capital calculation in Sect. 5.4.

5.3.1 PD Models

Similar to the loss forecast PD modeling, variable selection is also essential for capital PD modeling. There are some differences between loss forecast modeling and capital modeling in variable selection. First, loss forecasting modeling is more focusing on the accuracy of prediction in multiple periods, while capital modeling is focusing on the accuracy of one-time prediction used for either risk profile classification or PIT measures. Second, capital PD modeling (especially for regulatory purpose) commonly bases on large reference data, on which exhaustive variable selection algorithm may be too computationally expensive. Lastly, over the years of capital modeling practices, critical risk factors for various financial products have been leaned and used in the models with stable performance.

Due to these reasons, we split the pool of candidate variables into static and dynamic. The variables in the static pool are collected from previous models, line of businesses and model experts. These variables are always included in the candidate models. For variables in the dynamic pool, we assume less knowledge of their performance in the candidate models and go over the full processes of transformation and exhaustive selection. We control the size of this dynamic pool of variables and make sure the exhaustive selection is computationally feasible. The modified AEVS is still locally optimal and has the ability to learn what critical variables should be selected in the model.

We present this modified version of AEVS (called Smart AEVS) for capital PD modeling with generalized linear models used in capital PD modeling in the following:

GLM Modeling Process with Smart AEVS

- *Data Input Preparation*
  (a) Target risk event variable: (CDF_1Y): $Y = \begin{cases} 1, & \text{Event} \\ 0, & \text{Non} - \text{Event.} \end{cases}$
  (b) Select the static pool of risk factors X_{static} based on previous models, line of business (LOB) and capital model experts for the specific credit products. These risk factors will always be included in the candidate models.
  (c) Select the dynamic pool of risk factors, X_{dynamic}, including macroeconomic variables (MEV) chosen by line of business (LOB); calculate $T(X_{\text{dynamic}})$, the list of chosen meaningful transformations for each macro; calculate Lag (X_{dynamic}), Lag$(T(X_{\text{dynamic}}))$ meaningful over certain periods (e.g., quarters).
  (d) Determine sign(X_{static}), sign(X_{dynamic}) the meaningful relationship (positive or negative) between X_{static}, X_{dynamic} and Y.
  (e) N, the total number of X variables including X_{dynamic}, $T(X_{\text{dynamic}})$, Lag (X_{dynamic}), Lag$(T(X_{\text{dynamic}}))$ called four variations of X_{dynamic},select top $D = 10$ candidate variables from these N variables using each of the following methods:
    - Concordance Index (i.e., area under ROC) top D (10)
    - Standard stepwise (based on -2loglikelihood) top D (10)

  Combine variables selected from Concordance Index and stepwise.
  Determine TopN, the Combined list for selected variables derived above.

- *GLM Model Selection*
  (a) Generate possible subsets of TopN variables using exhaustive search, and combine with all variables in the static pool X_{static} to form Model $M_j, j = 1, \ldots, n \sim 2^{\text{Top}N}$.
  (b) Reject M_j if any regression coefficient doesn't satisfy sign(X_{static}) and sign (X_{dynamic}).

(c) Reject M_j if more than two variations of one X_{dynamic} are included in the model.
(d) Reject M_j if any GLM coefficient p value $>5\%$.

This determines the selected candidate GLM models.

- *Model Ranking*

  For the survived candidate models after step 2

  (a) Rank by SIC (the lower SIC, the higher Rank M_j has); optionally reject low-rank models.
  (b) Rank by AIC (the lower AIC, the higher Rank M_j has); optionally reject low-rank models.
  (c) Rank by v-fold cross-validation (the lower CV_v, the higher Rank M_j has) on survived models.
  (d) Calculate the average rank of the three performance rankings above; select the top 5 models based on the average rank. Note that we use v-fold cross-validation CV_v as the goodness-of-fit measure instead of the Leave-One-Out (LOO or CV_{nn}) due to computational efficiency for GLM models.

From the top five models, choose the final model based on LOB review and model assessment.

Example: Capital PD Modeling for a Subprime Mortgage Portfolio

In the following, we carry out the capital PD modeling process using our subprime mortgage data. As described in Sect. 5.2.3, the PD reference data are constructed from quarterly snapshots of the transaction data. The target variable is the 1-year cumulative default event. There are seven risk factors. The two risk factors aci and oltv form the static pool, and the rest five risk factors, DHPI, DUER, OHPI, OUER, and LoanSize, form the dynamic pool.

Model Selection

As we stated in loss forecast modeling, our model selection with AEVS tries to achieve a balance between automation with computational efficiency and adaptability with business and expert inputs. For capital PD modeling, one step further with AEVS is the learning ability over the time to catch the critical risk factors that have been showing stable performance in capturing important features in risk profiles, which is the focus of capital models. So, we introduced the Smart AEVS, which split the pool of candidate risk factors into static and dynamic.

With Smart AEVS, the risk factors in the static pool are collected from the previous capital PD models, critical inputs from line of business and modeling experts. These risk factors will always be added into the candidate models. The risk factors in the dynamic pool are less certain in capturing risk profile features and are subject to transformation and selection in the candidate models. The combination of all static risk factors and a selection of dynamic risk factors forms the candidate models. All candidate models are subject to the filtering process based on various model selection criteria. The final survival candidate models are ranked.

For our example with the subprime mortgage PD reference data, we have seven risk factors in our candidate variable pool. By our experience, the two risk factors, aci and oltv, are critical in capturing loan risk profiles. They are both measured at loan origination, aci is a risk index based on the loan acquisition information (alike the credit score), and oltv is the origination loan to value ratio as a measure of the equity of a loan. We treat the rest five risk factors DHPI, DUER. OHPI, OUER, and LoanSize as from the dynamic pool and will be subject to selection in the candidate models. We skip the transformation process for these dynamic risk factors for simplicity. In general, complex transformations (e.g., spline or other nonlinear transformations) could be applied for special nonlinear relationships between risk factors and the target risk event.

So, there are totally $2^5 = 32$ combinations of the dynamic risk factors (including the intercept only model). For 32 combinations, we can use the recursive algorithm with AEVS as described in Sect. 4.3.1.

Model Ranking

One may notice that we use the same ranking criteria in the Smart AEVS for capital PD modeling as we used for loss forecasting. The main reason is that SIC prefers parsimonious models, and we have a small pool of candidate risk factors. So, survived models are ranked by the three goodness-of-fit measures – SIC, AIC, and CV_v. For the v-fold cross-validation measure, due to computational efficiency for our relatively large data set, we choose $v = 11$. The negative 2loglikelihood (also called deviance) usually tends to select larger model and thus is not used as a goodness-of-fit measure in the model ranking. Note that with Smart AEVS, we also have a specific order for applying the three goodness-of-fit measures due to the optional further model rejection based on these measures. If these optional model rejections were not used, then the order of applying these measures on model ranking is not relevant.

The optional further model rejections based on these goodness-of-fit measures have several advantages than directly applying these measures to all survived models from the model selection step. First, it is computationally more feasible for AEVS with large data since some of these goodness-of-fit measures (e.g., measures from the cross-validation family) are computationally intensive, and computing such measures for large number of survival models from the model selection step may not be feasible in some cases. Secondly, the order defines a preference on different goodness-of-fit measures in the model ranking step since the preferred measure can be used first to further reject low-performance models based on this measure and less preferred measures only have the chance to do so on the survived models after the preferred measure. For the capital PD modeling, we put SIC as the first ranking criterion to make sure the selected models achieve a good balance between model complexity and goodness of fit on the reference data. Lastly, by adjusting the rejection threshold applied to difference measures, one can see how consistently difference measures behavior in the model ranking step. In our example, we can see SIC and AIC are largely consistent from the relatively stable ranks; these two measures present when adjusting the AIC rejection threshold; however, the ranks given by the CV_v measures are less consistent.

We select the top five models by SIC and obtain ranks of these models by AIC and CV_v with $v = 11$. These top five models with the three goodness-of-fit measures, as well as their ranks based on these measures and average ranks, are shown in Table 5.4.

Though ranks by AIC and CV_v are different from that of SIC, the first model has the top rank for all three measures. So, the top model selected by the average rank is the same top model for all measures, which is the best approximation to the model we used to simulate the synthetic data. Figure 5.5 displays the final top model selected.

Besides the two variables aci and oltv in the dynamic pool, there are four risk factors selected by the top model, DHPI, DUER, OHPI, and OUER. While the first four risk factors were used in the creation of the synthetic transaction data, OHPI and OUER are selected in the top model to explain the 1-year cumulative default risk due to vintage effects in the reference PD data.

5.3.2 LGD Models

For capital modeling, LGD is more accurately the economic loss given default or ELGD, which counts the net present values (NPV) of all recoveries and losses at default as well as the accrued missed interest payments as a percentage of the exposure at default. The actual ELGD for default loans in capital modeling can be calculated according to the formula (3.57) and the components descried in Sect. 3.1.4.1.

In capital modeling data processing, the actual ELGD calculation is a major step and requires high-level automation and maintenance. First, the reference default data for all assets of different products in an institution are required to be collected and processed in compliance of regulatory guidance. Such data are also required to be submitted and examined periodically in the form of various regulatory reports. So, over the years, financial institutions have built up dedicated information technology infrastructures for loss data collection, processing, and analytics and produced standard loss data models, which include some account characteristics at origination, all information about the account default (default date, exposure – EAD, default reason…), and loss and recovery transactions (accrued interests, costs and expenses, write-downs, recoveries, collateral disposals, collections…). As defined in Sect. 3.1.4.1, the economic loss is calculated as the sum of net present values of all these losses, deficiency of recoveries and collections, and missed interest payments prior default. These components need to be recorded correctly and updated promptly in the loss database. Ideally, the actual ELGD is automatically calculated and updated in this loss database with periodic validation and reconciliation.

Compared to PD reference data, the LGD reference default data traditionally have lower quality due to complex post-default data collection processes and high cost for

Table 5.4 Top five PD models selected for capital modeling

Model	SIC	AIC	CV11	Rank SIC	Rank AIC	Rank CV	AveRank
CDF_1Y ~ aci + oltv + DHPI + DUER + OHPI + OUER	295252.93	295180.05	0.308488	1	1	1	1.00
CDF_1Y ~ aci + oltv + DHPI + DUER + LoanSize + OHPI + OUER	295264.92	295181.64	0.308529	2	2	3	2.33
CDF_1Y ~ aci + oltv + DHPI + DUER + OUER	295343.36	295280.89	0.308643	3	4	5	4.00
CDF_1Y ~ aci + oltv + DHPI + DUER + OHPI	295350.39	295287.92	0.308577	4	5	4	4.33
CDF_1Y ~ aci + oltv + DHPI + DUER + LoanSize + OUER	295352.22	295279.34	0.308500	5	3	2	3.33

```
Top 1 Model: CDF_1Y ~ aci + oltv + DHPI + DUER + OHPI + OUER

Call:
glm(formula = frm_df, family = binomial(link = "logit"), data = RefData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8886  -0.8799  -0.7419   1.3201   2.3770

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.338e+00  3.713e-02  -36.03   <2e-16 ***
aci         -1.719e-03  2.902e-05  -59.23   <2e-16 ***
oltv         1.568e-02  2.973e-04   52.73   <2e-16 ***
DHPI        -1.541e-02  3.177e-04  -48.49   <2e-16 ***
DUER         1.292e-01  5.857e-03   22.07   <2e-16 ***
OHPI         8.803e-04  8.685e-05   10.14   <2e-16 ***
OUER        -3.431e-02  3.293e-03  -10.42   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 305945  on 245422  degrees of freedom
Residual deviance: 295166  on 245416  degrees of freedom
AIC: 295180

Number of Fisher Scoring iterations: 4

[1] "11 Fold Cross Validation"
[1] 0.3084878 0.3085145
```

Fig. 5.5 Top PD model selected for capital modeling

large-scale default data review and validation. Over the years with regulatory emphasis on data quality for capital modeling and enhancement on capital data technology infrastructures, reference default data quality has improved across the financial industry. Nevertheless, abnormal data are still commonly seen in the LGD reference default data, and such abnormality should be investigated and treated appropriately. Such data abnormality also should not be the motivation to seek complex LGD models to fit them.

For LGD modeling, as described in Sect. 3.1.4, there are three types of LGD models popularly used in credit risk modeling – the micro-structure dynamic models, regression models, and multiple resolutions models. Regression models are preferred for capital LGD modeling due to their capability of direct linking the ELGD and macroeconomic factors and less irregularity of the actual ELGD in the reference default data. The direct link between ELGD and macroeconomic factors simplifies the implementation of the through-the-cycle (TTC) concept for regulatory IRB LGD modeling, in which the model outputs are used for ranking individual loans based on their LGD risk profiles and the averages of actual ELGD are used for the final LGD parameter estimates.

Among the different types of regression models, ELGD is commonly modeled according to its observed shape. When a large portion of ELGDs are observed as extremes, which means either negative or far away from the full loss rate of 1, Tobit and inflated Beta (IBeta) models are preferred to catch these extremes. Very often, due to the pooling of large number of default loans in the default reference data for capital modeling, the observed ELGD is less irregular with small portion of extremes. These extremes are treated as outliers and are further investigated with their loss components. For example, some extreme large actual ELGD values are due to the small EAD values, which could be an accounting issue or data collection error. To handle these outliers, one can either impute them with the population average or truncate them at a fixed value. The actual ELGD after such outlier treatment looks much more regular, and a liner regression model is sufficient. The linear regression model is preferred due to its simplicity and transparency in relationships between the target ELGD and macroeconomic factors. For our subprime mortgage portfolio, we will show how we process the default reference data and build linear regression ELGD models.

Data Preparation

Corresponding to the PD reference data representing our subprime mortgages, we have the full historical loss data for defaults among these mortgages, which represent about 70% of our entire subprime mortgages. We take this full set of defaults as our reference default data for ELGD modeling. Here, we emphasize on the consistency between the PD reference data and the reference default data for LGD (ELGD). Both reference data should cover the target asset or product sufficiently. In addition, as we discussed in Sect. 5.2, both reference data should cover a full economic cycle as well as a downturn period for regulatory capital models.

Like the PD reference data, loans in the reference default data mostly originated between 2000 and 2008 and defaulted starting from 2002 till 2011. The reference default data include a rich set of risk factors measured at the default time. Table 5.5 presents a summary for a small portion of these risk factors.

The loan ID indicates there are 48,469 defaults in the reference default data (70% of 69,241 total loans on book). LoanSize and oltv are measured at the loan origination. The two macroeconomic variables chpi and cuer represent the changes of home price index and unemployment rate at default time since origination is measured as the ratio of current value against origination value. The rest three risk factors in Table 5.5, lien, OccType, and PropType, measure the number of liens, occupation type (Owner or NotOwner), and property type (planned unit development or condo, multifamily units, single family residential, and other).

Some other risk factors in the reference default data not listed in Table 5.5 are PMI_FLAG (mortgage insurance indicator), age (loan age at default), MOD (loan modification indicator), and TERM (loan term). It should be pointed out that some risk factors in the reference default data can't be used in the LGD model as predictor. For example, the default reason is a highly explainable factor with LGD; however, we do not observe this factor at the reporting time. For credit product with collateral

Table 5.5 Summary information for the subprime mortgage default data

ID	LoanSize	oltv	chpi	cuer	lien	LGD_Actual	OccType	PropType
Length: 48469	Min.: 4568	Min.: 1.528	Min.: 0.3515	Min.: 0.526	Min.: 1.000	Min.: 0.0000	Owner: 41146	SFR: 35030
Class: character	1st Qu.: 165600	1st Qu.: 70.851	1st Qu.: 0.6106	1st Qu.: 1.628	1st Qu.: 1.000	1st Qu.: 0.3994	Not owner: 7323	PUD: 8501
Mode: character	Median: 261000	Median: 79.449	Median: 0.7243	Median: 2.159	Median: 1.000	Median: 0.5624	NA	MULT: 4938
NA	Mean: 286797	Mean: 79.905	Mean: 0.7544	Mean: 2.115	Mean: 1.082	Mean: 0.5682	NA	NA
NA	3rd Qu.: 377600	3rd Qu.: 80.000	3rd Qu.: 0.8708	3rd Qu.: 2.536	3rd Qu.: 1.000	3rd Qu.: 0.7251	NA	NA
NA	Max.: 3920000	Max.: 460.309	Max.: 2.9624	Max.: 5.175	Max.: 2.000	Max.: 1.2500	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	NA

like mortgage, the resolution types are also highly explainable; however, we also do not observe this factor at the reporting time. The multiple resolution LGD model is set up to predict the resolution type based on the observed factors at reporting time as we presented in Sect. 4.3.2 for loss forecasting.

LGD_Actual in Table 5.5 is the actual ELGD calculated for each default loan. One can see its range is from 0 to 1.25. This is due to truncation of calculated values below 0 and above 1.25. The truncation happens in two cases with less than 1% of the total observations. The first case is with outliers we discussed earlier. Most of these outliers are due to recording errors with small balances at default and excess losses, which are results of accounting issues. The second case is with small negative ELGD values observed around 0. This is also the results of accounting issues with excess recoveries.

There are several approaches to deal with these cases. One approach is digging into these cases loan by loan to correct the errors in each of the data processing stages. Though this approach is straightforward, but backward data cleaning and correction are not efficient or even may not be feasible. So, most often a statistical approach is used. In statistics, extremes can be identified as outliers and handled appropriately. The simplest statistical approach is removing the identified outliers from the population. This can only be done when outliers are a very small portion of the population. Alternatively, we can treat these outliers as missing values and using statistical imputing to replace these outliers. Truncation is one way of imputing with fixed values. Lastly, robust models can be used to reduce the impact of these outliers to model outputs.

For our mortgage data, we take the approach of statistical imputing. For the first case, besides the truncation treatment, which replaces the outliers with a fixed value, there are other approaches (e.g., replace the outliers with population mean). For the second cases, the common treatment is truncating these small negative values at 0.

Model Selection and Ranking

As we discussed early in the section, for capital modeling, regression models are preferred. Among the various regression models, linear regression models are appropriate if the actual ELGD observations are distributed normally without significant truncation aggregation or asymmetric shape. Figure 5.6 presents the histogram of the actual ELGD from the reference default data. The empirical distribution shows that the actual ELGD largely follow a normal distribution within the truncated range [0, 1.25], so a linear regression should be sufficient.

With a linear regression, we would select proper risk factors in the final model. Variable selection for linear regression model can follow the Smart AEVS procedure we presented in Sect. 5.3.1 for capital PD modeling with GLM models. First, we split the pool of candidate risk factors into two pools – the static pool and dynamic pool. The static pool consists of the following risk factors:

- oltv, loan to value ratio at origination
- chpi, ratio of current HPI and HPI at origination
- LoanSize, loan amount at origination
- lien, number of liens
- PMI-FLAG, primary mortgage insurance indicator

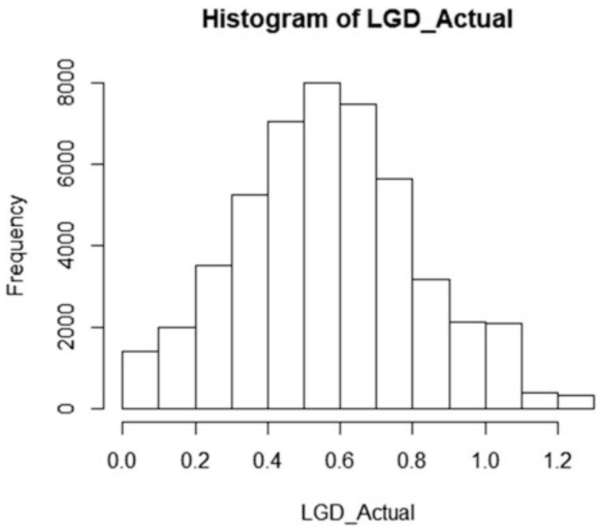


Fig. 5.6 Actual ELGD distribution with reference default data

And the dynamic pool consists of the following risk factors:

- Age, loan age at default time
- Cuer, ratio of current unemployment rate and unemployment at origination
- Term, loan term (30 years, 20 years, 15 years)
- Balloon, balloon loan indicator
- twoborr, two-borrower indicator
- OccType, occupation types, Owner or NotOwner
- PropType, property types (PUD, MULT, SFR)
- evermod, loan modification indicator

As a general rule, the static pool includes risk factors from experiences of modelers or line of business. In our subprime mortgage ELGD modeling, we have five risk factors in the static pool. The first four of these risk factors – oltv, chpi, LoanSize, and lien – are according to previous ELGD models for mortgage products, while the PMI flag is according to line of business.

The dynamic pool incudes eight risk factors that are potential candidates for the final selected model, and the full combinations of these risk factors will be searched and selected. There are $2^8 = 256$ combinations. We modify the Smart AEVS procedure by replacing some ranking criteria. We take AIC as the leading ranking criterion and downgrade SIC as the second one. We make this switch due the fact that for large data SIC could over-penalize per dimension and select less sufficient models. We also replace the CV criterion by the adjusted R-square due to expensive computing for CV with large data set.

Table 5.6 presents the top three models.

The ranks by AIC and adjusted R-square are the same, while SIC ranks the smaller model as better. For larger data, SIC tends to over-penalize per dimension

Table 5.6 Top three ELGD models selected for capital modeling

Model	AIC	SIC	Adj R2	Rank AIC	Rank SIC	Rank Adj-R2	AveRank
LGD_Actual ~ LoanSize + oltv + chi + cuer + lien + PMI_FLAG + OccType + PropType + twoborr	1468.981	1574.446	0.5190963	1	2	1	1.33
LGD_Actual ~ LoanSize + oltv + chi + cuer + lien + PMI_FLAG + OccType + PropType + twoborr + evermod	1470.953	1585.206	0.5190957	2	3	2	2.33
LGD_Actual ~ LoanSize + oltv + chi + cuer + lien + PMI_FLAG + PropType + twoborr	1471.869	1568.545	0.5174795	3	1	3	2.33

Top 1 Model
LGD_Actual ~ LoanSize + oltv + chpi + cuer + lien + PMI_FLAG + OccType + PropType + twoborr

```
Call:
lm(formula = frm_reg1, data = MData)

Residuals:
     Min       1Q   Median       3Q      Max
-1.11319 -0.10502  0.00119  0.10548  1.62001

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.060e-01  8.753e-03   46.38  < 2e-16 ***
LoanSize     -3.862e-07  4.824e-09  -80.06  < 2e-16 ***
oltv          2.127e-03  4.201e-05   50.62  < 2e-16 ***
chpi         -4.527e-01  4.813e-03  -94.05  < 2e-16 ***
cuer          4.197e-02  1.400e-03   29.97  < 2e-16 ***
lien          3.886e-01  3.586e-03  108.36  < 2e-16 ***
PMI_FLAG     -7.147e-02  4.363e-03  -16.38  < 2e-16 ***
OccTypeOwner -2.937e-02  2.294e-03  -12.80  < 2e-16 ***
PropTypePUD  -7.085e-02  3.149e-03  -22.50  < 2e-16 ***
PropTypeSFR  -2.164e-02  2.678e-03   -8.08 6.63e-16 ***
twoborr      -2.600e-02  1.709e-03  -15.22  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1727 on 48458 degrees of freedom
Multiple R-squared:  0.5192,    Adjusted R-squared:  0.5191
F-statistic:  5233 on 10 and 48458 DF,  p-value: < 2.2e-16
```

Fig. 5.7 Top ELGD model selected for capital modeling

and requires a risk factor reducing the residual deviance large enough to be allowed into the model. In our case, we further compared the first and third model in Table 5.6 and decide that the Occupation Type (OccType) is a significant risk factor. So, the first model is considered preferred to the third model. We also compare the first two models and find that the evermod indicator is not a balanced risk factor since our reference default data has less than 8% modified loans. So, we consider the effect due to this factor is not sufficient. The results for our top model are presented in Fig. 5.7.

5.4 Model Integration and Capital Calculation

As we discussed in Sect. 1.3, the usage of credit models for capital management is different from that for loss forecasting. For regulatory capital, under the internal rating-based (IRB) approach, credit models are used to rate and segment loans into grids such that loans within each grid are considered homogenous in credit risk profiles. The ratings are the outputs of credit models by applying the through-the-cycle (TTC) concept on economic conditions. For economic capital, credit models are used in the simulation of the loss generating process to obtain the full loss distribution by applying the point-in-time (PIT) concept on economic conditions.

In this section, we will integrate the models developed in Sect. 5.3 to show how these models are used in the regulatory capital RWA calculation and economic capital loss generating process simulation for the full loss distribution. Using our subprime mortgage portfolio, we will first demonstrate how to use the PD and LGD models developed in Sect. 5.3 to rate the loans in the corresponding reference data sets and segment them into proper grids. Then we will show how to map the loans at reporting time into these grids and obtain the grid-level risk parameters. Taking these parameters as inputs, we can obtain regulatory RWA from the formula (1.10).

One can see that the regulatory RWA heavily depends on the risk grids of the reference data. The risk grids depend on two important inputs. One is the reference data and the other is the credit model (PD or LGD). The reference data for both PD and LGD are subject to change due to new performance observations and default data added into these reference data sets, which will impact the credit models developed on these reference data as well as the final grids created based on these reference data. So, we need to monitor the stability of the created grids over the time for a stable regulatory capital process. Also, the procedure for creating the grids based on the ratings also needs to be appropriate for a stable regulatory capital process. We will show how to monitor each of these processes using our subprime mortgage portfolio.

For economic capital, we first simulate the conditional loss distribution based on the given macroeconomic condition at a reporting time for the snapshot portfolio of loans at that reporting time. As described by the EC simulation process in Fig. 5.2, the simulation is carried out by first obtaining PIT PD and LGD for each loan in the portfolio, and then risk events and LGD are generated in random for the aggregation of portfolio losses. A conditional distribution is obtained through a large number of simulations. In addition to the conditional loss distribution, we also obtain the marginal loss distribution by simulating macroeconomic factors from a joint distribution fitted through a t-copula.

5.4.1 Internal Risk Rating and Risk Grids

For regulatory capital, IRB requires rating loans in the reference data and grouping them into homogenous grids in respect to their risk profiles. While there are no explicit regulations on what rating method should be used, the common practice is using model-based estimates of risk parameters. For the PD reference data, model-based 1-year cumulative default probability estimates are used for the rating, while for LGD reference default data, the LGD estimates from the LGD model are used. For both of these risk parameters, the TTC measures are preferred to reduce the procyclical effect on capital management. Using the subprime mortgage portfolio, we will demonstrate how to generate these ratings and use them for the grids.

Table 5.7 Summary of PIT and TTC PDs by Top 1 model for PD reference data

ID	ActDate	PIT	TTC	CDF_1Y
Length: 245423	Min.: 2000-01-01	Min.: 0.04696	Min.: 0.04696	Min.: 0.0000
Class: character	1st Qu.: 2006-04-01	1st Qu.: 0.24506	1st Qu.: 0.24960	1st Qu.: 0.0000
Mode: character	Median: 2007-04-01	Median: 0.30222	Median: 0.30741	Median: 0.0000
NA	Mean: 2007-02-09	Mean: 0.31533	Mean: 0.31460	Mean: 0.3153
NA	3rd Qu.: 2008-01-01	3rd Qu.: 0.36738	3rd Qu.: 0.36992	3rd Qu.: 1.0000
NA	Max.:2013-10-01	Max.: 0.98628	Max.: 0.98628	Max.: 1.0000

Fig. 5.8 Tree splits by TTC PD

```
> rpart.rules(tree_fit1)
 CDF_1Y
    0.23 when TTC <  0.23
    0.29 when TTC is 0.42 to 0.59
    0.29 when TTC is 0.23 to 0.27
    0.36 when TTC is 0.27 to 0.42
    0.53 when TTC >=         0.59
```

PD Ratings and Grid

For the PD reference data, we apply the Top 1 model shown in Fig. 5.5 to estimate the 1-year cumulative PD. The Top 1 model has two macroeconomic factors DHPI and DUER, which measure cumulative changes of HPI and UER since origination. To apply the TTC concept, we should remove the effects of these two risk factors and only use the rest four risk factors to estimate the 1-year cumulative PD.

It should be pointed out that "removing the effects" does not simply mean removing these variables from the model. Instead, long-term averages of these MEVs should be plugged into the model. For our reference data, we use the averages of observed cumulative changes of HPI and UER for each loan when computing the TTC PD. Since most of the loans were originated before 2008, our reference data guarantee at least one economic cycle (boom-peak-down-recover). Table 5.7 presents the summary statistics for the PIT PD, TTC PD, and the actual observed PD based on the reference PD data. The PIT PD and TTC PD are very close.

Once we have the TTC PD, we use it to rate the loan snapshots in the PD reference data. A simple rating method is a binary regression tree, which can conveniently control the size of buckets by specifying the minimum number of elements in each bucket combined with the complexity parameter as described in Sect. 3.1.1.3.

Figure 5.8 presents a regression tree model with five leaves for the PD grids. One can see that the tree has crossovers between the TTC PD used as the splitter and the observed PDs (CDF_1Y) in each bucket. To achieve the monotonicity, we have to merge the middle tree buckets. More granular tree as shown in Fig. 5.9 does not make the monotonicity better, so we use three buckets for our final selected PD grid.

Fig. 5.9 Granular tree splits by TTC PD

```
> rpart.rules(tree_fit2)
 CDF_1Y
   0.23 when TTC <  0.23
   0.29 when TTC is 0.44 to 0.65
   0.29 when TTC is 0.23 to 0.27
   0.30 when TTC is 0.33 to 0.33
   0.32 when TTC is 0.35 to 0.35
   0.32 when TTC is 0.42 to 0.44
   0.34 when TTC is 0.27 to 0.32
   0.36 when TTC is 0.38 to 0.42
   0.36 when TTC is 0.33 to 0.34
   0.36 when TTC is 0.35 to 0.38
   0.37 when TTC is 0.34 to 0.34
   0.38 when TTC is 0.34 to 0.35
   0.40 when TTC is 0.32 to 0.33
   0.44 when TTC is 0.35 to 0.35
   0.48 when TTC is 0.38 to 0.38
   0.51 when TTC is 0.34 to 0.34
   0.54 when TTC is 0.34 to 0.34
   0.72 when TTC >=         0.65
```

Fig. 5.10 Tree splits by PIT PD

```
> rpart.rules(tree_fit3)
 CDF_1Y
   0.25 when PIT <  0.33
   0.38 when PIT is 0.33 to 0.46
   0.47 when PIT is 0.46 to 0.48
   0.48 when PIT is 0.53 to 0.54
   0.48 when PIT is 0.51 to 0.52
   0.49 when PIT is 0.48 to 0.50
   0.53 when PIT is 0.50 to 0.51
   0.54 when PIT is 0.52 to 0.53
   0.55 when PIT is 0.54 to 0.59
   0.66 when PIT >=         0.59
```

For comparison purpose, we also try to use the PIT PD as the tree splitter. Figure 5.10 shows the regression tree by using PIT PD, which also has crossovers, and merging the middle buckets will produce a simpler grid with monotonicity. This indicates that for our reference PD data, majority of the loan-snapshot observations cluster into a single bucket and have similar risk profiles.

LGD Ratings and Grid

For the LGD reference data, similarly we apply the Top 1 model shown in Fig. 5.7 to estimate the LGD given information at the default time. The Top 1 model has two

Table 5.8 Summary of PIT and TTC LGDs by Top 1 model for LGD reference data

ID	PIT	TTC	LGD_Actual
Length: 48469	Min.: −1.0450	Min.: −0.9664	Min.: 0.0000
Class: character	1st Qu.: 0.4596	1st Qu.: 0.4844	1st Qu.: 0.3994
Mode: character	Median: 0.5515	Median: 0.5474	Median: 0.5624
NA	Mean: 0.5682	Mean: 0.5682	Mean: 0.5682
NA	3rd Qu.: 0.6450	3rd Qu.: 0.6068	3rd Qu.: 0.7251
NA	Max.: 1.2730	Max.: 1.4415	Max.: 1.2500

Fig. 5.11 Tree splits by TTC LGD

```
> rpart.rules(tree_fit1)
 LGD_Actual
        0.37 when TTC <  0.45
        0.48 when TTC is 0.45 to 0.53
        0.56 when TTC is 0.53 to 0.59
        0.64 when TTC is 0.59 to 0.83
        0.99 when TTC >=         0.83
```

Fig. 5.12 Granular tree splits by TTC LGD

```
> rpart.rules(tree_fit2)
 LGD_Actual
        0.31 when TTC <  0.37
        0.40 when TTC is 0.37 to 0.45
        0.44 when TTC is 0.45 to 0.47
        0.49 when TTC is 0.47 to 0.53
        0.54 when TTC is 0.53 to 0.55
        0.58 when TTC is 0.55 to 0.59
        0.62 when TTC is 0.59 to 0.61
        0.66 when TTC is 0.61 to 0.83
        0.88 when TTC is 0.83 to 0.93
        1.01 when TTC >=         0.93
```

macroeconomic factors chpi and cuer, which measure ratios of home price index (hpi) and unemployment rate (uer) between default time and origination time, respectively. To apply the TTC concept, we use the averages of observed chpi and cuer for all default loans with the same origination date.

Table 5.8 presents the summary statistics of the PIT LGD, TTC LGD, and the actual LGD for the LGD reference data. The PIT LGD and TTC LGD are close.

Using the TTC LGD, we fit two regression trees. One is simpler with five buckets and the other is more granular with ten buckets. As shown in Figs. 5.11 and 5.12, both trees have monotonicity with respective to the splitter – TTC LGD. A further looking into the more granular tree found that it is just a further split of the buckets of the first tree as the minimum number of elements in the final buckets was set to a smaller number (1000 vs. 3000). For simplicity, we select the simpler tree with five buckets as the final LGD grid.

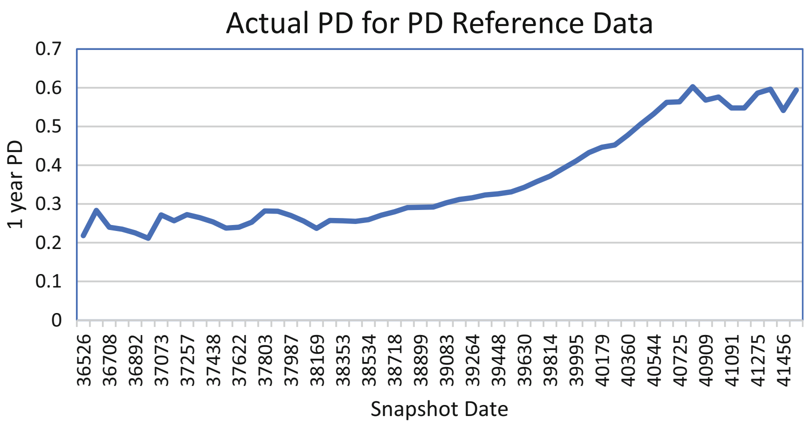


Fig. 5.13 Actual PD time series for reference data

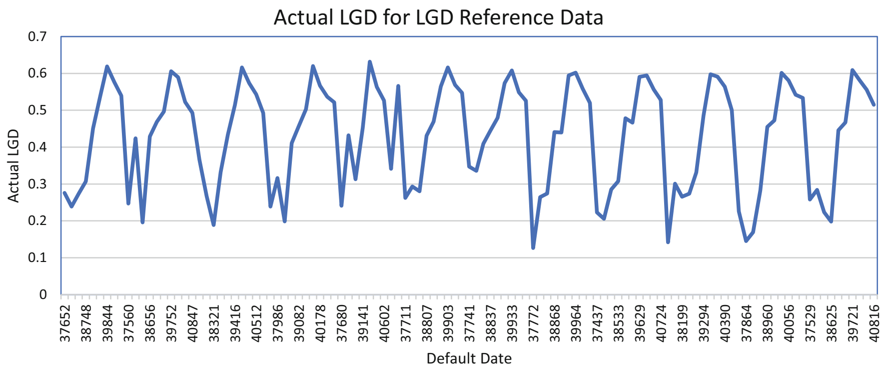


Fig. 5.14 Actual LGD time series for reference data

5.4.2 Regulatory Capital and Risk Weighted Asset Calculation

Once we have the PD and LGD grids decided, we can calculate the average actual PD and LGD within the buckets of the grids, and we call them Grid PD and Grid LGD. The LGD_Actual numbers in Fig. 5.11 show an example of the Grid LGD for the final LGD grid we selected.

For the regulatory capital, as described in Sect. 5.2.2, Basel II requires a downturn period identification and the corresponding downturn LGD calculation. For our subprime mortgage portfolio, although the 1-year cumulative PD peaks around 2011Q3 as shown in Fig. 5.13, it turns out that there is a large amount of loan modifications since 2010Q2. When excluding these loan modifications, the 1-year cumulative PD peaks around 2009Q2, and we select the period of 2008Q2–2010Q1 as the final downturn period. This period matches the period of high actual LGD of the portfolio as shown in Fig. 5.14.

Table 5.9 Actual LGD and downturn LGD by grid

Grid	Grid_lgd	DT_lgd
1	0.3710	0.4049
2	0.4825	0.5011
3	0.5697	0.5903
4	0.6467	0.6733
5	0.9886	1.0049

Table 5.10 2011Q1 portfolio summary

ID	LoanSize	oltv
Length: 585	Min.: 9800	Min.: 34.67
Class: character	1st Qu.: 181242	1st Qu.: 70.00
Mode: character	Median: 276000	Median: 79.97
NA	Mean: 319428	Mean: 80.10
NA	3rd Qu.: 417000	3rd Qu.: 80.00
NA	Max.: 1987500	Max.: 188.65

After the downturn period is identified, the downturn LGD is calculated as the average of actual LGD for the identified default loans defaulted during the downturn period in each bucket of the LGD grid. Table 5.9 shows the calculated downturn LGD as well as the original grid LGD for the selected final LGD grid. The downturn LGD is usually larger than the original Grid LGD as the identified downturn period is under more stressed economic condition.

To show the RWA calculation, we select the 2011Q1 snapshot as the portfolio. Table 5.10 presents the summary statistics for this portfolio. There are 585 loans in this portfolio, and the median loan size is $276,000, and the median origination lean to value (LTV) ratio is about 80%.

First, we calculate the TTC PD for these loans and then map each loan into a bucket of our PD grid based on the calculated TTC PD. Then we calculate the TTC LGD for each of these loans and map it into a bucket of LGD grid. Note that when calculating TTC PD and TTC LGD, we use the long-term MEV averages as we did for calculating the TTC PD and TTC LGD with the reference data. Once we have the mapped PD and LGD buckets for each loan, we have the corresponding Grid PD and Grid LGD (and downturn LGD) for each of these loans, and then we can plug these numbers into the RWA formulas:

$$\begin{aligned} \mathrm{RWA} = 12.5 * \sum_{i=1}^{N} \mathrm{LoanSize}(i) * \mathrm{LGD}_{\mathrm{Grid}}(i) \\ * \left(\Phi\left(\frac{\Phi^{-1}(\mathrm{PD}_{\mathrm{Grid}}(i)) + \sqrt{\rho}\,\Phi^{-1}(0.999)}{\sqrt{1-\rho}} \right) - \mathrm{PD}_{\mathrm{Grid}}(i) \right) \end{aligned} \tag{5.1}$$

which is a simplified version of (1.10) without the adjustments. The regulatory capital is just without the multiplier 12.5.

Table 5.11 presents the calculated RWAs for the 2011Q1 portfolio as well as the total loan balance. We calculated two versions of RWA: one is with the Grid LGD,

Table 5.11 2011Q1 portfolio RWA

RWA ($)	DT_RWA ($)	Balance ($)	RWA/balance	DT_RWA/balance
548,647,169	575,413,549	186,865,250	294%	308%

GRID	2007 Q1	2007 Q2	2007 Q3	2007 Q4	2008 Q1	2008 Q2	2008 Q3	2008 Q4	2009 Q1	2009 Q2	2009 Q3	2009 Q4	2010 Q1	2010 Q2	2010 Q3	2010 Q4
1	15%	15%	14%	13%	13%	13%	12%	9%	8%	6%	5%	4%	3%	3%	2%	2%
2	85%	85%	85%	86%	86%	87%	88%	90%	92%	93%	94%	95%	95%	96%	96%	96%
3	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	2%

Fig. 5.15 PD grid migration

and one is with the downturn LGD as shown in Table 5.9. As expected, the latter is relatively larger. We also calculated the ratio of the RWAs to the total loan balance. For our subprime mortgage portfolio, the RWA/balance ratios are about 300%, which is much larger than a regular mortgage portfolio (about 50%), though we use the same correlation parameter $\rho = 0.15$ as commonly used for regular mortgage portfolios. The relatively large RWAs for our subprime mortgage portfolio are due to the high default probability (Grid PD) compared to regular mortgage portfolio (less than 10% even around the 2008 housing crisis).

5.4.3 Risk Grid Migration and Monitoring

Through the RWA calculation, we can see that the PD and LGD grids play an important role in the final calculated RWA as the averages of actual PD and LGD within the buckets of these grids are the final PD and LGD parameters used in the RWA formula. With the Basel Internal Risk Based (IRB) approach, there is no fixed method specified for the grid creation. The method we presented earlier using the TTC PD and TTC LGD to create the grids follows the IRB guidance that grids should be created to reflect the classification of risk profile of the loans within the portfolio.

While using TTC PD and TTC LGD to create grids is a popular practice in RWA calculation, there are alternative methods for grid creation. When the RWA calculation covers a large number of portfolios, for example, the entire retail or wholesale loan population within large banks, carrying out grid creation by computing the TTC PD and TTC LGD for each portfolio is overwhelming given that the final grids for the portfolios are usually manually decided. So, for the purpose of efficiency, for large population RWA calculation, grids could be decided based on some loan characteristics – such as product type, current loan to value ratio, region, credit score, or rating.

Given the flexibility of grid creation, it is important to make sure the grids created are stable over time to avoid large variation of the calculated RWA. To monitor the grid stability, we calculate the bucket population percentage at each snapshot in the reference data. More specifically, for the PD reference data, we calculate the percentage of loans falling in each bucket of the PD grid at a snapshot time. Figure 5.15 presents the PD grid migration between snapshots 2007Q1 and

GRID	2008 Q1	2008 Q2	2008 Q3	2008 Q4	2009 Q1	2009 Q2	2009 Q3	2009 Q4	2010 Q1	2010 Q2	2010 Q3	2010 Q4	2011 Q1	2011 Q2	2011 Q3	2011 Q4
1	16%	15%	14%	14%	14%	16%	16%	14%	13%	12%	12%	14%	14%	12%	12%	14%
2	30%	30%	29%	29%	27%	26%	26%	25%	24%	24%	24%	23%	22%	24%	23%	24%
3	28%	28%	30%	29%	28%	27%	27%	26%	25%	26%	24%	26%	24%	23%	28%	27%
4	19%	19%	19%	20%	21%	21%	21%	25%	29%	30%	31%	30%	31%	31%	28%	27%
5	6%	7%	9%	8%	10%	11%	10%	9%	8%	8%	9%	7%	9%	10%	9%	9%

Fig. 5.16 LGD grid migration

2010Q4. We can see that the percentage in the low PD bucket is decreasing when passing the 2008 financial crisis and become relatively stable from 2010Q1. We consider the PD grid stable given that the loan proportions among the buckets do not change too much from 2010Q1, which indicate the population sizes within the different classified risk profile buckets are becoming relatively stable.

Figure 5.16 presents the LGD grid migration status with default date between 2008Q1 and 2011Q4. Population percentages are relatively stable during this period, with the second bucket population showing a slow decreasing and the fourth bucket population showing a slow increasing. We consider such migration magnitude is reasonable as the loan loss severity tends to increase past the financial crisis and become stable since 2011Q1.

When large deviations are found between the consecutive periods, especially the most recently periods, an alert should be issued in the grid stability monitoring, and further investigation on either model or data should be carried out.

5.4.4 Economic Capital Simulation

As described in Sect. 5.1.2, different from the regulatory capital and RWA calculation, economic capital is commonly computed through Monte Carlo (MC) simulation of the loss generating process (1.17). The goal is to track and measure the uncertainty of the generated losses, especially due to the macroeconomic conditions. If the macroeconomic condition is given, then under this condition, the simulation tracks the loss generating process together with other risk factors, and we can obtain the conditional economic capital. When the macroeconomic condition is given under some stress scenarios, the corresponding economic capital obtained can be used for specific purposes of stress testing. We can also consider the uncertainty of the macroeconomic factors using a joint statistical distribution, such as a t-Copula, to obtain the marginal economic capital.

In the following, using our subprime mortgage portfolio at a snapshot (2011Q1), we show how we simulate the loss distribution and calculate economic capital as the unexpected loss. Table 5.12 presents a summary of the snapshot portfolio, which is the same as that used for the regulatory capital calculation. Both PIT PD and PIT LGD are included here as these parameters are used in the simulation as descried in Fig. 5.2.

Conditional Economic Capital

For the conditional economic capital, we take the macroeconomic factors at the snapshot time as given; thus the risk parameters PIT PD and PIT LGD calculated at

Table 5.12 2011Q1 portfolio summary

ID	PIT_pd	PIT_lgd	LoanSize
Length: 585	Min.:0.1802	Min.: −0.07093	Min.: 9800
Class: character	1st Qu.: 0.4252	1st Qu.: 0.48045	1st Qu.: 181242
Mode: character	Median: 0.5089	Median: 0.56152	Median: 276000
NA	Mean: 0.5070	Mean: 0.58147	Mean: 319428
NA	3rd Qu.: 0.5919	3rd Qu.: 0.65346	3rd Qu.: 417000
NA	Max.: 0.8684	Max.: 1.16868	Max.: 1987500

Table 5.13 2011Q1 portfolio ECs with different LGD methods

	Loss distribution	
LGD method	99.9% VaR	99.97% VaR
Norm	8,166,486	9,019,035
LN	11,261,853	12,674,600
EDF	9,693,213	10,822,469

the snapshot time are used in the MC simulation to generate the loss for the i-th loan in the portfolio:

$$L_i = \mathrm{EAD}_i * \mathrm{LGD}_i * I\ (D_i = 1) \tag{5.2}$$

where $I\,(D_i = 1)$ is the random default event of a binary distribution with the PIT PD as its mean parameter. This is excised through a draw of uniform random number and then compared it with the PID PD parameter. The loss event is recorded when the uniform random number is less than the PIT PD parameter. For LGD_i, we assume three options:

- A normal distribution as we assume in the line regression model.
- A lognormal distribution with the PIT LGD as its mean and a standard deviation from the LGD reference data after the logarithm transformation.
- An empirical distribution based on the observed actual LGD in the LGD reference data.

The normal distribution follows the linear regression model we fitted in Sect. 5.3.2. The lognormal distribution corresponds to the FRR method discussed in Sect. 3.1.4.4 with the transformation function extended to the loglinear function. The empirical distribution function (EDF) method corresponds to the resampling method (without repeats). EAD_i is the loan balance at the snapshot.

Once the individual loan loss is calculated, their sum is considered as the portfolio loss. We execute 200,000 simulations, and the corresponding quantiles are the order statistics of the simulated loss distribution – 200th and 60th largest simulated portfolio losses.

Table 5.13 presents the ECs (unexpected losses) with both 99.9% and 99.97% quantile levels for the three LGD methods. The ECs with the lognormal LGD

Table 5.14 2011Q1 portfolio total losses with different LGD methods

LGD method	Loss distribution 99.9% VaR	99.97% VaR
Norm	56,157,725	57,010,273
LN	68,927,851	70,340,598
EDF	64,064,798	65,194,054

```
> summary(tc.ml <- fitCopula(tCopula(dim=2, dispstr="un"), smev, method="ml", start = c(0, 10)))
Call: fitCopula(tCopula(dim = 2, dispstr = "un"), data = smev, ... = pairlist(method = "ml", start = c(0,
    10)))
Fit based on "maximum likelihood" and 585 2-dimensional observations.
t-copula, dim. d = 2
       Estimate Std. Error
rho.1 -0.008719      0.046
df     6.308697      2.044
The maximized loglikelihood is 5.612
Optimization converged
Number of loglikelihood evaluations:
function gradient
      40       12
```

Fig. 5.17 Partial R-code for fitting a copula

distribution are larger than the ECs with normal and EDF LGD distributions. This is expected as the lognormal LGD distribution has heavier tails and may generate wider LGD draws in the simulation.

We also present the VaR numbers for the total losses (no deduction of the expected loss) in Table 5.14. We can see that the total loss VaR numbers are much larger than the ECs, which indicate that the total losses are dominated by the expected losses. We will further examine this phenomime in Sect. 5.5.

Marginal Economic Capital

When the uncertainty of the macroeconomic variables in the PD model (DHPI and DUER) are considered, we can fit a joint distribution of these two factors using the t-Copula. The following R-code in Fig. 5.17 is an example.

The data for the two macroeconomic variables (MEVs) are extracted from the PD reference data, and a two-dimension t-Copula is fitted. The R function fitCopula from the “copula” package estimates both the correlation and the degree of freedom of the copula. The correlation coefficient is negative, which indicates that the two macroeconomic variables have the opposite trend – higher home price appreciation usually happens in lower unemployment environment. The degree of freedom is around 6, which is high and indicates the two MEVs are very close to normal distribution in the tails of their joint distribution.

For simplicity (also being conservative), we take the fitted t-Copula with degree of freedom truncated to 6 in our MC simulation. The process is as follows:

- For each loan, we simulate the two MEVs from the t-Copula and recompute the PIT PD using the simulated MEVs while keeping the other factors unchanged.
- Using the recomputed PIT PD to decide the default event in the simulation.

Table 5.15 2011Q1 portfoliospiepr146 conditional andspiepr146 marginal ECs

VaR method	LGD method	Loss distribution	
		99.9% VaR	99.97% VaR
Conditional	EDF	9,693,213	10,822,469
Marginal	EDF	43,782,975	47,646,833

Table 5.16 2011Q1 portfolio conditional and marginal total losses

VaR method	LGD method	Loss distribution	
		99.9% VaR	99.97% VaR
Conditional	EDF	64,064,798	65,194,054
Marginal	EDF	98,077,236	101,941,094

Table 5.15 presents the conditional EC obtained earlier and the marginal EC with the recomputed PIT PD using simulated MEVs from the t-Copula, both using the EDF LGD method. The marginal ECs are significantly larger for both 99.9% and 99.97% quantile levels. Table 5.16 presents the VaR numbers for the total losses. Again, the marginal VaR numbers are much larger than the conditional VaR numbers.

The difference of these VaR numbers between conditional and marginal loss distributions is due to the consideration of default correlation among loans in the portfolio, though the expected losses do not change much. This is an example of concentration risk due to correlation. As a part of model validation, in Sect. 5.5, we will further analyze these results.

5.4.5 Capital Allocation

One of the common capital management practices is to allocate the total capital in a combined organization back into suborganizations, for example, legal entities or line of businesses. These allocated capital amounts will be used to assess the capital sufficiency and business operation efficiency with respect to their allocated capital.

For regulatory capital, due to the infinity granularity assumed in the Vasicek formulation, capital is additive and linear in exposure. So, allocation can simply be done by summation of corresponding capital in suborganizations. We would not say there is no diversification with regulatory capital, as loss rates could be different among different portfolios of an entity; thus portfolios with lower loss rates could offset portfolios with higher loss rates even with similar exposures. The diversification becomes simpler as the capital is linear in exposures of portfolios.

For economic capital, the linearity of capital in exposure is not guaranteed anymore, and very often the economic capital has a nonlinear relationship with the portfolio exposure – capital often increases in super-linear speed in exposure, and this leads to the surcharge of concentration in exposures. Such surcharges can be measured using the marginal economic capital as defined in Sect. 1.3.2.

In this section, we will focus on the economic capital allocation. We introduced two commonly used capital allocation methods – the Diversification Benefit-Based EC allocation and Contribution Based EC allocation.

Diversification Benefit Based EC Allocation

This method has been introduced in Sect. 1.3.2. The method allocates capital back to portfolios based on the Diversification Benefit (DB) allocation. The popular method to allocate the DB back to each portfolio is based on their marginal economic capital (MEC). Besides the inverse MEC method described in (1.24), one can also allocate the DB based on the following average MEC method:

$$\mathrm{EC}_{\mathrm{Allocated}}(P_i) = \mathrm{EC}(P_i) - \frac{\sum_{j \neq i}^{N} \mathrm{MEC}(P_j)}{(N-1)\sum_{j=1}^{N} \mathrm{MEC}(P_j)} \mathrm{DB} \tag{5.3}$$

Compared to the inverse MEC method of (1.24), the average MEC method (5.3) is more stable under MEC changes. Both these two DB-based allocation methods require extensive computation and may become infeasible for enterprise-wise capital allocation.

Contribution Based EC Allocation

When there is a large number of portfolios in the combined entity (e.g., $N > 1000$), the computation for the marginal EC for all portfolios could be expensive due to a large number of simulations required in the EC calculation. In such case, DB-based allocation method may not be feasible, and some alternative approaches could be used.

One such method is based on the EC contribution (vs. the Diversification Benefit above). For the Contribution-Based EC allocation method, first a portfolio-independent grid is designed to cover all portfolios. As an example, Fig. 5.18 presents a grid for wholesale credit portfolio based on Risk Rating (RR) and Maturity Tenor (MT):

		Risk Rating						
Tenor		1+	1	1-	2+	2	2-	...
	0.50	...	...	...	...	...	...	...
	1.00	...	...	...	...	...	...	...
	1.50	...	...	...	...	...	...	...
	2.00	...	...	...	...	...	...	...
	2.50	...	...	...	...	...	...	...
	3.00	...	...	...	...	...	...	...
	⋮	...	...	...	...	...	...	...

Fig. 5.18 Portfolio-independent segmentation

Each portfolio's exposure is mapped into the cells of this grid. Then each cell is considered as a portfolio, and we can focus on these much smaller numbers of "portfolios." These cell portfolios are first considered as diversified/independent, so the full grid population loss volatility is a weighted summation of individual volatilities:

$$\sigma_P = \sum_{i=1}^{N} w_i \sigma_i \tag{5.4}$$

Assuming EC is proportional to the volatility, then the diversified grid population EC is a weighted summation of individual portfolio EC, which is linear in the exposure E_i:

$$\mathrm{EC}^{\mathrm{Div}}(G) = \sum_{i=1}^{N} K\, w_i\, \mathrm{EC}(L_i) = \sum_{i=1}^{N} K\, [w_i\, \theta_i]\, E_i \tag{5.5}$$

$f_i = w_i\ \theta_i \sim w_i \sigma_i$ is the relative EC contribution from the i-th cell per unit of exposure and is defined as the normal intensity. K is the scalar to make sure grid population EC and the summation of cell ECs equal.

Based on the historical data of a diversified portfolio (taking as a snapshot of the target portfolio after removing large loans/accounts), we can simulate the loss distributions of all these cells and the combined grid portfolio loss distribution, then find the loss volatilities of these cell loss distributions, and calibrate the normal intensity f_i and the scalar K. The normal intensity table will be used for diversified capital allocation.

For EC contributions due to large exposures, we calibrate the Concentration Intensity coefficients based on an exponential EC growth rate with the exposure when the exposure exceeds a threshold:

$$f_j = (A - B)\, \frac{1 - e^{-\mathrm{DE}_j}}{1 + e^{-D\left(E_j - C\right)}} \tag{5.6}$$

A is the maximum EC intensity, and B is the normal/diversified EC intensity (calibrated based on the diversified portfolio); C the reference exposure and D the growth coefficient are parameters calibrated with each risk rating and tensor cell. E_j is the relationship exposure (or cell exposure). The calibration uses the full grid population while adding a series of large exposures into the cell and keeping all other cells unchanged. The marginal ECs corresponding to these series of large added exposures are considered EC surcharges and used for the parameter calibration. Commonly the maximum EC intensity A is subject to risk manager review and adjustments.

Once the parameters are calibrated, the concentration surcharges for all cells are added and scaled to equal to the final total grid population concentration surcharges:

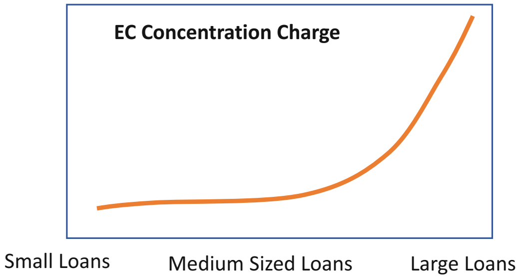


Fig. 5.19 Concentration surcharge

Table 5.17 Risk drivers for wholesale EC allocation

Risk drivers for credit EC allocation	
Obligor type	(C&I corporate, financial institution, sovereign)
Accounting treatment	(AFS, HFI, HFS, . . .)
Currency type	(local or foreign currency)

$$\mathrm{EC}^{\mathrm{Sur}}(G) = \sum\nolimits_{j=1}^{N} Kf_j\, E_j \tag{5.7}$$

After we have the grid normal intensity and concentration intensity matrices calibrated, for any portfolio, according to the exposure mapping, its allocated capital can be calculated as the combination of the diversified capital and the surcharge:

$$\mathrm{EC}(P) = \mathrm{EC}^{\mathrm{Div}}(P) + \mathrm{EC}^{\mathrm{Sur}}(P) \tag{5.8}$$

The contribution-based EC allocation divides the total EC into diversified EC and concentration surcharge and calibrates each part with a corresponding intensity table. This gives a clear separate measure of the concentration risk through the surcharge, which could be useful when concentration risk is a concern for any individual portfolio or aggregated portfolio.

The exponential surcharge (5.6) shown in Fig. 5.19 as a function of the exposure is adopted through empirical analysis based on historical portfolios.

In practice, different intensity tables can be calibrated based on different products and accounting treatments. For example, in wholesale credit, the following risk drivers in Table 5.17 can be considered for the intensity calibration.

5.5 Model Validation and Performance Monitoring

The general model risk management framework has been described in Sect. 4.5, and here we focus on the capital model validation and performance monitoring. Similarly, we will cover the following main components of the model validation procedure: model scope and usage, input and assumptions, theory and design, implementation and output analysis, and performance monitoring.

Capital models are subject to regulatory requirements. Periodically, regulators issue policies and guidance, which could have specific focuses on data, model, processes, and implementation. As a critical step, compliance assessment of the capital models and processes with the most recent regulatory policies and guidance is part of the model validation.

5.5.1 Scope and Usage

Model scope defines the model coverage on products, and model usage defines the applications in which the model can be used.

As discussed in Sect. 1.3, the usage of credit models for capital management is different from that for loss forecasting. For regulatory capital, under the internal rating-based (IRB) approach, credit models are used to rate and segment loans into grids such that loans within each grid are considered homogenous in credit risk profiles. The ratings are the outputs of credit models by applying the through-the-cycle (TTC) concept on economic conditions. For economic capital, credit models are used in the simulation of the loss generating process to obtain the full loss distribution by applying the point-in-time (PIT) concept on economic conditions. As capital management focuses on the 1-year performance measurement window, both regulatory capital and economic capital can share the same suite of credit models developed on the same reference data, which are prepared for both purposes under the consideration of convenience and cost of resources. However, the credit models developed for both regulatory and economic capital uses should meet both purposes. For example, the standard of model performance in model fitting should not be lowered significantly due to the TTC concept for regulatory capital, such that the PIT estimates loss sensitivity to macroeconomic factors in the economic capital application. It is totally acceptable to have two separate suites of credit models for regulatory capital and economic capital uses, respectively. In call cases, model usage should be clearly stated in the beginning of model development.

Credit models are commonly developed by credit products. For example, for retail credit products, credit models are commonly developed for mortgage, credit cards or qualified revolving exposures (QRE), auto loans and dealership financial services, and other retails. Models could be developed in more granular level within each of these products based on different portfolios and segments.

In our examples of the subprime mortgage portfolio, we use the same suite of credit models for both regulatory and economic capital. Although these models are used differently for regulatory and economic capital, we would have some comparison of the two applications based on the same suite of models.

5.5.2 Inputs and Assumptions

For capital modeling, input data are mainly observed. As discussed in Sect. 4.5.2, for observed input data, there could be both random and systematic errors during data collection and processing. For random input data errors, good data quality checks could identify them. Such checks could include outlier detection, hard limits, similarity tests, and machine learning-based data quality checks. Over the last decade, the use of machine learning techniques in data quality control has greatly improved the accuracy and efficiency in data quality checks, especially in automation of data quality control processes, and reduced false-positive rates. Input data systemic errors are hard to discover and could result to higher risks. A commonly used method is comparing with benchmark data. Large deviation from benchmark data will draw attention for investigation of potential systemic errors.

Assessing the impact of input data errors to model output is an essential part of model validation as sensitivity analysis. Under the reality that input data errors can't be avoided completely, sensitivity analysis on input data is a way to assess the range of model output deviation by assuming a certain degree of input data errors. Model output confidence intervals are commonly used to measure the impact of random input data errors under certain distributions.

Capital modeling heavily depends on the input data through the formation of reference data, which are used to represent the portfolio population by risk profile. The requirement of including at least one economic cycle in the reference data, especially for the regulatory capital, commonly leads to large reference data, and proper sampling methods are used to form the final model data. Capital modeling sensitivity to the initial input data, sampling methods, and economic cycle coverage should be assessed as part of the model validation.

For our subprime portfolio, input data errors to the actual LGD calculation are common due the complex post-default processes and manual recording. Over the years, mortgage servicers have been improving their loss recording systems. However, there are abnormal LGD values from the reference loss data, and the simplest way is putting a threshold (e.g., 3) to filter out the larger LGD values to investigate whether this is due to a small exposure at default or some errors. As show in Table 5.5, the largest actual LGD is 1.25, which is within a reasonable limit.

Model assumption makes the model conceptually sound, while specific assumptions on model implementation and model use should all be assessed for potential risks if these assumptions break. Statistics tests are popular tools used for the assessment and measurement of the assumptions. As we discussed in earlier sections of this chapter, simpler linear and generalized linear models (GLM) are commonly used in capital modeling.

For the GLM models we used for risk event modeling, we assume the risk event probability meets the proportional odds ratio assumption with the logit link to the risk drivers. The goodness-of-fit statistics in Sect. 3.1.1.1 are the common measurements of how this assumption is satisfied based on the observed data. So, such statistics must be provided by default as a part of the model validation.

Specific assumptions related to model choice and implementation must be independently and continuously assessed. In our model selection process with AEVS, we used various criteria to measure the final selected model as the best model.

Robust controls and processes must exist to ensure the completeness, timeliness, and integrity of key data inputs. This may include data reconciliations, independent review of manually input data, or other data quality controls. Developers should be able to demonstrate that data inputs are suitable for the model and consistent with model methodology. Any data proxies used must be identified, justified, and documented. All key assumptions underlying the model should be supported by initial and ongoing analysis and documentation, so that users are aware of any model limitations.

5.5.3 Theory and Design

Model theory and design are strongly linked to the concept soundness. Model development should provide sufficient evidence to demonstrate that the proposed models are mathematically and statistically correct before building the model. Model validation should confirm the correctness of model theory and identify any theoretical deficiency. Very often, model failure and model rejection are due to insufficient theoretical support or obvious deficiency in the design of modeling processes. Such model risks are classified as lack of concept soundness and are often hard to be mitigated. The best mitigation approach for such risks is making sure there are no errors in model theory and design before the model development.

Identifying and assessing model errors in theory and design require understanding in depth of the specific modeling areas and practices and sufficient training in model theories. For credit modeling, in Chap. 3, we provide most commonly used models with fundamental theories and sound practical uses. For model validation purposes, the modeling background provided in Chap. 3 is a foundation. In capital modeling, the outputs of those credit models described in Chap. 3 are not the final results. For regulatory capital, these model outputs are taken as inputs in the Vasicek distribution tail value (or VaR) calculation, so the sensitivity of Vasicek tail values to these inputs is important. We will present some sensitivity analysis on the Vasicek distribution tail values in this section.

A more appealing approach for validating model theory and design is comparison with alternative theories and approaches, since such comparison directly shows the advantages and disadvantages of proposed and alternative models. Further, the comparison may demonstrate if the proposed modes work as intended, are appropriate for the intended business purpose, and are conceptually sound and mathematically and statistically correct. Using our subprime reference default data, we show the comparison of regulatory and economic capital calculations and explore some insights when these estimated losses can be benchmarks for each other.

At the end, we present some sensitivity analysis. Regulatory capital is sensitive to the granularity of the grids. By fixing the PD grid, we test the impact of more LGD granularity on regulatory capital based on our subprime mortgage data.

Vasicek Distribution Tail Sensitivity

The Vasicek α-quantile $\Phi\left(\frac{\Phi^{-1}(\mathrm{PD}_{\mathrm{Grid}}(i))+\sqrt{\rho}\,\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right)$ in (5.1) corresponds to the following Vasicek probability density function:

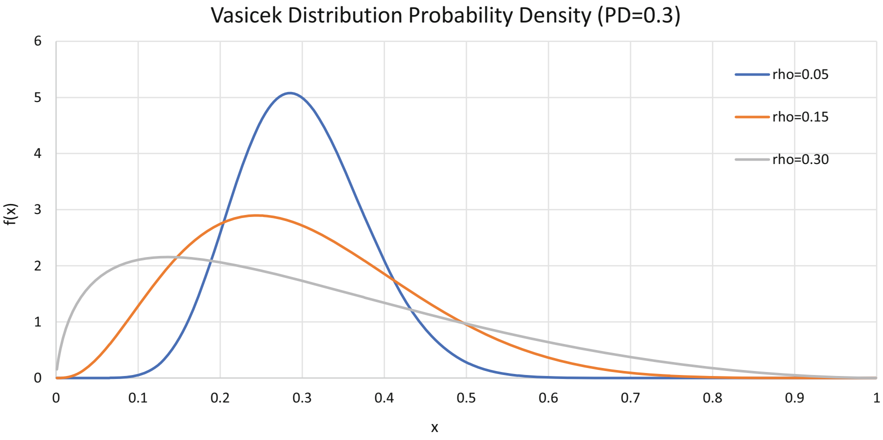


Fig. 5.20 Vasicek probability density

$$f_{p,\rho}(x) = \sqrt{\frac{1-\rho}{\rho}} \exp\left(\frac{1}{2}\left\{\Phi^{-1}(x)^2 - \left(\frac{\sqrt{1-\rho}\,\Phi^{-1}(x) - \Phi^{-1}(p)}{\sqrt{\rho}}\right)^2\right\}\right) \tag{5.9}$$

where p is the grid probability, ρ is the Asset Value Correlation (AVC), and $x \in [0, 1]$. Figure 5.20 presents the density function for three sets of parameters $p = 0.3$ and $\rho = 0.05, 0.15, 0.30$. Vasicek distribution has heavier tails on the right, especially when ρ becomes larger. This indicates when the portfolio has larger correlation among accounts, the portfolio loss has larger VaR.

As the core of the regulatory capital calculation, the Vasicek unexpected loss (UL) rate (per unit of net loss) is the difference between the 99.9% quantile and its mean:

$$K = \Phi\left(\frac{\Phi^{-1}(p) + \sqrt{\rho}\,\Phi^{-1}(0.999)}{\sqrt{1-\rho}}\right) - p \tag{5.10}$$

To understand the sensitivity of regulatory capital to the credit model output parameters, we present plots of K with respect to p and ρ in Figs. 5.21 and 5.22, respectively.

Figure 5.21 shows that the Vasicek UL is not monotone in the grid PD. As the grid PD increases from 0, the UL increases quickly, especially for larger AVC, and it peaks and then decreases toward 0. This indicates that the Vasicek loss dominates when PD is small, and then when PD becomes larger, the expected loss part dominates.

For our subprime mortgage portfolio, we use the required AVC of 0.15. As shown in Fig. 5.21, with AVC = 0.15, Vasicek UL is over 0.5 when grid PD falls in [25%, 30%], and the peak Vasicek UL (corresponding to the largest regulatory capital) is about 0.51 when grid PD is around 25%. The majority of loans in the subprime mortgage portfolio fall in a grid with grid PD close to 30%. These are the major factors why the final regulatory capital and RWA for our subprime mortgage portfolio are big.

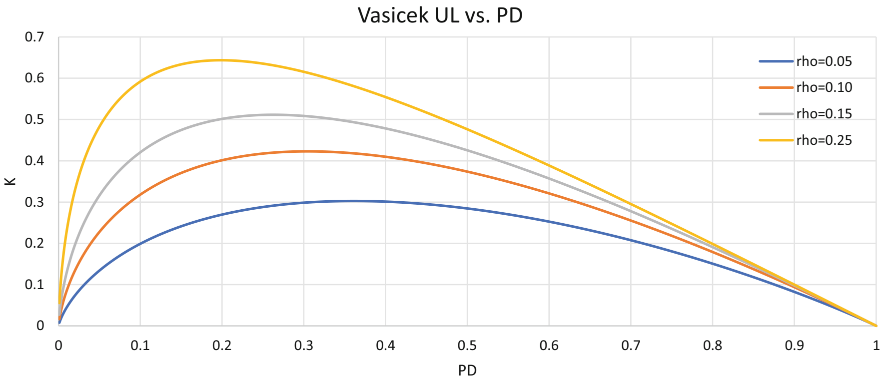


Fig. 5.21 Vasicek unexpected loss vs. grid PD

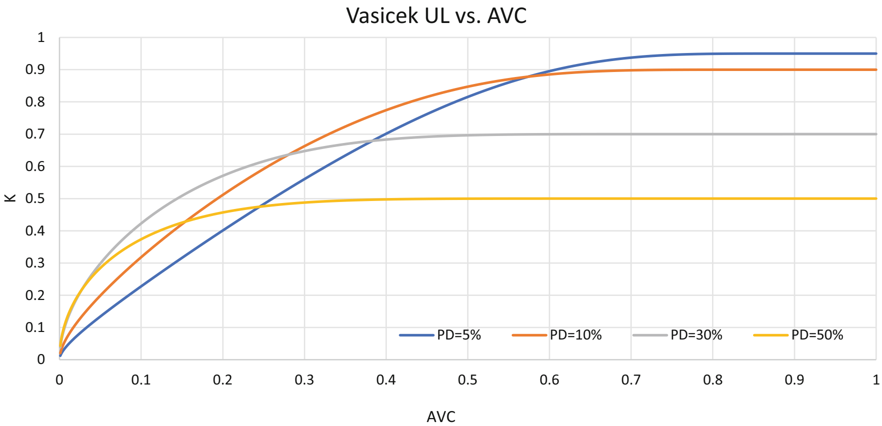


Fig. 5.22 Vasicek unexpected loss vs. AVC

Figure 5.22 presents the relation between the Vasicek UL and the AVC. We can see that the Vasicek UL is monotone in AVC – it increases quickly when AVC increases from 0 and then gradually becomes flat with larger AVC. This indicates that regulatory capital will always increase with the asset value correlation parameter but will be stable with the increasing of the correlation, especially with larger grid PD. For regular portfolios, the grid PD is less than 10% for majority of the accounts, so AVC does play a significant role in portfolio regulatory capital.

For our subprime mortgage portfolio, majority of the loans fall in the grid with gird PD about 30%, so the AVC plays a significant role before it reaches 0.3. We take the required AVC = 0.15 for mortgage portfolio; the Vasicek UL is about 0.51 (vs. the maximum 0.70) as shown in Fig. 5.22.

Table 5.18 2011Q1 portfolio regulatory capital (RC)

RC ($)	RC ($) (DT LGD)	Balance ($)
43,891,774	46,033,084	186,865,250

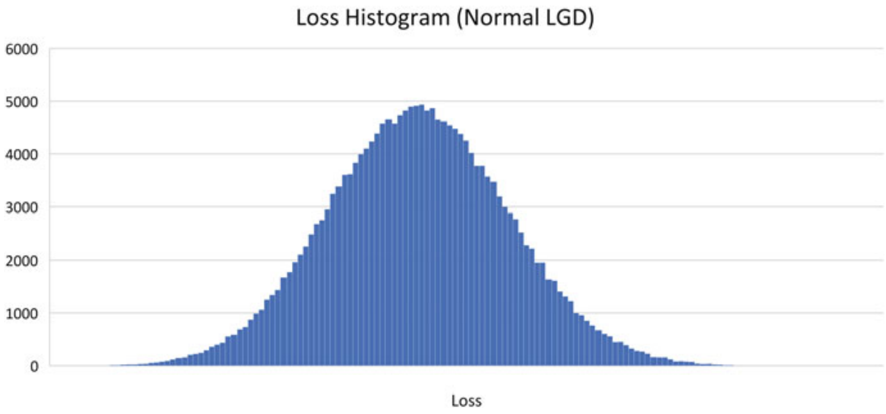


Fig. 5.23 Simulated losses for EC calculation with normal LGD

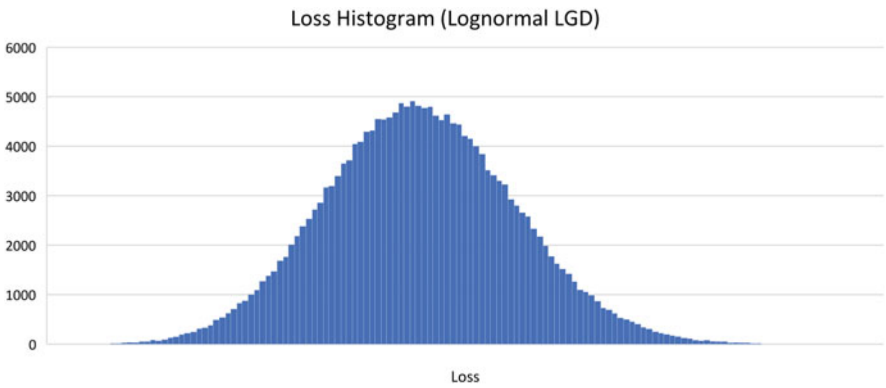


Fig. 5.24 Simulated losses for EC calculation with lognormal LGD

Comparison of Vasicek Distribution Tail and MC Simulated Loss Distribution Tail

In Sect. 5.4.2, we present the RWA for the snapshot portfolio. The corresponding regulatory capital (without the multiplier 12.5) is shown in Table 5.18:

Compared to the conditional ECs with the same confidence level (99.9%) in Table 5.13, the EC numbers are much smaller. This indicates that the loss distributions from RC and conditional EC are quite different. Figures 5.23 and 5.24 present the histograms of the simulated losses in conditional EC calculation with normal LGD and lognormal LGD, respectively. The histograms have much thinner tails compared to the Vasicek distributions as shown in Fig. 5.20, which lead to small VaRs.

The thinner tails of the simulated loss distributions in the conditional EC calculation are due to the binary approach used in the conditional EC loss generating process (5.2).

Assume a pool of a large number (N) of loans with similar risk profile. With the binary approach, the default rate of this pool is

$$\mathrm{DR} = \frac{1}{N}\sum_{i=1}^{N} I\ (D_i = 1) \tag{5.11}$$

The binary approach has the following properties: $E(\mathrm{DR}) = \mathrm{PD}$ and

$$\mathrm{var}(\mathrm{DR}) = \frac{1}{N}\mathrm{PD}(1-\mathrm{PD}) \sim 0 \tag{5.12}$$

which indicates that if default events were independent, for large pools of borrowers with similar credit risk profiles, the variation of observed default rates over time would be small. This leads to the thinner tails of the simulated loss distributions and thus smaller ECs. The problem is with the assumption of independence of loans in the portfolio embedded in the binary approach simulation:

$$E\big(I\ (D_i = 1) * I\big(D_j = 1\big)\big) = \mathrm{PD}_i * \mathrm{PD}_j \tag{5.13}$$

which is contradictory to the observed credit cycles.

The Vasicek default rate derived from the ASRF of (1.4) with the specified correlation to the single risk factor has the following properties: $E(\mathrm{DR}) = \mathrm{PD}$ and

$$\mathrm{var}(\mathrm{DR}) = \frac{1}{N}\mathrm{PD} - \mathrm{PD}^2 + \frac{N-1}{N}\Phi_2\big(\Phi^{-1}(\mathrm{PD}), \Phi^{-1}(\mathrm{PD}), \rho\big) \tag{5.14}$$

where Φ_2 is the two-dimensional normal cumulative distribution function.

Since $\Phi_2(\Phi^{-1}(\mathrm{PD}), \Phi^{-1}(\mathrm{PD}), \rho) > \mathrm{PD}^2$ for $\rho > 0$, thus $\lim_{\mathrm{N}\to\infty} \mathrm{var}(\mathrm{DR}) > 0$ for $\rho > 0$. This means that the Vasicek default rate can depict cycle volatility for pools of a large number of loans with similar risk profiles.

To incorporate the default correlation among loans within a portfolio, the marginal EC approach introduced in Sect. 5.4.4 assumes that default probability depends on the same macroeconomic variables. Such dependence is based on the final credit models selected in the EC simulation. Even the dependence is not so strong, the difference between the conditional and marginal EC VaRs could be significant as shown in Table 5.15. Figure 5.25 presents the histograms of the simulated losses for both the conditional (in red color) and marginal (in green color) EC approaches. We can see that the marginal EC simulated losses have much heavier tails than that of the conditional EC simulated losses.

The marginal EC approach is similar to the Vasicek ASRF approach with the MEVs taking the role of common risk factors. The Vasicek ASFR approach links the individual loan default probability with its credit worthiness as linear function of the asymptotic single risk factor through the probit function, while the marginal EC approach links the individual loan default probability with its credit worthiness as

Fig. 5.25 Simulated loss distributions for conditional and marginal EC

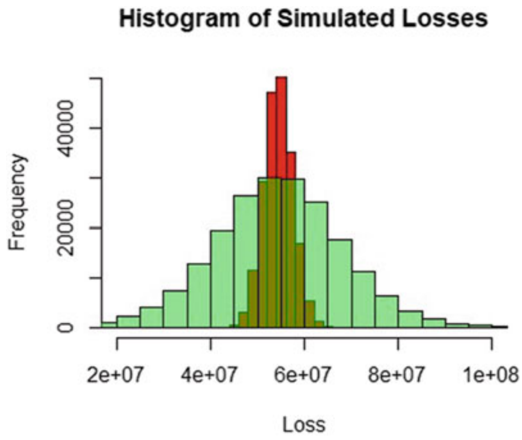


Table 5.19 2011Q1 portfolio marginal EC with different DF of MEV t-Copula

MEV method	VaR method	LGD method	Loss distribution	
			99.9% VaR	99.97% VaR
t-Copula(df = 6)	Marginal	EDF	43,782,975	47,646,833
t-Copula(df = 3)	Marginal	EDF	51,903,016	53,971,942

linear function of MEVs through the logit function. This is the foundation that for any portfolio the calculated regulatory capital (RC) and marginal economic capital (EC) should not be significantly different and can be used as benchmarks for each other. On the other hand, if the calculated RC and EC are significantly different, then some investigation should be carried out to examine the cause of the difference, which could lead to insightful information about the portfolio and models.

In the next section, we will further examine factors that impact the variability of the simulated losses in the EC calculation.

MC Simulated Loss Distribution Tail Sensitivity to MEV Joint Distribution

The marginal EC is based on default probability calculated from the PD model and the MEV random draws from the t-Copula. So, the variability of the simulated losses will depend on the PD model as well as the simulated MEVs. We would check the impacts from both the model dependences on MEVs and MEVs' variability.

First, we change the copula used in Sect. 5.4.4 for MEV by replacing its degree of freedom from 6 to 3 with the purpose to increase MEV large tail randomness. The more frequent MEV large tail events act as more frequent extreme MEV scenarios, which certainly could lead to more frequent simultaneous default events for all loans within the portfolio, thus heavier tails of the aggregated losses.

Table 5.19 shows the simulated ECs with different degrees of freedom of the t-Copula. These EC numbers confirm our previous reasoning – heavier tail MEV t-Copula leads to higher ECs.

Table 5.20 2011Q1 portfolio marginal EC with different PD models

PD model	VaR method	LGD method	Loss distribution	
			99.9% VaR	99.97% VaR
Original	Marginal	EDF	43,782,975	47,646,833
Doubling MEV effects	Marginal	EDF	52,673,120	54,436,697

Fig. 5.26 More granular LGD grid

```
> rpart.rules(tree_fit2)
 LGD_Actual
        0.31 when TTC <  0.37
        0.40 when TTC is 0.37 to 0.45
        0.44 when TTC is 0.45 to 0.47
        0.49 when TTC is 0.47 to 0.53
        0.54 when TTC is 0.53 to 0.55
        0.58 when TTC is 0.55 to 0.59
        0.62 when TTC is 0.59 to 0.61
        0.66 when TTC is 0.61 to 0.83
        0.88 when TTC is 0.83 to 0.93
        1.01 when TTC >=         0.93
```

Table 5.21 2011Q1 portfolio RWA with different LGD grid granularity

LGD grid granularity	RWA ($)	DT_RWA ($)	Balance ($)
5 buckets	548,647,169	575,413,549	186,865,250
10 buckets	544,205,846	571,508,218	186,865,250

Second, we increase the dependence of the PD model on the MEVs by doubling the standard deviation of these MEVs, which is equivalent to doubling the MEVs' coefficients in the PD logistic regression model. Table 5.20 compares the impact to the simulated ECs.

Similarly, doubling the MEV effects in the PD model leads to higher simulated ECs, which confirms our reasoning.

Grid Granularity (LGD)

We examine the impact of grid granularity to the final regulatory capital and RWA using the 2011Q1 snapshot from our subprime mortgage portfolio. Figure 5.26 presents a more granular LGD grid. The number of buckets (10) doubles that (5) we used originally in Sect. 5.4.1.

Keeping all other factors unchanged, we only change the LGD grid and run the RWA calculation. Table 5.21 presents the calculated RWA with both 99.9% and 99.97% confidence levels. We can see that the more granular LGD grid (with 10 buckets) leads to relatively smaller RWA number for both confidence levels.

It is commonly true that more granular LGD grid will lead to smaller RWA and regulatory capital with all other factors unchanged. This is largely due to the negative correlation between actual LGD and loan size as shown in Fig. 5.7. So, the result in Table 5.21 is a confirmation of this negative correlation.

5.5.4 Implementation and Output Analysis

Conceptually, a model is independent of the system or platform in which it resides. However, in practice, the performance of a model critically depends on the correct configuration and implementation in the system or platform.

A full modeling cycle includes model development and redevelopment, user acceptance testing (UAT), and the production delivery. Usually, these processes are carried out in different computing platforms and environments; however, in recent years the concepts of model continuous integration and continuous deployment or delivery (CI/CD) have become more and more popular. The CI/CD model implementation brings both efficiency in model production and new challenges for model implementation and validation. In the following, we will cover different stages of modeling and implementation processes to address potential issues model validation could embrace.

Dev Environment

Model development is carried out in the model development environment, which commonly called the Dev environment. The model Dev environment is the computing platform built for model developers to carry out all model development task, including data analysis, initial model design, and various testing related to model development. In recent years, the concept of integrated development environment (or IDE) has become popular. The IDE commonly defines a software platform including capabilities of data loading and visualization, user-friendly code editor which can provide smart programming, build and execution, debugging and profiling for integration, version control, deployment, and delivery. There are both open-source and commercial IDEs. In recent years, open-source IDEs have grown quickly, for example, in modeling, RStudio, Jupiter Notebook, Apache NetBeans, Apache Spark, and many others.

For independent model validation and auditing, model validators and auditors are often granted the same level of accesses as model developers with the Dev environment. Such requirements are not just for independently replicating model results; they are also critical for model validation and auditing to assess the model development environment for potential model risks.

UAT Environment

The User Acceptance Test environment is the next level of software development platform after th model development. For large and complex model, the UAT layer is necessary for model usage testing before model production. The UAT environment requires efficient ETL processes and friendly interfaces for user to test their own data with the deployed model. A data lake is usually a great help for this purpose.

UAT is a safe buffer between model development and production, which allows model users to test and learn the model implementation. Any issues discovered in the UAT level can be sent back to model development for further investigation and testing. Since UAT is a less restricted environment comparing to the production environment, this can speed up the model redevelopment and testing.

For high-quality model production, the UAT layer should not be skipped. Very often, model users could miss the UAT for some model releases and cause auditing failure. In recent years, automation of the UAT layer has become popular for model production efficiency.

PROD Environment

Model production environment requires higher restriction for access and data security, since very often the production is delivered to clients (internal or external) and has exposure security requirements. Model production also has much less error tolerance, and any implementation issues could result in critical model risks.

Monitoring system is built on production environment for continuous assessment of the model production, which is a critical part of model validation and will be discussed further later.

CI/CD Design

In recent years, to enhance the efficiency of model production process, the concept of continuous integration, deployment, and delivery (CI/CD) has become popular, and CI/CD implementation has become a trend in new software platforms, especially in the cloud computing platforms. From the model development side, this is an automation of the coding, building, packaging, and delivery process and greatly shortens the model production timelines. However, for the validation and auditing sides, there could be some transitions with less transparency.

To overcome the transparency issue with CI/CD, one option is using the layered model development processes as we described with Dev, UAT, and PROD for the initial model release and only use CI/CD for model redevelopment and production update. The other option is adding intermediate testing within CI/CD as a UAT layer. Either way, CI/CD should not become the reason for a less transparent modeling process, especially for large complex models.

Model Output Analysis

A basic requirement for output analysis is the ability to replicate the model outputs on appropriate platforms, especially in the production platform. For outputs with random number generation, a random seed has to be set for result replication. Output replication also presents validator a chance to review the model development and implementation logic, design, and coding. It also helps to check the model documentation consistency with model development and implementation. A full model output results replication should be done periodically to check gaps which could be created by ongoing model updates.

Variation analysis is an essential part of model output analysis. For model validation, variation analysis can be useful to access model behaviors and discover any model issues contradictory to business intuitions, because, very often, line of business requires to assess and explain the variation of capital numbers from quarter to quarter as a way to understand portfolio dynamics and take proper business actions for regulatory requirements.

Sensitivity analysis is used to assess how sensitive a model's outputs are to the change of model inputs, which could be data and assumptions. Models should have proper sensitivity to corresponding inputs. In general, capital models should not be too sensitive to input data, since it is assumed that models are built on data from different economic cycles. Models show extreme sensitivity to data changes seen in different business and economic environments could indicate instabilities. Further investigation should be carried out to check if models have fundamental issues or the modeling data do not cover sufficient business and economic scenarios. Such issues could lead to model rejection.

Models could be sensitive to model assumptions. The assessment of model assumptions should be a part of the model concept soundness. Critical model assumptions should have been included in model limitations. Model outputs should be assessed for some minor deviations from critical model assumptions.

5.5.5 Performance Monitoring

Model performance monitoring is critical. Ongoing performance assessment (OPA), back testing, and benchmarking are common practices for capital model performance assessments.

Ongoing performance assessment monitors whether production models continue to perform with the time advancing. With the changes in markets, products, exposures, activities, clients, or business practices, production models not updated promptly could deviate from the business trend and results in poor performance as measured by both statistical and business criteria. There is also the possibility that model assumptions could be broken and new model limitations are needed. Models on which business decisions depend require close performance monitoring in a timely manner; otherwise flawed and costly business decisions could be made before deterioration in model performance becomes apparent.

For capital models, reference data are subject to change due to new data availability or portfolio changes. Risk parameters as outputs of the models built on these reference data, as well as the intermediate results such as the PD and LGD grids in the regulatory capital process, require closed monitoring such that the capital changes can be tracked and interpreted. In Sect. 5.4.3, we present an example of grid migration monitoring, which is commonly used in the capital model monitoring process for grid stability. There is also the requirement that when new data and information inputs are added, the production model should be competitive in the model selection process as required by the model selection criteria. So, both the performance measured by those statistical criteria in the model selection process and the model ranking in the model selection process should not deteriorate significantly.

Benchmarking is the comparison of a given model's inputs and outputs to estimates from alternative internal or external data or models. It can be incorporated in model development as well as in ongoing monitoring. For credit risk models, examples of benchmarks include models from different methodologies, vendor firms or industry consortia, and data from retail credit bureaus. Whatever the source, benchmark models should be rigorous, and benchmark data should be accurate and complete to ensure a reasonable comparison.

Discrepancies between the model output and benchmarks should trigger investigation into the sources and degree of the differences and examination of whether they are within an expected or appropriate range given the nature of the comparison. The results of the analysis may suggest revisions to the model. However, differences do not necessarily indicate that the model is in error. The benchmark itself is an alternative prediction, and the differences may be due to the different data or methods used. If the model and the benchmark match well, that is evidence in favor of the model, but it should be interpreted with caution so we do not get a false degree of comfort. In Sect. 5.4.3, we explain why regulatory capital and marginal economic capital can be used as benchmarks for each other.

5.5.6 Model Governance

On top of the previous components of model risk management is the model governance, which sets an effective framework with defined roles and responsibilities for clear communication of model limitations and assumptions, as well as the authority to restrict model usage.

The model risk management framework as shown in Fig. 4.18 is set up by the model governance through policies and procedures. The common practice is that an institute has an overall model risk policy, which covers all aspects of model risk management, including model and model risk definitions; assessment of model risk; acceptable practices for model development, implementation, and use; appropriate model validation activities; and governance and controls over the model risk management process. Then, within different legal entities, line of businesses, or modeling areas, there may be different model risk management policies and procedures. These policies and procedures cover specific model risk management activities, for example, specific policies for certain legal entities or specific procedures for model development, validation, and monitoring and reporting for a specified modeling area. Procedures usually provide more detailed guidance on the required activities.

Capital models should comply to these policies and procedures. Additional model risk policies may be added to capital models due to periodic regulatory requirements.

Stress Test and CCAR

6

This chapter covers credit risk and capital modeling in stress test. While stress test is a broad topic, we will focus on some practical stress test frameworks – the regulatory stress test framework and the systematic stress test framework. As introduced in Chap. 1, the regulatory stress test framework was created due to the DFAST and implemented in the annual CCAR process for participant institutions. The systematic stress test framework is used by some larger institutions for their internal risk management purposes. In addition, a bottom-up risk integration framework like the conditional economic capital framework described in the previous chapter can also be used for stress test purpose, especially for reverse stress test. For all these frameworks, credit risk is one component, most often one critical component. We will illustrate how credit risk modeling is carried out in each of these frameworks, as well as how these modeling results are used in risk management and reporting.

We first introduce the three different stress test frameworks. In Sect. 1.4, we have introduced the regulatory stress test framework, so here we will more focus on the systematic stress test framework and the relationship among these stress test frameworks and also how they fit into different risk management practices, such as Risk Appetite (RA), Internal Capital Adequacy Assessment Process (ICAAP), CCAR, and business risk management limits. Then, we will focus on the credit model applications on stress test. Although credit models for stress test are implemented similarly as for accounting loss forecasting, there are situations that credit modeling needs special treatment for stress test or some alternative models and approaches are required. Aligning with the modeling methodologies used by regulators in the CCAR process, we review both the models in Chap. 3 and some alternative models commonly used in the stress test practices.

We introduce some synthetic data sets, which will be extensively used in this chapter as examples for various stages of stress test modeling process. Again, programming is first provided in the prototype format for the purpose that readers can practice using their own preferred languages. Then R or Python code is partially provided as examples of implementation of these prototypes.

C. Chen, *Practical Credit Risk and Capital Modeling, and Validation*, Management for Professionals, https://doi.org/10.1007/978-3-031-52542-1_6

Model validation is a key component of the stress test frameworks, especially with different scenarios within the frameworks. Based on the full model validation framework introduced earlier in Chap. 4, we carry out a model validation based on the synthetic data, on which we can manipulate for clear insights.

6.1 Introduction to Stress Test Frameworks

Stress test has been a topic in the financial industry for decades. However, it has not been extensively used before the 2008 financial crisis. Part of the reason is the lacking of probability interpretation of the stress scenarios expected by practitioners as in the VaR-based methods. There were efforts to combine subjective probability with the stress scenarios using Bayesian approach[1]; however, it is not easy to implement and assess in practice. The general argument is that stress test may not be a useful tool for early warning, but it can be useful for crisis management and resolution.[2] Based on these arguments, the regulatory stress test framework built up by the regulators post the 2008 financial crisis has become the main required risk management practices in recent years, and institutions, especially large financial institutions, start to align their internal risk management frameworks to the regulatory stress test framework.

The Systematic Stress Test (SST) framework is the internal stress testing framework used by large institutions for internal risk management purpose. Just like using Economic Capital (EC) for internal risk management to align with Regulatory Capital (EC), systematic stress test aligns with the regulatory DFAST as implemented in CCAR. The goal is to use SST results as risk measures in Risk Appetite (RA), Internal Capital Adequacy Assessment Process (ICAAP), as well as business risk management limits such that the internal risk measures would be consistent with the regulatory risk measures in CCAR. Figure 6.1 presents a description of such consistency.

As SST has extended uses beyond capital management, there would be more scenarios needed, and the stress testing is carried out in a more systematic approach. In the following sections, we will give more details on how to build a systematic stress test framework.

The bottom-up risk integration framework, like the conditional economic capital framework described in Chap. 5, can carry out scenario-based risk integration analysis. Given a stress scenario defined by macroeconomic factors and market shocks, loss and liquidity can be calculated based on direct inputs or models and can be integrated in any specified levels. As we discussed in Sect. 5.4.4, such risk

[1] Rebonato, Riccardo (2010), Coherent stress testing: a Bayesian approach, New York: John Wiley & Sons.

[2] Borio, C., Drehmann, M. and Tsatsaronis, K. (2012), Stress-testing macro stress testing: Does it live up to expectations? BIS Working Paper 369, Bank for International Settlements, Basel, Switzerland.

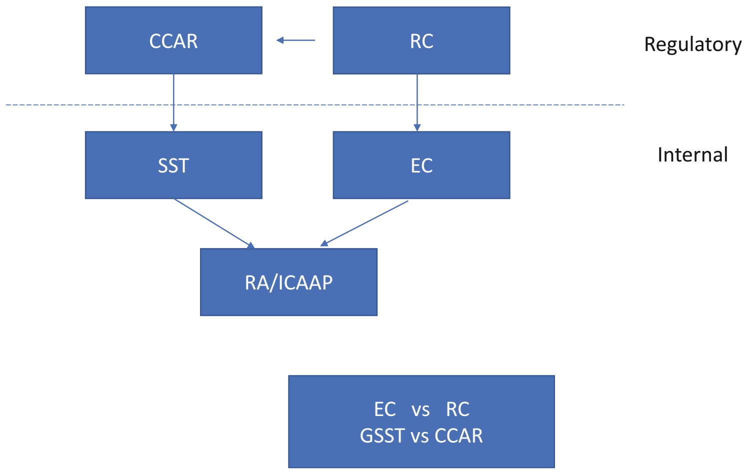


Fig. 6.1 Regulatory and internal risk management frameworks

integration takes into consideration the correlation among portfolios through dependence on the same risk factors and corresponds to the regulatory capital framework of ASRF when the risk factors and market shocks are considered random with their mean values taking the stressed values. The integrated loss or liquidity distributions at a specified portfolio aggregation level can be used for risk measurement limits with a specified confidence level under the stress scenario. For a given risk measurement limit, one can also scale the stress scenario defined by macroeconomic factors and market shocks by parallelly scaling the means of these factors to identify stress scenarios that could lead to breaking these prespecified limits. Such process is defined as reverse stress testing. In the following sections, we describe how these frameworks are implemented in different risk management programs.

6.1.1 Regulatory Stress Test

The regulatory stress test framework was defined by regulation (like Dodd-Frank Act in USA) and has been implemented by the participant banks over the last decade. As described in Sect. 1.4, the regulatory stress test framework includes two main components – Stress Test Scenarios and Stress Test Methodologies.

As a summary, a stress scenario in the regulatory stress testing is not a forecast of macroeconomic and financial conditions but a coherent set of conditions designed specifically to assess the resilience of banks when facing a deterioration in global economic conditions. In the USA, FRB defined the Scenario Design Framework[3] through the Dodd-Frank Act stress test rules. In addition, CCAR participating firms

[3] 12 C.F.R. Appendix A to Part 252 - Policy Statement on the Scenario Design Framework for Stress Testing. Feb 28, 2019.

are also required to develop their own stress scenarios as part of the stress testing results in the CCAR submission.

The FRB Scenario Design Framework consists of two components for stress conditions, the macroeconomic scenarios, and the market shocks. While the macroeconomic scenarios focus on the general stress factors, market shocks focus on instantaneous events which immediately affect the market value of the companies' trading assets and liabilities. So, market shocks apply only to companies with significant trading activity. FRB designs the three scenarios – baseline, adverse, and severely adverse scenarios for these two components correspondingly. In Sect. 1.4, general guidelines are provided on how these scenarios are created in the FRB Scenario Design Framework. FRB creates and distributes the macroeconomic scenario data each year in February to CCAR participants. FRB only provides scenario data for the 28 macroeconomic variables (MEVs) listed in Sect. 2.5.1 and provide direction and intensity for other MEVs in which participant institutes can develop for their own uses. FRB also provides the historical values for these 28 MEVs in its distributions.

FRB provides the stress test methodologies used in its annual supervisory stress test by its annual release of Supervisory Stress Test Methodology, typically published at the end of the first quarter. In this publication, FRB provides details about the models and methodologies used in the supervisory stress test. While these models and methodologies provide some guidance for participant banks in their own CCAR implementation, there is no restriction to strictly follow the supervisory models and methodologies. In Sect. 6.2, we will go over these models and methodologies aligning with the models in Chap. 3.

While FRB carries its own regulatory stress testing using its own models and methodologies and the combined data submitted by the participant banks, individual banks are required to carry out their independent stress testing based on their own data and submit the stress test results through the Y-14A report as the quantitative part of the annual CCAR exercise. In additional to the annual CCAR exercise, large participant banks start to use the CCAR results for internal risk management purposes, for example, the cumulative first five quarter losses (called stress loss usage or SLU). In the following, we will discuss how these stress test results are used in the risk management practices, including both the regulatory risk management exercises and internal risk management exercises.

CCAR

To meet the Dodd-Frank Act, the Comprehensive Capital Analysis and Review (CCAR) program was designed to assess bank's capital strength under stressed scenarios. CCAR consists of two components, the quantitative part based on the Dodd-Frank stress test results (FRB, 2020b)[4] and a qualitative assessment of firms' capital plans. In addition, capital surcharges for the Stress Capital Buffer (SCB) and

[4]FRB, 2020. Dodd-Frank Act Stress Test 2020: Supervisory Stress Test Methodology. March 2020. Board of Governors of the Federal Reserve System, Washington, DC.

Global Systemically Important Banks (GSIB) are assessed for each participating bank from 2020 as a simplification of the minimal capital requirements under both normal and stress scenarios.

The full quantitative part of the CCAR process was described in Fig. 1.8 in Sect. 1.4.2. In summary, the quantitative part calculates the quarterly changes in regulatory capital under the five scenarios.

In each February, FRB delivers the supervisory baseline and severely adverse scenarios. FRB does not deliver the supervisory adverse scenario but instead provides three methods to generate the supervisory adverse scenario – scaling (scale the differences between severely adverse and baseline (by half or two-thirds), probabilistic (take a quantile of the macroeconomic variable distribution developed from the baseline), and stable (take stable value of macroeconomic variables) – and banks have the option to choose one. In addition to the three supervisory scenarios from FRB, banks are required to come up with their internal adverse and severely adverse scenarios. Largely, the bank internal adverse and severely adverse scenarios are modified versions of the corresponding supervisory stress scenarios by deepening the downside of certain risk factors, which the firm may have more related exposures, or by adding more risk factors into the stress scenarios due to exposures not covered by the regulatory stress scenarios. The modification is generally considered more conservative.

Once the five scenarios are ready, the center of CCAR within a bank delivers these scenarios to all units handling the different components of CCAR process as described in Fig. 1.8. These units will apply the models and calculations for each of the components, mainly the PPNR and loss projections. All units send back the results at the center of CCAR for aggregation. The final regulatory capital changes for each of the nine quarters are used to project the quarterly regulatory capital amounts and capital ratios by dividing RWA under all five scenarios. A pass in the annual CCAR exam quantitative part requires the capital ratio for each of the nine quarters and under all scenarios, especially those severely adverse scenarios, above bank-specific limits. FRB carries out the same CCAR quantitative components, but more in a combined way for all participant banks. The final pass of the CCAR exam requires both pass of the quantitative part and qualitative part. The qualitative part more focuses on the bank's capital plan and the annual supervisory exams.

Stress Loss Usage

Since CCAR has become one of the most important regulatory requirements for banks, some participant banks start to use the CCAR results as their internal risk measures for risk management purposes. One of the derived quantities from the CCAR quantitative part is the cumulative losses offset by PPNR for the first five quarters, defined as Stress Loss Usage (SLU). The selection of the first five quarters is due to that, under severely adverse scenario, banks usually are subject to the worst cumulative losses (offset by PPNR) in the first five quarters.

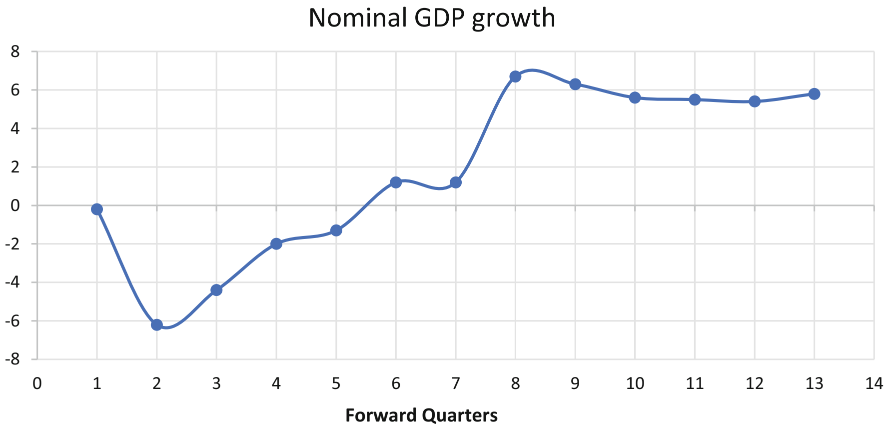


Fig. 6.2 Nominal GDP growth – 2022 supervisory severely adverse

Figure 6.2 shows the nominal GDP growth for the 13 quarters in the FRB 2022 CCAR supervisory severely adverse scenario. We can see that the nominal GDP growth are all negative for the first five quarters and become positive in the sixth quarter. This is consistent with the selection of the first five quarters' cumulative losses in SLU.

SLU has been used for risk management limits by line of businesses in some banks. It can also be used for tangible common equity (TCE) attribution for the Stress Capital Buffer (SCB) part, which is the capital buffer required by the regulator due to CCAR. Besides the alignment with CCAR, SLU has certain advantages over other risk measures traditionally used in banks. First, it considers both loss and income, alike traditional loss-only risk measures such as those capital-based measures. Secondly, it has an easier interpretation compared to traditional VaR-based risk measures. Certainly, SLU has its shortcomings. As derived from CCAR results, it doesn't represent a forecast and thus can't be back-tested. Also, depending on the CCAR implementation, it may not align well with line of business risk management in segmentation, frequency, and granularity. So, converting the CCAR results into SLU may not be direct.

6.1.2 Systematic Stress Test

The Regulatory Stress Test Framework was designed specifically for CCAR purposes and has limitations to be extensively used in business risk management practices. First, the limited scenarios in CCAR are not sufficient to cover the scenarios banks may face in their daily risk management. So, a proper set of scenarios should be built as a stress scenario inventory. A popular method to construct the stress scenario inventory is through the Structural Scenario Analysis (SSA) as shown in Fig. 6.3.

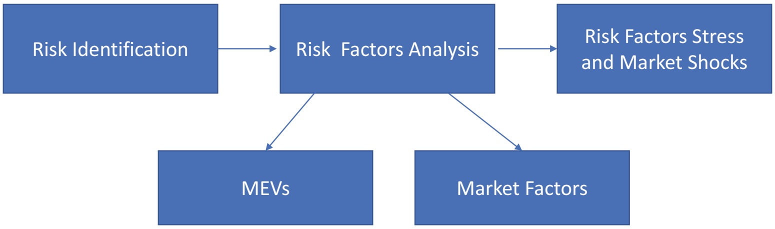


Fig. 6.3 Structural scenario analysis

The structural scenario analysis starts from the risk identification. Similar to the Risk Control Self-Assessment (RCSA) process popularly used in Operational Risk Management Framework (ORMF) for operational risk identification, for a comprehensive SSA, bank should carry out a bottom-up risk assessment based on risk exposure, frequency, and severity to identify possible scenarios that will lead to material stress. Forward looking is the main attention of the process; however, historical events can be the starting point of the identification process. The initial risk identification starts from the bottom, e.g., line of business at some level; the identified risk scenarios require classification and aggregation.

The risk scenarios identified from the bottom-up process should be combined with the top-down what-if analysis carried out by senior management based on a periodic top-risk analysis in the Risk Appetite program. The top-risk analysis can be qualitative or quantitative to cover all current top risks related to the bank and reviewed by senior management. The final scenarios selected and approved by senior management should form the scenario inventory. The creation of the scenario inventory is the result of the risk identification process, which should be a continuous process and subject to periodic audit to make sure the process is running efficiently.

Once the scenario inventory is created, the next step is the scenario factor analysis for each scenario. This is the process to identify all macroeconomic and financial market factors which may have an impact on the exposure, frequency. and severity of the scenario. Both empirical analyses based on historical data and subject matter experts may be required to identify all risk factors. Among these risk factors, drivers are the risk factors that "drive" the scenario to stress. For each scenario, one or more drivers are required.

A scenario can be stressed through restricting the driver observations or adding shocks. This can be done using the Vector Autoregression (VAR) model introduced in Sect. 1.4.1. In the VAR model, by either restricting the observed driver values or adding specific shocks, the full set of risk factors can be stressed.[5] This can also be a convenient way to adjust the degree of stress for each scenario.

Due to large numbers of risk factors included in each scenario and frequent update of the stress scenarios, a computational system is required to maintain the

[5]Kilian, L. and Lütkepohl, H. (2017), Structural Vector Autoregressive Analysis, Chapter 4. Cambridge University Press.

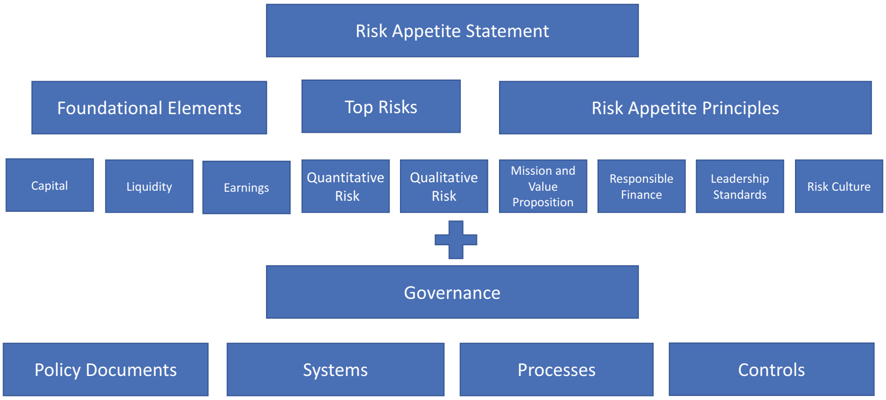


Fig. 6.4 Risk appetite program

systematic stress scenario framework. We call such a system Scenario Manager (SM). SM is also used to create specific scenarios for ad hoc and what-if scenario analysis. In the following, we introduce some important internal risk management programs and show how the systematic stress test framework is used in these risk management programs.

Risk Appetite

Risk appetite describes the aggregate level and types of risk that a firm is willing to take to achieve its strategic objectives and business plan. Risk appetite is mechanized through a set of carefully calibrated boundaries and monitored through robust reporting as shown in Fig. 6.4.

The risk appetite program includes two components – Risk Appetite Statement and Risk Appetite Governance. Risk Appetite Statement articulates the risk appetite and consists of both quantitative boundaries and qualitative principles to guide behavior. The Risk Appetite Statement expresses the firm's risk appetite relative to capital, liquidity, earnings, and top risks. It incorporates risk appetite principles that describe how the firm identifies, accesses, measures, monitors, and controls risks and thus supports a safe and sound risk culture that focuses on customers, creating economic value and maintaining the integrity of the global financial system.

Risk Appetite Governance, accompanying Risk Appetite Statement, includes the policies, processes, controls, and systems through which risk appetite is established, communicated, and monitored.

The application of the systematic stress test framework in the Risk Appetite program is with the Earnings Power as defined by the following two measures:

Risk Appetite Ratio (RAR) is a risk/reward measure (ratio) assessing earnings adequacy, and Risk Appetite Surplus (RAS) is a measure of excess earnings:

- Risk Appetite Ratio is a risk reward metric that tests pretax earnings (EBT) against a one in ten unexpected stress losses over a 1-year period and is used to identify business activities which may not return enough to support their risk profile.

$$\text{RAR} = \frac{\text{Revenue} - \text{Expense} - \text{Expected Loss}}{\text{Unexpected 1 in 10 Loss}} \tag{6.1}$$

- Risk Appetite Surplus is the surplus of pretax earnings (EBT) against a one in ten unexpected stresses loss over a 1-year period.

$$\text{RAS} = \text{Revenue} - \text{Expense} - \text{Expected Loss} - \text{Unexpected 1 in 10 Loss} \tag{6.2}$$

where Revenue and Expense are measured in 1 year (12 months) and can be Forward Looking or Trailing view or Mixed; Expected Loss is for retail and wholesale credit measured in a base scenario or long-term view; Unexpected 1 in 10 stress loss (or UL) is the difference between the Total 1 in 10 Stress Loss (or TL) from SST and the Expected Loss (or EL) over the 1-year horizon. So, RAR and RAS depend on SST.

RAR and RAS are risk appetite metrics that commonly used to represent Board-Level commitments that businesses should have no more than one in ten likelihood of incurring negative earnings in any given year. So, the target level RAR should be greater than 1, and RAS should be positive. Results below the target level could be acceptable in certain circumstances, such as during the “start-up” phase, restructuring of a business, or as a part of managing the Risk/Return trade-off of a portfolio.

Banks without the SST framework have used the stress test results from their internal adverse scenarios under the Regulatory Stress Test framework instead of the more explicit (such as the 1 in 10 likelihood) stress scenarios specially designed for the Risk Appetite program. The scenarios under the Regulatory Stress Test framework are limited and may not cover the desired stress tests businesses face related to their products and operations. So, multiple stress scenarios should be designed in the SST, and the maximum risk assessment among these scenarios should be considered in the Risk Appetite program.

ICAAP

The internal capital adequacy assessment process (ICAAP) is a part of Pillar 2 within the Basel framework and represents a financial institution’s own assessment of the capital needed to run the business. This capital may differ from the minimum regulatory capital requirement since, for instance, a financial institution may include risks that are not formally subject to the minimum regulatory capital (e.g., liquidity risk, reputational risk, business risk, or interest rate risk in the banking book) or may use different parameters or methodologies for credit risk, market risk, or operational risk. ICAAP is commonly executed in the middle of the year to avoid the CCAR submission in the early of the year.

In recent years, to align with the regulatory capital requirement in CCAR, some large US banks start to develop stress scenarios specifically for ICAAP in the SST framework. Banks without the SST framework may use their internal severely adverse scenario in the Regulatory Stress Test framework for ICAAP purpose. Again, the scenarios under the Regulatory Stress Test framework are limited and may not cover the desired stress tests businesses face related to their products and operations. So, multiple stress scenarios should be designed in the SST for ICAAP.

6.1.3 Reverse Stress Test

While the Systematic Stress Test framework is designed to answer the what-if questions, the Reverse Stress Test framework is designed to identify specific scenarios as the most plausible scenarios that lead to breaks of the prespecified risk limits at a certain level of aggregation. Although the two frameworks can be implemented in a single system, the reverse stress test goes one step further on scenario identification and closer to the intension of achieving some early warning.

Figure 6.5 presents the full process of the reverse stress test framework popularly implemented in the large financial institutions. For a financial institution, the three major risk areas are liquidity, capital, and earnings as described in the Risk Appetite program in Sect. 6.1.2.

Liquidity risk is the most critical risk for a bank as a quick and severe bank-run can easily lead to closing of the bank and a take-over by the deposit insurance agency. The regional bank crisis happened in early 2023, leading to the bankruptcy of the Silicon Valley Bank and the Signature Bank, is largely due to liquidity

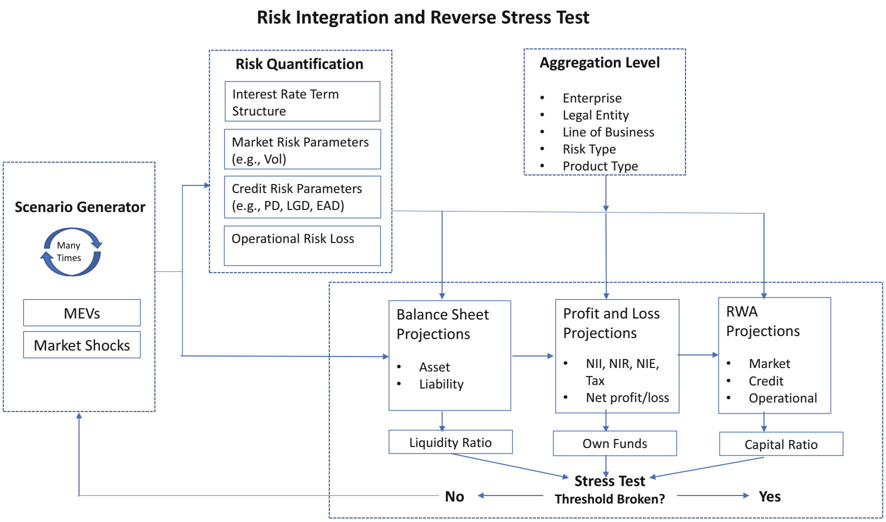


Fig. 6.5 Risk integration and reverse stress test

mismatching embedded deeply in these banks. To better manage the liquidity risk, the bank Asset and Liability management (ALM) should be able to promptly classify the asset (e.g., HTM or AFS) and liability and project the balance sheet under both base and stress scenarios.

Based on the projected balance sheet, incomes which include the Net Interest Income (NII), Noninterest Revenue (NIR), and Noninterest Expense (NIE) as defined in the PPNR can be projected. Also, various losses, including the credit charges and provisions as the focus of this chapter, can be projected. Then, the final profit and loss (PnL) can be projected. Here, the stress test can be applied directly to specific losses at any aggregated levels as preferred by line of businesses in their risk management practices.

The projected PnL can be aggregated at higher levels for capital management together with the projected RWA (standard or advanced). As described in Sect. 6.1.1, regulatory capital changes can be projected with the projected PnL under multiple scenarios for multiple periods, and thus the regulatory capital ratio can be calculated by dividing the projected RWA. Stress test checks the capital ratios to specified thresholds for potential breaks.

Any break from the specific thresholds indicates the desired stress is achieved. If not, the initial scenario can be further stressed, commonly through linear scaling or parallel shocks. The reverse stress test would like to search the most possible stress scenario that first hit the break. Due to computation expense, it is not efficient to find the exact most possible stress scenario that first hit the break; instead we would pre-select a few stress scenarios or a few stress scenarios in a grid setup for the most possible one. Commonly these prespecified scenarios already have some desired interpretation.

Combined with the Systematic Stress Test framework and Economic Capital framework, the Reverse Stress Test framework could be a very useful internal risk management tool. Given the complexity of modern financial institutions, especially large financial institutions, what-if type of risk management scale is not sufficient, and some systematic risk identification and quantification are on the demand. In the following, we present some cases where the presented reverse stress test frameworks could be useful.

Stress Scenario Identification

By design, the reverse stress test framework searches for the most possible scenario that first hit prespecified risk threshold. Identifying a scenario that only impacts a single risk measure with a simple portfolio is usually easier due to the direct and simple sensitivity relationship between some driver risk factors in the scenario. At a highly aggregated level with multiple risk measures, stress scenario identification could be complex. Under such situation, we need to clarify the goal and our stress test and simplify the stress scenario identification.

The first step is to identify the driver risk factor (or a few driver risk factors). The driver risk factor acts as the key concern of the stress test, for example, the Secured Overnight Financing Rate (SOFR) rate as the main driver of various interest rates.

Then, an interest rate scenario can be constructed by the Structural VAR model discussed in Sect. 6.1.2. Through this process, we can adjust the stress scenario by only scaling the SOFR rate. When there are multiple driver risk factors, the same process can be applied by scaling all multiple driver risk factors. Since the impacts from multiple risk factors could be interacted, the most possible scenario identified in this way is an approximation. In practice, mostly the scenario identified by such independent scaling of multiple driver risk factors is sufficient for stress testing purpose.

Concentration Risk Management

Concentration risk is defined as the risk due to significant nonlinearity of large exposures. As we discussed in Sect. 5.4 for economic capital, tail loss could have significant nonlinearity when there is correlation among exposures, so EC is commonly used to measure concentration risk. Although conditional mean used in the stress test is additive for aggregation, the correlation among these conditional means is embedded in the risk parameters, which are inputs to risk measures and depend on the same risk factors in a scenario. When the exposures are also correlated with these same risk factors, the aggregated risk measures at different aggregation levels could have significant nonlinearity on exposures and thus significant concentration risk. One example is the wrong-way risk (WWR), which is commonly seen when exposures increase with the default rate.

As implemented in a general risk integration framework, the reverse stress test framework designed here can also be used to test the concentration risk by the flexibility of being able to aggregate risk measures at any given level.

6.2 Credit Model Applications in Stress Testing

In this section, we cover how credit risk models are used in stress testing, including the three stress test frameworks we introduced in the previous section. In all these frameworks, credit risk models are used for loss forecasting. So, they may share the same models, though the model segmentation and forecasting horizon could be different. These credit risk models could be integrated with the models used in CECL as we discussed in Sect. 4.1.2. However, depending on implementation, largely CCAR models were developed in a rush to fulfill the regulatory requirement and may not have the sufficient granularity and execution frequency to meet the systematic stress test or CECL. If the high-level CCAR credit risk models can't meet the granularity and frequency requirements in the more demanding systematic stress test framework or portfolio-based CECL framework, model redevelopment could be the only choice.

While we will focus on the high-level model segmentation used in the regulatory stress test framework as the common starting point, we will also pay attention to more granular segmentation and suitable model methodology for all applications. As for all stress test, the focus is on the relationship between credit loss and macroeconomic factors. So, proper segmentation may help on discovering such relationships. However, model segmentation should not be selected only based on such relationship significance without a comprehensive review and validation.

6.2.1 Credit Models for CECL and CCAR

Here we somewhat repeat our discussion in Sect. 4.1.2, since the integration of models between CECE and CCAR is an inevitable topic for model application and implementation. For most institutions, CCAR was implemented in a high-level model-based process instead of a portfolio-based process for the accounting purpose as with CECL. However, as we discussed earlier, to align the CCAR results more on business risk management, the model segmentation granularity and execution frequency are not sufficient, and model redevelopment becomes an opportunity to integrate the two risk frameworks if different platforms were used. The main platform integration overloading will fall on a consistency integration of data, portfolio segmentation, and, most importantly, the modeling methodologies as discussed in Sect. 4.1.2. In the following sections, we will focus on the segmentation and modeling methodology.

6.2.2 Model Segmentation

In the regulatory stress test framework, credit models are first segmented by product as wholesale and retail credit models. Under both the wholesale and retail, the credit models for PD, LGD, and EAD have their separate model segmentation according to the properties of these models.

Figure 6.6 presents an example of the wholesale credit model segmentation commonly used in stress test. We include primary segmentation factors, as well as secondary and other factors could possibly be used as segmentation factors depending on the modeling strategy.

Wholesale PD model is usually segmented according to the obligator type due to the common practice that wholesale credit risk rating is based on obligator type. The obligors are classified largely into the following types;

Large Corporation – Financial Institutions: Bank and insurance obligors belong to similar industries and have low default rates.

Large Corporation – Other: Large corporation other than financial institutions, including a diverse set of industries with average to good credit quality.

Middle Market: Covers obligors in the middle market, which are larger than small business, but smaller than large corporation.

Real Estate: Real estate covers all commercial real estate client type/industry.

Individuals: Individuals in the wealth management business with FICO scores.

Models	Segmentation Factors: Primary	Secondary	Other
PD	**Obligor Type:** - Large Corporation - Finance Institutions - Large Corporation - Other - Middle Market - Real Estate - Individuals	**Product Type:** - Traditional Credit Products - Cash Securities - Derivatives - Repos - CVA Portfolios	*** Region** *** Line of Business**
LGD	**Collateral Type:** - Cash in Procession - Securities - Accounts Receivable and Inventory - All Asset Lien - Asset Based Lending - Equipment - Real Estate - Derivatives - Prime Brokerage - Government Lending	*** Line of Business**	*** Region**
EAD	**Facility Type:** - Committed - Advised	*** Line of Business**	*** Region**

Fig. 6.6 Wholesale credit model segmentation for stress test

Figure 6.7 presents an example of mapping of the various industries into the obligor type used as the primary segmentation factor in the wholesale PD model.

PD model can be further segmented by the product type, which is commonly used for models based on specific products. Region and line of business are used for more granular segmentation usually for large banks.

Wholesale LGD model is commonly segmented based on collateral type, which is a key risk factor for the recovery rate after default. Line of business (LOB) can be used as the secondary segmentation factor if LOBs have different loss recovery practices.

Wholesale ED model is commonly segmented by facility type based on which the facility is committed or advised.

In general, wholesale models can be complex due to segmentation, and regulatory stress test prefers simple segmentation; thus high-level segmentation by combining those segments described above is commonly observed in CCAR modeling.

Retails models have relatively simpler segmentation. For PD, LGD, and EAD models, the segmentation is commonly done by product type as shown in Fig. 6.8. Large banks with international businesses can further segment the retail portfolios into domestic and foreign. Line of business is less used as the retail portfolios are likely under the same line of business. Specific portfolios may be treated independently in the stress test.

Energy & Utilities	Large Corporation
Finance	
Funds	
Hedge Fund	
Large Corporate Private	
Large Corporate Public	
Media	
Money Manager	
Retail	
Telecom & Technology	
Municipal Government (US)	
Telecom & Technology (Non US)	
Bank (Non US)	Large Corporation-Financial Institutions
Bank (US)	
Broker Dealer	
Insurance (Non US)	
Insurance (US)	
Agricultural Production	Middle Market
Commercial-General Industries	
CRMS Commercial - Agricultural Production	
CRMS Commercial - General Industries Scorecard	
CRMS Commercial Restaurant	
CRMS CSA - Agricultural Production	
CRMS CSA - General Industries Scorecard	
CRMS CSA Restaurant	
Dealer Finance	
For Profit Healthcare	
Mortgage Bank	
Not Elsewhere Covered	
Not For Profit	
Not For Profit Education	
Not For Profit Hospital	
Not For Profit Religion	
Mapping - Leasing SBRI/CCS	
Restaurant	
Commercial Homebuilders	Real Estate
Commercial Mortgage Backed Securities	
Commercial Real Estate Developers	
Commercial Real Estate Guarantor	
Community Development Lending	
Large Corporate Homebuilders	
REITs	

Fig. 6.7 Obligor industries mapping to obligor type

	Segmentation Factors		
Models	**Primary**	**Secondary**	**Other**
PD **LGD** **EAD**	**Product Type:** - First Lien Mortgages - Home Equity Loans - Home Equity Line of Credit - Credit Cards - Auto Loans - Small Business Lending - Retail Other	**Region:** - Dometic - Foreign	*** Line of Business** *** Portfolio**

Fig. 6.8 Retail credit model segmentation for stress test

6.2.3 Model Methodology Choice

In this section, we describe the model methodology applied to the wholesale and retail credit model segments as described in the previous section. We will start with the model methodology for the regulatory stress test framework and illustrate whether the same methodology can be naturally extended to the other two stress test frameworks, or some alternative methods or further improvements are needed.

The model segments presented in the previous section are a general segmentation, and different stress test frameworks could adopt more or less granularity on segmentation based on some common practices under these frameworks. There could be some intentions on why such practices are adopted. For example, one intention under the stress test frameworks is aggregating segments into less granular segments to better catch the relationship between credit loss and macroeconomic factors when granular segments don't have sufficient data for such a purpose. On the other hand, with sufficient data, more granular segments may show higher sensitivity to some macroeconomic factors. Either way, a compressive review and validation should be carried out as required by the model risk management.

While multiple model methods could be applied for the same segment, we try to select the champion model method for the segment and list possible model methods, which could be taken as challenger or benchmarking models.

Wholesale Credit Models

Under the regulatory stress test framework, wholesale credit loss projection is executed under two segments – the commercial real estate segment and all others are combined into the corporate loan segment. While FRB simplifies the wholesale credit model segmentation due to its aggregated modeling strategy for all participant banks under the regulatory stress test framework, individual banks likely have more granular segmentation for the wholesale credit loss projection in their CCAR execution. In addition, different from CECL, prepayment is usually not considered as a competing risk in stress testing for wholesale credit, so default is the only target event. We will focus the main methods used for the credit models, regardless of the segmentation granularity.

PD Models

For wholesale credit models, risk rating from either internal rating system or external agency rating is the main risk factor which directly relates to the obligor long-term default probability (like the FICO score for individuals). So, this long-term default probability can be estimated through a logistic regression model based on risk rating system:

$$\operatorname{logit}\left(\mathrm{PD}\left(r_i, s_j\right)\right) = \beta_0 + \sum_{j=1}^{J} \beta_j r_i I\left(s_j = j\right) \tag{6.3}$$

where $\mathrm{PD}(r_i, s_j)$ is the long-term unit time period default probability for an obligor which belongs to the s_j rating system and has numeric rating r_i and $I(s_j = j)$ is the indicator function for the obligor falling in the j rating system. β_0 is the intercept, and $\beta_j,\ j = 1, \ldots, , J$ are coefficients for different rating systems. As we pointed out in the previous section, the rating systems may depend on granular obligor types.

The estimated long-term unit time period default probability $\widehat{\mathrm{PD}}\left(r_i, s_j\right)$ can be used as the initial default probability for an obligor at the starting point of projection. Under the regulatory stress test framework, for corporate loans, the PD changes over the projection periods of $t - 1$ to t which is modeled as a function of the macroeconomic factors:

$$\Delta\mathrm{PD}\left(r_i, s_j, t\right) = f(\Delta Z(t)) \tag{6.4}$$

where $\mathrm{PD}(r_i, s_j, t)$ is the default probability over period of $t - 1$ to t with obligors falling in the (r_i, s_j) segment and $\mathrm{PD}(r_i, s_j, 0) = \widehat{\mathrm{PD}}\left(r_i, s_j\right)$. $\Delta Z(t)$ are measures of the changes of the macroeconomic factors $Z(t)$, which can take the form of direct difference or difference in ratio. $\Delta\mathrm{PD}(r_i, s_j, t)$ usually takes the logarithm difference $\Delta\mathrm{PD}\left(r_i, s_j, t\right) = \log\left(\frac{\mathrm{PD}\left(r_i, s_j, t\right)}{\mathrm{PD}\left(r_i, s_j, t-1\right)}\right)$, and $f(\Delta Z(t)) = \gamma' \Delta Z(t)$ takes the linear function of the macroeconomic factors changing $\Delta Z(t)$ for all segments, which corresponds to the parallel multiplier for all (r_i, s_j) segments. Under such setting, we call this corporate loan PD model under the regulatory stress test framework "Loglinear model." More general settings can lead to general nonlinear models.

The parallel multiplier in (6.4) leads to parallel stressing with macroeconomic factors, which has a clear and easier interpretation of the stress impact to default probability from the macroeconomic factors. In addition, the stress impact is directly applied to the default probability instead of the logit of the default probability as commonly used in the logistic regression:

$$\operatorname{logit}\left(\mathrm{PD}\left(r_i, s_j, t\right)\right) = \beta_0 + \sum_{j=1}^{J} \beta_j r_i I\left(s_j = j\right) + \gamma' Z(t) \tag{6.5}$$

For non-parallel impact from the macroeconomic factors on different ratings, a linear interaction $\alpha' Z(t) r_i$ is commonly added to the logistic regression as:

$$\text{logit}\big(\text{PD}\big(r_i, s_j, t\big)\big) = \beta_0 + \sum_{j=1}^{J} \beta_j r_i I\big(s_j = j\big) + \gamma' Z(t) + \alpha' Z(t) r_i \tag{6.6}$$

This is popularly used in the systematic stress tests, where more comprehensive sensitivity is required to be measured for the impact of macroeconomic factors.

When random effects are considered in the default probability, a generalized linear mixed model (GLMM) extends the fixed effects model of (6.6) to:

$$\text{logit}\big(\text{PD}\big(r_i, s_j, t\big)\big) = \beta_0 + \sum_{j=1}^{J} \beta_j r_i I\big(s_j = j\big) + \gamma' Z(t) + \alpha' Z(t) r_i + \sigma_t \tag{6.7}$$

where σ_t is the standard deviation of the random intercept and taken as a parameter and ϵ is the standard normal random variable. The GLMM is commonly used in stress test with a few periods and focusing on high volatility. σ_t can be estimated from the GLMM model, but often is adjusted according stress scenarios.

For the commercial real estate (CRE) segment in the wholesale credit model, besides common loan characteristics, there are property characteristics and seasoning effects, so the PD model for this segment under all stress frameworks prefers directly using these factors instead of the risk rating system:

$$\text{logit}(\text{PD}(i, t)) = \beta_0 + \beta' X(i, t) + \gamma' Z(t) + \lambda(i, t) \tag{6.8}$$

where $X(i, t)$ represents the loan and property characteristics for i-th loan, $\lambda(i, t)$ represents its seasoning effect and is taken as a loan age function, and $Z(t)$ represents macroeconomic factors.

While the projected default probability for the corporate loan segment is based on the ratings, the projected default probability is applied for all loans falling in the same rating segment. For the CRE segment, the projected default probability is loan level and directly applied to each individual loan.

LGD Models

As in CECL, LGD models in stress tests are based on the collateral types, which are classified as secured and unsecured under the regulatory stress test framework for simplicity. Under the systematic stress test framework, such classification may require more granular segmentation based on collateral type as shown in Fig. 6.6 if recovery processes for these different types of collaterals show significant different results.

Under the regulatory stress test framework, FRB adopts a LGD modeling strategy similar to that of the PD model for corporate loans – specifically model the LGD change from the initial LGD at the projection date. In general, use the following model:

$$\text{LGD}(i, t) = f(\text{LGD}(i, t-1), \text{PD}(i, t), \text{PD}(i, t-1)) \tag{6.9}$$

where $\mathrm{LGD}(i, t)$ is the loss given default for i-th loan default over time period t and $\mathrm{PD}(i, t)$ is the default probability for the i-th loans over time period t. $\mathrm{LGD}(i, 0)$ is the initial LGD at the beginning of the projection and can take the long-term average of actual LGD from historical data. In (6.9), the LGD stress by macroeconomic factors is through the stress of PD under the assumption that wholesale loan loss severity is highly sensitive to the current and previous default rate.

If the general function in (6.9) takes a multiplicative format with a power function of the ratios of PD, then a parametric format can be used as:

$$\mathrm{LGD}(i,t) = \lambda * \mathrm{LGD}(i,t-1)\left(\frac{\mathrm{PD}(i,t)}{\mathrm{PD}(i,t-1)}\right)^{\beta} \tag{6.10}$$

or a simple linear regression in logarithm:

$$\log\left(\frac{\mathrm{LGD}(i,t)}{\mathrm{LGD}(i,t-1)}\right) = \alpha + \beta * \log\left(\frac{\mathrm{PD}(i,t)}{\mathrm{PD}(i,t-1)}\right) \tag{6.11}$$

If PD takes the loglinear model as we described earlier, the LGD also takes the loglinear model in this case with the intercept $\alpha = \log(\lambda)$ and a modified slope with the parameter β. These parameters can be estimated from the historical LGD and PD data.

The loglinear model takes the macroeconomic factor impact in the multiplicative format of their changes. It is different from the linear regression LGD models we introduced in Chap. 3. Commonly the loglinear model is more sensitive to large macroeconomic factor changes. Such high sensitivity should be checked to follow common sense to avoid extreme model results under severely adverse stress scenarios.

For the commercial real estate (CRE) segment, similar to the PD model, LGD is directly modeled using the loan and property characteristics $X(i)$ and macroeconomic factors $Z(t)$:

$$\mathrm{LGD}(i,t) = f(X(i), Z(t)) \tag{6.12}$$

These models can take the classical LGD models as presented in Chap. 3.

Under the systematic stress test framework, classical LGD models as presented in Chap. 3 are also preferred, especially the regression models, which assume simpler relationship between LGD and the macroeconomic factors.

EAD Models

Under the regulatory stress test framework, the EAD is treated separately according to wholesale loan type. For closed-end corporate loans, the EAD is the outstanding balance. For standby letters of credit and trade finance credits, EADs are conservatively assumed to equal the total commitment since typically these types of credits are fully drawn when they enter default status. For lines of credit and other revolving commitments, the EAD equals the outstanding balance plus a portion of the unfunded commitment (i.e., the difference between the committed exposure and outstanding balance), which reflects the amount that is likely to be drawn down by the borrower in the event of default.

The Federal Reserve calibrates the amount that is likely to be drawn down to the historical drawdown experience for defaulted US syndicated revolving lines of credit that are in the Shared National Credit (SNC) database. The model sets the EAD for a line of credit or other revolving product as follows:

$$\text{EAD}(i) = \text{OB}(i) + \text{LEQ} * (C(i) - \text{OB}(i)) \tag{6.13}$$

where i represents the revolving product or line of credit, EAD(i) represents the EAD, OB(i) represents the line's outstanding balance at the start of the projection horizon, LEQ represents the calibrated drawdown rate, and $C(i)$ represents the line's committed balance at the start of the projection horizon. The LEQ model in Sect. 3.1.5.4 can be used to calibrate the LEQ factor.

The EAD model assumes EAD for CRE loans equals the total committed exposure amount, which is the outstanding balance of the loan plus any remaining undrawn committed amount at the start of the projection horizon.

Retail Credit Models

Retail credit models are based on the product types as different retail credit products are quite different in the definition of default, loss resolution, and recovery processes, as well as collateral, credit usage, and exposures. In the following we will cover models used in the regulatory stress test framework and popular models used in other stress test frameworks. We focus on the three segments – mortgages, credit cards and auto loans. The rest of the retail segments including small business lending (SBL) and other retails (line of credits) adopt similar models as these three segments.

PD and PP Models

Under the regulatory stress test framework, Probability of Default (PD) and Probability of Prepayment (PP) models are developed for first-lien residential mortgages, Home Equity Loan (HeLoan) and Home Equity Line of Credit (HELOC), auto loans, credit cards, and retail other. Although the general PD and PP model introduced in Sect. 3.1.2 under the competing risk framework can all be applied to these segments, there are some specifics for each of these segments, and we will go over each of them one by one.

For the *first-lien residential mortgages*, under the regulatory stress test framework, besides the common loan status of current, default, and paid-off, a fourth loan status of impaired is added, which is defined as 90–179 days of delinquency (180 day or more delinquency is defined as default). So, in total there are five transitions: current to impaired, current to paid-off, impaired to current, impaired to paid-off, and impaired to default. These transitions are measured quarterly instead of monthly. Each of these transitions is modeled by a binary logit model as:

$$\operatorname{logit}(P(i,t)) = \alpha + \beta' X(i,t) + \gamma' Z(t) \tag{6.14}$$

where $P(i,t)$ is the conditional probability of one of the five transitions for i-th loan at time t; $X(i,t)$ is the vector of loan, property, and borrower characteristics as risk factors; and $Z(t)$ is the vector of macroeconomic factors. α, β, and γ are parameters. These binary logit models are taken the format of binary approximation under the competing risk framework as we described in Sect. 3.1.2.

The addition of the impaired status as an intermediate state increases the complexity of the PD/PP models. As an intermediate state, it does not involve loss directly. Loan delinquency status can be easily added into the model as a risk factor.

For *Home Equity Loans* (HeLoan) and *Home Equity Line of Credit* (HELOC), the same model (6.14) is adopted for the PD modeling under the regulatory stress test framework. For each of these products, the conditional probability for each of these same five transitions can be projected using the binary logit model, though different sets of risk factors and macroeconomic factors are used. For example, for HELOC, the utilization rate is one critical risk factor in the PD model. Collectively, these models project a probability of default, conditional on product type, initial payment status, loan and borrower characteristics, and economic conditions over the projection horizon.

The HELOC PD model contains an additional feature to account for the fact that, for most lines of credit, the borrower may draw on the line for a fixed period, known as the "draw period," during which repayments of principal are not required. At the end of the draw period, the outstanding line balance either becomes immediately payable or converts to a fully amortizing loan. Borrowers holding these products after the draw period ends must make higher payments than were required during the draw period. The PD model assumes HELOCs that reach the end-of-draw period pay off and default at higher rates than HELOCs that are still in their draw period.

Under the regulatory stress test framework, the default of *credit cards* is defined as 120 days or more delinquency, in bankruptcy, or charged off. This is different from the Basel default definition for credit cards (180 days or more delinquency, in bankruptcy, or charged off). In addition, FRB defined three segments for the credit cards:

- Delinquent – Payment is between 30 and 119 days past due.
- Current active – Payment is no more than 29 days past due and have had activity in the past 12 months.
- Current inactive – Payment is no more than 29 days past due and have had no activity in the past 12 months.

Default on credit cards heavily depends on the payment behavior and account status. Under the systematic stress test framework, more granular segmentation is usually required. Figure 6.9 presents an example for a more granular PD model segmentation of credit cards:

Under the contractual account status, there are closed, new, and current accounts. Closed accounts are contractually closed, and new accounts are accounts opened

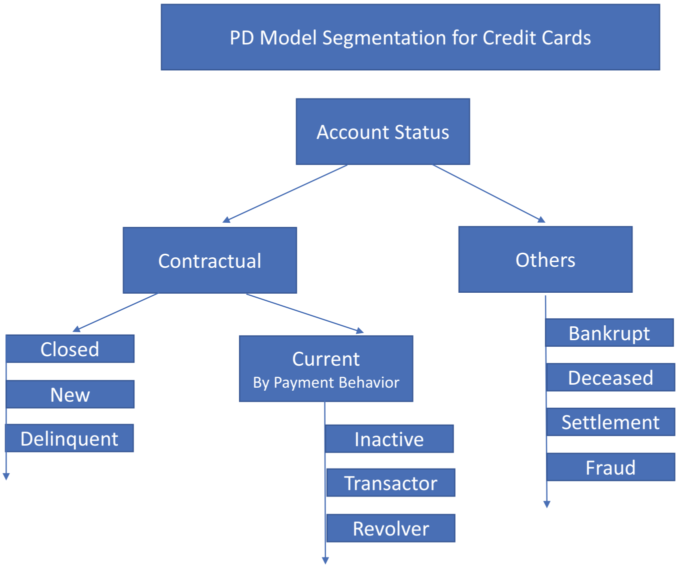


Fig. 6.9 PD model segmentation for credit cards

with a specified period of time, e.g., 6 months. Delinquent accounts are defined as over 30 and more days past due. The rest are current accounts which can be further segmented by the payment behavior. Inactive accounts are no more than 29 days past due and have had no activity in the past 12 months as in the regulatory stress test framework. The active accounts are further segmented into transactor and revolver accounts as the former always pay off the balance and the later only pay the portion of minimum required current balance and roll the unpaid portion in an interest-bearing balance.

Other non-contractual account status include bankrupt, deceased, under settlement, and fraud accounts. These accounts have some specific characteristics in default treatment.

The binary logit model (6.14) is the popular model used for PD and PP of credit cards, although the propositional hazard model (3.17) described in Sect. 3.1.1.2 is also commonly used under some assumptions on the baseline hazard.

A compromise between these two models is the binary logit model with specified parametric baseline hazard:

$$\operatorname{logit}(P(i,t)) = \alpha + f(t) + \beta' X(i,t) + \gamma' Z(t) \tag{6.15}$$

where $f(t)$ is related to the baseline hazard and usually is specified as a linear combination of b-splines with given knots. Due to less significance, a single PP model is commonly fitted for the combined segment.

Under the regulatory stress test framework, the default of *auto loans* is defined as 120 days or more delinquency, in bankruptcy, or charged off. This is same as the Basel default definition. FRB defined two segments for the PD/PP model with auto loans:

- Delinquent – Payment is between 30 and 119 days past due.
- Current – Payment is no more than 29 days past due.

This is also the segments used under the systematic stress test framework within a more granular level of segmentation with products, for example, auto loans, specialty (marine, motorcycle, and recreation vehicle) loans, and aircraft loans.

Similar to credit cards, the binary logit model (6.14) is the popular model used for PD and PP of auto loans. In addition, the propositional hazard model (3.17) described in Sect. 3.1.1.2 and the compromise model (6.15) are also commonly used.

LGD Models

The LGD models for retail credit products are more diversified compared to that for wholesale credit products due to the verities of retail products and richness of the retail default data. In the following, we go over the LGD models used in the regulatory stress test framework, as well as popularly used LGD models used for other stress test frameworks.

Under the regulatory stress test framework, for *First-Lien Residential Mortgages*, FRB uses two models for LGD. The first model is a loglinear model used to project the post default resolution timeline:

$$\log(T(i)) = \alpha + \beta' X(i, t) + \gamma' Z(t) \tag{6.16}$$

where $T(i)$ is the length between the default time t and the post default resolution time for the i-th loan. $X(i, t)$ is the vector of loan, property, and borrower characteristics as risk factors, and $Z(t)$ is the vector of macroeconomic factors. α, β, and γ are parameters.

Once the resolution timeline $T(i)$ is projected, one of the regression models presented in Sect. 3.1.4.4 can be used with $T(i)$ as an extra risk factor:

$$\mathrm{LGD}(i, t) = f(X(i, t), T(i), Z(t + T(i))) \tag{6.17}$$

where i represents the i-th loan; t represents the time of default; $\mathrm{LGD}(i, t)$ represents the accounting loss severity rate of loan i that enters default at time t; $X(i, t)$ represents a set of loan, borrower, and property characteristics; $T(i)$ represents the liquidation timeline for loan i; and $Z(t + T(i))$ represents one or more of the macroeconomic variables included in the stress scenarios at the time of liquidation. The LGD model is fitted separately for Prime, Alt-A, and Subprime segments.

The LGD model with resolution timeline can be considered as a simplified version of the micro-structure dynamic model presented in Sect. 3.1.4.3. The micro-structure dynamic model includes both the resolution event probability models and resolution severity models and requires both the post-default resolution transaction information and the gross loss severity and macroeconomic factors information at the resolution, while this simplified LGD model with resolution timeline only requires resolution time and macroeconomic factors information at the resolution. If we have the projected resolution event probabilities – probability of gross loss $\mathrm{PGL}(i, \tau)$ and probability of paid-off $\mathrm{PPO}(i, \tau)$ – we can easily estimate the resolution timeline:

$$T(i) = \sum\nolimits_{\tau=1}^{T_{\mathrm{L}}} \tau * (\mathrm{PGL}(i, \tau) + \mathrm{PPO}(i, \tau)) \tag{6.18}$$

where T_{L} is the maximum resolution time and can be specified or estimated from historical data for specific product as described in Sect. 3.1.4.3.

Under the systematic stress test framework, when the required information for the micro-structure dynamic model is available, it is preferred as it is more accurate compared to alternative models.

FRB also uses the projected time of liquidation to allocate estimated losses between credit losses on the defaulted loans and net losses arising from the eventual sale of the underlying property in PPNR. This explains why the LGD model with resolution timeline is preferred under the regulatory stress test framework.

For *Home Equity loans and Home Equity Line of Credit*, the regulatory stress test framework uses the LGD model for the primary lien above to project total recovery at default and the excess recovery beyond the prime lien exposure as the recovery for these second liens. If there is no excess recovery in the projection, the LGD is taken as 1 for these secondary liens.

For other stress test frameworks, independent LGD models for these secondary liens are preferred when loss severity data on these liens are available. Due to the large number of observed LGDs for these secondary liens being either 0 or 1, inflated beta regression model is commonly used.

For *credit cards,* under the regulatory stress test framework, the LGD model assumes that LGD for credit cards is a fixed percentage of EAD. This percentage is calculated separately for bank cards and charge cards based on historical industry data of gross charge-offs and recoveries.

However, under other stress test frameworks, LGD model is still preferred when account level loss data are available. Simple linear regression models are commonly used for direct and transparent relationship between LGD and macroeconomic factors. Usually, a separate linear regression model is fitted for the open accounts and closed accounts, respectively.

For *Auto Loans,* the regulatory stress test framework uses a pool-level LGD model and assumes that LGD takes the format of

$$\mathrm{LGD}(k, t) = f(X(k), Z(t)) \tag{6.19}$$

where k represents the poll and t can be a single time or a period of time when the pool includes loans default within the period. $\mathrm{LGD}(k, t)$ is the loss severity rate for this pool at (or during) time t. $X(k)$ represents characteristics of the pool k, such as product type, LTV, and pool segment age, and $Z(t)$ represents one or more of the macroeconomic variables included in the stress scenarios. The model then projects LGD by applying coefficient estimates from Eq. (6.19) to pool-level data.

The LGD model uses projected values of a national used car price index in addition to unemployment rates and house prices that are projected on the state level.

For other stress test frameworks, loan-level LGD model is still preferred. Like for credit cards, loan-level linear regression model is fitted for the open accounts and closed accounts, respectively.

EAD Models

The EAD models for retail products are relatively simpler for stress tests. Under the regulatory stress test framework, Federal Reserve assumes EAD for *First-Lien Residential Mortgages* and *Home Equity Loans (HeLoan)* to be the UPB of the loan at the start of the projection horizon. *Home Equity Line of Credits* (HELOCs) that have been permanently closed or have reached the end-of-draw period are essentially closed-end loans. For these HELOCs, the Federal Reserve assumes EAD to equal the UPB at the start of the projection horizon. For all other HELOCs, the Federal Reserve sets EAD to the higher of the UPB at the start of the projection horizon and the original credit limit.

For *credit cards,* FRB estimates EAD with the sum of the amount outstanding on the account (i.e., UPB) and the estimated amount of the credit line that is likely to be drawn down by the borrower between the beginning of the projection horizon and the time of default. The model calculates EAD for an account that defaults at a specific time as

$$\mathrm{EAD}_{iT} = \mathrm{UPB}_{it} + \mathrm{LEQ}_{it} * \mathrm{Line}_{it} \tag{6.20}$$

where i represents the account, t represents projection time, EAD_{iT} represents the EAD, UPB_{it} represents the reported unpaid balance of account i at the start of the projection horizon, Line_{it} represents the reported credit line of account i at the start of the projection horizon, and LEQ_{it} represents the utilization factor.

For the utilization factor, FRB only estimates LEQ_{it} as a function of account and borrower characteristics:

$$\mathrm{LEQ}_{it} = \alpha + \sum_{j=1}^{p} \beta_j x_{jit} + \epsilon_{it} \tag{6.21}$$

without the macroeconomic factors, but with the consideration of payment status of the account and time to default. So, FRB uses separate models to estimate the drawdown amount for current and delinquent accounts and for accounts with short-, medium-, and long-term transitions to default. For accounts that are current, FRB estimates separate models for segments with credit lines of different sizes. FRB adjusts estimated EAD to exclude delinquent interest and fees. The LEQ model

(6.21) is neutral to macroeconomic factors, and the full LEQ model (3.82) described in Sect. 3.1.5.4 may be preferred for other stress test frameworks for credit cards.

The EAD model (6.20) FRB adopted corresponds to the EAD model (3.81) for blocked and highly drawn accounts we described in Sect. 3.1.5.4. This is certainly a conservative consideration. In systematic stress test framework, more granular segmentation and corresponding models as described in Sect. 3.1.5.2 are commonly used.

For *auto loans*, under the regulatory stress test framework, aligning with the pool-level LGD model FRB used, EAD is estimated based on the pattern of amortization of loans that ultimately defaulted in the data provided by a major credit bureau, as reflected in the following equation:

$$\mathrm{EAD}(k,t) = \mathrm{UPB}(k) * \mathrm{PR}(k,t) \tag{6.22}$$

where k represents the loan segment, t represents time, $\mathrm{EAD}(k,t)$ represents the EAD, $\mathrm{UPB}(k)$ represents the unpaid principal balance for loan segment k, and $\mathrm{PR}(k,t)$ represents a paydown ratio for loan segment k at period t. $\mathrm{PR}(k,t)$ is estimated as a function of loan characteristics.

Compared to the exposure calculation for committed credit products described in Sect. 3.1.5.1, the FRB paydown EAD model (6.22) is simpler. It is also preferred when the accuracy of prepayment and amortization ratios are challenged. The extra required information is the historical paydown amount for default auto loans, which certainly makes more sense in the aggregate pool level than loan level.

Although the pool-level EAD model of (6.22) is attractive, it is neutral to the macroeconomic factors and may not be preferred under other stress test frameworks. For example, under the reverse stress test framework, macroeconomic factors are tuned to break the assumed thresholds, and a macroeconomic factor neutral model like (6.22) would not count the impact of EAD stress. Also, in practice, exposures do change when macroeconomic conditions change and ignorance of the EAD stress may not be expected by the line of business. So, under the systematic stress test framework commonly requested by risk management purposes from line of business, the full macroeconomic sensitive EAD model like (3.83) is till preferred.

6.3 Model Data for Stress Testing

In this section, we will use similar synthetic data sets created in Sect. 4.2 for CECL modeling for illustration purposes of stress testing. In Sect. 4.2, we simulate the loan transaction data sets using both a static and a dynamic model. While the static model was used for illustration purpose for CECL modeling, it is less informative for stress testing given that the risk events probabilities do not depend on macroeconomic factors and thus will not be impacted under different macroeconomic scenarios. So, we will focus on the dynamic model. As an effort to show how alternative sampling methods work on the full simulated transaction data, we select different sample sizes of loans and demonstrate their impact to our model selection results.

6.3.1 A Sample of Mortgage Portfolio

As described in Sect. 4.2, our subprime mortgage portfolio consists of 69,241 loans. For comparison purposes, we use the same 10% sample randomly selected from this portfolio with 6924 loans as in Sect. 4.2.1. Refer to Table 4.1 for a summary of this sample with some attributes in the sample. For the CECL modeling in Chap. 4, we simulated two transaction data sets using two different models. For stress testing purpose, we only use the dynamic transaction data set, which was simulated using risk event models with the two macroeconomic factors DHPI and DUER. Refer to Table 4.3 for a summary of this simulated transaction data set.

For fitting risk event models in CECL modeling, we used the full simulated transaction data set instead of any sampling methods in Chap. 4. The full simulated transaction data trace the transactions from either the starting booking date for an existing loan or the origination date of a new loan. If we treat the starting booking date also as "some" origination date for those existing loans, the full simulated transaction data are more like panel data for origination. If the loan risk event dynamics only depends on the origination information and macroeconomic factors as our prespecified risk event models, using such origination transaction data should be sufficient for the risk event model fitting. However, when risk events dynamics also depends on the current loan, borrow, and property characteristics, such as the ongoing delinquent status, the origination type of transaction data will not be sufficient to catch those dynamics. Models built on the origination type of transaction data will also not be sufficient for the forecasting purpose for a snapshot portfolio, which is usually the target portfolio for stress testing.

To catch the dependence of risk event dynamics on current information, various panel data sampling methods as described in Sect. 2.4 can be used. In this chapter, for illustration purpose, we focus on the snapshot sampling method on panel data due to its simplicity and popularity. Although our prespecified risk event models do not depend on loan-specific current information, we demonstrate how to execute the snapshot sampling for model fitting and forecasting in stress testing. When rich current loan-specific information is available, we recommend the Full Observations Stratified Sampling (FOSS) method with further segmentation based on such current loan-specific information (e.g., current account delinquent status) on the explored panel data.

To demonstrate different snapshot sampling frequency, we execute annual, semiannual, and quarterly snapshot sampling on the full simulated transaction data. Refer to Table 4.3 for a summary of the full simulated transaction data and Sect. 4.2.1 for the simulation process.

The annual snapshot sampling takes a fixed month snapshot in each year (e.g., January) of the full simulated transaction data and takes all following transactions of these loans in these snapshots as panel data. Table 6.1 shows a summary of the annual snapshot panel data with some loan attributes.

Compared to the original full simulated transaction data (Table 4.3), the annual snapshot panel data sample has similar size (86,999 vs. 83,095). This is due to that the average loan life of the original transaction data is about a year. In general, when

Table 6.1 Summary information for the annual snapshot panel data sample

ID	SnapDate	Age	aci	oltv	DHPI	DUER	DF_Event	PP_Event
Length: 86999	Min.: 2000-01-01	Min.: 1.00	Min.: 300.0	Min.: 10.30	Min.: −212.302	Min.: −13.767	Min.: 0.00000	Min.: 0.00000
Class: character	1st Qu.: 2006-01-01	1st Qu.: 11.00	1st Qu.: 411.0	1st Qu.: 73.00	1st Qu.: −75.710	1st Qu.: −4.044	1st Qu.: 0.00000	1st Qu.: 0.00000
Mode: character	Median: 2007-01-01	Median: 20.00	Median: 545.0	Median: 79.99	Median: −48.263	Median: −3.017	Median: 0.00000	Median: 0.00000
NA	Mean: 2007-02-28	Mean: 23.82	Mean: 550.7	Mean: 80.89	Mean: −54.674	Mean: −3.248	Mean: 0.04662	Mean: 0.03322
NA	3rd Qu.: 2008-01-01	3rd Qu.: 32.00	3rd Qu.: 680.0	3rd Qu.: 80.00	3rd Qu.: −27.573	3rd Qu.: −2.049	3rd Qu.: 0.00000	3rd Qu.: 0.00000
NA	Max.: 2013-01-01	Max.: 114.00	Max.: 850.0	Max.: 363.46	Max.: −1.252	Max.: −0.120	Max.: 1.00000	Max.: 1.00000

the average loan life in the original transaction data is much longer than a year, the annual snapshot panel data sample should be much larger. The SnapDate attribute shows the range of the snapshots. Compared to the original full simulated transaction data, the age attribute has larger values due to fact that the loan ages at the snapshot dates are larger (seasoning). For other attributes, we can see that the average default rate becomes larger while the average prepayment rate becomes smaller – these are all due to that we include more seasoned loans in the panel data due to the snapshot sampling.

The semiannual snapshot sampling takes two fixed month snapshots in each year (e.g., January and July) of the full simulated transaction data and takes all the following transactions of these loans in these snapshots as panel data. Table 6.2 shows a summary of the semiannual snapshot panel data with some loan attributes.

The semiannual snapshot panel data sample almost double the annual snapshot panel data sample, while the summary statistics of those attributes do not change much. For snapshot sampling, the higher snapshot frequency is mainly used to increase sample size and make sure we have sufficient samples when further segmentation is executed based on current loan-specific factors. We fix the dates in a year for snapshot sampling in our example. As discussed in Sect. 2.4.2, specific dates and frequencies can be used if there is a support for such specifics.

Table 6.3 shows a summary of the quarterly snapshot panel data, which is nearly double the size of the semiannual snapshot panel data sample, while similarly the attribute summary statistics do not change much.

Based on these sampled transaction data sets, following the similar process in CECL modeling, we build generalized linear models for default and prepayment risks and carry out the AEVS model selection procedure illustrated in Sect. 3.3.5 to show the impact of different sampled transaction data sets to the final selected models.

6.3.2 MEV Scenarios

For stress testing, MEV projection is within the scenario design, which is more comprehensive and commonly handled by a dedicated group. In our case, for simplicity, we only address the two MEVs used in our model.

Following the reasoning in Sect. 4.2.2 for the CECL MEV forecasting using the mean reversion approach, we design the three scenarios for MEV forecasting from the reporting date – baseline, adverse, and severely adverse. While our scenario design is simple and only for illustration purpose, it is required to meet some general principles, e.g., the degree of stress follows the order that adverse scenario is more stress than the baseline scenario and severely adverse scenario is more stress than the adverse scenario. Also, we apply the common observation that the stress reversion starts from the sixth quarter (or 16th month).

For the baseline scenario, we assume the HPI has a 3% annual appreciation over the next 13 quarters, which cover the regulatory nine quarter forecasts and the extra

Table 6.2 Summary information for the semiannual snapshot panel data sample

ID	SnapDate	Age	aci	oltv	DHPI	DUER	DF_Event	PP_Event
Length: 172071	Min.: 2000-01-01	Min.: 1.00	Min.: 300	Min.: 10.30	Min.: −212.302	Min.: −13.767	Min.: 0.00000	Min.: 0.00000
Class: character	1st Qu.: 2006-01-01	1st Qu.: 11.00	1st Qu.: 411	1st Qu.: 73.00	1st Qu.: −75.977	1st Qu.: −4.061	1st Qu.: 0.00000	1st Qu.: 0.00000
Mode: character	Median: 2007-01-01	Median: 20.00	Median: 545	Median: 80.00	Median: −48.519	Median: −3.025	Median: 0.00000	Median: 0.00000
NA	Mean: 2007-03-01	Mean: 23.97	Mean: 550	Mean: 80.94	Mean: −54.895	Mean: −3.264	Mean: 0.04709	Mean: 0.03322
NA	3rd Qu.: 2008-01-01	3rd Qu.: 33.00	3rd Qu.: 678	3rd Qu.: 80.00	3rd Qu.: −27.852	3rd Qu.: −2.070	3rd Qu.: 0.00000	3rd Qu.: 0.00000
NA	Max.: 2013-07-01	Max.: 114.00	Max.: 850	Max.: 363.46	Max.: −1.252	Max.: −0.115	Max.:1.00000	Max.: 1.00000

Table 6.3 Summary information for the quarterly snapshot panel data sample

ID	SnapDate	Age	aci	oltv	DHPI	DUER	DF_Event	PP_Event
Length: 343217	Min.: 2000-01-01	Min.: 1.00	Min.: 300.0	Min.: 8.103	Min.: −212.302	Min.: −13.767	Min.: 0.00000	Min.: 0.00000
Class: character	1st Qu.: 2006-04-01	1st Qu.: 11.00	1st Qu.: 411.0	1st Qu.: 73.103	1st Qu.: −76.072	1st Qu.: −4.064	1st Qu.: 0.00000	1st Qu.: 0.00000
Mode: character	Median: 2007-04-01	Median: 20.00	Median: 545.0	Median: 80.000	Median: −48.681	Median: −3.026	Median: 0.00000	Median: 0.00000
NA	Mean: 2007-03-04	Mean: 23.99	Mean: 549.8	Mean: 80.937	Mean: −54.991	Mean: −3.267	Mean: 0.04722	Mean: 0.03331
NA	3rd Qu.: 2008-01-01	3rd Qu.: 33.00	3rd Qu.: 678.0	3rd Qu.: 80.000	3rd Qu.: −27.934	3rd Qu.: −2.070	3rd Qu.: 0.00000	3rd Qu.: 0.00000
NA	Max.: 2013-10-01	Max.: 114.00	Max.: 850.0	Max.: 363.459	Max.: −1.245	Max.: −0.115	Max.: 1.00000	Max.: 1.00000

1-year (four quarters) forecasts for loss reserve purpose. For unemployment rate, we assume a constant of 5%.

For the severely adverse scenario, we assume the HPI has a 24% annual depreciation (corresponding to 2% monthly depreciation) over the next five quarters and then recovers to the normal 3% appreciation for the rest eight quarters. We assume unemployment rate annually to increase 120% (corresponding to 10% monthly increase) in the next five quarters and then decrease in half of that rate (5% monthly).

Baseline

$$\mathrm{HPI}(t) = \mathrm{HPI}(t-1) * 1.0025, \quad 1 \leq t \leq 39 \tag{6.23}$$

$$\mathrm{UER}(t) = 0.05, \quad 1 \leq t \leq 39 \tag{6.24}$$

Severely Adverse

$$\mathrm{HPI}(t) = \begin{cases} \mathrm{HPI}(t-1) * 0.98, & 1 \leq t \leq 15 \\ \mathrm{HPI}(t-1) * 1.0025, & 16 \leq t \leq 39 \end{cases} \tag{6.25}$$

$$\mathrm{UER}(t) = \begin{cases} \mathrm{UER}(t-1) * 1.10, & 1 \leq t \leq 15 \\ \mathrm{UER}(t-1) * 0.95, & 16 \leq t \leq 39 \end{cases} \tag{6.26}$$

The adverse scenario is simply assumed as the middle point of the baseline and severely adverse.

Adverse *Average of Baseline and Severely Adverse.*

The three scenarios for the home price index and unemployment rate are shown in Figs. 6.10 and 6.11.

For the severely adverse (SA) scenario, HPI cumulatively depreciates about 27% over the next five quarters, which is close to the HPI depreciation magnitude during the 2008 housing crisis, though we have a much shorter duration of five quarters instead of 5 years. The corresponding adverse (AD) scenario represents a milder housing market down with the maximum cumulative HPI depreciation of 11% and a recovery to the 5% HPI depreciation at the end of the 13 quarters.

For the unemployment rate, under the severely adverse scenario, we assume the maximum jump at about 400%, which is close to what was experienced during the 2020 pandemic (from 3.5% February 2020 to 14.7% April 2020 as shown in Fig. 6.12), though we use a longer duration of five quarters instead of just 2 months. The corresponding adverse (AD) scenario represents a milder job market downturn of 250% jump during the next five quarters and return to the long-term average at the end of the 13 quarters.

The magnitude of our stress scenarios may be a little extreme; however, we would use these scenarios as an illustration that any scenario design should have good interpretation.

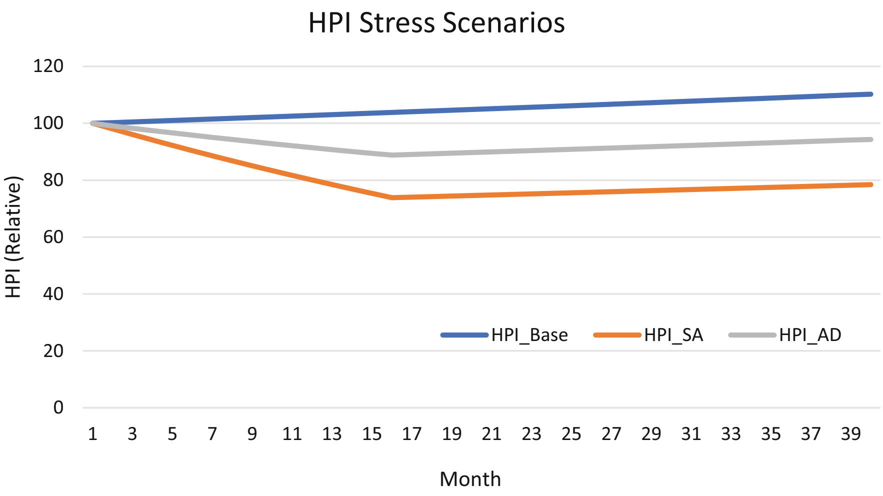


Fig. 6.10 Home price index stress scenarios

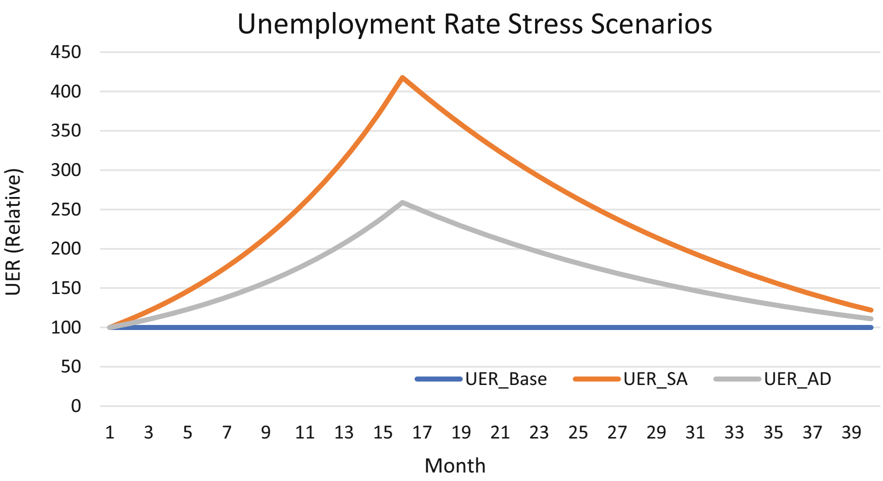


Fig. 6.11 Unemployment rate stress scenarios

6.4 CCAR Models

In this section, similar to the CECL modeling process in Sect. 4.3, we build the component models for stress testing based on the transaction data sets described in Sect. 6.3.1. With the same logic, for the default and prepayment risks, we adopt the competing risk modeling framework and use the binary approximation to estimate the default and prepayment probabilities due to its computing efficiency.

Fig. 6.12 National unemployment rate

As in CECL modeling, model selection is important since good model performance in loss forecasting accuracy is commonly required in stress testing. Under the competing risk framework, model selection for multinomial models is complex and lacks a good procedure. So, our solution is carrying out model selection under the target risk framework as illustrated in Sect. 3.1.1 for individual risk event independently using the AEVS procedure. As a further investigation on how AEVS perform on different transaction data samples presented in Sect. 6.3.1, we carry out AEVS on multiple transaction sets and make recommendation on the final selected models. Following the competing risk framework, such selected models for both default and prepayment risks are recalibrated using the binary approximation to obtain the unbiased parameter estimates under the competing risk framework for loss forecasting.

For LGD models, different from the multi-resolution approach used for CECL LGD modeling in Sect. 4.3.2, we follow the LGD modeling approach used in the regulatory stress test framework given that the reference default data set for our subprime mortgage portfolio includes resolution time information. So, we first build a resolution time model as in (6.16) and then include the projected resolution time in a LGD regression model as in (6.17). As in CECL modeling, EAD is calculated according to the formulas given in Sect. 3.1.5.1 for committed credit products.

These component models are integrated with the MEV scenarios described in Sect. 6.3.2 in the following section to produce the loss forecasts under different scenarios for snapshots of our sample subprime mortgage portfolio.

6.4.1 PD and PP Models

As loss forecasting models, PD and PP models in stress testing follow the same process as described in Sect. 4.3.1 for CECL, which include data preparation, model selection, and model ranking. For the purpose of comparison, we go over the same process for stress testing by showing some differences between CECL and stress testing. Same as CECL, we start from the target risk framework, in which a binary variable marks the target risk event (default or prepayment) as 1 if it occurs and as 0 else. In the transaction data sets, two binary variables PD-Event and PP-Event are used for marking default or prepayment, respectively. They act as the dependent variable in the binary logit model we will build. We focus on the default model, as the prepayment modeling fits in the same modeling process.

Data Preparation

The transaction data sets presented in Sect. 6.3.1 include the loan ID (ID), snapshot date (SnapDate), loan transaction date (ActDate, now shown), loan age at each transaction date (age), as well as the three loan characteristic measures at origination – acquisition index (aci), loan to value ratio at origination (oltv), and loan amount at origination (LoanSize, now shown). The performance of the loan in the transaction data is recorded by the default risk event variable DF-Event, which is a binary variable taking value 1 when there is a default event at the transaction date and 0 otherwise.

For macroeconomic variables, we include four candidate variables. The house price index and unemployment rate at origination (OHPI and OUER) are static, i.e., not changing over the transaction path. The cumulative home price index change since origination (DHPI) and cumulative unemployment rate change since origination (DUER) are dynamic, i.e., changing over the transaction path. The two dynamic macroeconomic variables (DHPI and HUER) and the two loan origination characteristics (aci and oltv) are used in the prespecified PD/PP models. So, the synthetic risk events are simulated based on these four variables.

The three transaction data sets are based on annual, semiannual, and quarterly snapshots on the full simulated synthetic transaction data set.

Model Selection

Our pool of candidate variables include these seven variables – aci, oltv, DHPI, DUER, OHPI, OUER, and LoanSize. There are $2^7 = 128$ variable combinations, which we consider relatively small and choose the recursive AEVS algorithm described in Sect. 4.3.1.

Model Ranking

Survived models from the recursive AEVS algorithm are ranked by the three goodness-of-fit measures – SIC, AIC, and CV_ν. We choose $\nu = 11$ for the ν-fold cross-validation measure due to computational efficiency. For the three goodness-of-fit measures, we prefer SIC and use it as the first ranking criteria. Top five models based on SIC are selected, and their ranks based on AIC and CV_v are used to compute the full average ranks.

The goodness-of-fit measures and their ranks for the top five selected models based on SIC are presented in Tables 6.4 and 6.5 for the annual snapshot panel data set and the semiannual snapshot panel data set, respectively. Note that default models may have the same CV_v measure (ties for some of the top models), which means these models have the same classification error defined by the cost function (see Fig. 4.3 R-code) for the binary outcome.

Table 6.5 shows that AIC tends to select models with more variables, which causes the average rank picks the second model as the top model. This happens more often with larger panel data set when the deviance part becomes more dominant compared to the penalty part in the AIC criterion. Even with SIC, this could be the case as shown in Table 6.6. For this quarterly snapshot panel data, the extra penalty for an additional variable included is $\log(N) = \log(343217) = 12.75$, which is less than the deviance difference between the true model and the top model selected by SIC (with LoanSize added in the model). The deviance difference between some other large models and the true model is even larger than 2 and 3 times of the single extra dimension penalty, such that the true mode only ranks as the fourth best model by SIC. This explains that when data noise increases, even the more conservative selection criterion like SIC tends to overfit the model. In such case, a further analysis with business inputs is required.

Figure 6.13 presents the model fitting results for the top model using the annual snapshot panel data set. This is the model we used to simulate the synthetic dynamic transaction data. This example demonstrates that AEVS performs well on annual snapshot panel data set as model fitting data.

Figure 6.14 presents the model fitting results for the top model using the semiannual snapshot panel data set. This is also the model we used to simulate the synthetic dynamic transaction data. This example demonstrates that AEVS performs well on semiannual snapshot panel data set as model fitting data.

Figure 6.15 presents the model fitting results for the top model using the quarterly snapshot panel data set. This is not the model we used to simulate the synthetic dynamic transaction data. The model has the additional factor LoanSize added. Note that LoanSize doesn't have high correlation with other factors in the model. As we explained early, this is due to the dominance of deviance in the SIC criterion when data noise increases with data set size. After checking the LoanSize factor, it turns out that it is not a significant risk factor, and we drop it from the model.

6.4.2 LGD Models

Similar to CECL, LGD in stress testing is based on total net accounting loss (including costs) as a percentage of the exposure at default, and the actual LGD for default loans can be calculated according to the formula (3.58) and the components descried in Sect. 3.1.4.1.

Table 6.4 Top five default models selected (annual snapshot panel data)

Model	SIC	AIC	CV11	Rank SIC	Rank AIC	Rank CV	AveRank
DF_Event ~ aci + oltv + DHPI + DUER	32424.11	32377.24	0.046621225	1	3	1	1.67
DF_Event ~ aci + oltv + DHPI + DUER + LoanSize	32430.13	32373.89	0.046621226	2	2	3.5	2.50
DF_Event ~ aci + oltv + DHPI + DUER + OHPI	32434.05	32377.81	0.046621226	3	4	3.5	3.50
DF_Event ~ aci + oltv + DHPI + DUER + OUER	32434.15	32377.91	0.046621226	4	5	3.5	4.17
DF_Event ~ aci + oltv + DHPI + DUER + LoanSize + OHPI	32437.92	32372.30	0.046621226	5	1	3.5	3.17

Table 6.5 Top five default models selected (semiannual snapshot panel data)

Model	SIC	AIC	CV11	Rank SIC	Rank AIC	Rank CV	AveRank
DF_Event ~ aci + oltv + DHPI + DUER	64541.90	64491.62	0.047091026	1	4	2	2.33
DF_Event ~ aci + oltv + DHPI + DUER + LoanSize	64545.11	64484.77	0.047091026	2	2	2	2.00
DF_Event ~ aci + oltv + DHPI + DUER + OUER	64551.11	64490.77	0.047091026	3	3	4.5	3.50
DF_Event ~ aci + oltv + DHPI + DUER + OHPI	64552.55	64492.21	0.047091026	4	5	4.5	4.50
DF_Event ~ aci + oltv + DHPI + DUER + LoanSize + OHPI	64552.94	64482.55	0.047091026	5	1	2	2.67

Table 6.6 Top five default models selected (quarterly snapshot panel data)

Model	SIC	AIC	DEV	CV11	Rank SIC	Rank AIC	Rank CV	AveRank
DF_Event ~ aci + oltv + DHPI + DUER + LoanSize	128987.51	128923.04	128911.04	0.0472179	1	4	5	3.33
DF_Event ~ aci + oltv + DHPI + DUER + LoanSize + OHPI	128988.46	128913.24	128899.24	0.0472179	2	2	1.5	1.83
DF_Event ~ aci + oltv + DHPI + DUER + LoanSize + OHPI + OUER	128988.83	128902.86	128886.86	0.0472179	3	1	1.5	1.83
DF_Event ~ aci + oltv + DHPI + DUER	128990.77	128937.04	128927.04	0.0472179	4	5	3.5	4.17
DF_Event ~ aci + oltv + DHPI + DUER + LoanSize + OUER	128993.60	128918.38	128904.38	0.0472179	5	3	3.5	3.83

```
Call:
glm(formula = frm_df, family = binomial(link = "logit"), data =
EPData2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.5779  -0.3276  -0.2937  -0.2659   2.9439

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.5153722  0.0939116 -37.433  < 2e-16 ***
aci         -0.0009462  0.0001054  -8.973  < 2e-16 ***
oltv         0.0091631  0.0010074   9.096  < 2e-16 ***
DHPI        -0.0101983  0.0005472 -18.638  < 2e-16 ***
DUER         0.1007436  0.0127724   7.888 3.08e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 32789  on 86998  degrees of freedom
Residual deviance: 32367  on 86994  degrees of freedom
AIC: 32377

Number of Fisher Scoring iterations: 6
```

Fig. 6.13 Top default model selected (annual snapshot panel data)

```
Call:
glm(formula = frm_df, family = binomial(link = "logit"), data =
EPData1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.6952  -0.3293  -0.2950  -0.2671   2.9506

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.523e+00  6.577e-02  -53.57   <2e-16 ***
aci         -9.507e-04  7.487e-05  -12.70   <2e-16 ***
oltv         9.445e-03  7.000e-04   13.49   <2e-16 ***
DHPI        -1.029e-02  3.840e-04  -26.78   <2e-16 ***
DUER         1.035e-01  8.899e-03   11.63   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 65339  on 172070  degrees of freedom
Residual deviance: 64482  on 172066  degrees of freedom
AIC: 64492

Number of Fisher Scoring iterations: 6
```

Fig. 6.14 Top default model selected (semiannual snapshot panel data)

```
Call:
glm(formula = frm_df, family = binomial(link = "logit"), data =
EPData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7724  -0.3299  -0.2959  -0.2678   2.9529

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.567e+00  4.882e-02 -73.077  < 2e-16 ***
aci         -9.297e-04  5.290e-05 -17.575  < 2e-16 ***
oltv         9.161e-03  4.953e-04  18.497  < 2e-16 ***
DHPI        -1.011e-02  2.720e-04 -37.156  < 2e-16 ***
DUER         9.903e-02  6.295e-03  15.731  < 2e-16 ***
LoanSize     1.888e-07  4.686e-08   4.029  5.6e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 130588  on 343216  degrees of freedom
Residual deviance: 128911  on 343211  degrees of freedom
AIC: 128923

Number of Fisher Scoring iterations: 6
```

Fig. 6.15 Top default model selected (quarterly snapshot panel data)

The LGD data processing for stress testing is the same as for CECL on the default reference data as described in Sect. 3.1.4.2. For LGD modeling, Sect. 3.1.4 presents three types of LGD models popularly used in credit risk modeling – the micro-structure dynamic models, regression models, and multiple resolution models. For CECL in Chap. 4, we adopt the multiple resolution models. Here, following the LGD model used in the regulatory stress test framework, we adopt the two-stage model approach with the first stage to build the resolution time model for our subprime mortgage portfolio. In the second stage, we include the resolution time in the LGD regression model. The stress LGD forecast combines the resolution time model and the LGD regression model.

Data Preparation

We use the same reference default data as prepared for the CECL LGD model in Sect. 4.3.2 with the addition of the resolution times since default date. Table 6.7 presents the summary statistics of the reference default data which corresponds to Table 4.6 in Sect. 4.3.2.

The reference default data include the mortgage loan ID, default date (DefDate), resolution timeline in months since default (Resol_TL), and some loan characteristics – loan to value at origination (oltv), occupation types (OccType) and the loan resolution types (Reso_Type), and actual LGD in accounting (LGD Acc).

Table 6.7 Summary information for a sample reference default data

ID	DefDate	oltv	chpi	OccType	Reso_Type	LGD_Acc	Resol_TL
Length: 4847	Length: 4847	Min.: 1.528	Min.: 0.3545	Length: 4847	Length: 4847	Min.: 0.0000	Min.: 1.00
Class: character	Class: character	1st Qu.: 71.339	1st Qu.: 0.6059	Class: character	Class: character	1st Qu.: 0.3850	1st Qu.: 4.00
Mode: character	Mode: character	Median: 79.300	Median: 0.7217	Mode: character	Mode: character	Median: 0.5926	Median: 9.00
NA	NA	Mean: 80.155	Mean: 0.7531	NA	NA	Mean: 0.6016	Mean: 11.96
NA	NA	3rd Qu.: 80.000	3rd Qu.: 0.8675	NA	NA	3rd Qu.: 0.7897	3rd Qu.: 17.00
NA	NA	Max.: 328.828	Max.: 2.8278	NA	NA	Max.: 1.5000	Max.: 89.00

Macroeconomic variables are created separately as described in Sect. 6.3.2 for all stress scenarios and then merged with the reference default data. In Table 6.7, we have the cumulative home price index (chpi), which is defined as the ratio of the current hpi and the hpi at origination at MSA (metropolitan statistical area) level. The chpi in the reference default data is calculated at the default time for each loan. The chpi forecasts under stress scenarios are derived from the HPI forecasts in Sect. 6.3.2. For our subprime portfolio, chpi is assumed the most important macroeconomic driver in the LGD models.

Model Selection

For the two-stage LGD model, model selection becomes relatively simple. Since the resolution time model assumes the resolution time is driven by macroeconomic variables, we need only focus on selection of the proper relationship between the resolution time and the key macroeconomic variable. An easy choice is the loglinear model commonly used for a positive response variable like the time.

The resolution time is highly correlated to the resolution type – as the short sales have the shortest resolution time in average, the REO loans have the longest resolution time in average, and the third-party sales (TPS) are in the middle. So, to include the resolution time into the LGD regression model, we categorize the resolution time into three buckets, within 6 months, longer than 6 months but within 12 months, and longer than 12 months. This categorization of the resolution time largely maps the resolution time to resolution types, and that is why we claim that the two-stage model with the first stage of a resolution time model is a special case of multiple resolution models.

Once the resolution time is included in the LGD model as a categorical variable, we fit simple linear regression models for loss severity with this categorical variable and other loan characteristics at origination and macroeconomic variables observed at the default time or better at the resolution time. The model is equivalent to assuming a base LGD for each resolution time category adjusted by the same suite of loan characteristics at origination and macroeconomic variables. So, it is a simplified version of the multiple resolution model, which can include different suite of risk factors under different resolution types.

Model Results and Performance

For the resolution time model, we choose a loglinear model, which takes the impact of risk factors to the resolution time in multiplication format. Figure 6.16 presents the bivariate plot of Resol_TL and chpi, which shows a clear pattern of negative relationship between these two variables. For the variable selection with this loglinear regression, we adopt the backward stepwise selection procedure – we start with the full model having all variables in the candidate pool and then removing variables with less contribution to the F-statistics. Since we have a small pool of candidate variables, the backward stepwise selection procedure is reasonable in modeling complexity. The best model includes two variables, loan to value at origination (oltv) and cumulative home price index (chpi). Figure 6.17 presents this loglinear regression model results.

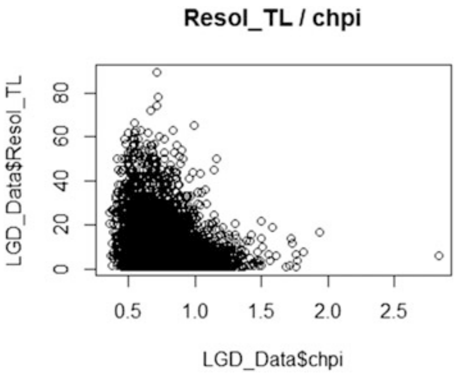


Fig. 6.16 Scatter plot of Resol_TL vs. chpi

```
> summary(lm(log(Resol_TL)~ chpi + oltv, data = LGD_Data  ))

Call:
lm(formula = log(Resol_TL) ~ chpi + oltv, data = LGD_Data)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4952 -0.5826  0.1052  0.6946  2.6953

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.930679   0.069102  42.411  < 2e-16 ***
chpi        -1.421000   0.070543 -20.144  < 2e-16 ***
oltv         0.002454   0.000666   3.685 0.000231 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9525 on 4844 degrees of freedom
Multiple R-squared:  0.0773,    Adjusted R-squared:  0.07692
F-statistic: 202.9 on 2 and 4844 DF,  p-value: < 2.2e-16
```

Fig. 6.17 Resolution timeline initial model

A further looking into this model shows that the oltv factor does not have the same level of significance compared to the macroeconomic variable chpi, and it is a weak predictor. So, we decided to select a simpler model with only the macroeconomic variable chpi as the predictor to be the final model. Figure 6.18 presents the final resolution time line model results.

Figure 6.16 shows that there could be some outliers in both directions of resolution timeline and chpi. To reduce the impact from these outliers, alternatively we fit a robust loglinear regression model using the R package ***robustbase***. Figure 6.19 presents the model fit results for this robust version, which is not very different from the final model. So, we stick with the selected final model.

```
> summary(lm(log(Resol_TL)~ chpi, data = LGD_Data  ))

Call:
lm(formula = log(Resol_TL) ~ chpi, data = LGD_Data)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4775 -0.5914  0.1043  0.7012  2.5795

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.08996    0.05398   57.24   <2e-16 ***
chpi        -1.37129    0.06933  -19.78   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9537 on 4845 degrees of freedom
Multiple R-squared:  0.07471,   Adjusted R-squared:  0.07452
F-statistic: 391.2 on 1 and 4845 DF,  p-value: < 2.2e-16
```

Fig. 6.18 Resolution timeline model selected

```
> library(robustbase)
> summary(lmrob(log(Resol_TL)~ chpi, data = LGD_Data  ))

Call:
lmrob(formula = log(Resol_TL) ~ chpi, data = LGD_Data)
 \--> method = "MM"
Residuals:
     Min       1Q   Median       3Q      Max
-2.54970 -0.63249  0.06371  0.65487  2.74494

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.20679    0.05975   53.67   <2e-16 ***
chpi        -1.47111    0.07736  -19.02   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Robust residual standard error: 0.9319
Multiple R-squared:  0.08256,   Adjusted R-squared:  0.08237
Convergence in 9 IRWLS iterations
```

Fig. 6.19 Alternative resolution timeline model

```
> summary(lm(LGD_Acc ~ chpi + oltv + resol_len_class, data = LGD_Data  ))

Call:
lm(formula = LGD_Acc ~ chpi + oltv + resol_len_class, data = LGD_Data)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.88358 -0.15547 -0.03108  0.12760  1.06714 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)       0.7668279  0.0177935   43.10   <2e-16 ***
chpi             -0.6281468  0.0175026  -35.89   <2e-16 ***
oltv              0.0025145  0.0001592   15.79   <2e-16 ***
resol_len_class2  0.0954397  0.0083755   11.39   <2e-16 ***
resol_len_class3  0.2377415  0.0078560   30.26   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2273 on 4842 degrees of freedom
Multiple R-squared:  0.3914,    Adjusted R-squared:  0.3909 
F-statistic: 778.4 on 4 and 4842 DF,  p-value: < 2.2e-16
```

Fig. 6.20 Stress test LGD model

The next step in the two-stage model is to fit the LGD model with the resolution timeline variable. As mentioned early, we fit a linear regression model with the resolution timeline as a categorical variable, plus loan characteristics and macroeconomic variables. The macroeconomic variables are observed at the resolution time. Similar backward stepwise variable procedure is used. The results of the selected final model are presented in Fig. 6.20.

Besides the resolution timeline variables, oltv and chpi are strong predictors of the LGD linear regression model. This is expected as in the CECL LGD model presented in Sect. 4.3.2, and the interpretation of these risk factors should be the same as for the CELC LGD model. The two resolution time classes shown in Fig. 6.20 are indicators for time being in the second class (more than 6 months but within 12 months) and the third class (more than 12 months). The positive coefficients for these two classes indicate that longer resolution time leads to higher LGD.

One difference between the CECL LGD model and the LGD model here for stress tests is the observation time for the macroeconomic variable chpi. Here the LGD model takes chpi observed at the predicted resolution time when doing forecast, while the CECL LGD model takes the chpi observed at the assumed default time. Though both need to be projected for the loss forecast, the former requires forward time projection and thus adds model complexity in forecasting.

6.5 Model Integration and Loss Forecasting

As pointed in CECL modeling, loss forecasting is a process consisting of several critical components, including portfolio identification, portfolio forecasting data processing, model integration and implementation, and forecasting result analysis and reporting. Similarly as in Sect. 4.4 for CECL, for stress test loss forecasting, we will also briefly go over these components, especially the model integration, for which proper modeling frameworks are required to meet the concept soundness of the loss forecasting process as a part of the requirement by model validation in the next section.

For stress test, the forecasting implementation also needs to take care of incorporating various scenarios represented by different sets of macroeconomic factor projections together with loan characteristics and other dynamic drivers needed. In the following, using the subprime mortgage portfolio, we demonstrate a concise flow of forecasting process for stress test with an intention of automation.

6.5.1 Portfolio Identification and Forecasting Data Processing

Portfolio identification is always an important initial step in loss forecasting implementation as it relates to an important issue in model governance – model use. Models could be developed based on data from multiple portfolios or even third-party data. Model use coverage is a top issue to be addressed at the beginning of model development. Only after the model coverage for the specified portfolio is certified, the forecasting implementation can start to process the portfolio data.

For stress test, we carry out loss forecasting for the same portfolio from our subprime mortgage loans as used for the CECL loss forecasting examples in Sect. 4.4.1. To reflect the impact of macroeconomic factors as in stress test, we assume the portfolio transition dynamically depends on macroeconomic factors as the dynamic portfolio presented in Table 4.8. Also, for the purpose of comparison, we use the same portfolio in CECL loss forecasting exercises for stress test loss forecasting.

6.5.2 Model Implementation and Integration

The stress test loss forecast is the aggregation of predicted periodic (monthly) losses over the specified future period and under a specified scenario for each loan in a portfolio:

$$\mathrm{Loss}_{\mathrm{CCAR}}(p, s) = \sum_{i=1}^{N_{\mathrm{p}}} \sum_{t=1}^{T_i} \mathrm{UPB}_i(t, s) * \mathrm{PD}_i(t, s) * \mathrm{LGD}_i(t, s) \tag{6.27}$$

where p indicates the target portfolio with total N_{p} loans. $\mathrm{UPB}_i(t, s)$, $\mathrm{PD}_i(t, s)$, and $\mathrm{LGD}_i(t, s)$ are unpaid balance, default probability, and loss severity of the i-th loan under scenario s and at period t of the prespecified periods of T_i.

Given the conditional periodic default and prepayment probabilities projected from the models, as well as the amortization rate, $\mathrm{UPB}_i(t, s)$ can be computed from (3.78). It should be pointed out that the models we fit in Sect. 6.4.1 are under the independent risk event framework. To apply (3.78), the default and prepayment probabilities under the competing framework should be projected, for which we use the binary approximation as described in Sect. 3.1.2.2. Figures 6.21 and 6.22 show

```
Call:
glm(formula = frm_df, family = binomial(link = "logit"), data =
EPData2_df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.5681  -0.3324  -0.3000  -0.2731   2.9062

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.4624469  0.0939693 -36.847  < 2e-16 ***
aci         -0.0008768  0.0001055  -8.308  < 2e-16 ***
oltv         0.0085893  0.0010100   8.504  < 2e-16 ***
DHPI        -0.0096600  0.0005487 -17.607  < 2e-16 ***
DUER         0.0946710  0.0127996   7.396  1.4e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 32508  on 84108  degrees of freedom
Residual deviance: 32134  on 84104  degrees of freedom
AIC: 32144

Number of Fisher Scoring iterations: 6
```

Fig. 6.21 Binary approximation default model for stress test

```
Call:
glm(formula = frm_pp, family = binomial(link = "logit"), data =
EPData2_pp)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8885  -0.3015  -0.2445  -0.1911   3.1027

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.5875699  0.1252860  -20.65   <2e-16 ***
aci          0.0021743  0.0001233   17.63   <2e-16 ***
oltv        -0.0208960  0.0013909  -15.02   <2e-16 ***
DHPI         0.0216544  0.0009043   23.95   <2e-16 ***
DUER        -0.2284562  0.0146625  -15.58   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 25081  on 82942  degrees of freedom
Residual deviance: 24056  on 82938  degrees of freedom
AIC: 24066

Number of Fisher Scoring iterations: 6
```

Fig. 6.22 Binary approximation prepayment model for stress test the binary approximation model results corresponding to the results in Fig. 6.13 using the annual snapshot panel data.

The two macroeconomic variables, HPI and UER, are projected according to (6.23) to (6.26) described in Sect. 6.3.1. The macroeconomic variables DHPI, DUER, and chpi used in the PD, PP, and LGD models are derived from the HPI and UER projection.

The amortization factor is calculated from the monthly interest rate $I_i = r_i/12$ of the loan as:

$$\text{Amor}_i = \frac{I_i}{1-(1+I_i)^{-360}} \tag{6.28}$$

across all periods for 30-year mortgages.

$\text{LGD}_i(t)$ is predicted according to the linear regression model presented in Fig. 6.20 with the resolution time projected according to the loglinear model presented in Fig. 6.18.

6.5.3 Forecasting Result Analysis

Table 6.8 shows the summary of the stress loss forecasts for the selected portfolio. We aggregate two cumulative loss forecasts for the portfolio, one for the cumulative 15 months (5 quarters) and the other for the 27 months (9 quarters). The former is commonly used in SLU (Stress Loss Usage) application for the maximum cumulative stress loss (offset by PPNR), while the latter is for the regulatory CCAR report.

For this subprime portfolio with total 1328 loans, with high default and prepayment rates, the majority of loss happens at the early periods of the forecast timeline. We can see the aggregated losses between the two applications (SLU and CCAR) are not that big for all three scenarios given that there is a 12-month difference in loss cumulation time.

Among the three designed stress scenarios, the aggregated loss forecasts for this portfolio increase from baseline to adverse to severely adverse. For CCAR, the increment of the projected cumulative losses is almost double from baseline to adverse and triple from baseline to severely adverse. Although the loss projection process is not liner in MEVs under these different stress scenarios, the projected losses among these three scenarios are close to liner – projected loss under adverse scenario is close to the average of that under baseline and severely adverse.

Compared to the projected (lifetime) loss under CECL, the projected loss under the severely adverse scenario for CCAR has a higher loss ratio of 42.45% (vs. 35.5% as reported in Sect. 4.4.3), which indicates that the severely adverse scenario is more stress than the CECL scenario as we designed. Further investigation shows that the CECL loss ratio is higher than that under the adverse scenario even extending the cumulative time from 27 months to lifetime. This is largely due to the fact that our

Table 6.8 Summary of stress loss projections

Stress loss projection under three scenarios							
	Loss ($)			UPB ($)	Loss/UPB percentage		
	Base	Adverse	Severely adverse		Base	Adverse	Severely adverse
SLU	45,003,247	82,312,693	136,171,872	381,196,136	11.81%	21.59%	35.72%
CCAR	53,328,642	104,189,901	161,833,992	381,196,136	13.99%	27.33%	42.45%

portfolio has high PD and PP rates, and major losses happen in the early forecast periods during which CECL assumes relatively more stress MEV conditions than the adverse scenario.

6.6 Model Validation and Performance Monitoring

The general model risk management framework has been described in Sect. 4.5, and here we focus on the stress test model validation and performance monitoring. Similarly, we will cover the following main components of the model validation procedure: model scope and usage, input and assumptions, theory and design, implementation and output analysis, and performance monitoring.

There is a higher flexibility in using variant of models in stress test compared to other regulatory fields. FRB provides details about the models and methodologies used in the supervisory stress test (called Regulatory Stress Test Framework in this book) by its annual release of Supervisory Stress Test Methodology. While these models and methodologies provide some guidance for participant banks in their own CCAR implementation, there is no restriction to follow the supervisory models and methodologies. However, no matter what models used by participant banks, the requirements of comprehensive model validation and review, as well as a sound governance process, have been put on high standards by the regulators. This is especially important for large banks since models may be developed, derived, adopted, or even shared among portfolios and under different stress test frameworks. Given the complexity of line of businesses and products within large banks, models in use could easily have gaps in development, validation and review, and usage and other risks. Regulators have acknowledged that such risks need to be sorted out, and proper controls should be put in place. Further, periodic assessments of how such controls perform have also become regulatory focus.

By going through the model validation components and performance monitoring process for stress test, we would identify potential issues leading to the risks mentioned above and propose controls and assessments in practice.

6.6.1 Scope and Usage

Model scope defines the model coverage on products, and model usage defines the applications in which the model can be used. Very often model usage should cover model limitations and use restrictions. The importance of these two factors has been illustrated in Sect. 4.5.1, and here we further summarize the potential risks related to these factors and suggest some controls to mitigate such risks, especially for stress tests.

Model scope and usage should be clearly defined and communicated by model sponsor with all related participants as the initial step in model development and validation. They should be clearly documented in the model documents and validation reports. Potential risks related to model scope and usage are the following:

- Incorrect product coverage by the proposed model
- Inaccurate model exposure
- Wrong line of business involvement
- Wrong model development design
- Wrong model validation requirement applied
- Wrong model implementation
- Wrong model use
- Model development failure
- Rejection from model validation and review
- Violation of regulatory requirements

To avoid these risks, some controls should be put in place:

- Sufficient BRD (Business Requirement Document) with model scope and usage clearly described and approved by higher management
- Model development policy and procedure with a periodically updated model development document template including specific chapter/section on model scope and usage
- Model validation policy and procedure with a periodically updated model validation and review document template including specific chapter/section on model scope and usage

For stress test models, the usage of credit models is defined as credit loss projection for specified future periods of the products or portfolios under some scenarios. However, models should be clearly specified for what applications and under what stress test frameworks, e.g., models used for systematic stress testing may be different from that used for CCAR. As for scope, portfolio identification as we described in Sect. 6.5.1 is a good starting point. Model scope extension after model has been approved requiring both business judgment and model risk reassessment or even model revalidation.

For our subprime portfolio, we focus on the loss forecast under the regulatory stress test framework and developed a probability of default model based on snapshot sample, which is considered a closer approximation of the portfolio at reporting time. The two-stage LGD model with resolution time modeled in the first stage follows that from the supervisory stress test by FRB. We have discussed the credit models developed for our subprime portfolio show some uncommon behaviors with fast default and prepayment rates, so these models may be specific for our subprime portfolio and may not be extended to other portfolios.

6.6.2 Inputs and Assumptions

For model inputs and assumptions, Sect. 4.5.2 provides some general risks and risks related to loss forecasting. As for stress test loss forecasting, these risks generally apply, and we summarize these risks here:

- Random error for observed data
- Systemic error for observed data
- Incorrect use of proxy data
- Inconsistency of assumed data
- Improper data sampling
- Implied model assumptions
- Improper model assumptions
- Insufficient assessment of model assumptions

The following controls should be in place to avoid these risks:

- Data quality control, including outlier detection, hard limits, similarity tests, and machine learning-based data quality checks
- Model data sensitivity analysis
- Data reconciliation and benchmarking
- Proxy data analysis
- Data risk management related to assumed data
- Data sampling analysis
- Model assumption identification
- Model assumption sensitivity analysis
- Statistical tests on critical model assumptions

Data risk, which usually was considered a part of the IT management, is starting to become an independent risk taxonomy, especially when its downstream consumers are model-related applications. Still as model inputs, data risk related to models is considered part of the model risk and requires formal review and validation through some of the controls listed above.

As mentioned in Sect. 4.5.2, robust controls and processes must exist to ensure the completeness, timeliness, and integrity of key observed data inputs. This may include data reconciliations, independent review of manually input data, or other data quality controls. Developers should be able to demonstrate that data inputs are suitable for the model and consistent with model methodology. Any data proxies used must be identified, justified, and documented.

For stress test, models are built on observed data, so assessing the impact of observed input data errors to model output is still an essential part of model validation as sensitivity analysis. Model output confidence intervals are commonly used to measure the impact of random input data errors under certain distributions. When proxy data are used in the model, to avoid systemic input data errors, a population similarity analysis should be executed either qualitatively or quantitatively depending on the data availability.

For credit models in stress tests, in addition to the portfolio-related observed data, the assumed data related to various stress scenarios are important part of the input data. These assumed data represent the stress scenarios with future projections of macroeconomic factors and market shocks and require periodical update and constant maintenance. The common practice is using a specified system (e.g., Scenario

Manager) and dedicated supports. The Scenario Manager could become complex when it includes all stress scenarios under various stress test frameworks, and formal data risk management with large-scale database including data stewardship, metadata, data risk governance, and control should be applied to make sure the data quality, consistency, integrity, and performance efficiency.

Model assumption risk has been a topic not being fully addressed since the initial regulatory model risk management setup. A critical reason is the difficulty in identifying critical model assumptions and how to assess and measure the risk of these assumptions. Most often, a model is used in a specific area due to the modeling tradition without careful investigation of what assumptions have already been embedded in the model development process. Further, model assumptions could be convolved into the model concept soundness. So, a sound model validation practice should have a specific section on model assumption identification and assessment. Assumptions make the model conceptually sound; however, specific assumptions on model implementation and model use should all be assessed for potential risks if these assumptions break. Statistics tests are popular tools used for the assessment and measurement of the assumptions. All key assumptions underlying the model should be supported by initial and ongoing analysis and documentation, so that users are aware of any model limitations.

For the GLM models we used for risk event modeling in stress test, we assume the risk event probability meets the proportional odds ratio assumption with the logit link to the risk drivers. The goodness-of-fit statistics in Sect. 3.1.1.1 are the common measurements of how this assumption is satisfied based on the observed data. So, such statistics must be provided by default as a part of the model validation.

Specific assumptions related to model choice and implementation must be independently and continuously assessed. In our model selection process with AEVS, we used various criteria to measure the final selected model as the best model.

6.6.3 Theory and Design

In Sects. 4.5.3 and 5.5.3, we discussed some general model risks related to theory and design, which also apply to stress test. Here we present a summary of these risks:

- Models with no or insufficient theoretical support
- Models with mathematic or statistical errors
- Hard rejection due to errors in theory and design
- Difficulty in mitigation due to model theory and design errors

To avoid such risks, we suggest the following controls in model theory and design:

- A periodic compliance review with regulatory requirements
- A comprehensive review of theory and methodology in the related modeling area

- Expertise of the specific modeling areas and practices and sufficient training in model theories
- Simulation testing
- Comparison with alternative theories and approaches

First of all, as for capital modeling, stress test implementation such as CCAR needs to comply with the regulatory requirements issued periodically issued by the regulators, so a periodically compliance review is necessary.

For stress test, similar to CECL, credit models are used for loss forecasting. These credit models have been well studied, and corresponding theories behind these models have become mature in some areas, for example, the credit models introduced in Chap. 3. More efforts are put on how to correctly implement these models in specific areas, which include model input data creation, variable selection, and parameter calibration as we present in Sect. 6.4.

Identifying and assessing model errors in theory and design require understanding in depth of the specific modeling areas, for which a comprehensive review of theory and methodology in the related modeling area is a good start as we carried out in Sect. 6.2.3. In general, one can refer to our Chap. 3 for most popularly used models in credit modeling, including stress test loss forecasting. For model validation purposes, the modeling background provided in Chap. 3 is a foundation. On top of this foundation, simulation testing is a common practice to validate model theory and design, though it is considered conservative due to the strong conditions required in simulations.

A more appealing approach for validating model theory and design is comparison with alternative theories and approaches, since such comparison directly shows the advantages and disadvantages of proposed and alternative models. Further, the comparison may demonstrate if the proposed modes work as intended, are appropriate for the intended business purpose, and are conceptually sound and mathematically and statistically correct. We have introduced this approach for both CECL and capital model validation. Here, again using our subprime mortgage default data, we show the comparison of LGD two-stage model with resolution timeline adopted in the regulatory stress test framework with the multiple resolution model by multinomial logit regression used in Sect. 4.3.2 for CECL LGD modeling.

It should be pointed out that, although both the multiple resolutions model developed in Sect. 4.3.2 and the two-stage model developed in Sect. 6.4.2 are based on the same reference default data set, we reran the post default transaction simulation in Sect. 6.4.2 with a different random seed, which results in a slight difference between the target response variable LGD_Acc as shown in Tables 4.6 and 6.7. We consider this simulation-based difference is minor and would only generate a minor change in model parameters of the multiple resolution model if the model is redeveloped on the reference default data in Sect. 6.4.2.

We acknowledge that even we assume the two reference default data sets from Sect. 4.3.2 and 6.4.2 are the same, the comparison of the multiple resolutions model and the two-stage model is not data independent due to the fact that the two approaches use different observed data as intermediate response variables – the

Table 6.9 Comparison of 5Q LGD projections between two-stage model and multiple resolutions model

LGD_Base	LGD_AD	LGD_SA	alt_LGD_Base	alt_LGD_AD	alt_LGD_SA
Min.: 0.0304	Min.: 0.1281	Min.: 0.1937	Min.: 0.2024	Min.: 0.2189	Min.: 0.2232
1st Qu.: 0.3510	1st Qu.: 0.4407	1st Qu.: 0.5295	1st Qu.: 0.3155	1st Qu.: 0.3572	1st Qu.: 0.3978
Median: 0.3773	Median: 0.4649	Median: 0.5545	Median: 0.3359	Median: 0.3890	Median: 0.4590
Mean: 0.3754	Mean: 0.4647	Mean: 0.5602	Mean: 0.3429	Mean: 0.3971	Mean: 0.4654
3rd Qu.: 0.3962	3rd Qu.: 0.4832	3rd Qu.: 0.5744	3rd Qu.: 0.3551	3rd Qu.: 0.4239	3rd Qu.: 0.5245
Max.: 0.7694	Max.: 0.8571	Max.: 1.0023	Max.: 0.7463	Max.: 0.8337	Max.: 0.9469

multiple resolutions model uses the observed resolution types, while the two-stage model uses the observed resolution times. So, the following model comparison results may not be automatically extended to any modeling exercises.

Using both models, we project the LGD for the same target portfolio used in Sect. 6.5.3 over the rest lifetime of each loan. Table 6.9 shows a summary of the projected LGD over the first five quarters for these two models under the three scenarios. These projections are used in the SLU application.

For the three scenarios, the two-stage model produces a little higher LGD projections and wider spreads among the scenarios compared to the multiple resolution model. A further investigation on the LGD calculation with these two models and the portfolio data points to two main drivers. The first driver is based on the model property. The two-stage model as shown in Fig. 6.19 is highly sensitive to the HPI change (chpi), and this impact is direct to the final LGD projection. Although the multiple resolution model (refer to Fig. 4.8, Figs. 4.11 to 4.13) can be sensitive to the HPI change on the projected probabilities of different resolution types, the final LGD projection is a probability weighted average of the projected LGDs under all resolution types, and this average reduces the sensitivity of the final projected LGD on HPI change. The second driver is due to the portfolio we selected. As described in Sect. 4.4.1, we select the portfolio which is from the January 2006 snapshot, for which the current value of chpi at the forecasting date is very high. So, the absolute chpi change under the different scenarios is relatively high and results in higher LGD projection for the more sensitive model. This can also be seen from the wider spreads among the three scenarios for the more sensitive two-stage model. These two drivers lead to higher LGD projections by the two-stage model across the three scenarios.

Table 6.10 shows a summary of the projected first nine-quarter LGD for these two models under the three scenarios. It is similar to the first five-quarter LGD summary.

The impact of the different LGD models on the projected losses is shown in Table 6.11. We use the same PD/PP models and amortization factor for the selected

Table 6.10 Comparison of 9Q LGD projections between two-stage model and multiple resolutions model

LGD_Base	LGD_AD	LGD_SA	alt_LGD_Base	alt_LGD_AD	alt_LGD_SA
Min.: 0.001882	Min.: 0.1281	Min.: 0.1937	Min.: 0.1922	Min.: 0.2189	Min.: 0.2232
1st Qu.: 0.339281	1st Qu.: 0.4318	1st Qu.: 0.5244	1st Qu.: 0.3074	1st Qu.: 0.3662	1st Qu.: 0.4361
Median: 0.366400	Median: 0.4568	Median: 0.5482	Median: 0.3267	Median: 0.4003	Median: 0.5084
Mean: 0.364961	Mean: 0.4571	Mean: 0.5577	Mean: 0.3344	Mean: 0.4055	Mean: 0.4989
3rd Qu.: 0.387970	3rd Qu.: 0.4765	3rd Qu.: 0.5710	3rd Qu.: 0.3472	3rd Qu.: 0.4312	3rd Qu.: 0.5549
Max.: 0.769419	Max.: 0.8571	Max.: 1.0023	Max.: 0.7463	Max.: 0.8337	Max.: 0.9469

portfolio in the loss projection; thus the difference in projected losses is purely due to the LGD model.

Consistent with the higher LGD projection by the two-stage model, the projected losses for both SLU (cumulative five quarters) and CCAR (cumulative nine quarters) are higher than that by the multiple resolution model across all three scenarios.

In summary, the two LGD models produce close LGD projection, which is a validation of the two-stage LGD model we used. The two-stage LGD model is relatively more sensitive to the macroeconomic factor than the multiple resolution model. The multiple resolution LGD model is based on the projected probabilities of resolution types and the projected LGD under each resolution types through a weighted average. Such average may reduce the sensitivity on risk factors and depends more on historical observations. For the portfolio we selected, the two-stage LGD model is more conservative due to high HPI appreciation at the forecasting date. It should be pointed out that the conservativeness among these two models is data dependent and one can't draw the conclusion that the two-stage LGD model is always a more conservative LGD model.

6.6.4 Implementation and Output Analysis

Conceptually, a model is independent of the system or platform in which it resides. However, in practice, the performance of a model critically depends on the correct configuration and implementation in the system or platform.

A full modeling cycle includes model development and redevelopment, user acceptance testing (UAT), and the production delivery. Usually, these processes are carried out in different computing platforms and environments; however, in recent years the concepts of model continuous integration and continuous deployment or delivery (CI/CD) have become more and more popular. The CI/CD model implementation brings both efficiency in model production and new challenges for

Table 6.11 Comparison of loss projections with two-stage LGD model and multiple resolution LGD model

Projected loss ($)	Two-stage LGD model			Multiple resolution LGD model			
	Base	Adverse	Severely adverse	Base	Adverse	Severely adverse	UPB
SLU	45,003,247	82,312,693	136,171,872	37,421,397	63,893,147	99,504,483	381,196,136
CCAR	53,328,642	104,189,901	161,833,992	44,355,419	82,457,241	120,830,151	381,196,136

model implementation validations. In the following, we will cover different stages of modeling and implementation processes to address potential issues model validation could embrace.

Dev Environment

Model development is carried out in the model development environment, which is commonly called the Dev environment. The model Dev environment is the computing platform built for model developers to carry out all model development task, including data analysis, initial model design, and various testing related to model development. In recent years, the concept of integrated development environment (or IDE) has become popular. The IDE commonly defines a software platform including capabilities of data loading and visualization, user-friendly code editor which can provide smart programming, build and execution, debugging, and profiling for one or more programming languages. More powerful IDEs can even carry out unit-testing, code integration, version control, deployment, and delivery. There are both open-source and commercial IDEs. In recent years, open-source IDEs have grown quickly, for example, in modeling, RStudio, Jupiter Notebook, Apache NetBeans, Apache Spark, and many others.

For independent model validation and auditing, model validators and auditors are often granted the same level of accesses as model developers with the Dev environment. Such requirements are not just for independently replicating model results; they are also critical for model validation and auditing to assess the model development environment for potential model risks.

UAT Environment

The User Acceptance Test environment is the next level of software development platform after the model development. For large and complex model, the UAT layer is necessary for model usage testing before model production. The UAT environment requires efficient ETL processes and friendly interfaces for user to test their own data with the deployed model. A data lake is usually a great help for this purpose.

UAT is a safe buffer between model development and production, which allows model users to test and learn the model implementation. Any issues discovered in the UAT level can be sent back to model development for further investigation and testing. Since UAT is a less restricted environment comparing to the production environment, this can speed up the model redevelopment and testing.

For high-quality model production, the UAT layer should not be skipped. Very often, model users could miss the UAT for some model releases and cause auditing failure. In recent years, automation of the UAT layer has become popular for model production efficiency.

PROD Environment

Model production environment requires higher restriction for access and data security, since very often the production is delivered to clients (internal or external) and has exposure security requirements. Model production also has much less error tolerance, and any implementation issues could result in critical model risks.

Monitoring system is built on production environment for continuous assessment of the model production, which is a critical part of model validation and will be discussed further later.

CI/CD Design

In recent years, to enhance the efficiency of model production process, the concept of continuous integration, deployment, and delivery (CI/CD) has become popular, and CI/CD implementation has become a trend in new software platforms, especially in the cloud computing platforms. From the model development side, this is an automation of the coding, building, packaging, and delivery process and greatly shortens the model production timelines. However, for the validation and auditing sides, there could be some transitions with less transparency.

To overcome the transparency issue with CI/CD, one option is using the layered model development processes as we described with Dev, UAT, and PROD for the initial model release and only use CI/CD for model redevelopment and production update. The other option is adding intermediate testing within CI/CD as a UAT layer. Either way, CI/CD should not become the reason for a less transparent modeling process, especially for large complex models.

Model Output Analysis

A basic requirement for output analysis is the ability to replicate the model outputs on appropriate platforms, especially in the production platform. For outputs with random number generation, a random seed has to be set for result replication. Output replication also presents validator a chance to review the model development and implementation logic, design, and coding. It also helps to check the model documentation consistency with model development and implementation. A full model output results replication should be done periodically to check gaps which could be created by ongoing model updates.

Variation analysis is an essential part of model output analysis, especially for loss forecasting models. For model validation, variation analysis can be useful to access model behaviors and discover any model issues contradictory to business intuitions. Because, very often, line of business requires to assess and explain the variation of loss projections from quarter to quarter as a way to understand projection dynamics and take proper business actions on top of these projections.

Sensitivity analysis is used to assess how sensitive a model's outputs are to the change of model inputs, which could be data and assumptions. Models should have proper sensitivity to corresponding inputs. In general, stress test models should have proper sensitivity to macroeconomic data and other input as designed. Either too high or too low sensitivity would result in further investigation and model adjustments or overlay. Under such cases, the initial step should be checking if models have fundamental issues or the modeling data are abnormal. Such issues could lead to model rejection.

Models could be sensitive to model assumptions. The assessment of model assumptions should be a part of the model concept soundness. Critical model assumptions should have been included in model limitations. Model outputs should be assessed for some minor deviations from critical model assumptions. For example, time series models are built under the assumptions that the target time series variables are stationary as discussed in Sect. 3.3.1 and need to be tested. Time series models rejected by a strict stationarity test with a lower test level may pass a less strict test with a higher test level, and these models could still be used for stress test loss forecasting if they pass all other tests.

6.6.5 Performance Monitoring

Model performance monitoring is critical, especially for loss forecasting models which require continuous assessments. Although loss forecasting in stress test focuses on hypothetic scenarios, the credit models used for loss forecasting were developed with observed data, and model performance still requires continuous checking. So, for these models, ongoing performance assessment (OPA), back testing, and benchmarking are common practices for model performance assessments.

Ongoing performance assessment monitors whether production models continue to perform with the time advancing. With the changes in markets, products, exposures, activities, clients, or business practices, production models not updated promptly could deviate from the business trend and result in poor performance as measured by both statistical and business criteria. There is also the possibility that model assumptions could be broken and new model limitations are needed. Models on which business decisions depend require close performance monitoring in a timely manner; otherwise flawed and costly business decisions could be made before deterioration in model performance becomes apparent.

For forecasting models, there are too many factors that could be subject to change, and the loss forecast results require timely checking with the actual observations. Large deviation of the forecasts from the actuals indicates the production model is not performing at least with the current observation, and model enhancements or adjustments are required. There is also the requirement that when new data and information inputs are added, the production model should be competitive in the model selection process as required by the model selection criteria. So, both the performance measured by those statistical criteria in the model selection process and the model ranking in the model selection process should not deteriorate significantly.

For stress test, loss forecasting under hypothetic scenarios is not observable and thus cannot be back tested. For these forecasts under hypothetic scenarios, benchmarks are commonly used as the defending bars. We discuss these techniques next.

Back-testing involves the comparison of actual outcomes with model forecasts during a sample time period not used in model development and at an observation frequency that matches the forecast horizon or performance window of the model. The comparison is generally done using expected ranges or statistical confidence intervals around the model forecasts. When outcomes fall outside those intervals, one should analyze the discrepancies and investigate the causes that are significant in terms of magnitude or frequency. The objective of the analysis is to determine whether differences stem from the omission of material factors from the model, whether they arise from errors with regard to other aspects of model specification such as interaction terms or assumptions of linearity or whether they are purely random and thus consistent with acceptable model performance. Analysis of in-sample fit and of model performance in holdout samples (data set aside and not used to estimate the original model) is an important part of model development but is not a substitute for back-testing.

One can extend the one-period back test to a more comprehensive *walk-through test.* In a walk-though test, starting from a specified historical timeline with a specified forecast period (e.g., quarterly or yearly), the target model being tested (usually the production model) is refitted on the data before the selected historical timeline, and its projection on the next period is compared to that period's actual observations; then these actual observations are added to the historical data to refit the target model, and its projection for the next period is used to compare with the actual observation in the corresponding period. Continue this process to measure the target model's projection performance, while walk through all these periods to the most recent period. The walk-through test continuously measures the target model's performance over several periods to see its pertinent strength in performance instead of only the most recent period. It is also often used to detect whether the model catches the trend of the underlying changes.

Benchmarking is the comparison of a given model's inputs and outputs to estimates from alternative internal or external data or models. It can be incorporated in model development as well as in ongoing monitoring. For credit risk models in stress test, examples of benchmarks include models from different methodologies, vendor firms or industry consortia, and data from retail credit bureaus. Whatever the source, benchmark models should be rigorous, and benchmark data should be accurate and complete to ensure a reasonable comparison.

Discrepancies between the model output and benchmarks should trigger investigation into the sources and degree of the differences and examination of whether they are within an expected or appropriate range given the nature of the comparison. The results of the analysis may suggest revisions to the model. However, differences do not necessarily indicate that the model is in error. The benchmark itself is an alternative prediction, and the differences may be due to the different data or methods used. If the model and the benchmark match well, that is evidence in favor of the model, but it should be interpreted with caution so we do not get a false degree of comfort.

Benchmarking is especially important for stress testing given that the forecasting under hypothetic scenarios in stress test can't be back tested. Comparison of these hypothetic forecasting results with those alternative results under the same or similar hypothetic scenarios provides a reliability checking.

The multiple resolution model for LGD used in Sect. 6.6.3 is a good example of a benchmark model. It is a simpler and more popular model with an intermediate model on types of resolutions. The mapping between resolution type and resolution time connects the multiple resolution LGD model and the two-stage LGD model, and they can be benchmarks for each other.

6.6.6 Model Governance

On top of the previous components of model risk management is the model governance, which sets an effective framework with defined roles and responsibilities for clear communication of model limitations and assumptions, as well as the authority to restrict model usage.

The model risk management framework as shown in Fig. 4.18 is set up by the model governance through policies and procedures. The common practice is that an institute has an overall model risk policy, which covers all aspects of model risk management, including model and model risk definitions; assessment of model risk; acceptable practices for model development, implementation, and use; appropriate model validation activities; and governance and controls over the model risk management process. Then, within different legal entities, line of businesses, or modeling areas, there may be different model risk management policies and procedures. These policies and procedures cover specific model risk management activities, for example, specific policies for certain legal entities or specific procedures for model development, validation and monitoring, and reporting for a specified modeling area. Procedures usually provide more detailed guidance on the required activities.

Stress test models should comply to these policies and procedures. Additional model risk policies may be added to stress test models due to the periodic new regulatory requirements, new models adopted, extended model use, and critical data risks.

Underwriting and Credit Scoring

7

In the previous three chapters, we deal with credit risk modeling for loans and portfolios already in house. For financial institutions in the origination business, there is a critical credit risk at the door when doing underwriting, since the quality of the loans approved and funded will decide those risks we discussed in the previous chapters or the price if these loans are sold. The credit risk in underwriting is traditionally treated independently from the credit risks in the common credit risk management frameworks we presented early given that not all financial institutions carry origination businesses for products in their portfolios and credit risk embedded in underwriting is more considered business strategic risk or market risk from the underwriting line of businesses. Nevertheless, such risks are measured by credit events, and a similar credit risk modeling technique called credit scoring is dominantly used.

We include the credit risk modeling for underwriting in this chapter to demonstrate how credit risk models, such as those presented in Chap. 3, are used in credit scoring for underwriting. While credit underwriting covers a broad scope of techniques (called credit intelligence), we follow the recent developments in credit scoring models and focus on some efficient credit scoring applications, especially in small business lending led by some FinTech companies.

We first give a brief introduction to financial underwriting and the credit underwriting business and how risk rating, which includes credit scoring, is used in the underwriting process as a critical tool. Then we introduce how credit modeling is used in risk rating by focusing on the recent modeling development in credit scoring. The details on the credit scoring modeling process are presented step by step with a focus on building standard scorecards. We apply the modeling process on a research data set from small business lending by showing how to execute each of the steps. Again, programming is first provided in the prototype format for the purpose that readers can practice using their own preferred languages. Then R or Python code is partially provided as examples of implementation of these prototypes.

Model validation is a necessary component of the credit scoring models given the flexibility of various modeling approaches being adopted, especially with different

C. Chen, *Practical Credit Risk and Capital Modeling, and Validation*, Management for Professionals, https://doi.org/10.1007/978-3-031-52542-1_7

approaches for population selection, sampling, and model choice. Based on the full model validation framework introduced earlier in Chap. 4, we carry out a model validation based on the research data for small business lending and show that some of the arbitrary sampling method recommended in machine learning practices may not work.

7.1 Introduction to Underwriting and Risk Rating

Financial underwriting is the process of customer information collection and gathering, customer segmentation and risk classification, and decision-making and actions related to the sale of the financial product. So, financial underwriting is largely a customer assessment and selection process for specified financial products. These customers have been collected by the marketing operation, and some preselection process has been executed, for example, the popular four Rs – Response, Risk, Retention, and Revenue. However, the real customer assessment starts from the underwriting process.

The two biggest financial underwriting areas are credit underwriting and insurance underwriting. While both of these areas have a long and ancient history, it is accepted that the former may be much earlier, given the evidence that our ancestors took intelligent selection for giving and sharing with expected returns. The modern underwriting for both these areas took the form of professionals as credit managers or insurance agents. The fast-growing computing technology starting from the middle of the last century changed these areas differently. While the credit underwriting, especially with consumers, evolved from the manual process by credit managers to algorithm and model-based process by computers, the insurance underwriting largely still sticks to the field underwriting by the insurance agent (combined with more automated company underwriting). Such divergence may change in the future, but there are some fundamental differences between these two areas as shown in Table 7.1. Besides these differences, one of the main reasons is probably due to the complexity of the insurance policy, which requires face-to-face communications executed as trainings (called representations).

However, for both there is the same core of the underwriting process, the customer segmentation and risk classification, as there is risk related to the financial product. Such risk should be assessed not only due to the opponent sides of buy and sale; it should also be well communicated between both sides as required by regulations. This is essential as shown by the irresponsible mortgage underwriting, which led to the 2008 financial crisis.

In recent years, machine learning (ML) under the umbrella of artificial intelligence (AI) has exploded in credit underwriting led by some Financial Technology (FinTech) companies. Largely these applications call for complex non-parametric ML models (including those tree-based ML models presented in Chap. 3). As we discussed, such ML models have a big disadvantage in stability, and the most popular use is for their power to find patterns unknown (especially for new data) or validate and track patterns as we discussed in previous chapters. However, such

Table 7.1 Simple comparison of credit and insurance underwriting

Characteristics		Credit underwriting	Insurance underwriting
Difference	Product	Loan or credit product	Insurance policy
	Financial transaction	Lending	Premium payment
	Key measure	Creditworthiness	Mortality
	Factors	No physical hazard	Physical hazard (age, health)
	Type of underwriters	Company	Field and company
	Regulation	National	State
	Decision	Approval/reject	Eligibility
Similarity	Information collection	Application, internal and external data	Application, internal and external data
	Decision process	Segmentation and risk rating	Segmentation and risk rating
	Decision delivery	Explain decision results	Explain decision results

usage requires domain expertise to verify since often some "outstanding" patterns discovered by naïve ML turn to be invalid due to data or sampling issues. In addition, even valid ML models are often difficult to explain to the management and regulators, and their opacity causes issues where rejected customers need decision reasons in underwriting. There are also concerns that using extensive data (including alternative, non-financial data) by ML models may violate fair lending laws (the Student Borrower Protection Center's investigation on Upstart Holdings and other FinTech companies).[1] This is a warning that regulation weakness exists within underwriting for FinTech companies, and these companies should enhance their risk management, especially their model risk management, which doesn't exist for most of these FinTech companies. Even worse, some FinTech companies only focus on some smart selection processes under the ML and AI title and lack the rigorous implementation of a complete underwriting process. This has been reported in large number of frauds and criminal filings by courts with the PPP (Paycheck Protection Program during the COVID-19 pandemics) loan applications processed by FinTech companies without some basic underwriting sanity checks.

In the following, we present a brief introduction to the credit underwriting process, including the application and data collection, customer segmentation and risk classification, and decision-making and actions. We would like to show the relationship between underwriting and risk and further the importance of risk assessment by risk rating. Regulations for credit underwriting are also discussed with some recommendations on closing the current gaps.

[1] Student Borrower Protection Center (February, 2020), Educational Redlining

7.1.1 Credit Underwriting Process

Credit underwriting has a long history from the ancient time as an act of selection in sharing or giving with certain expected return. Over this long history, some insights and experiences have been collected, for example, the five Cs (character, capacity, collateral, capital, and conditions), which have been commonly used by credit managers in credit underwriting. Most of these factors are hard to measure (especially the first two) while partial information could be obtained through in-person interviews and other relationship channels. Although such manual underwriting process by credit managers could have high quality, the cost of heavy labor and slow processing speed could not meet the large volume of credit demands with the economy growth. This became a bottleneck in the underwriting business, especially for consumer lending, until the modern computing power enabled the execution of statistical models from the second half of the last century. In USA, regulations accelerated the adoption of such model-based nonjudgmental underwriting processes by issuing the Fair Credit Reporting Act (FCRA) in 1970 and the Equal Credit Opportunity Act (ECOA) in 1974.

The requirement for efficient underwriting by the credit managers is a tool of risk rating which measures the risk of borrowers' potential inability or unwillingness to honor their obligations. Such a tool should be built on techniques that can efficiently covert all proper information obtained (especially related to the five Cs) into the target risk measure and help on the decision-making. Algorithms and especially statistical models naturally fit in the task, and the broad scope of techniques are called credit intelligence by some practitioners.[2] As for the tool, FICO (from Fair, Isaac, and Company) becomes the dominant one as it almost covers the full population in USA and has international extensions. While the FICO score based on credit bureau data (from one of the three credit reporting agencies or CRAs – Experian, Equifax, and Transunion) is treated as a base for credit scoring, for specific consumer lending, lenders still prefer to develop product-specific credit scores and proprietary scoring systems. In the following, we show how such scoring systems fit into the underwriting process. We classify the underwriting process into three phases and explain their roles and relationships.

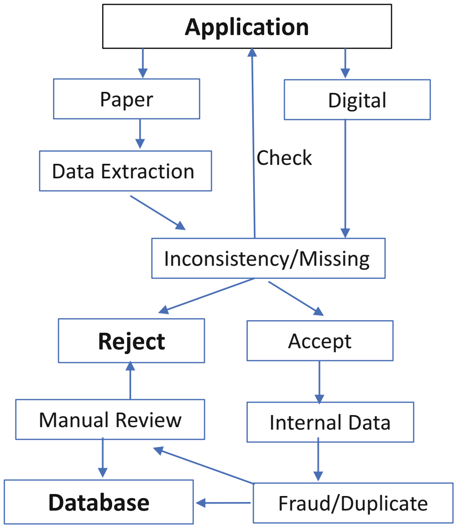
Loan Application and Data Processing

The direct result of marketing or solicitation for credit products, especially on consumer loans (credit card, auto loan, mortgage, personal loan, and small business loan), is the loan application. Figure 7.1 presents a loan application flowchart.

Given the fast growing of digital access to loan application, large portion of applications are submitted through online websites and cellphone apps, while there are applications through paper files. Paper applications require data extraction, which require manual entry and are more error-producing. So, in recent years,

[2] Anderson, A. R. (2022), Credit Intelligence & Modelling: Many Paths through the Forest of Credit Rating and Scoring. Oxford University Press, New York

Fig. 7.1 Loan application and data collection



banks encourage digital application by either adding promotions or providing better products with digital application. Also, banks have started to apply in-time error-checking with digital application, which greatly reduces application errors, repeated applications, and fraud applications. In the future, paper application may only be allowed for specific groups which could not have digital access given the growing ESG standard implementation in the financial industry.

Applications need go through the inconsistency and missing value check. Applications which do not pass this check will be returned and require re-application. For digital application, this check is commonly executed in time, and indications for such inconsistency and missing values are provided promptly, so applicants can fix such errors more easily and quickly.

Once the application is error-free, the application is accepted and goes through internal data for fraud and duplicate check. While it is relatively easy for the latter, fraud detection is more involved. In recent years, machine learning (ML) has been extensively applied in this area given that frauds are much rarer than credit events and fraud patterns have been changing. The popular fraud detection implementation in loan application is a combination of a set of rules and ML models. An alert is reported if any of the rules is broken or the ML model indicates a high probability of fraud. When the application fails the Fraud/Duplicate check indicated by a fraud alert or a duplication hit, the application needs a manual review. A further review note is commonly issued, and depending on the review result, extra documents and proofs may be required to resolve the alert. Otherwise, the application including all data is pushed into a database for storage purpose. This step may also include some sanity checks, for example, some identity checks with name, address, SSN, or EIN. An alert will be issued if any of such sanity checks fails and further review is required. Some powerful fraud detection tools incorporate such sanity checks. The application process is complete, and the next phase of segmentation and risk classification starts.

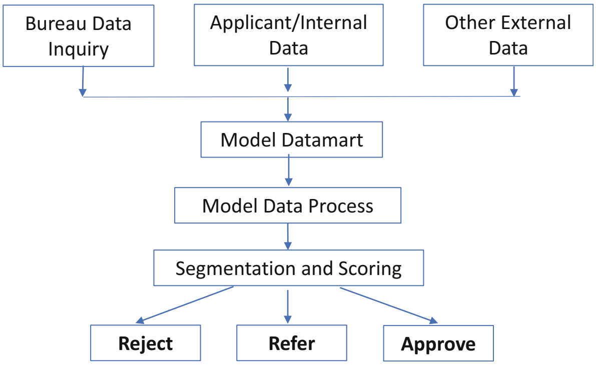


Fig. 7.2 Segmentation and risk classification

Segmentation and Risk Classification

This phase is the core part of the underwriting. It starts from information and data collection related to the applicant, the application, the product, as well as the economic environment. Figure 7.2 presents a flowchart of this phase.

The mark of the phase is a bureau credit data inquiry (called hard inquiry), which is a reported credit event in the bureau data collection and has an impact on the credit score when run on the bureau data. The bureau data inquiry is classified as hard inquiry and soft inquiry. A hard inquiry is defined as inquiry related to credit application initialized by the applicant, while the soft inquiry is defined as inquiry related to lender initialized events, e.g., marketing campaigns. Soft credit inquiry doesn't impact the credit score. The bureau data inquiry commonly include a FICO score based on the bureau data, as well as other bureau reported data, e.g., identity variables (name, address, phone for applicant life history), bankruptcy records, payment records for all credit liens (cards, auto, mortgage, any other loans, and credit lines), collection records, fraud and dispute records, and credit inquiries. The bureau data inquiry costs a fee paid to the Credit Reporting Agency (CRA), which could be considered expensive by the lender for a large number of applicants if not covered by the application fee. So, some lenders may carry out the bureau data inquiry for only some specific segments (usually applicants on the lower side of credit worthiness or new customers with less information) based on its underwriting policy.

The second block of data is the applicant data collected from the application phase, plus any other internal data on the same customer existing with other accounts or products. In recent years, banks increase cross-sale, and very often the applicant already had some other products with the bank in the past. Such internal data are valuable in customer identity, credit worthiness evaluation, payment habit, and financial conditions. Large financial institutions have started building large systems to efficiently use all customer information and create some internal risk rating called Customer Risk Rating (or CRR). Such CRRs can be measured for credit worthiness, fraud risk, profitability, and other matrices.

For smaller lenders, especially start-ups and FinTech companies, due to lacking customer information from internal data, they turn to sources of external data, which largely are non-financial data and are often called alternative data. Such alternative data are often easy and cheap to obtain, and they are abounded, e.g., customer shopping data, online and cellphone usage data, education, and other social public data. This is where big data processes play a significant role.

All related data are pulled from appropriate databases to form the model Datamart, which can efficiently store and carry ETL functions on these data. These data are also required to be efficiently processed to fit in as inputs to the underwriting segmentation and credit scoring models.

Customer segmentation is probably the most variant part in credit underwriting. The reason is that there are various flexible rules which can be applied on segmentation based on the applicant characteristics, application channels, credit products, collateral types, business units, business strategies, and even economic environments. Combining with credit scoring, segmentation can enhance the underwriting efficiency, for example, lender can make a decision on a segment based on risk of the entire segment instead of individual applicant.

Credit scoring is the main risk rating tool for consumer underwriting. The credit scoring model takes in the credit score drivers as inputs and produces a score based on the target risk measurement, which we will discuss later in this chapter. The credit scoring model should be able to identify the main drivers that impact the credit score, since a decision based on the credit score requires an interpretation based on the risk drivers. So, the risk drivers are required to be interpretable and compliant with regulation, and that is the most significant difference between credit scoring models and other credit risk models. For example, although some variables (e.g., race, religion, national origin, disability, education level) can be very significant in measuring the target risk, they are not allowed in the credit scoring model by regulation. We will further discuss this in Sect. 7.1.3.

The decision process based on segmentation and credit scoring defines the underwriting strategy, which could be complex and highly correlated to the business strategy. For example, segmentation can depend on customer age (which is allowed by regulation), and different credit score cut-offs could be used for different age segments – credit score cut-offs for younger applicants may be lower for their longer lifetime value. Another example is that credit score cut-offs could be higher for some geographical areas due to higher customer population already in bank's portfolios from those areas and the concern of high concentration risk on geographical areas, while the opposite underwriting strategy could be used when banks compete on new markets. Overrides are commonly used based on information not available in the system, largely from business side.

The decision based on the segmentation and credit scoring can be Approve, Reject, or Refer, and we will follow the corresponding actions from each of these decisions.

Decision and Action

Among the three decisions, Reject and Approve are self-clear, while Refer usually means that a different credit product is referred to instead of the original one, commonly a less risky credit product by reduced loan amount or lower credit limit or higher collateral. This is also called down-sell in the credit underwriting business. The applicant may take the referred product or not with or without a response. If the reference deadline passes without a response, another notice with a rejection will follow. So, the Refer decision will fall into one of the other two decisions ultimately. Figure 7.3 presents the decision flowchart.

When the underwriting decision is Rejection, credit underwriting requires interpretation for the decision it makes. Regulation requires that the reason for a rejection of the credit application needs to be explained to the applicant and the decision process should be compliant with the fair lending laws. For example, if the rejection is due to some negative items in your credit report from any of the credit reporting agencies (or credit bureaus), the lender needs to tell you about your right and the contact to get a free copy of your credit report from the credit reporting company that provided it within 60 days of your adverse action notice, so you can verify and correct these items if any error through working with the reporting agencies or filing a dispute. Such rejection reasons usually come from applicant underwriting segmentation and easy to communicate. Table 7.2 lists some commonly used rejection reasons collected from the three credit bureaus with corresponding reason code. It should be pointed out that these rejection reasons may not be the full list and the corresponding reason code may be subject to changing over time.

The implementation of proper rejection reason requires a mapping between the rejection decision and rejection reason. For decisions made by rules, the mapping is usually direct since the most suitable reason can be selected from all relative reasons. If the decision is due to a model or an algorithm, the mapping requires a ranking of the credit scoring drivers. From most models (e.g., generalized linear models or tree-based models), there is the variable importance factor (VIF), and the credit scoring drivers can be ranked by VIF. Then, map the rejection reasons to the top credit scoring drivers, and a selection process can be built based on both the VIF ranking and the reason-drivers mapping.

Fig. 7.3 Decision and action

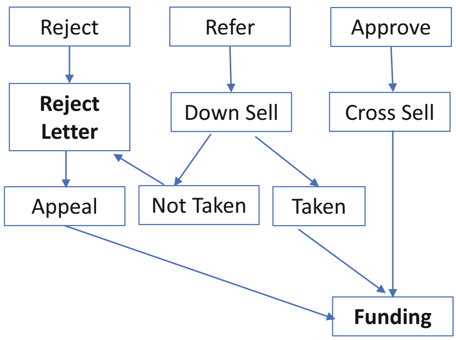


Table 7.2 Adverse action statement and code based on bureau credit score

Credit score reason for adverse action	Reason code by credit bureau		
Statement	Equifax	TransUnion	Experian
Amount owed on accounts is too high	1	1	1
Level of delinquency on accounts	2	2	2
Too few bank revolving accounts	3	N/A	3
Too many bank or national revolving accounts	4	N/A	4
Too many accounts with balances	5	5	5
Too many consumer finance company accounts	6	6	6
Account payment history is too new to rate	7	7	7
Too many recent inquiries last 12 months	8	8	8
Too many accounts recently opened	9	9	9
Proportion of balances to credit limits is too high on revolving accounts	10	10	10
Amount owed on revolving accounts is too high	11	11	11
Length of time revolving accounts have been established	12	12	12
Time since delinquency is too recent or unknown	13	13	13
Length of time accounts have been established	14	14	14
Lack of recent bank revolving information	15	15	15
Lack of recent revolving account information	16	16	16
No recent non-mortgage balance information	17	17	17
Number of accounts with delinquency	18	18	18
Date of last inquiry too recent	N/A	19	N/A
Too few accounts currently paid as agreed	19	27	19
Length of time since derogatory public record or collection is too short	20	20	20
Amount past due on accounts	21	21	21
Serious delinquency, derogatory public record or collection filed	22	22	22
Number of bank or national revolving accounts with balances	23	N/A	23
No recent revolving balances	24	24	24
Number of revolving accounts	26	N/A	26
Number of established accounts	28	28	28
No recent bankcard balances	N/A	29	29
Time since most recent account opening too short	30	30	30
Too few accounts with recent payment information	31	N/A	31
Lack of recent installment loan information	32	4	32
Proportion of loan balances to loan amounts is too high	33	3	33
Amount owed on delinquent accounts	34	31	34
Serious delinquency and public record or collection filed	38	38	38
Serious delinquency	39	39	39
Derogatory public record or collection filed	40	40	40
Payments due on accounts	N/A	N/A	46
	N/A	98	N/A

(continued)

Table 7.2 (continued)

Credit score reason for adverse action	Reason code by credit bureau		
Statement	Equifax	TransUnion	Experian
Length of time consumer finance company loans have been established			
Lack of recent auto finance loan information	98	N/A	N/A
Lack of recent auto loan information	N/A	97	98
Lack of recent consumer finance company account information	99	99	99

The applicant can appeal a rejected application, especially when some information in the original application has been updated in favor of the application (e.g., income information). Lenders commonly provide guidance on how to appeal an application. The appeal process is usually long since the process is largely manual. A successful appeal may lead to an approval, but likely with some reduced risk.

For credit underwriting, the action following the Approve decision is relatively easy. It leads to the release of the funds through some prespecified channels (e.g., a check to the settlement company in mortgage underwriting or a credit card issuing with decided credit limit). The approved applicant could be recommended for other products (cross-sale).

Following the underwriting process, there is the documentation. Regulation requires a minimum record retention period (e.g., 25 months by Consumer Financial Protection Bureau or CFPB for consumer credit applications).[3] Besides the application, there are support documents related to customer contact and identification, financial records on income and account balances, and proofs of asset and insurance if collateral required. In recent years, both legal "Know Your Customer" (KYC) requirement and internal customer risk management add the importance of customer documentation, not only to protect against credit risk but also to meet legal requirements to protect against money laundering, identity theft, and terrorist financing.

7.1.2 Risk Rating, Credit Scoring, and Scorecards

As described in the underwriting process, customer segmentation and risk classification are core processes which define the underwriting strategy. These are also the processes more relying on quantitative approaches like statistical models, especially risk classification.

Risk Rating

Risk classification depends on risk rating, which is defined as a ranking mechanism. This ranking mechanism can either be based on risk grades (nonnumeric letters) or

[3] CFPB, 12 CFR Part 1002 (Regulation B), §1002.12 Record retention

risk scores (often number in some scale). The former is commonly used in business risk rating (or corporate rating) by a Rating Agency (e.g., Moody, Fitch, S&P), while the latter is more focused on consumer credit risks (e.g., FICO, Vantage). Both ratings measure the credit risk, which is defined as lenders' risk due to borrowers' inability or unwillingness to honor their obligations. Business risk grades are mainly used for bond pricing, and consumer credit scores are mainly used for underwriting decision in consumer lending. Large banks may build their own Internal Risk Rating (IRR) systems for business lending based on their proprietary data and information on business, with the agency risk ratings for business taken as the starting point for such systems. However, for consumer risk rating, FICO score dominates, while the Vantage score by the three credit bureaus (Equifax, Experian, and Transunion) takes a small portion.

Even with FICO's dominance in consumer credit score, credit underwriting directly based on FICO score is not common due to the fact that FICO score is based on the credit bureau reports, and those reports may not include all credit information about a consumer and may not be updated promptly. Also, there is always a delay between the FICO score and the credit bureau reports. So, financial institutions would still like to develop their own underwriting risk rating. Such well-known risk rating systems in the mortgage industry have been created by Fannie Mae and Freddie Mac, the two mortgage giants (Government Supported Enterprises or GSEs) for the MBS market, called ACI (Acquisition Credit Index) by Fannie as we used in the examples of previous chapters. Fannie's underwriting system is based on this ACI score.

Like bank's Internal Risk Rating system for businesses taking agency risk ratings (grades) as inputs, ACI also takes the FICO score as inputs, but also with other factors related to mortgage as inputs. The goal is to take more industry-specific factors into the underwriting decision process as well as to make up the FICO information lags.

Risk rating in underwriting based on ACI type of credit scoring becomes more and more popular given the fast-growing FinTech companies, which expand quickly to cover various industries in credit underwriting. Next, we will discuss more about credit scoring in the history of credit underwriting.

Credit Scoring

According to Anderson[4] (Chap. 8), the history of credit scoring can be split into three periods based on how statistics was used – pre-statistics, statistical experiments, and predictive statistics. It is not a surprise these periods align with the developments of computing technology and computational statistics involving large data.

The pre-statistics period for credit scoring is defined as largely before the 1930s. Over this long history of period, credit decisions were guided by the hard-earned experience of individuals. There was no evidence of scoring or ranking borrowers,

---

[4] Anderson, A. R. (2022), Credit Intelligence & Modelling: Many Paths through the Forest of Credit Rating and Scoring. Oxford University Press, New York

although those experiences handed down by predecessors may have some underlying unintentional rating thinking based on some of the five Cs we mentioned before, especially the "character" of the borrower through assessments based upon personal interviews or manual review of submitted applications. Such experiences were later codified as rules, which we can now call them as "expert models." These "expert models" got popular during World War II, as those credit managers with such experiences were called up to serve and they codified their rules of thumb for use by housewives employed in their stead for the war's duration.

The period of statistical experiments spans over the next 20 years from earlier of 1940s. This is marked by the first-ever empirical credit-scoring model by Durand,[5] a researcher at the USA's National Bureau for Economic Research (NBER), using Linear Discriminant Analysis (LDA) on car loans. Although Durand's research more focused on the earlier expert models and even he doubted that the model would ever replace human judgment, he did apply LDA to a data set sampled from 7000 car loans obtained from banks and finance house, with risk factors like age, gender, time at residence and employer, occupation, industry, and holding of a bank account, real estate, and/or life insurance. Of course, as pointed out by Anderson, the major shortcoming with his model is that the model didn't include any risk factor representing customer's credit history. During this period, there were other unsuccessful experiments using statistical models for credit decisions. E. F. Wonderlic, better known for the Wonderlic Intelligence test developed in 1936, became director of personnel at Household Finance Corporation (HFC) in the mid-1940s. Using his knowledge of statistics, Wonderlic developed HFC's "The Credit Guide Score" in 1946 and gave instructions to analysts on the score calculation. But in 1948, the score was disused despite substantial evidence that it worked. Credit analysts did not trust it and made their decisions without the score and then dutifully calculated the score afterward.

The success of predictive statistics in credit scoring started from the 1960s onward, with the continuing resistance from credit managers, who considered the produced scores as mistrusted correlations that contradicted to their long-held assumptions and intuitions. Given the current AI booming threats to some career and a lookback of historical machine-human fights, it is not surprised to see such response in the earlier stage of a new technology revolution. Even with such resistance to accept credit scoring as a predictive tool, scorecard vendors' effort to develop scoring systems with automated office and record-keeping functions gradually won the battle with credit scoring initially sold as a supplement to judgment, not a replacement. The situation fundamentally changed after the 1970s when credit requests boomed with low credit losses, high volumes, and a lack of skilled analysts as well as the continued improvements of credit scoring systems in their soundness and stability by credit score vendors, like FICO.

---

[5] Durand, David (1941), Risk elements in consumer instalment financing. In Studies in Consumer Instalment Financing. New York: NBER

Fair, Isaac & Co. (FICO) was founded in 1956 by Engineer William Rodden Fair (1922–1996) and mathematician Earl Judson Isaac (1921–1983), two ex-employees of the Stanford Research Institute in California. The company's first contract was to develop a billing system for Carte Blanche in 1957, a credit card offered by Conrad Hilton's hotel chain. FICO developed its first predictive mode using linear programming (LP). Although nowadays LP is not considered a good technique for predictive models, its general format to fit a function of predictors with flexible constraints on the fitted function and its computation breakthrough in the 1970s could be the reason why FICO selected LP for its credit scoring model. FICO has been criticized for not making its model public, but it has been successful over so many years in giving enough information about the score to avoid anger but still selling their proprietary model, by publishing hints, but not explanations, about what makes their scores work.

Of course, besides LP, there are other techniques for predictive model, e.g., linear regression, linear discriminate analysis, generalized linear models, and nonparametric models. Over the years, the likelihood-based models as those presented in Chap. 3 are winners for classification problems in the banking and finance world – at least where model risks, governance, and regulatory oversight are high concerns, such as for application and behavioral scoring. The major exception is those institutions using FICO's proprietary models or methodology. Alternative credit scores (e.g., Vantage score) have been developed with different predictive models and made effort to replace FICO score at least in some fields (e.g., mortgages), but it is not easy due to FICO's current market dominance.

In Sect. 7.2, we will cover some new trends on credit scoring development, including how FICO score is treated as a basic information summary from credit reports in building specific credit scoring systems.

Scorecards

The successful use of credit scoring in credit underwriting has been depending on the invention and continued development of scorecards. Different for the back-office risk management, underwriting requires quick and easy to understand tools in risk rating. For business underwriting risks (mainly for corporate bond), risk grades have been developed led by the rating agencies and implemented by banks in their internal risk rating systems for commercial products. For consumer underwriting risks, over the history, scores have been preferred. The choice of the score system using integers between 1 and 999 meets the human's preference on numbers measuring something a little more complex – two digits between 0 and 100 may be with less accuracy for a credit measurement (though that is the common score system for student exams!), while four digits are a little overwhelmed for our brains. The three-digit score system evolved since the 1960s when tabulated by hands and has been preserved so far with no hints of changing.

In Sect. 7.2.7, we will explain why it is nature to scale the credit scoring model outputs to the three-digit system when the creditworthiness and log odd ratio are linked in the logistic regression when used for credit scoring. Such scaling process defines the scorecards. Over the time, the FICO score scaling method has been

Table 7.3 Scorecard example for small business lending

Characteristics	Attributes								Points
Constant									570
Personal Credit Score	<500	500-575	**576-625**	626-660	661-720	>720			
	0	12	**24**	39	54	63			24
Debt to Income Ratio %	0-25	26-50	**51-75**	76-100	>100				
	27	21	**12**	3	0				12
Time in Business (month)	36 or more	25-35	**24**	13-23	12	1-11	0		
	60	51	**45**	30	24	12	0		45
Industry Risk	High	Medium	**Low**						
	0	10	**30**						30
Corporate Risk	High	Medium	**Low**						
	0	6	**15**						15
Cash Flow Revenue Ratio %	76-100	51-75	**26-50**	16-25	6-15	1-5	<=0		
	30	24	**18**	12	9	6	0		18
Revenue ($1,000)	0-50	51-250	**251-500**	501-1,000	1,001-2,500	2,501-5,000	5,001-10,000	>10,000	
	0	12	**18**	30	36	42	48	55	18
Final Score									732

popularly adopted, since it has not only a fixed-odd increment but also a benchmark odd and score.

Table 7.3 presents a scorecard example for small business lending. The scorecard has seven characteristics used from the underwriting data. All attributes are bucket-based and are binary for all characteristics. So, the calculation of score points is simply the sum of all corresponding attribute contributions. The model has a constant attribute (the intercept), and its point corresponds to the default buckets for all characteristics, which are marked by the 0 point with the corresponding attributes of these characteristics. We selected the third bucket for each characteristic and got the score 732 shown by the final column.

The final scorecard as shown in Table 7.3 is often not the direct result from the model, but with manual adjustments from risk managers. Some adjustments are intentional to make the scores more evenly distributed without significantly changing the risk ratings from the model outputs. The scores are ranged from 570 to 850. However, 570 is only the lower bound by the score drivers listed on the table, and some other risk drivers, like bankruptcy and legal events, may drive the score further down.

7.1.3 Regulations on Credit Underwriting

Over the time, there are quite some laws in the credit-related operations – credit reporting, making loans and lending, debt collection and recording, and data privacy and security. We focus on regulations more related to credit underwriting, largely the first two categories, as credit reporting is related to the input side of credit underwriting, while regulations on lending directly guide the underwriting process. We would not dive into the details of these regulations, but more focus on their impact to the credit underwriting evolution. We mainly discuss regulations within the USA, as other countries and regions may have similar ones.

TILA

The Truth in Lending Act (TILA) could be the first law in modern time on credit product and lending. It was passed on 1968 during a time of credit booming. Most of the specific requirements imposed by TILA are found in Regulation Z, so a reference to the requirements of TILA usually also refers to the requirements contained in Regulation Z. The main purpose of the act is transparency about credit products and clear communications to borrowers. So, this Act requires lenders help customers to understand what they are agreeing to in a credit transitions. TILA also regulates how credit providers can advertise their products. It took a long time to achieve such goals partially, given that the credit product transparency may not be the top priority of the lenders if not for intentional confusions.

From TILA's inception, the authority to implement the law by issuing regulations was given to the Federal Reserve Board (FRB). However, from 2011, TILA's general rule making authority was transferred to the Consumer Financial Protection Bureau (CFPB) due to the Dodd Frank Act.

TILA has been frequently re-enforced and enhanced, even after the 2008 financial crisis. Over the time, TILA introduced the APR (annual percentage rate) calculation to distinguish the amount financed (or sales price) and finance charges and required clear statements on minimum payment, monthly finance charges, payment due dates, current balance, and how late charges are assessed and how much they are. The introduction of the APR calculation was used to bar the "Zero Percent APR" financing advertised by the auto loan industry back in the 1980s. By bundling the price of the car and its financing charges, auto makers were able to shift money between sales price and finance charges or even eliminating the financing charge entirely, and thus "Zero Percent APR" was born. The minimum payment amount was introduced to relieve credit card holders from lofty late fees charged by credit card companies, and other statement requirements also would increase the transparency of borrower's obligations to avoid any hidden ones.

TILA's impact to credit underwriting is on the transparency of the credit product. It provides help for consumers to compare and select different products, although such help can be significant or minor for different industries depending on the product diversity.

FCRA

The Fair Credit Reporting Act is one of the most critical legislations in the credit landscape, which regulates the collection of consumers' credit information and access to their credit reports in the USA. It is the federal legislation enacted to promote the accuracy, fairness, and privacy of consumer information contained in the files of consumer reporting agencies. It was intended to shield consumers from the willful and/or negligent inclusion of erroneous data in their credit reports. To that end, the FCRA regulates the collection, dissemination, and use of consumer information, including consumer credit information. It was passed in 1970. FCRA is implemented by Regulation V, originally written by the Federal Reserve Board. As TILA, its general rule making authority was transferred to the Consumer Financial Protection Bureau due to the Dodd Frank Act. It is enforced by the US Federal Trade Commission and the Consumer Financial Protection Bureau.

FCRA's major role is on regulating the consumer information collection through the consumer reports, also commonly referred to as credit reports. The regulation prohibits the use of investigative reporting, i.e., collection of information on a person's character, general reputation, personal characteristics, or mode of living obtained through personal interviews with neighbors, friends, associates, or others with such knowledge as commonly used by risk managers before the regulation. The information collected in a consumer report can contain information about your bill payment history and the status of your credit accounts. This information includes how often you make your payments on time, how much credit you have, how much credit you have available, how much credit you are using, and whether a debt or bill collector is collecting on money you owe. Credit reports also can contain rental and utility payment information if you are a property renter or pay utility bills. It also can contain public records such as liens, judgments, and bankruptcies that provide insight into your financial status and obligations.

The completeness and correctness of the consumer reports depend on the cooperation of information furnishers (creditors) and Credit Reporting Agencies (CRA), which both are under this regulation and subject to penalty if any violations. While the correctness is much the focus of FCRA, the completeness is more on the biased (often more negative) report instead of the types of credit information in the report, e.g., most often rental and utility payment information is not included in the credit report, which has been the case since the enactment of the regulation. This is one of the main reasons that in credit underwriting, credit scores based on bureau credit reports are used as a summary of the credit report and an input to the credit scoring models.

FCRA also regulates the use of consumer reports and consumers' right regarding the use of their credit reports.[6] Most importantly, the consumer should keep an eye on the accuracy of their credit report and request correction for any reporting errors, as well as obtain free credit score and report from any of the CRAs annually.

No doubt, FCRA takes a critical role in the growth of coverage of consumer reports over the population in the USA. The regulation defines not only the credit reporting industry (the CRAs) but also the behaviors of creditors and consumers. Nowadays, it is hard for anyone on almost all financial activities without credit report and credit score. The popularity of credit reports among the entire population has changed the credit underwriting, especially consumer credit underwriting, to more automated processes. In addition, the impact of FCRA to credit underwriting is

[6] Consumer Financial Protection Bureau, (2018). A Summary of Your Rights Under the Fair Credit Reporting Act

direct, not only through the direct use of the credit score produced by CRAs as input of credit scoring models but also the quality and richness of the consumer information included in the credit report if credit scoring modelers decide to directly use the information in the credit reports.

ECOA

The Equal Credit Opportunity Act (ECOA) was implemented in 1974, 4 years later after FCRA. It requires that credit companies not discriminate against an applicant based on race, color, religion, national origin, age, sex, marital status, eligibility for public assistance, the exercise of any rights under the specific consumer protection laws, or the so-called protected groups. The only justifiable bases for declining credit are things like the applicant's financial status (earnings and savings) and credit record defined as credit creditworthiness. The following is a more accurate definition of the act from the regulator, which also clarifies the two sources of liability under ECOA – disparate treatment and disparate impact:[7]

> The statute provides that its purpose is to require financial institutions and other firms engaged in the extension of credit to "make credit equally available to all creditworthy customers without regard to sex or marital status." Moreover, the statute makes it unlawful for "any creditor to discriminate against any applicant with respect to any aspect of a credit transaction (1) on the basis of race, color, religion, national origin, sex or marital status, or age (provided the applicant has the capacity to contract); (2) because all or part of the applicant's income derives from any public assistance program; or (3) because the applicant has in good faith exercised any right under the Consumer Credit Protection Act." The ECOA has two principal theories of liability: disparate treatment and disparate impact. Disparate treatment occurs when a creditor treats an applicant differently based on a prohibited basis such as race or national origin. Disparate impact occurs when a creditor employs facially neutral policies or practices that have an adverse effect or impact on a protected class unless it meets a legitimate business need that cannot reasonably be achieved as well by means that are less disparate in their impact.

ECOA is implemented by Regulation B, originally written by the Federal Reserve Board. Like TILA and FCPA, its general rule making authority was transferred to the Consumer Financial Protection Bureau due to the Dodd Frank Act. In December 2011, the CFPB restated the Federal Reserve's implementing regulation at 12 CFR Part 1002 (76 Fed. Reg. 79442) (December 21, 2011). In January 2013, the CFPB amended Regulation B to reflect the Dodd-Frank Act amendments requiring creditors to provide applicants with free copies of all appraisals and other written valuations developed in connection with all credit applications to be secured by a first lien on a dwelling.

ECOA is probably the most controversial credit regulation in its implementation. Since in its early implementation, Regulation B provided an escape clause. It considered discrimination fair if based upon a model that is (1) empirically derived;

[7] Consumer Financial Protection Bureau, (2013). CFPB Consumer Laws and Regulations: Equal Credit Opportunity Act (ECOA)

(2) credit-focused; (3) statistically sound; and (4) regularly updated. Certain characteristics were banned outright, such as race and religion, and since then, the list has expanded (until now, age is the only one acceptable). This had a big impact on the credit underwriting operations. Since under this clause, statistical models were okay, but credit men's long-held right to judgmental decisions was put on question. As a result, the credit analysts' role in the consumer economy was massively curtailed, as statistical models came to dominate.

Under the restated Regulation B by CFPB, ECOA neither requires nor endorses any particular method of credit analysis. Creditors may use traditional methods, such as judgmental systems that rely on a credit officer's subjective evaluation of an applicant's creditworthiness, or they may use more-objective, statistically developed techniques such as credit scoring. However, the trend of using credit scoring systems based on statistical models is almost not reversible.

Regulation B prescribes the standards that a credit scoring system must meet to qualify as an "empirically derived, demonstrably and statistically sound, credit system." All forms of credit analysis that do not meet the standards are automatically classified as "judgmental" systems. This distinction is important because creditors that use a "demonstrably and statistically sound" system may take applicant age directly into account as a predictive variable, whereas judgmental systems may do so only to determine a pertinent element of creditworthiness or to favor an elderly applicant.

While the focus has been shifted on developing credit scoring systems, validating the ECOA compliance with such a system has not been designed from the beginning, especially with the second source of liability – disparate impact, which has been a grey area. Credit scoring system developers may feel they had done their job best with ECOA by paying attention to the disparate treatment during the model developing cycle and they can either get away from the liability of disparate impact by using some modeling tricks or leave that to the corporate lawyers or with the last resort of a business need.

Recently, such "who knows what is important or what can we get away with ignoring" attitude has started changing with a few lawsuits to some FinTech companies, including the Student Borrower Protection Center's investigation on Upstart Holdings and other FinTech companies mentioned early. ECOA is going to have a more and more significant impact on the credit scoring systems, especially those using new modeling technologies like machine learning. Fair Lending testing for all credit scoring systems to make sure the underlying models do not discriminate requires not only the post-model (production model) checking but also a pre-model (development model) analysis on current portfolios, potential populations, potential reject populations, and correlations.

While the direct impact on credit underwriting from FCRA is on the input side, the impact of the Equal Credit Opportunity Act (ECOA) on credit underwriting is on its process and results. We will discuss some recent developments on ECOA compliance with credit scoring systems in later sections.

7.2 Credit Model Applications in Credit Scoring

In this section, we cover how credit models are used in credit scoring and scorecard building. Like in other applications as we presented in the previous chapters, credit models in credit scoring are used to assess performance of accounts in a new pool based on performance of the historical pool by assuming these pools have the same performance under the same condition. However, one difference from the previous applications is that a large portion of accounts from the historical pool in credit scoring do not have the opportunity to have their performance recorded as either not approved or not taken. So, credit models built on the performance-only population could be biased when applying to a new pool assumed from the full population. To avoid such model bias, model population should cover the full population by including those not approved or not taken ones. So, we need to take care of the model population by incorporating the so-called reject inference.

Once we have the model population decided, as with other credit model applications, the first thing for modeling is model segmentation within the population. For credit score modeling, the model segmentation should follow the customer segmentation we discussed early in Sect. 7.1.1, as the decision can be directly made on the top level of customer segmentation.

Next, as we discussed in the previous section, under the Equal Credit Opportunity Act (ECOA), credit scoring systems and scorecards are allowed to only use an applicant's creditworthiness to rate and make decisions, so we need a proper definition of creditworthiness, which requires proper selection of the target credit event for the scoring model. Although credit scores generated from the scoring model are for the risk rating purpose, the impact of using different types of target credit event in the scoring model could result in different scoring systems and scorecards.

In addition, ECOA enforces two sources of liability for a credit scoring system, disparate treatment and disparate impact. While disparate treatment may be relatively easy to avoid in a scoring model, e.g., by removing any score drivers related to prohibited bases from the model, testing and avoiding disparate impact require more works to be done in the scoring model development cycle. To achieve a successful fair lending testing to pass model review and validation, scoring model developers should consider proper model design at the start of the modeling cycle. The concept of fair lending has a basic factor embedded in it – the reference population. Commonly the full application population is considered the reference population when evaluating fair lending compliance; however, largely the application population would not be recorded completely, and some sampling methods are needed for fair lending testing.

In general, different from other applications as we presented in the previous chapters, for which regulations either are directly on methodology or provide guidance for implementation, credit scoring systems and scorecards are subject to regulations which only give conceptual guidance and do not help much on reducing the complexity of the credit model application. The impact of such conceptual guidance has two sides. On one side, it encourages flexibility on adoption of different methodology and implementation, while on the other side, there is a lack of standard on such methodology and implementation, and it makes comparison hard on conclusions. Other considerations could be that credit scoring and the corresponding systems are only assistance to business decisions and regulators would rather focus on a lender's broader underwriting business practices. However, modelers have to consider regulatory compliance from the design of such models and reduce the model rejection rate.

In the following we address the above items in the credit model application for credit scoring and propose some improvements over the current treatments.

7.2.1 Model Population

For any predictive model built on historical population, there is the assumption that the prediction by the model is carried out on similar population because it is required that the underlying relationship between the model inputs and outcome should not change among populations. In reality, this assumption is loosely tolerated with business continuation and no strategic changes. Thus, the so-called model population is generally defined as all accounts are either already on the book or in future with the business continuing.

However, for credit underwriting, there is a fundamental difference between the historical population having performance recorded and can be used for model building and the application population on which the predictive model is used for scoring purpose, since the historical population having performance recorded only includes applicants approved and on the book, but not all applicants. To apply the credit model on any new applicants without bias, the model population is required to also cover those applicants who did not get the approval as well as those who did not take the offer. We define all these applicants (or accounts) belong to the non-performance population.

Once the model population with its coverage has been defined, model data processing can be started by collecting all account-level information of the model population, which include both the observational data collected during the underwriting process as shown in Figure 7.2 and the outcomes of the underwriting decision. The workload of such data processing could be huge depending on the volume of the application population. However, a large part of the data processing should have been completed during the underwriting, and majority of the data process at this phase should be ETL from the underwriting databases. The processed data should be ready for model segmentation in the next phase of modeling.

7.2.2 Model Segmentation

Model segmentation for credit scoring is likely more complex than for other applications. One of the main reasons is that for consumer lending, besides the dimension of operations, there is also the flexibility of business strategic decisions. In addition, for easy use of scorecards, the requirement of linearity between scores

Table 7.4 General credit scoring segmentation

	Credit scoring segmentation factors		
Dimension	Primary	Secondary	Other
Operational	*Customer and product type*:	*Collateral type*:	Region
	Residential mortgage	Secured	
	Home loans	Unsecured	
	Home equity line of credit		
	Auto loans		
	Cards		
	Personal loans		
	Small business lending		
	C&I loans		
	Commercial real estate		
	Commercial paper		
Strategic	*Business strategy*:		
	Market		
	Profit		
	Risk		
	Capital		
	Reputation		
	Short term		
Driver-based	*Score driver*:		
	Interaction		
	Accuracy		

and drivers and model accuracy adds a third dimension of segmentation – driver-based segmentation to interpret interactions among drivers or enhance model performance. Table 7.4 shows these three dimensions.

Segmentation based on operational factors is the most popular as in other credit model applications. The main segmentation factors are customer and product types. For example, credit models used in underwriting for retail and wholesale products are different as the former focuses on the large volume of individual consumers, for which credit scoring is mainly used and the focus of this chapter, while the later more focuses on the one-to-one client service with high exposure and internal or external risk ratings (by model-based risk grades) are used. Different types of products generally have different underwriting systems, and segmentation for credit scoring could be more granular. For example, first lien residential mortgage, home loans, and home equity line of credit generally have different underwriting and credit scoring systems. Collateral type could be a segmentation factor, for example, credit cards and debit cards are commonly different segments under cards. Region is a common segmentation factor in credit underwriting, since regulation and customer base are different in different regions and different credit scoring systems should be used in underwriting.

Credit scoring segmentation is more likely based on some business strategy than in other model applications. Businesses commonly create some special portfolios based on business strategic decisions, for example, on market share, business expansion, profit, risk mitigation, regulatory risk, and some short-term effects. For example, in credit cards, when some subprime borrowers were identified as high profitable in the 1990s, credit card issuers launched special products for this group with different credit scoring and underwriting systems. The segmentation based on strategic decisions more aligns with customer segmentation in the underwriting process, when a special group of customers are identified for special treatment. For example, students and military service personnel are always treated as special groups and usually have different credit scoring systems in credit underwriting.

The more granular segmentation for credit scoring models is based on score drivers, which is not preferred in other credit model applications. Credit scores measure the creditworthiness and prefer a linear relationship with score drivers, such that corresponding scorecards can have simple monotone relationship with score drivers. When there are interactions among score drivers in the credit scoring models, the monotone relationship will not be guaranteed. So, it is preferred to break the interaction by the related score drivers such that the monotone relationship is maintained within those subregions (e.g., break a concave function at the top point to guarantee the monotone relationship on the two sides of the top point). An alternative to handling the interaction is using linear approximation with less accuracy. Another motivation for the segmentation by score drivers is to enhance the model accuracy. Basically, this is equivalent to use multiple piecewise linear functions to approximate a more complex nonlinear function for a higher accuracy.

Searching the optimal split points for the related score drivers in credit scoring is sometimes called segmentation mining, which was popular in the early times of scorecard development. However, given the cost of model maintenance, especially the increasing requirements on credit scoring model validation, driver-based segmentation is gradually losing its attraction. With the increased computing power and requirement of higher automation in underwriting, nonlinear models in credit scoring are not forbidden anymore. In addition, some nonparametric models, e.g., tree-based models, have been popularly used to identify the nonlinear relationships and optimal split points such that segmentation mining becomes simple. So, the driver-based segmentation is more likely replaced by nonlinear scoring models identified by more accurate nonparametric methods, including some machine learning methods, especially for underwriting with new populations.

Besides the three dimensions discussed, there is always some practical consideration with model segmentation, for example, data richness. When data is thin for some segments, merging is considered with some population similarity analysis.

7.2.3 Target Definition

Under ECOA, creditworthiness should be the only measure to be used for credit scoring. However, alike in other credit model applications, where the target credit event (such as default) has clear definition under regulation, target credit event for credit underwriting does not have a standard definition from regulation and has flexibility according to business decision. So, we present some general rules per Anderson[8] for a proper definition of credit event in the binary format of good and bad:

- *Relevance* – The target event (good or bad) should be defined appropriately for the risk rating purpose in credit scoring. The cutoff should be clear, and the target event be classified with sound business acknowledgement. For example, in the mortgage lending business, payment past due 90 or more days is a clear cutoff since it is well known for the borrowers that such an event will trigger the initial foreclosure process. Although in other lending businesses, there might not be such a clear cutoff, the target event definition still should be clear and polarized.
- *Focus* – The target event should focus on the quantities related to borrower's responses and behaviors so to properly represent the creditworthiness regarding to the borrower, but not something else, like something from the observers. For example, returned checks may not only represent borrower's bad credit but also banks' operations.
- *Transparency* – The defined risk event should be easily understood by the management, regulators, and others. The underlying calculation should not be too complex such that it is implementable and should enable performance monitoring to make sure the model works.
- *Adequacy* – The risk event definition should be robust, little affected by minor changes, with a lifespan longer than the scorecards, appropriate for all cases within the sample, and able to provide enough subjects in each category (good and bad).
- *Quality* – The risk event definition should be based on accurate, consistent, and recent data as at the performance dates and monitored over the history of available data.

For consumer lending, largely the target event is defined based on the borrower's payment behavior observed from the transaction data as follows:

Table 7.5 only shows some general target event definition in consumer lending. There could be more underlying details, for example, accounts with very small balance written-off (e.g., <\$10) are usually categorized into G instead of B. The more involved quantity is DPD, which could be defined differently based on the payment flow. In the following, we present one approach to calculate DPD with daily, weekly, bimonthly, and monthly payment schedules.

[8] Anderson, A. R. (2022), Credit Intelligence & Modelling: Many Paths through the Forest of Credit Rating and Scoring. (Chapter 18). Oxford University Press, New York

Table 7.5 General target events for consumer lending

Payment behavior	Target event
Fraud, deceased	X
Paid off	G
Written off	B
DPD $>=90$	B
DPD >0 and bankruptcy	B
Otherwise	G

where "X" represents exclusion and "G" and "B" represent good and bad, respectively. DPD represents days past due

$$\mathrm{DAA} = \frac{\mathrm{RTR}}{\mathrm{Term} * \frac{365.25}{12}} \tag{7.1}$$

$$\mathrm{DPD} = \frac{\mathrm{DAA} * (\text{Deal Age in days} - B) - \text{Cumulative Actual Payment}}{\mathrm{DAA}} \tag{7.2}$$

where

DPD = Days past due
DAA = Daily accrual amount. This is the daily expected payment if paid daily
RTR = Right − to − receive amount. This is the funded amount∗Interest Rate
$B = 1, 7, 14, 30$ for daily, weekly, bimonthly, monthly deals, respectively
Term = Deal term in month with minimum 6 months

The approach by (7.2) is more suitable for small business lending where payments could be frequent but with smaller amounts as cash inflows for such businesses. For less frequent payment schedule, one can count the days the scheduled payment past and directly use that as DPD. The difference between (7.2) and the counter is that counter does not take the past due amount as a factor – as long as the full scheduled payment not received (even with a small deficit), it is counted as past due. So, the counter method may not fully reflect the borrower's payment ability when frequent payments are made and thus (7.2) is preferred in such case.

An alternative approach to define the target event, which is in a similar sprint of (7.2), is the Payment Made vs. Due measure or PMvD. PMvD is defined as the ratio between the cumulative payment made and the due amount:

$$\mathrm{PMvD} = \frac{\text{Cumulative Actual Payment}}{\text{Scheduled Due Amount}} \tag{7.3}$$

A PMvD of 100% for a merchant equates to full repayment as per schedule, while lower values constitute weaker repayment performance. A risk event is defined as PMvD crossing a lower bound, e.g., 85%.

The risk event is observed over a performance observation window, which usually takes 1 year or longer (credit bureaus take 2-year window for the generic FICO scores). Also, the target event is commonly defined as the worst over this observation window that means once the event happens during this window it marks, although such event could be cured later in the window. This definition takes the more conservative side of the measure of creditworthiness and is consistent with most of credit policies, e.g., it will take a long time (7 or more years) to remove missed payment records from the bureau credit report (unless a single one due to some technique or other reasonable issues). The length of the performance window has more flexibility, which could depend on the data availability, business decision cycle, and business trends. So, the length of the performance observation window should keep a balance among the following factors:

- Align with business decision cycle to keep the creditworthiness measure relevant to business decision. If the window is too short, then the observed performance may not be stable yet, and if the window is too long, then the observed performance may be not so relevant to the business decision.
- Adjust slightly due to data availability to make sure there are sufficient risk events for the credit scoring modeling purpose. Such adjustments should not be the cure of data sparsity, which should be addressed fundamentally for the data side.
- Follow the business trend such that the observed performance will not change dramatically in the expected future business environment. If the length of the window is too long, then the observed performance may not represent the current business trend any more.

Given that most underwriting businesses are subject to annual review, the 1-year window is popularly used. Of course, there are businesses that would like to have quick business decision cycles, e.g., 6 months, likely use short performance observation windows. In such cases, target event definitions based on more sensitive measure of creditworthiness like PMvD are preferred.

7.2.4 Reject Inference and Missing Target

As discussed in Sect. 7.2.1, the model population incudes the non-performance population, for which the performance-based target event can't be defined and takes missing value. For a segment, if there are a significant portion of data records with such missing target events from the non-performance population (e.g., more than 5%), we need to carry out some special treatments before building scoring models. Such missing value treatment is known as part of the Reject Inference and is preferred to be segment-wise.

There are generally two approaches to treat the missing target values from the non-performance population. The first approach is operational and may depend on extra data and information obtained for the non-performance population. The second approach is statistical and purely based on statistical assumption and applies imputation to fill the missing target values from the non-performance population.

Operational Treatment
Based on management decisions or extra information obtained for the non-performance population, the following missing target value treatments are suggested:

- All are treated as bad – This can be applied if the non-performance population in the segment is minor ($<1\%$) and the underwriting only applied hard rejection rules.
- Extra information obtained on credit performance of the non-performance population (e.g., data from bureaus) – The performance on other credit products can be used as surrogate for the current product.
- Internal information obtained on credit performance of the non-performance population for credit products provided before – Existing customers with good performance are unlikely rejected unless there were significant credit events (e.g., bankruptcy).

Statistical Treatment
Based on statistical assumptions on missing values, Missing in Random (MIR) is the commonly used assumption when doing statistical imputation for missing values. It assumes that the missing pattern is purely probabilistic with some prespecified probabilities or based on conditional probabilities derived from a model. For the non-performance population, the missing target value in a segment can be imputed:

- With prespecified probabilities of good and bad for the segment – Treat the non-performance population in the segment equally with prespecified probabilities of good and bad, which can be based on the scaled probabilities of good and bad by the performance population in the same segment. Then randomly draw good/bad based on these probabilities.
- With probabilities of good and bad derived from a model for the segment – Use a model built on the performance population in the same segment to estimate the probabilities of good and bad for the non-performance population. Then randomly draw good/bad based on the estimated probabilities.

One can also combine the operational and statistical approaches. In a segment, one can first apply the operational approach when either internal or external information is available for some portion of the non-performance population and then apply the statistical approach for the rest of the non-performance population. Operational approaches with additional information either external or internal are preferred. In practice, the non-performance population should not dominate in a segment, given that either the operational or statistical approach approximates the consumer performance using some surrogate information or statistical assumptions.

7.2.5 Score Drivers

In credit scoring modeling, the independent variables or score drivers are the most sensitive to use due to the ECOA. As we mentioned, it is not only the disparate treatment (as a source of fair lending liability) that prevents the use of certain prohibited variables; there is also the disparate impact, which requires broader investigation and testing to meet the fair lending regulations. On the other hand, traditional model validation requires that the final developed models follow the common good model properties, e.g., goodness-of-fit, optimization, stability, and sensitivity.

Over the years, model developers have experienced the difficulty to meet both requirements on the selection of proper score drivers and likely sacrifice the later by only using some pre-decided score drivers based on management decision or previous modeling experiences. Such a “conservative” approach might not be so conservative in respect to the regulatory requirement, since the pre-decided drivers could fail the fair lending test, while the final developed models are also subject to the critics of model validation that certain regular model properties are not kept.

Here, based on our Adaptive and Exhaustive Variable Selection (AEVS) procedure described in Chap. 3, we develop a new variable selection procedure for credit scoring (AEVS_CS). The new variable selection procedure not only inherits the adaptiveness of AEVS by pre-selection of a pool of score drivers provided by SME, management, and previous model experiences to avoid the liability of disparate treatment but also expands the adaptiveness to incorporate the Prohibited Correlation Index (PCI), which is defined to measure the correlations between any potential score drivers and the available prohibited bases collected for fair lending tests, as a defense to the more demanding disparate impact liability.

In the following, we first introduce the Prohibited Correlation Index (PCI) and then explain how it can be incorporated into AEVS naturally to defend the disparate impact liability. Then we will present the AEVS_CS procedure.

Prohibited Correlation Index (PCI)

ECOA requires an assessment of disparate impact for score drivers used in credit scoring models. As a measure of the disparate impact for a single score driver, we define the Prohibited Correlation Index for a single score driver X as one of the following correlation measures:

$$\mathrm{PCI}_{\max}(X) = \max(|\mathrm{corr}(X, Y)|, Y \in PB) \tag{7.4}$$

$$\mathrm{PCI}_{\mathrm{avg}}(X) = \frac{1}{M} \sum_{Y \in PB} |\, \mathrm{corr}(X, Y)\,| \tag{7.5}$$

where PB is the set of prohibited bases (including race, color, religion, national origin, sex, marital status) and M is its size. corr is the correlation function. (7.4) takes the maximum correlation in absolute values between the score drivers and factors in the prohibited bases set, while (7.5) takes the average.

For a scoring model M, correspondingly, we define the model PCI as the maximum PCI for all score drivers used in the model or take the grand average of PCI for all score drivers:

$$\mathrm{PCI}_{\max}(M) = \max\{\max(|\mathrm{corr}(X, Y)|, Y \in PB), X \in M\} \tag{7.6}$$

$$\mathrm{PCI}_{\mathrm{avg}}(M) = \frac{1}{P}\frac{1}{M}\sum_{X \in M}\sum_{Y \in PB} |\mathrm{corr}(X, Y)| \tag{7.7}$$

$\mathrm{PCI}_{\max}(M)$ is preferred as a stricter measure than $\mathrm{PCI}_{\mathrm{avg}}(M)$ when restricted by an upper bound.

Credit Scoring Modeling Process with AEVS

1. *Data Input Preparation*
   (a) Target risk event variable $Y = \begin{cases} 1, & \text{Event} \\ 0, & \text{Non} - \text{Event} \end{cases}$
   (b) Reject inference and missing target value treatment
   (c) Sampling (Synthetic Minority Oversampling Technique or SMOTE)
   (d) Identify the pool of prohibited drivers
   (e) Identify the initial pool of candidate score drivers
   (f) Calculate PCI for each candidate score drivers
   (g) N, the total number of candidate score drivers, select top D = 10 candidate score drivers from these N variables using each of the following methods:
   - Concordance Index (i.e., area under ROC) top D (10)
   - Standard stepwise (based on −2loglikelihood) top D (10)

   Combine score drivers selected from Concordance Index and stepwise.
   Add score drivers chosen by business intuitions.
   Determine TopN, the combined list for selected score drivers derived above.
2. *GLM Model Selection*
   (a) Generate M_j GLM models from possible subset of TopN score drivers using exhaustive search, $j = 1, \ldots, n \sim 2^{\mathrm{TopN}}$.
   (b) Reject M_j if any regression coefficient doesn't satisfy $sign(X)$.
   (c) Reject M_j if any GLM coefficient p value $>5\%$.
   (d) Reject M_j if model PCI > 0.95.

   This determines the selected candidate GLM models.
3. *Model Ranking*
   For the survived candidate models after Step 2
   (a) Rank by SIC (the lower SIC, the higher Rank M_j has); optionally reject low-rank models.
   (b) Rank by AIC (the lower AIC, the higher Rank M_j has); optionally reject low-rank models.
   (c) Rank by PCI (the lower PCI, the higher Rank M_j has).

(d) Rank by v-fold cross-validation (the lower CV_v, the higher Rank M_j has) on survived models.
(e) Calculate the average rank of the three performance rankings above; select the top five models based on the average rank. Note that we use v-fold cross-validation CV_v as the goodness-of-fit measure instead of the Leave-One-Out (LOO or CV_{nn}) due to computational efficiency for GLM models.

From the five models, choose the final model based on LOB review and model assessment.

AEVS_CS is an extension of the general AEVS on credit scoring by considering some specifics on credit scoring, e.g., disparate impact and missing with the target. The final selected model should meet both ECOA on fair lending, as well as the local optimality for model validation purpose.

7.2.6 Fair Lending Test

For credit scoring, once the scoring model is built, it is subject to the Fair Lending Test (FLT) to make sure that the model and the resulting scorecards comply with the laws in ECOA. The execution of the laws in ECOA including FLT is currently under weak controls overall. On the mortgage side, FHFA issued rules on the approval of credit score models for the enterprises it regulates.[9] The final rule sets forth standards and criteria for the process an enterprise must establish to validate and approve any credit score model that produces any credit score that the enterprise requires in its mortgage purchase procedures and systems. The validation and approval process for a credit score model includes the following phases: solicitation of applications, submission of applications, and initial review, credit score assessment, and enterprise business assessment. In the part of §1254.6 Submission and Initial Review of Applications, it requires the fair lending certification and compliance:

> Fair lending certification and compliance. Each application must address compliance of the credit score model and credit scores produced by it with federal fair lending requirements, including information on any fair lending testing and evaluation of the model conducted. Each application must include a certification that no characteristic that is based directly on or is **highly correlated solely with a classification prohibited** under the Equal Credit Opportunity Act (15 U.S.C. 1691(a) (1)), the Fair Housing Act (42 U.S.C. 3605(a)), or the Safety and Soundness Act (12 U.S.C. 4545(1)) was used in the development of the credit score model or is used as a factor in the credit score model to produce credit scores.

Although the rule does not provide an exact number defined as high correlation, we consider 0.95 in our AEVS_CS as the upper bound.

[9] FHFA, (2019), PART 1254 – Validation and Approval of Credit Score Models. Final Rule by the Federal Housing Finance Agency

By implementing PCI as a model selection criterion, we successfully incorporate the fair lending testing into the model building and selection procedure and guarantee that the final models selected by AEVS_CS automatically pass the fair lending test.

Fair lending test beyond mortgage has not been popularly implemented due to the weak regulation of ECOA, especially for FinTech companies.

7.2.7 Target Projection, Scaling, and Scorecards

For credit scoring, the focus of the output of the credit scoring model is the (log) odds ratio for a new applicant, which is also called creditworthiness. However, it is a number not so friendly for our brain to take in. So, a scaling of this number is helpful, and the successful use of credit scoring in credit underwriting has been depending on the invention and continued development of scorecards. Over the time, the three digits between 1 and 999 meet the human's preference on numbers measuring something a little more complex – a scorecard. The three-digit score system evolved since the 1960s when tabulated by hands and has been preserved so far with no hints of changing.

The scaling of the credit scoring model outputs to the three-digit system gets more support when the creditworthiness and log odds ratio are linked in the logistic regression when used for credit scoring:

$$\log\left(\frac{P_G}{1-P_G}\right) = -\log\left(\frac{P_B}{1-P_B}\right) = -[\beta_0 + \beta_1 * x] \tag{7.8}$$

where P_G and P_B are the probabilities of "Good" and "Bad." Here "Bad" is defined as the general inability or unwillingness to pay back a loan (we will discuss more of the target variable in the next section). x is the risk characteristic/attribute (or its WoE transformation), and β_0 and β_1 are the coefficients in the logistic regression model as described by (3.5) in Chap. 3 for default modeling. $CW = -[\beta_0 + \beta_1 * x]$ is the creditworthiness in our logistic regression model setup. Since the logit function log $(P/(1-P))$ is a strict monotone function of the probability P, a rating on the creditworthiness (or log odds ratio) is equivalent to a rating on the probability of being "Good." There are other interpretations of the creditworthiness CW; however, the log odds ratio through logistic regression is the most popular one, and the impact of creditworthiness in scaled points (credit scores) to the odds ratio is more transparent than other interpretations.

Once the logistic regression model is fitted, we have the estimates of the coefficients, and the creditworthiness can be calculated for any customer with their characteristics given. Scorecards make scaling on these creditworthiness numbers, which are hard to understand. An idea scorecard should have the following properties:

- All individual point values are positive.
- Points for all characteristics are monotone.
- The final score must lie with the range of [1,999].
- The total points are always positive.
- There is a reference score associated with a specific credit quality.
- The difference between scores implies a specific change in credit quality.

While it may not be necessary to meet all these properties, the last three are considered a must for good scorecards. Over the time, the FICO score scaling method has been popularly adopted, since it has not only a fixed-odds increment but also a benchmark odd and score. The creditworthiness CW is scaled linearly and rounded to the integer values by:

$$S = \text{int}(F_0 + F_\Delta * CW) \tag{7.9}$$

To set the parameters F_0 and F_Δ, the FICO score scaling defines the triple (also called $X \mid Y \mid Z$)

$$\text{Odd}_0 \mid S_0 \mid S_\Delta \tag{7.10}$$

where S_0 is the reference score corresponding to the odds ratio Odd_0 and S_Δ is the score points needed to double the odds ratio. So, we have the following relationships between F_0, F_Δ, and the triple $\text{Odd}_0 \mid S_0 \mid S_\Delta$:

$$F_\Delta = S_\Delta / \ln(2) \tag{7.11}$$

$$F_0 = S_0 - S_\Delta * \ln(2) * \ln(\text{Odd}_0) \tag{7.12}$$

Once we have these scaling parameters, we can convert a calculated creditworthiness (or odds ratio) to a score. For example, given the scaling system (32 | 660 | 30), we have the scaling parameters $F_\Delta = \frac{30}{\ln(2)} = 43.28$ and $F_0 = 660 - 43.28 * \ln(32) = 510$. Then a customer with an odds ratio of 20 will have a credit of $510 + 43.28 * \ln(20) = 640$.

The creditworthiness CW is a linear function of the attributes with an intercept. To see the impact of the attributes, we can first separate the impact of the intercept by a constant score:

$$S_{\text{constant}} = \text{int}(F_0 + F_\Delta * \alpha_0) \tag{7.13}$$

where $\alpha_0 = -\beta_0$ if we rewrite $CW = -[\beta_0 + \beta_1 * x] = \alpha_0 + \alpha_1 * x$ according to (7.1), and the unit (WoE) impact of the attribute x will be:

$$S_x = F_\Delta * \alpha_1 \tag{7.14}$$

Score increase or decrease due to the attribute changes can also be easily calculated; and due to the scaling linearity, all these changes are additive to the final score for any attribute. So, the FICO score scaling can assess the attribute impact to the scaled score immediately, which adds transparency of the scorecards on the attribute impacts. For example, if we get $\alpha_1 = -0.5$ from the model estimates for the risk factor x, then a single unit increase in x will contribute to $43.28 * -0.5 = -21.64$ or 22 point decrease. So, attribute contribution in unit is commonly marked in a scorecard table for easy interpretation and calculation.

In Sect. 7.1.2, Table 7.3 presents a scorecard example for small business lending. The final selected scoring model with logistic regression has seven characteristics used from the underwriting data, thus the seven characteristics for the scorecard. All attributes are bucket based and are binary for all characteristics. So, the calculation of score points is simply the sum of all corresponding attribute contributions. The model has a constant attribute (the intercept), and its point corresponds to the default buckets for all characteristics, which are marked by the 0 point with the corresponding attributes of these characteristics. We set the default buckets corresponding to the highest risky ones, so the lower bound of the scores by these drivers is 570 at the intercept. The highest score is 850. The lower bound could be broken if there are other risk events outside the table, for example, bankruptcy or legal events. For an applicant with all seven characteristics falling in the third bucket of each characteristic, the scorecard presents the score 732 shown by the final column.

7.3 Model Data for Credit Scoring

In this section, we introduce the research data set for small business lending. The research data set is by no means the complete data set for a production scorecard; however, it can be used to demonstrate the credit scoring modeling process and scorecard building.

The small business lending data set includes both the performance population and the non-performance population, and we consider it represents the full application population. The population is considered for the small business lending program focusing on subprime borrowers with various repayment schedules (daily, weekly, biweekly, and monthly). The target event is defined as the repayment risk depending on the measure of Payment vs. Due (PMvD) as defined in (7.3) in the repayment transaction history. For the non-performance population, we use a preliminary credit model built on the performance population to project the performance by simulating the target event as part of the reject inference.

We remove records with missing risk drivers, and we have the complete data set once the target event for non-performance part is filled. This is our model data, and we start the model building based on this data set. Due to the relatively small data set size and high percentage of minority group (repayment failure rate), we skip the sampling process, in which the Synthetic Minority Oversampling Technique (or SMOTE) is popularly used for credit score modeling. In Sect. 7.6 for credit score model validation, we will present some examples on how to correctly use SMOTE and its benefits for credit scoring.

For fair lending purpose, we assume a correlation matrix between the potential score drivers in the model data set and the prohibited bases, in which we assume there are five factors. Once the model data and the PCIs are ready, we can start to build the credit scoring model by executing the model selection procedure (AEVS_CS).

7.3.1 The Research Data Set for Small Business Lending

The research data set for small business lending is a partially synthetic data set from DSAC (Data Science and Analytics Consultants[10]). It was designated for subprime small business lending research purpose based on a random sample of a loan population. The raw data set includes 7000 applicants with most of the attributes populated from original records with missing values (except DTI and business age). We assume the data set includes both performance and non-performance populations. For the latter, the target event attribute RRisk (or repayment risk) is missing.

Table 7.6 presents a summary of the research data set with some of its attributes. Among these attributes, PCS is the personal credit score of the applicant; DTI is the debt-to-income ratio in percentage points; age is the business age since its inception; IndRisk and CorpRisk are classifications of the industry and corporation structure (e.g., sole proprietorship, corporation, and other); CashMargin is the difference between revenue and operating expense (not shown) in percentage points of revenue; and RRisk is the repayment risk defined by PMvD as we described in Sect. 7.2.3. There are missing values with these attributes.

For credit scoring modeling purpose, the first step is to deal with the missing values for both potential score drivers and the target RRisk. For potential score drivers, to be simple, we just remove all records with missing values. These missing values are dominated by the two drivers, IndRisk and CorpRisk (others not shown in Table 7.6). After removing these records with missing values, the total data set size reduces to 5356, among which there are 3214 records from the performance population. We carry out some preliminary data analysis on the performance data, including the candidate pool of potential score drivers based on their univariate analysis, on which we develop a preliminary credit model for the purpose of reject inference in the next section. Table 7.7 presents a summary of the performance population data (RRisk renamed RR in Table 7.7)

7.3.2 Rejection Inference and PCIs

In Sect. 7.2.4, we mentioned several approaches for the missing target treatment with the non-performance population. For our synthetic small business lending data set, a

[10] A subdivision of American Chinese Media and Services Inc.

Table 7.6 Research data for small business lending

Appid	PCS	DTI	Age	IndRisk	CorpRisk	CashMargin	Revenue	RRisk
Length: 7000	Min.: 0.0	Min.: −1.00	Min.: 0.00	Min.: 0.00	Min.: 0.0000	Min.: 0.000	Min.: 0.000e+00	Min.: 0.0000
Class: character	1st Qu.: 578.0	1st Qu.: 96.00	1st Qu.: 25.00	1st Qu.: 0.00	1st Qu.: 0.0000	1st Qu.: 0.000	1st Qu.: 2.076e+05	1st Qu.: 0.0000
Mode: character	Median: 635.0	Median: 100.00	Median: 56.00	Median: 1.00	Median: 1.0000	Median: 0.000	Median: 5.373e+05	Median: 0.0000
NA	Mean: 624.1	Mean: 90.21	Mean: 87.48	Mean: 1.08	Mean: 0.9105	Mean: 7.751	Mean: 2.495e+06	Mean: 0.4969
NA	3rd Qu.: 691.0	3rd Qu.: 100.00	3rd Qu.: 114.00	3rd Qu.: 2.00	3rd Qu.: 2.0000	3rd Qu.: 3.000	3rd Qu.: 1.500e+06	3rd Qu.: 1.0000
NA	Max.: 850.0	Max.: 100.00	Max.: 600.00	Max.: 2.00	Max.: 2.0000	Max.: 100.000	Max.: 3.508e+09	Max.: 2.0000
NA	NA	NA	NA	NA's: 1291	NA's: 1291	NA	NA	NA's: 1188

Table 7.7 Performance population data for small business lending

Appid	PCS	DTI	Age	IndRisk	CorpRisk	CashMargin	Revenue	RR
Length: 3214	Min.: 0.0	Min.: −1.00	Min.: 0.00	Min.: 0.000	Min.: 0.0000	Min.: 0.000	Min.: 0	Min.: 0.0000
Class: character	1st Qu.: 567.0	1st Qu.: 99.00	1st Qu.: 25.00	1st Qu.: 0.000	1st Qu.: 0.0000	1st Qu.: 0.000	1st Qu.: 207632	1st Qu.: 0.0000
Mode: character	Median: 632.5	Median: 100.00	Median: 54.00	Median: 1.000	Median: 1.0000	Median: 0.000	Median: 492382	Median: 0.0000
NA	Mean: 624.7	Mean: 97.43	Mean: 87.73	Mean: 1.106	Mean: 0.8998	Mean: 2.311	Mean: 1532756	Mean: 0.4188
NA	3rd Qu.: 691.0	3rd Qu.: 100.00	3rd Qu.: 113.00	3rd Qu.: 2.000	3rd Qu.: 2.0000	3rd Qu.: 1.000	3rd Qu.: 1332349	3rd Qu.: 1.0000
NA	Max.: 847.0	Max.: 100.00	Max.: 600.00	Max.: 2.000	Max.: 2.0000	Max.: 100.000	Max.: 54024002	Max.: 1.0000

```
Call:
glm(formula = frm_df, family = binomial(link = "logit"), data =
Perf_pop)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1176  -1.0825  -0.4231   1.1352   4.7914

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.7705919  0.2587924   6.842 7.82e-12 ***
PCS         -0.0029021  0.0003995  -7.265 3.73e-13 ***
Age         -0.0011347  0.0004013  -2.828  0.00469 **
IndRisk      0.2193013  0.0454348   4.827 1.39e-06 ***
CorpRisk    -0.1148524  0.0492335  -2.333  0.01966 *
CashMargin  -0.2905104  0.0266088 -10.918  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4391.2  on 3213  degrees of freedom
Residual deviance: 3959.8  on 3208  degrees of freedom
AIC: 3971.8

Number of Fisher Scoring iterations: 7
```

Fig. 7.4 Preliminary credit model based on performance population

natural selection is the model-based statistical treatment since we don't have any additional information about the non-performance population.

Based on the performance population, we develop a credit model using the defined target event (based on PMvD cut at 85%, 1-year observation window) and several risk factors as described in Table 7.7. The two factors, DTI and Revenue, are not significant based on our performance data. DTI should be a significant score driver in small business lending; however, our DTI data has poor quality, and it is highly correlated with the CashMargin factor (correlation = −0.9). Further looking into the performance data, we see that DTI is likely the difference between 100 and CashMargin with a few (10) values either −1 or 0. So, we prefer to include CashMargin in our credit model.

Figure 7.4 shows the preliminary credit model by the binary logit regression. PCS, Age, and CashMargin are all significant (at 1% level) with a negative sign, which is expected as the higher of these drivers leads to lower repayment risk. Both IndRisk and CorpRisk are categorical drivers; however, here we treat them as continuous to maintain their risk level impact monotone to the repayment risk. CorpRisk is not significant at the 1% level, but significant at the 5% level.

The preliminary credit model is applied to the non-performance population to estimate the probability of repayment failure, and the repayment risk event is simulated from these estimated probabilities. Table 7.8 presents the summary of the complete data after filling in the RR event missing values for the

Table 7.8 Complete research data for small business lending

Appid	PCS	DTI	Age	IndRisk	CorpRisk	CashMargin	Revenue	RR
Length: 5356	Min.: 0.0	Min.: −1.00	Min.: 0.00	Min.: 0.000	Min.: 0.0000	Min.: 0.000	Min.: 0.000e+00	Min.: 0.0000
Class: character	1st Qu.: 567.0	1st Qu.: 199.00	1st Qu.: 25.00	1st Qu.: 0.000	1st Qu.: 0.0000	1st Qu.: 0.000	1st Qu.: 2.010e +05	1st Qu.: 0.0000
Mode: character	Median: 631.0	Median: 100.00	Median: 54.00	Median: 1.000	Median: 1.0000	Median: 0.000	Median: 4.866e +05	Median: 0.0000
NA	Mean: 623.9	Mean: 97.37	Mean: 87.05	Mean: 1.093	Mean: 0.9113	Mean: 2.434	Mean: 1.668e+06	Mean: 0.4255
NA	3rd Qu: 689.2	3rd Qu.: 100.00	3rd Qu.: 111.00	3rd Qu.: 2.000	3rd Qu.: 2.0000	3rd Qu.: 1.000	3rd Qu.: 1.320e +06	3rd Qu.: 1.0000
NA	Max.: 847.0	Max.: 100.00	Max.: 600.00	Max.: 2.000	Max.: 2.0000	Max.: 100.000	Max.: 1.000e+09	Max.: 1.0000

Table 7.9 Correlation (in absolute value) between score drivers and prohibited bases

Correlation	Potential score drivers						
Prohibited bases	PCS	DTI	Age	IndRisk	CorpRisk	CashMargin	Revenue
PH1	0.7863	0.8591	0.2727	0.8221	0.4839	0.6985	0.2862
PH2	0.2672	0.4466	0.5357	0.2917	0.8280	0.3543	0.7202
PH3	0.4572	0.4416	0.5108	0.5448	0.7808	0.7864	0.4440
PH4	0.8381	0.3066	0.7065	0.9137	0.2963	0.2542	0.4645
PH5	0.7391	0.2755	0.2751	0.7792	0.7770	0.4715	0.5924
PCI_MAX	0.8381	0.8591	0.7065	0.9137	0.8280	0.7864	0.7202
PCI_AVG	0.6176	0.4659	0.4602	0.6703	0.6332	0.5130	0.5015

non-performance population. We can see that the averages of the RR events between the performance and the complete populations are close (0.4188 vs. 0.4255), which indicates that the non-performance population doesn't deviate from the performance population too much with respect to the repayment failure rate.

If there is a big deviation on the target risk event rate between the populations, one should investigate the cause of such deviation. If such deviation can be explained by the risk drivers, then these drivers should be checked for their PCIs to avoid fair lending violations. Table 7.9 presents the correlation table between the potential score drivers and the simulated prohibited bases. Ideally, the prohibited bases should be observed for the complete population or a sample of the complete population. Here, for illustration purpose we obtained these factors through simulation to represent the five prohibited factors of race, color, sex, marital status, and religion. Both PCIs measured in maximum and average are calculated. If any of the measure exceeds the threshold, then the corresponding driver needs to be removed from the model.

With our simulated PCIs, the highest correlation in maximum is between IndRisk and the fourth prohibited factor (0.9127). If we use the threshold of 0.95, then all potential score drivers are immune to the fair lending violation.

Certainly, the PCIs in Table 7.9 depend on the sample data selected to calculate these correlation coefficients. Such sample data should be properly designed in the fair lending tests. It may not be appropriate to collect the prohibited bases for the entire population due to regulation and cost, but the designed sample should be representative for the population of all potential applicants based on geography and demography.

7.4 Credit Score Models and Scorecards

The focus of credit score modeling has shifted from the probability of repayment risk to the creditworthiness. Such shift follows the extensive use of scorecards as well as the regulatory requirements on the use of proper score drivers to comply with fair lending laws. So, the once popular tools on modeling probability of repayment risk for credit scoring, including the linear discriminant analysis and linear probability models, look out-of-date. Instead, logistic regression and other generalized linear models, which explicitly link the probability of repayment risk to the creditworthiness through the logit (or other) link function, become dominant, even with the overwhelming machine learning techniques claiming smarter ways to figure out special segments or subpopulations with distinct creditworthiness. In addition, the high stability requirements for credit score models also prefer the logistic regression type of models.

With creditworthiness, the credit score model is the same as the Probability of Default (PD) model as in other credit model applications. However, different from other credit model applications, the final goal of the credit score modeling is to build scorecards using the credit score model outputs, which more focus on the creditworthiness – a linear combination of the model coefficients and the score drivers. As scorecards prefer all score drivers binned, the final credit score model could fit on the binned score drivers. However, for score driver selection as described in the AEVS_CS procedure, the unbinned score drivers should be used either as continuous or factor variables.

In this section, we first run the proposed AEVS_CS procedure for score driver selection based on the complete data we prepared in the previous section. We treat all potential score drivers as continuous variables even with some of the factor variables. We code the factor variables in integers such that the different factor levels are sorted in numerical level by the order of risk. In this way, we assume the log odds ratio are proportional equally among the factor levels, which is a strong assumption for the model but should be reasonable for the variable selection purpose given the order of risk levels among these factor levels. The main difference between AEVS_CS and AEVS we applied in other chapters is the inclusion of PCIs.

The selected models are ranked and subject to the final review. The score drivers in the final selected model will be used to construct scorecards. We explain how to use the WoE (Weight of Evidence) for binning of score driver as discussed in Sect. 2.6.3. In practice, WoE-based binning only provides some binning guidance, and the exact binning is still an art, which is based on business requirement and modeler preference. Again, due to the data limitation, the scorecards built on our research data are just for illustration purpose, and they may not look like the regular scorecard as we present early in the chapter (Table 7.3).

7.4.1 Score Driver Selection and AEVS_CS

The score driver selection process for the credit score model as described in AEVS_CS includes data preparation, model selection, and model ranking. Based on the research data for small business lending, we go over these modeling components to illustrate how score drivers are selected through our automatic variable selection process designed for credit scoring.

Data Preparation

Data preparation starts from the complete research data post the reject inference (see Table 7.8 for a summary). The data is cross-sectional in the sense that each row of record corresponds to an applicant with potential score drivers and a repayment risk event indicator observed in a fixed window (e.g., 1 year).

The model data set has 5356 observations with 7 potential score driver as we introduced early – PCS, DTI, Age, IndRisk, CorpRisk, CashMargin, and Revenue. RR is the repayment risk indicator defined by PMvD as we described in Sect. 7.2.3. As we pointed out early, the two categorical variables IndRisk and CorpRisk are treated as continuous variables together with the rest five variables in variable selection. The values for these two variables are 0, 1, and 2 corresponding to low-, middle-, and high-risk levels. DTI and CashMargin are highly correlated. RR has a high event rate of 42.55%, which indicates oversampling of the minority group is not necessary.

Model Selection

Due to limited candidate score drivers, we skip the dimension reduction steps in AEVS_CS. Our pool of candidate variables includes these seven variables – PCS, DTI, Age, IndRisk, CorpRisk, CashMargin, and Revenue. There are $2^7 = 128$ variable combinations, which we consider relatively small and apply the recursive algorithm in the AEVS_CS procedure as described in Sect. 7.2.5. We applied the PCI matrix shown in Table 7.9 and use the threshold 0.95, so all score drivers are immune to the fair lending test in our case.

Model Ranking

Survived models from the recursive AEVS_CS algorithm are ranked by the three goodness-of-fit measures – SIC, AIC, and CV_v. We choose $v = 11$ for the v-fold cross-validation measure due to computational efficiency. For the three goodness-of-fit measures, we prefer SIC and use it as the first ranking criteria. Top five models based on SIC are selected, and their ranks based on SIC, AIC, and CV_v are used to compute the full average ranks.

The goodness-of-fit measures and their ranks for the top five selected models based on SIC are presented in Table 7.10. Note that the top five models are ranked by the first criteria SIC in the table, and rank = 1 indicates the top rank. Model name is defined by the model formula in the generalized liner model fit, which includes both response variable and the score drivers as independent variables.

Table 7.10 shows that SIC and AIC select the same model as the top model, which includes five score drivers. This is the same model as the preliminary model we used for the reject inference (with different coefficients). This model also ranked second by CV11. Since the data are not simulated from a true model, we can only assess these selected models by the criteria. Further looking into the pool of the seven candidate score drivers as we did for the preliminary model, we confirm that DTI and CashMargin are highly correlated as we see in the performance population. So, models including these two score drivers simultaneously are likely rejected by the 5% significant level for a single score driver. Similarly, as in the preliminary

Table 7.10 Top five score models selected

ModelName5	SIC5	AIC5	CV11	rankSIC5	rankAIC5	rankCV	AveRank
RR ~ PCS + Age + IndRisk + CorpRisk + CashMargin	6653.83	6614.31	0.3393	1	1	2	1.3333
RR ~ PCS + Age + IndRisk + CashMargin	6658.61	6625.68	0.3404	2	4	4	3.3333
RR ~ PCS + DTI + Age + IndRisk + CorpRisk	6661.27	6615.17	0.3491	3	2	5	3.3333
RR ~ PCS + Age + CorpRisk + CashMargin	6662.36	6616.26	0.3389	4	3	1	2.6667
RR ~ PCS + DTI + Age + IndRisk	6665.86	6626.34	0.3402	5	5	3	4.3333

```
Call:
glm(formula = frm_df, family = binomial(link = "logit"), data =
MFITData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1521  -1.0625  -0.5184   1.1353   4.6712

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.8200617  0.1975006   9.215  < 2e-16 ***
PCS         -0.0030042  0.0003040  -9.883  < 2e-16 ***
Age         -0.0016548  0.0003126  -5.294 1.19e-07 ***
IndRisk      0.2538476  0.0353994   7.171 7.45e-13 ***
CorpRisk    -0.1376566  0.0376967  -3.652 0.000261 ***
CashMargin  -0.2562990  0.0188681 -13.584  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7305.7  on 5355  degrees of freedom
Residual deviance: 6602.3  on 5350  degrees of freedom
AIC: 6614.3

Number of Fisher Scoring iterations: 7
```

Fig. 7.5 Top score model selected

model, revenue is also not significant and is likely rejected by the 5% significant level.

Figure 7.5 presents the model fitting results for the top model using the complete research data set. This is the model we will use to build scorecard.

7.4.2 Score Driver Binning and Scorecards

Scorecards prefer to use the binned score drivers, which put the applicants with the score driver values falling in a range in a single group and lead to the same odds ratio, thus the same credit score. So, the scorecards tend to group applicants into limited number of groups sharing the same score to simplify the credit scoring process. The binning process is basically a process to transfer continuous or categorical variables into grouped (or binned) variables, which only takes values in a limited number of bins.

As introduced in Sect. 2.6.3, WoE (Weight of Evidence) is the most popular binning approach for score drivers in the binary scoring models. It optimizes the bins through the Information Value (or IV) defined by (2.9) in Sect. 2.6.3, which is a function of some parameters including the bins. The recursive splitting and joining algorithms commonly used to solve the optimization bins are equivalent to the binary decision tree based on minimizing the impurity function (3.45) in Sect.

3.1.1.3 with the Gini index. As with all tree-based algorithms, the solution is always an approximation to the optimal solution, and the stability of the solution depends on the data and the implementation of the algorithm. So, the optimized bins heavily depend on the score driver data as well as the WoE-based binning algorithms. There are cases that such optimized bins don't exist or look too trivial to be used in practice. So, very often, bins are specified by risk managers or modelers based on their preference and experience.

In the following, we present some examples with binning of the score drivers in our final selected scoring model and illustrate how optimal binning based on Information Value can be used for scorecard building in practice. We use the R-package SMBinning Version 0.9 by Herman Jopia, which provides both the WoE-based binning optimization and score scaling tools. The following code shows how we execute SMBinning functions to bin each of the five score drivers and then fit the binary logit model on the binned score drivers.

```
## Final Binning ##
cuts_PCS<-c(570, 625, 680, 770)
result_PCS=smbinning.custom(df=MFITData, y="RR", x="PCS", cuts=cuts_PCS)
MFITData_Bin=smbinning.gen(MFITData,result_PCS,"PCS_Bin")

cuts_Age<-c(11, 35, 119)
result_Age<-smbinning.custom(df=MFITData, y="RR", x="Age", cuts=cuts_Age)
MFITData_Bin=smbinning.gen(MFITData_Bin,result_Age,"Age_Bin")

result_CM=smbinning(df=MFITData, y="RR", x="CashMargin")
MFITData_Bin=smbinning.gen(MFITData_Bin,result_CM,"CashMargin_Bin")

result_IndRisk=smbinning.factor(df=MFITData, y="RR", x="IndRisk")
MFITData_Bin=smbinning.factor.gen(MFITData_Bin,result_IndRisk,"IndRisk_Bin")

result_CorpRisk=smbinning.factor(df=MFITData, y="RR", x="CorpRisk")
MFITData_Bin=smbinning.factor.gen(MFITData_Bin,result_CorpRisk,"CorpRisk_Bin")

## Change Target Variable ##
Good<-data.frame(1-MFITData$RR)
names(Good)<-c("Good")
MFITData_Bin<-cbind.data.frame(MFITData_Bin, Good)

## Fit on Binned Score Drivers ##
frm_Bin <- as.formula(paste("Good ~", "PCS_Bin", "+", "Age_Bin", "+",
"IndRisk_Bin",  "+", "CorpRisk_Bin", "+", "CashMargin_Bin"))
glmmodel_Bin<-glm(formula=frm_Bin, family=binomial(link='logit'),
data=MFITData_Bin)
summary(glmmodel_Bin)
```

We start from PCS. The direct SMBinning suggests the four bins when we execute the following code as shown in Table 7.11:

```
## Inital Binning ##
result_PCS=smbinning(df=MFITData, y="RR", x="PCS")
result_PCS$ivtable
```

Table 7.11 shows that SMBinning suggests the following three cut points with PCS (627, 681, 769) with the corresponding counts (total, good, and bad) within each bin, as well as odds, log(odds), WoE, and IV values. The first bin takes about 48% of total counts, and WoE values are monotone with PCS increasing. By default,

Table 7.11 Direct SMBinning results with PCS

Cutpoint	CntRec	CntGood	CntBad	PctRec	GoodRate	BadRate	Odds	LnOdds	WoE	IV
< = 627	2594	1314	1280	0.4843	0.5066	0.4934	1.0266	0.0262	0.3264	0.0524
< = 681	1251	496	755	0.2336	0.3965	0.6035	0.657	−0.4201	−0.1199	0.0033
< = 769	1240	422	818	0.2315	0.3403	0.6597	0.5159	−0.6619	−0.3616	0.0292
>769	271	47	224	0.0506	0.1734	0.8266	0.2098	−1.5615	−1.2613	0.0658
Missing	0	0	0	0	NA	NA	NA	NA	NA	NA
Total	5356	2279	3077	1	0.4255	0.5745	0.7407	−0.3002	0	0.1507

SMBinning only has the restriction that each bin must have over 5% of total counts. Given these observations, we would test a further split with the first bin by 570 and prefer to keep the WoE monotone. For other cut points, we would round to the nearest integers ending in 0's or 5's. So, we custom SMBinning using the following cut points with PCS (570, 625, 680, 770). This can be executed with the *smbinning.custom* function:

```
cuts_PCS<-c(570, 625, 680, 770)
result_PCS=smbinning.custom(df=MFITData, y="RR", x="PCS", cuts=cuts_PCS)
result_PCS$ivtable
```

Table 7.12 shows the new binning results with the custom cut points. The counts among these bins are more evenly distributed while we keep WoE monotone. The total IV increases a little. These are the final bins we use for the scorecard.

Similar binning analysis is carried out for all other score drivers. For Age, the direct SMBinning suggests only a single cut point at 11 months. To add granularity on Age for the scorecard, we added two more cut points, and the final cut points are (11, 35, 119). For the third continuous score driver CashMargin, SMBinning suggests the cut points as (0, 1, 3). Looking into the four bins and the distribution of CashMargin values, we consider the cut points, and the resulting bins are reasonable for the scorecard. So, we did not custom the cut points further.

For the two categorical score drivers "IndRisk" and "CorpRisk," SMBinning explores the optimal grouping with the level of the categorical driver in respect to the Information Value. This can be executed by the function *smbinning.factor:*

```
result_IndRisk=smbinning.factor(df=MFITData, y="RR", x="IndRisk")
result_CorpRisk=smbinning.factor(df=MFITData, y="RR", x="CorpRisk")
```

The results show that SMBinning suggests no further grouping for these two categorical score drivers and keep the original three levels (marked as 0, 1, and 2). It should be pointed out that IndRisk has a reverse order of risk levels (higher level means riskier).

The newly created binned score drivers corresponding to the original score drivers can be created by the *smbinning.gen* or *smbinning.factor.gen* functions for continuous and categorical drivers correspondingly and added to the original model data set.

With all the binned score drivers available, we can fit the binary logit model on the binned score drivers to obtain odds for each cell of the cross bins. Before that, we redefine the target from risk event indicator "RR" to "Good," Figure 7.6 presents the binary logit model results with the binned score drivers. Note that all binned score drivers are considered factor variables with the corresponding bins as the level of the factor. Model coefficients are not shown for the default level of the binned score drivers.

Table 7.12 Custom SMBinning results with PCS

Cutpoint	CntRec	CntGood	CntBad	PctRec	GoodRate	BadRate	Odds	LnOdds	WoE	IV
< = 570	1396	730	666	0.2606	0.5229	0.4771	1.0961	0.0918	0.392	0.0407
< = 625	1166	567	599	0.2177	0.4863	0.5137	0.9466	−0.0549	0.2453	0.0133
< = 680	1268	505	763	0.2367	0.3983	0.6017	0.6619	−0.4127	−0.1125	0.003
< = 770	1260	433	827	0.2353	0.3437	0.6563	0.5236	−0.6471	−0.3468	0.0273
>770	266	44	222	0.0497	0.1654	0.8346	0.1982	−1.6185	−1.3183	0.0697
Missing	0	0	0	0	NA	NA	NA	NA	NA	NA
Total	5356	2279	3077	1	0.4255	0.5745	0.7407	−0.3002	0	0.154

```
Call:
glm(formula = frm_Bin, family = binomial(link = "logit"), data = MFITData_Bin)

Deviance Residuals:
    Min        1Q    Median        3Q       Max
-3.0968   -1.0689   0.2706   1.0401   1.5380

Coefficients:
                      Estimate Std. Error z value             Pr(>|z|)
(Intercept)           -2.98379    0.18083 -16.500 < 0.0000000000000002 ***
PCS_Bin02 <= 625       0.23484    0.08964   2.620             0.008798 **
PCS_Bin03 <= 680       0.62057    0.08923   6.955      0.00000000000353 ***
PCS_Bin04 <= 770       0.93239    0.09180  10.157 < 0.0000000000000002 ***
PCS_Bin05 > 770        1.88301    0.18906   9.960 < 0.0000000000000002 ***
Age_Bin02 <= 35        2.69732    0.16891  15.969 < 0.0000000000000002 ***
Age_Bin03 <= 119       2.71411    0.16494  16.455 < 0.0000000000000002 ***
Age_Bin04 > 119        2.78464    0.17141  16.246 < 0.0000000000000002 ***
IndRisk_Bin02 = '1'   -0.13545    0.08163  -1.659             0.097063 .
IndRisk_Bin03 = '2'   -0.47469    0.07643  -6.211      0.00000000052685 ***
CorpRisk_Bin02 = '1'  -0.03475    0.07437  -0.467             0.640318
CorpRisk_Bin03 = '2'   0.35444    0.08212   4.316      0.00001587328164 ***
CashMargin_Bin02 <= 1  0.39280    0.11663   3.368             0.000757 ***
CashMargin_Bin03 <= 3  2.67636    0.19918  13.437 < 0.0000000000000002 ***
CashMargin_Bin04 > 3   3.30906    0.18100  18.282 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7305.7  on 5355  degrees of freedom
Residual deviance: 5879.7  on 5341  degrees of freedom
AIC: 5909.7

Number of Fisher Scoring iterations: 5
```

Fig. 7.6 Top score model with binned score drivers

Figure 7.6 shows the binary logit model estimates corresponding to binned score drivers. Totally, there are 480 = 5*4*4*3*3 cells with these five binned score drivers. We have the following notes on the binned binary logit model:

- The parameter estimates for a binned score driver, which correspond to the log odds of the bins of the score driver, are not guaranteed monotone even the WoE of the binned score driver is monotone on the score driver. This can be seen with the CorpRisk driver, which has parameter estimates 0, −0.03475, and 0.3544 with the three factor levels 0, 1, and 2. Table 7.13 shows that the WoE is monotone on these three factor levels. This indicates that there are interactions among the score drivers.
- Not all parameter estimates for the binned score drivers are statistically significant, especially when cut points are customed. For the purpose of scorecards, though statistical significance is not a critical requirement, these non-significant bins tend to have non-significant scores.
- The binary logit model with binned score drivers fits the data better than the original binary logit model (non-binned binary logit model; see Fig. 7.5) as shown by the smaller AIC in Fig. 7.6 due to more parameters used in the binned

Table 7.13 SMBinning results with CorpRisk

Cutpoint	CntRec	CntGood	CntBad	PctRec	GoodRate	BadRate	Odds	LnOdds	WoE	IV
= '0'	1922	866	1056	0.3588	0.4506	0.5494	0.8201	−0.1984	0.1019	0.0037
= '1'	1987	884	1103	0.371	0.4449	0.5551	0.8015	−0.2213	0.0789	0.0023
= '2'	1447	529	918	0.2702	0.3656	0.6344	0.5763	−0.5512	−0.251	0.0166
Missing	0	0	0	0	NA	NA	NA	NA	NA	NA
Total	5356	2279	3077	1	0.4255	0.5745	0.7407	−0.3002	0	0.0226

```
> glm_scaled$logitscaled # Scaled model
[[1]]
   Characteristic Attribute Coefficient      Weight WeightScaled Points
1     (Intercept)           -2.98378731 -129.140855      0.00000      0
2         PCS_Bin 01 <= 570  0.00000000    0.000000     98.73049     99
3         PCS_Bin 02 <= 625  0.23484421   10.164257    108.89474    109
4         PCS_Bin 03 <= 680  0.62056959   26.858780    125.58927    126
5         PCS_Bin 04 <= 770  0.93239245   40.354739    139.08522    139
6         PCS_Bin  05 > 770  1.88301070   81.498306    180.22879    180
7         Age_Bin  01 <= 11  0.00000000    0.000000     98.73049     99
8         Age_Bin  02 <= 35  2.69732000  116.742306    215.47279    215
9         Age_Bin 03 <= 119  2.71410818  117.468912    216.19940    216
10        Age_Bin  04 > 119  2.78464153  120.521656    219.25214    219
11    IndRisk_Bin  01 = '0'  0.00000000    0.000000     98.73049     99
12    IndRisk_Bin  02 = '1' -0.13544828   -5.862317     92.86817     93
13    IndRisk_Bin  03 = '2' -0.47469089  -20.545026     78.18546     78
14   CorpRisk_Bin  01 = '0'  0.00000000    0.000000     98.73049     99
15   CorpRisk_Bin  02 = '1' -0.03474784   -1.503916     97.22657     97
16   CorpRisk_Bin  03 = '2'  0.35443753   15.340358    114.07084    114
17 CashMargin_Bin   01 <= 0  0.00000000    0.000000     98.73049     99
18 CashMargin_Bin   02 <= 1  0.39279823   17.000642    115.73113    116
19 CashMargin_Bin   03 <= 3  2.67635741  115.835027    214.56551    215
20 CashMargin_Bin    04 > 3  3.30906027  143.218945    241.94943    242
```

Fig. 7.7 The scorecard built on the top score model

model. An alternative binary logit model fits the WoE corresponding to each score driver, which can guarantee the monotone of log odds of the bins as long as WoE is monotone on the score driver. We will explore this alternative later in model validation.

The R package SMBinning also provides score scaling tools. Using the same approach as we described in Sect. 7.2.7, the following code with SMBinning executes the FICO type of score scaling:

```
glm_scaled=smbinning.scaling(glmmodel_Bin,pdo=30,score=770,odds=30)
glm_scaled$logitscaled # Scaled model
glm_scaled$minmaxscore # Expected minimum and maximum Score
```

In the function *smbinning.scaling*, *pdo* defines the points to double the odds, and *score* defines the reference score corresponding to the specified *odds*. The corresponding scores for all bins of the model score drivers can be shown by the *logitscaled* component of the scaled model as shown in Fig. 7.7. SMBinning sets equally a score corresponding to each of the default bin of the score driver (note that IndRisk default bin has the highest score instead of the lowest score as with other score drivers.)

One can check the scaled score range by the *minmaxscore* component of the scaled model, which is [472, 854].

The scorecard shown in Fig. 7.7 is the initial scorecard and subject to adjustments. Very often some bins are subject to further splitting and some to merging. For example, based on modeling experience, the three bins for Age (Age_Bin 02, 03, and 04) in Fig. 7.7 have very close points and should be merged. Risk managers may prefer more granular bins than the initial scorecard, and modelers and risk managers need several rounds of interactions to decide the final scorecard. As all score models (and scorecards) are used to rate the repayment risks, the relative size of scores may be adjusted by prior knowledge from risk managers and subject experts. Risk managers may also adjust the points for a particular score driver based on their forecast of the coming portfolio characteristics. We call all such analysis and adjustments the post model adjustments or model overlay. Post model adjustments and model overlay should be documented in detail with proper reasons for each of these adjustments. Such documentation is required for model validation purpose.

With the scaled binary logit model, we can scale any new data of applicants given the five score drivers by executing the *smbinning.scoring.gen* function in the SMBinning package:

```
NewData_score=smbinning.scoring.gen(smbscaled=glm_scaled, dataset=NewData_Bin)
```

Note that *smbinning.scoring.gen* requires the new data must have the binned risk drivers instead of the original risk drivers.

SMBinning package can even generate the SQL code for the scorecard:

```
smbinning.scoring.sql(glm_scaled)
```

as shown in Fig. 7.8.

7.5 Scorecard Use

The usage of scorecards has been extensively extended in recent years, especially with the faster growth of FinTech companies. The extension has been both on the horizontal and vertical directions. On the horizontal direction, underwriting is not the only area for scorecards; marketing also extensively use scorecards for quick decisions. On the vertical direction, multiple scorecards are created and used jointly for decision-making. In the following, we simply introduce such trends on scorecard usage as some evidence on why scorecards become more and more popular on decision-making.

Single Scorecard for Multiple Purposes

Most scorecards are built for the purpose of underwriting of some products in the scope. However, given the extensive benefits of cross-sale and lack of data for new products, scorecards built for one purpose are commonly extended to the usage of other purposes, not just for underwriting, even for marketing purpose.

```
-- Replace 'TableName' with your table in SQL
update TableName set
PCS_BinPoints=(
 case
 when PCS_Bin = '01 <= 570' then 99
 when PCS_Bin = '02 <= 625' then 109
 when PCS_Bin = '03 <= 680' then 126
 when PCS_Bin = '04 <= 770' then 139
 when PCS_Bin = '05 > 770' then 180
 else Null end
 ),
Age_BinPoints=(
 case
 when Age_Bin = '01 <= 11' then 99
 when Age_Bin = '02 <= 35' then 215
 when Age_Bin = '03 <= 119' then 216
 when Age_Bin = '04 > 119' then 219
 else Null end
 ),
IndRisk_BinPoints=(
 case
 when IndRisk_Bin = '01 = ''0''' then 99
 when IndRisk_Bin = '02 = ''1''' then 93
 when IndRisk_Bin = '03 = ''2''' then 78
 else Null end
 ),
CorpRisk_BinPoints=(
 case
 when CorpRisk_Bin = '01 = ''0''' then 99
 when CorpRisk_Bin = '02 = ''1''' then 97
 when CorpRisk_Bin = '03 = ''2''' then 114
 else Null end
 ),
CashMargin_BinPoints=(
 case
 when CashMargin_Bin = '01 <= 0' then 99
 when CashMargin_Bin = '02 <= 1' then 116
 when CashMargin_Bin = '03 <= 3' then 215
 when CashMargin_Bin = '04 > 3' then 242
 else Null end
 ),
Score=(PCS_BinPoints + Age_BinPoints + IndRisk_BinPoints + CorpRisk_BinPoints +
CashMargin_BinPoints)
```

Fig. 7.8 SQL code for the scorecard

For cross-product use, the extension of scorecard use should be focused on the difference of the products to make sure the difference should be within the product domain. For example, scorecards built for credit cards may not work well for debit cards.

For cross-area use, the extension of scorecard use should be taken as more of a reference purpose on the decision-making given that the existing scorecard likely has a different target from that of the cross-area use. For example, a scorecard for the purpose of underwriting focuses on the repayment risk, while marketing on related products may focus of the response rate.

Multiple Scorecards for a Single Purpose

Use multiple scorecards for a single decision-making may complicate the decision process. However, for the purpose of multiple underwriting targets, multiple scorecards are natural selection. One popular practice in the small business lending area is using the scorecard developed under the repayment risk target for the initial underwriting and then using the scorecard developed under the total return target for further underwriting. The scorecard under the total return target uses a different target event on the same model population. One example of the total return target is the Internal Rate of Return (IRR), which is calculated based on the cash flows of the performance population. A binary target event is defined based on the required minimum IRR (bad as the ones with IRR lower than this minimum).

Using multiple scorecards for a single purpose has become more and more popular for FinTech companies, for which the cost of scorecard regulation is less compared to larger financial institutions. However, this will lead to the suspicion that the underwriting process will be a "cherry picking" and potential violation of the fairing lending laws and the requirement of extensive fair lending tests.

7.6 Model Validation and Performance Monitoring

The general model risk management framework has been described in Sect. 4.5, and here we focus on the score model validation and performance monitoring. Similarly, we will cover the following main components of the model validation procedure: model scope and usage, input and assumptions, theory and design, implementation and output analysis, and performance monitoring.

The scope of credit models in applications presented in the previous chapters is mostly on the portfolio level. Different from those credit model applications, credit scoring likely covers a broad scope of populations and multiple products. Although the small business lending example we used in this chapter has relatively small sample size and we have not applied sampling techniques, in practice for computation efficiency with big data, sampling is necessary as a part of the input data operations. We will review some sampling techniques for score models and point out risks with certain sampling techniques, especially the synthetic sampling techniques (e.g., SMOTE) borrowed from their popular uses in machine learning.

For modeling methodology, there is a higher flexibility for score models due to more business requirements than regulatory standards. Although the chorus has gradually settled down on logistic regression models, there are alternatives on score driver selection, transformation and binning, and scorecard building. We compare an alternative model using WoE (weight of evidence) with the model using general binned score drivers as we presented in Sect. 7.4.2 based on our small business lending data and summarize the pros and cons of different types of models for credit scoring.

As we pointed early in the chapter, score models and scorecards heavily depend on the model population. Besides the standard population stability index (PSI) checking for score models, back-testing is also popularly used in score model performance monitoring. Fair lending tests as the main regulatory requirement for score models mark the difference of score model review and validation from other credit models. Although conceptually simple, regulations around fair lending tests have been under developing, and new standards may come in practice any time. We implement the PCI concept fundamentally into the score driver and model selection procedure and would approach a solution from the start of the modeling process rather than from the post modeling process.

Lastly, although score models are subject to less regulatory frameworks, to properly manage the risk of models critically important to a firm's business, the requirements of comprehensive model validation and review, as well as a sound governance process should be put on high standards by the management. This is especially important for large banks since models may be developed, derived, adopted, or even shared among different divisions and legal entities. Given the complexity of line of businesses and products within large banks, models in use could easily have gaps in the development, validation and review, and usage and other risks. The management should acknowledge that such risks need to be sorted out and proper controls should be put in place.

By going through the model validation components and performance monitoring process for score models, we would identify potential issues leading to the risks mentioned above and propose controls and assessments in practice.

7.6.1 Scope and Usage

As described in the previous chapters, model scope defines the model coverage on products, and model usage defines the applications in which the model can be used, and model usage should cover model limitations and use restrictions. Score models usually cover a relatively broad range of products, though sub-models may focus on some specific products or line of businesses. The main model usage for score models is for underwriting but could extend beyond that as we discussed in Sect. 7.5.

Such model structure complexity requires model scope and usage be clearly defined and communicated by model sponsor with all related participants as the initial step in model development and validation. They should be clearly documented in the model documents and validation reports in addition to the user guide.

For model risk control purpose, we list some potential risks related to score model scope and usage:

- Incorrect product coverage by the proposed model
- Wrong line of business involvement
- Wrong model development design
- Wrong model validation requirement applied
- Wrong model implementation
- Wrong model use
- Model development failure

- Rejection from model validation and review
- Violation of regulatory requirements

Score models are mostly developed per business requests rather than risk management requirements. Although the fairing lending laws put some regulatory requirements for these models, over a long time, business has been the sole player on model development, testing, and use. Model risk management hasn't been on the same standards compared to models presented in the previous chapters, which are commonly under regulatory review and exam. Such differences are more obvious when looking into the model risk management in FinTech companies on score models, for which model risk is purely on fairing lending violations. One effort on score model risk management enhancement has been shown on the mortgage industry, where FHFA pushed out some model risk review guidance on score models used for underwriting purpose. Such effort should be expanded to other industries involving score model use in underwriting.

As an initial step for such model risk management enhancements, we suggest the following controls focusing on model scope and usage:

- Sufficient BRD (Business Requirement Document) with model scope and usage clearly described and approved by higher management
- Model development policy and procedure with a periodically updated model development document template including specific chapter/section on model scope and usage

7.6.2 Inputs and Assumptions

Score models and scorecards heavily depend on the model input data and assumptions. Section 4.5.2 provides some general risks related to inputs and assumptions for loss forecasting models. For score models, these risks generally apply. In addition, input data for score models are commonly collected from different sources and more often subject to incompleteness with large portion data missing. Score models also have strong assumptions to meet business intuitions and requirements. So, based on the characteristics of score model inputs and assumptions, we summarize these risks as follows:

- Random error for observed data
- Systemic error for observed data
- Insufficient data coverage
- Incomplete data with missing values
- Incorrect use of proxy data
- Inconsistency of synthetic data
- Improper data sampling
- Implied model assumptions

- Improper model assumptions
- Insufficient assessment of model assumptions

Correspondingly, for score models, we recommend the following controls should be in place to avoid these risks:

- Data quality control, including outlier detection, hard limits, similarity tests, and machine learning-based data quality checks
- Data coverage analysis
- Model data sensitivity analysis
- Proxy data analysis
- Synthetic data analysis
- Data sampling analysis
- Model assumption identification
- Model assumption sensitivity analysis
- Statistical tests on critical model assumptions

Data risk, for score models, is still a part of the IT management, especially for FinTech companies. In other modeling areas, data risk is starting to become an independent risk taxonomy, especially when its downstream consumers are model-related applications. For score models, data risk related to models should be considered part of the model risk and require formal review and validation through some of the controls listed above.

Over the past years, synthetic sampling methods (e.g., SMOTE) have been extensively used in imbalance data mining and machine learning by oversampling the minority groups. While such synthetic sampling methods have been shown successful in improving classification accuracy with imbalanced data, these methods are also boldly introduced into score models. Here, we would point out a caution of using such techniques in score models. The fundamental difference for imbalanced learning and credit scoring is that score models are required to be interpretable with score drivers in the model. The inclusion of synthetic data may destroy the proper relationship between the creditworthiness and score drivers and make the score model uninterpretable. In the following, using our small business lending data, we demonstrate that the synthetic sampling method (SMOTE) leads to uninterpretable score drivers in the score model, while classical stratified sampling methods keep the proper relationship between the creditworthiness and score drivers and produce interpretable score models.

For synthetic sampling, we use the SMOTE function implemented in the DMwR R-library, which allows synthetical sampling for data with both continuous and categorical score drivers. Our synthetic data analysis is based on the performance data as shown in Table 7.7. First, we extract only the five score drivers and the repayment risk indicator used in the preliminary credit model shown in Fig. 7.4. Then, we call the SMOTE function in the DMwR R-library to generate two model data sets with the mixture of synthetic and original data. The following R code loads the DMwR library and generates the first model data set.

```
> library(DMwR)
> SData1= SMOTE(RR ~ ., SData, perc.over = 100, perc.under=200)
> table(SData1$RR)

   0    1
2760 2760
```

SData is the extracted data set. The SMOTE function requires the model formula with the repayment risk indicator *RR* as the response variable and the rest variables in the data set (represented by the ".") as risk drivers. The *perc.over* parameter specifies extra synthetic data created for the minority group (in percentage of original minority group data), and the *perc.under* parameter specifies how much data generated in the majority group with respect to the new data generated in the minority group (also in percentage). So, *perc.over=100* and *perc.under=200* require doubling the minority group data by generating one time of synthetic minority data joining the original minority data and also sampling majority data with size of two times of the new synthetic minority data; thus both groups have the same sample size as twice of the original minority group size. The majority group sampling will be with replacement if only the required majority group sample size exceeds the original majority group data size.

To test the impact of SMOTE, we run the binary logit model on the newly created *SData1*. Compared to the model estimates based on the original data set as shown in Fig. 7.4, the model estimates in Fig. 7.9 change dramatically. First of all, the sign of

```
> glmmodel_SData1<-glm(formula=frm_df, family=binomial(link='logit'), data=SData1)
> summary(glmmodel_SData1)

Call:
glm(formula = frm_df, family = binomial(link = "logit"), data = SData1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1468  -1.1828   0.2104   1.0322   5.5191

Coefficients:
              Estimate Std. Error z value            Pr(>|z|)
(Intercept)  2.0871030  0.2097703   9.949 < 0.0000000000000002 ***
PCS         -0.0025373  0.0003283  -7.729   0.0000000000000109 ***
Age         -0.0020884  0.0003280  -6.368   0.0000000001914920 ***
IndRisk      0.2019926  0.0366594   5.510   0.0000000358863025 ***
CorpRisk     0.1090874  0.0386811   2.820               0.0048 **
CashMargin  -0.3822875  0.0236611 -16.157 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7652.3  on 5519  degrees of freedom
Residual deviance: 6761.1  on 5514  degrees of freedom
AIC: 6773.1

Number of Fisher Scoring iterations: 7
```

Fig. 7.9 Model estimates based on SMOTE data set (1)

```
> glmmodel_SData2<-glm(formula=frm_df, family=binomial(link='logit'), data=SData2)
> summary(glmmodel_SData2)

Call:
glm(formula = frm_df, family = binomial(link = "logit"), data = SData2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2716  -1.2453   0.7997   0.9074   5.5620

Coefficients:
              Estimate Std. Error z value            Pr(>|z|)
(Intercept)  2.1641056  0.2017845  10.725 < 0.0000000000000002 ***
PCS         -0.0023292  0.0003106  -7.499   0.0000000000000642 ***
Age         -0.0026620  0.0003083  -8.634 < 0.0000000000000002 ***
IndRisk      0.1544956  0.0316485   4.882   0.0000010522023667 ***
CorpRisk     0.0601800  0.0323857   1.858               0.0631 .
CashMargin  -0.3954563  0.0219080 -18.051 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 9287.6  on 6899  degrees of freedom
Residual deviance: 8234.8  on 6894  degrees of freedom
AIC: 8246.8

Number of Fisher Scoring iterations: 7
```

Fig. 7.10 Model estimates based on SMOTE data set (2)

the CorpRisk score driver has changed from negative to positive. We know that the risk level for CorpRisk decreases with the increasing of CorpRisk values, so the positive parameter for CorpRisk doesn't make sense. In addition, except IndRisk, parameter estimates for all other score drivers have significant relative changes.

We tried to increase the synthetic sample size for the minority group of the original data and generate the second model data set *SData2*.

```
> SData2= SMOTE(RR ~ ., SData, perc.over = 200, perc.under=100)
> table(SData2$RR)

   0    1
2760 4140
```

SData2 doubles the synthetic sample size for the minority group, thus totally including three times of the original minority group sample while maintaining the same size for majority group as *SData1*. We run the same binary logit model on *SData2* and present parameter estimates in Fig. 7.10.

Increasing the synthetic sample size for the minority group does correct the parameter estimates. The parameter estimate for CorpRisk still has the wrong sign, and parameter estimates for other score drivers also have small relative changes.

Looking into the two synthetic data sets *SData1* and *SData2* and compare them with the original data set *SData*, we found that the synthetic data sets change the data structure for the score drivers, especially for categorical score drivers. This is due to the fact that SMOTE randomly selects a level for the categorical score driver between the original minority data point and its nearest neighbor when generating the synthetic data point, which may destroy the existing relationship between this score driver and the creditworthiness. This impact could be significant if the existing relationship is weak and more nearest neighbors have different level values from that of the original minority data point. The impact will be less significant if the relationship between the score driver and creditworthiness is strong, which means that the minority data point has close or similar values in the score driver with its nearest neighbors and the randomly selected values for the synthetic data point also has close or similar values in the score driver; thus the data structure in the score driver with the generated synthetic data can be maintained.

With our small business lending performance data set, CorpRisk is a categorical score driver and has a relatively weak relationship with the repayment risk indicator. That is the fundamental reason the synthetic data changes the sign of its parameter estimate and generates erroneous results.

As we know sample size for credit scoring could be big and proper sampling is necessary for computational efficiency. In the following, we demonstrate how stratified sampling can be used for this purpose with less risk to distort the relationship between score drivers and creditworthiness.

For the small business lending data with performance, the sizes of the minority and majority groups are not hugely different as shown by the following code:

```
> table(SData$RR)

   0    1
1834 1380
```

For stratified sampling, first we calculate the weight for each data points in the two groups. The weight for the minority group is 1, and the weight for the majority group is *weight=1834/1380= 1.328986.* We randomly select 1380 data points from the majority group (without replacement) to match the size of the minority group and then combine with the minority group to form the stratified sample data *SData3*:

```
> table(SData3$RR)

   0    1
1380 1380
```

Then we run the weighted binary logit model and present parameter estimates in Fig. 7.11.

We see that the parameter estimates with the stratified sampling have correct signs as the original model parameter estimates in Fig. 7.4 and changes with these parameters are relatively small. This confirms that the stratified sampling doesn't change the data structure with score drivers too much to change the relationship between score drivers and creditworthiness and is a safe sampling method for score models.

```
> glmmodel_SData3w<-glm(formula=frm_df, family=binomial(link='logit'),
weights=weight, data=SData3)
> summary(glmmodel_SData3w)

Call:
glm(formula = frm_df, family = binomial(link = "logit"), data = SData3,
    weights = weight)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4254  -1.2144   0.2274   1.1653   4.7405

Coefficients:
              Estimate Std. Error z value              Pr(>|z|)
(Intercept)  1.7778533  0.2608569   6.815    0.000000000009398 ***
PCS         -0.0029711  0.0004042  -7.351    0.000000000000197 ***
Age         -0.0010754  0.0004076  -2.639              0.00833 **
IndRisk      0.2323930  0.0454457   5.114    0.000000316002722 ***
CorpRisk    -0.0990749  0.0494958  -2.002              0.04532 *
CashMargin  -0.2841491  0.0264529 -10.742 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4391.2  on 2759  degrees of freedom
Residual deviance: 3957.9  on 2754  degrees of freedom
AIC: 3502.6

Number of Fisher Scoring iterations: 7
```

Fig. 7.11 Model estimates based on stratified sampling data set (3)

Data sampling has risk as part of the model input risk and requires careful analysis before its adoption. Our example shows that the classical stratified sampling method is safer in handling big data for score models. While this observation is generally true, there are cases that direct stratified sampling may also lead to model input risk and further investigation is required. Once the risk is understood, model sponsor and developer can make their decisions on whether such risk can be tolerated or select alternative approaches. The message is that adoption of some techniques successfully applied in other applications without careful investigation and treatment may result in critical model risk.

Model assumption risk, as pointed out in previous chapters, has been a topic not being fully addressed for most models due to the difficulty in identifying critical model assumptions and how to assess and measure the risk of these assumptions. Score models are used to rate the creditworthiness of potential customers, and thus the models could have various assumptions on the relationship between score drivers and creditworthiness.

The binary logit model we used for the score model assumes that the log odds of the risk event measured by repayment risk (or other risk events) are a linear function of the score drivers, which is defined as the creditworthiness in credit scoring. The goodness-of-fit statistics in Sect. 3.1.1.1 are the common measurements of how this assumption is satisfied based on the observed data. So, such statistics must be provided by default as a part of the model validation.

Specific assumptions related to model choice and implementation must be independently and continuously assessed. For example, most score models and scorecards assume the monotone relationship between odds and score drivers, which may not be supported by the observed data and requires the business decision on adoption of alternative models to meet this assumption. In the next section, we will introduce an alternative score model based on WoE, which usually can better fulfill this model assumption.

7.6.3 Theory and Design

In Sects. 4.5.3 and 5.5.3, we discussed some general model risks related to theory and design. In Sect. 6.6.3, for stress test, we present a summary of those risks. Score models have more flexibility in adopting broad model theory and design; thus those risks generally apply. In addition, we add some more specific risks to score models.

- Models with no or insufficient theoretical support
- Models with mathematic or statistical errors
- Models without considering compliance with fair lending laws
- Hard rejection due to errors in theory and design
- Difficulty in mitigation due to model theory and design errors

To avoid such risks, we suggest the following controls in model theory and design:

- A periodic compliance review with regulatory requirements on fair lending
- A comprehensive review of theory and methodology in the related modeling area
- Expertise of the specific modeling areas and practices and sufficient training in model theories
- Comparison with alternative theories and approaches

First of all, score models and scorecards are required to comply with the fair lending laws. Although there is no standard modeling framework for this regulatory requirement, the PCI concept and its implementation in the automated variable selection procedure AEVS_CS are trying to alleviate this regulatory liability in the model design stage instead of the post model stage, in which model risk mitigation may not be easy any more.

Score models as tools of risk rating took off with the predictive statistics, which have been built on the modern statistical theory. Although some preliminary score models (e.g., the linear probability model) have fundamental theoretical shortcomings for risk rating, over the time such models have been moved out of the candidate pool, and more theoretically sound models like Linear Discriminant Analysis (LDA), Linear Program (LP), and generalized linear models (GLM) as the most popular predictive statistical models dominate. These models have been well studied, and corresponding theories behind these models have become mature, for example, the credit models introduced in Chap. 3. More efforts are put on how to correctly implement these models in specific areas, which include model input data creation, variable selection, and parameter calibration as we presented in Sect. 7.4.

Identifying and assessing model errors in theory and design require understanding in depth of the specific modeling areas, for which a comprehensive review of theory and methodology in the related modeling area is a good start as we carried out in Sect. 7.1.2. In general, one can refer to Chap. 3 for most popularly used models in credit modeling, including credit scoring. For model validation purposes, the modeling background provided in Chap. 3 is a foundation. On top of this foundation, simulation testing is a common practice to validate model theory and design, though it is considered conservative due to the strong conditions required in simulations.

A more appealing approach for validating model theory and design is comparison with alternative theories and approaches, since such comparison directly shows the advantages and disadvantages of proposed and alternative models. Further, the comparison may demonstrate if the proposed modes work as intended, are appropriate for the intended business purpose, and are conceptually sound and mathematically and statistically correct. We have introduced this approach for credit model applications in the previous chapters. Here, again using the small business lending data, we show the comparison of WoE-based score models with general binned score models.

The optimal binning based on WoE described in Sect. 2.6.3 not only provides the optimal bins but also the WoE transformation. Tables 7.11 and 7.12 show the

```
Call:
glm(formula = frm_WOE, family = binomial(link = "logit"), data = MFITData_WOE)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.1421  -1.0699   0.2732   1.0391   1.5616

Coefficients:
               Estimate Std. Error z value             Pr(>|z|)
(Intercept)     0.31037    0.03275   9.478 < 0.0000000000000002 ***
PCS_WOE        -1.11606    0.08586 -12.999 < 0.0000000000000002 ***
Age_WOE        -1.58594    0.09306 -17.042 < 0.0000000000000002 ***
IndRisk_WOE    -0.97465    0.15083  -6.462       0.000000000103 ***
CorpRisk_WOE   -1.05049    0.21212  -4.952       0.000000733161 ***
CashMargin_WOE -1.30291    0.06147 -21.195 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 7305.7  on 5355  degrees of freedom
Residual deviance: 5892.5  on 5350  degrees of freedom
AIC: 5904.5

Number of Fisher Scoring iterations: 5
```

Fig. 7.12 Model estimates on transformed WoE score drivers transformed WoE values for PCS corresponding to different binning options. Table 7.13 shows the transformed WoE values for CorpRisk with the final adopted binning. The WoE transformation for the score drivers has some advantages compared to the original score drivers as discussed in Sect. 2.6.3. Besides those advantages, we also notice that the transformed WoE values are monotone with respect to the values or levels of score drivers, especially for CorpRisk. As we show in Fig. 7.6, the odds estimated from the general binned score model, which directly use the binned score drivers for CorpRisk, are not monotone with respect to the risk levels of CorpRisk. Using the WoE transformation will guarantee the monotone property with CorpRisk in scorecards. To demonstrate this impact, we first create all transformed WoE score drivers and add them into the model data. The following code shows we add the transformed PCS and Age into the model data, and the rest three score drivers are treated similarly.

```
PCS_WOE<-read.csv("PCS_WOE.csv")
PCS_WOE$PCS_Bin<-as.factor(PCS_WOE$PCS_Bin)
MFITData_WOE<-merge.data.frame(MFITData_Bin, PCS_WOE, by="PCS_Bin")

Age_WOE<-read.csv("Age_WOE.csv")
Age_WOE$Age_Bin<-as.factor(Age_WOE$Age_Bin)
MFITData_WOE<-merge.data.frame(MFITData_WOE, Age_WOE, by="Age_Bin")
```

Then we fit the binary logit model on the transformed data:

```
> frm_WOE <- as.formula(paste("Good ~", "PCS_WOE", "+", "Age_WOE", "+",
"IndRisk_WOE",  "+", "CorpRisk_WOE", "+", "CashMargin_WOE"))

> glmmodel_WOE<-glm(formula=frm_WOE, family=binomial(link='logit'),
data=MFITData_WOE)
> summary(glmmodel_WOE)
```

Figure 7.12 shows the model fit and parameter estimates for all transformed score drivers. Compared to the general binned score model shown in Fig. 7.6, the score model using the transformed WoE score drivers fits the data better (with respect to AIC) even with much smaller number of parameters. Compared to the unbinned model shown in Fig. 7.5, this model fits the data much better (with respect to AIC) with the same number of parameters. All parameter estimates in the model are highly significant.

To further show the impact on the odds and scores, we create a scorecard from this model based on the same score scaling used on the general binned score model as shown in Fig. 7.7. The following code extracts the odds from the binary logit model fit and scales these odds to scores.

Table 7.14 Score comparison between binned model and WoE model

Appid	Score	Score_WOE	Score_diff
Length: 5356	Min.: 472.0	Min.: 479.0	Min.: −12.00000
Class: character	1st Qu.: 613.0	1st Qu.: 612.0	1st Qu.: −4.00000
Mode: character	Median: 631.0	Median: 630.0	Median: −1.00000
NA	Mean: 644.4	Mean: 644.3	Mean: 0.07076
NA	3rd Qu.: 657.0	3rd Qu.: 655.0	3rd Qu.: 4.00000
NA	Max.: 851.0	Max.: 858.0	Max.: 17.00000

Table 7.15 AUC and AIC for the three models

	Models		
Criterion	Unbinned	Binned	WOE
AUC	0.7213	0.7664	0.7654
AIC	6614.3	5909.7	5904.5

```
Odds<-data.frame(exp(predict.glm(glmmodel_WOE)))

F1<- 30/log(2)
F0<- 770 - F1*log(30)

Score_WOE<-data.frame(round(F0+F1*log(Odds)))
```

The scores from the WoE-based model are compared to the scores obtained from the general binned model. Table 7.14 presents a summary of these two sets of scores and their differences. The differences between these two sets of scores are relatively small, with the maximum difference of 17. The averages of the two sets of scores are almost the same, and the medians have 1-point difference.

Given the relatively small difference in scores and the monotone property with transformed WoE score drivers, the WoE transformed model is preferred.

The better model fit on the data using binned score drivers or WoE score drivers derived from the bins than the unbinned score drivers is due to the fact that the optimal bins based on WoE have already used information from the response variable – the repayment risk indicator; thus the optimal bins and the derived WEO based on these bins have higher correlations compared to the original score drivers. This can be further confirmed by the AUC values listed in Table 7.15.

As we know better model fit on the training data doesn't mean the model fits the current data better, unless the current data and the training data are from the same population. To check the current data do not deviate much from the population represented by the model training data, the Population Stability Index (PSI) based on the Kullback divergence statistics of data frequency is a popular tool. It is commonly used on individual score driver X:

Table 7.16 PSI for performance and non-performance populations on PCS

Feature	Bins	NonPerf	Perf	NonPerf_pct	Perf_pct	PSI_i	PSI
PCS	01.(-Inf,514]	216	324	10.10%	10.10%	0	0.001
PCS	02.(514,549]	245	328	11.40%	10.20%	0.001	0.001
PCS	03.(549,587]	217	312	10.10%	9.70%	0	0.001
PCS	04.(587,608]	202	324	9.40%	10.10%	0	0.001
PCS	05.(608,632]	210	325	9.80%	10.10%	0	0.001
PCS	06.(632,654]	199	313	9.30%	9.70%	0	0.001
PCS	07.(654,677]	211	329	9.90%	10.20%	0	0.001
PCS	08.(677,702]	204	321	9.50%	10%	0	0.001
PCS	09.(702,738]	220	318	10.30%	9.90%	0	0.001
PCS	10.(738, Inf]	218	320	10.20%	10%	0	0.001

$$\mathrm{PSI}(X)=\sum\nolimits_{j=1}^{J}\left(\frac{f_j^T}{\sum f_j^T}-\frac{f_j^C}{\sum f_j^C}\right)\left[\log\left(\frac{f_j^T}{\sum f_j^T}\right)-\log\left(\frac{f_j^C}{\sum f_j^C}\right)\right] \tag{7.15}$$

where f_j^T and f_j^C are the data frequency observed in the jth bin for the score driver X, $j = 1, \ldots, J$ with the training data and current data, respectively. The number of bins J is commonly taken from 5 to 20 based on the data sizes.

A traffic-light approach is generally used for the PSI benchmarks: green, for values below 0.10 meaning little or no difference; yellow, from 0.10 to 0.25, some change, but not serious; and red, 0.25 upward, the change is sufficient to require some attention.

Violating these standard thresholds does not mean that a predictor should not be used or the model is invalid, only that some investigation may be required to understand if there is a strategic change with the score driver. If such a strategic change is confirmed, then model sponsor and developer should make a decision on whether model training data should be updated. If values over 1.00 are encountered, however, the situation is serious – either resulting from massive changes to the population or processes or a mistake in the exclusion criteria.

We use the *get_psi* function in the R-package "creditmodel" to calculate the PSI for features between the performance and non-performance populations. Table 7.16 shows the calculated PSI for the PCS score driver. Note that the last column in the table is the sum of all PSIs for the ten bins. The PSI of 0.001 for PCS between the two populations indicates PCS is stable between these two populations. We also carried out the PSI test for all other four score drivers used in the model, and they are all stable between these two populations.

```
> library(creditmodel)
> get_psi(dat=Perf_pop, dat_test = NonPerf_pop, x="PCS")
```

For credit scoring, PSI is also often applied to score model results – population scores. PSI can be calculated based on populations classified by time, and see if the population scores are stable over the time. When PSI indicates that population scores are not stable anymore, then an investigation on score model update should be carried out. We present more details on PSI on scores in the next section on model output analysis.

7.6.4 Implementation and Output Analysis

As for other credit models, a full modeling cycle for score model includes model development and redevelopment (Dev), user acceptance testing (UAT), and the production delivery (PROD). Usually, these processes are carried out in different computing platforms and environments due to information security and access control.

In recent years the concepts of model continuous integration and continuous deployment or delivery (CI/CD) have become more and more popular, especially for credit scoring due to its relatively simple modeling process with only one type of model. However, the CI/CD model implementation brings both efficiency in model production and new challenges for model implementation validations.

Output analysis on credit score models more focuses on the scores and scorecards. Using the small business lending data, we carry out the PSI test on scores from the performance and non-performance populations.

CI/CD Design

In recent years, to enhance the efficiency of model production process, the concept of continuous integration, deployment, and delivery (CI/CD) has become popular, and CI/CD implementation has become a trend in new software platforms, especially in the cloud computing platforms. From the model development side, this is an automation of the coding, building, packaging, and delivery process and greatly shortens the model production timelines. However, for the validation and auditing sides, there could be some transitions with less transparency.

To overcome the transparency issue with CI/CD, one option is using the layered model development processes as we described in the previous chapters with Dev, UAT, and PROD for the initial model release and only use CI/CD for model redevelopment and production update. The other option is adding intermediate testing within CI/CD as a UAT layer. Either way, CI/CD should not become the reason for a less transparent modeling process, especially for large complex models.

Model Output Analysis

A basic requirement for output analysis is the ability to replicate the model outputs on appropriate platforms, especially in the production platform. For outputs with random number generation, a random seed has to be set for result replication. Output replication also presents validator a chance to review the model development and implementation logic, design, and coding. It also helps to check the model documentation consistency with model development and implementation. A full model

Table 7.17 PSI for performance and non-performance populations on scores

Feature	Bins	NonPerf	Perf	NonPerf_pct	Perf_pct	PSI_i	PSI
Score	01.(-Inf,591]	222	324	10.40%	10.10%	0	0.009
Score	02.(591,607]	244	321	11.40%	10%	0.002	0.009
Score	03.(607,616]	198	330	9.20%	10.30%	0.001	0.009
Score	04.(616,623]	191	309	8.90%	9.60%	0.001	0.009
Score	05.(623,632]	274	392	12.80%	12.20%	0	0.009
Score	06.(632,638]	142	253	6.60%	7.90%	0.002	0.009
Score	07.(638,650]	252	348	11.80%	10.80%	0.001	0.009
Score	08.(650,672]	175	299	8.20%	9.30%	0.001	0.009
Score	09.(672,758]	212	315	9.90%	9.80%	0	0.009
Score	10.(758, Inf]	232	323	10.80%	10%	0.001	0.009

output results replication should be done periodically to check gaps which could be created by ongoing model updates.

PSI test is a common tool to check if the models developed on the training data population continue to work on testing or new data populations. For credit scoring, this is especially important because score models and scorecards developed overtime require frequent checking on its stability on the new applicants. Scores from training data populations failing the PSI test on the new population could indicate population shift, which may be confirmed with PSI tests on score drivers as we show in Sect. 7.6.3. If that can be confirmed, the model sponsor and model user need to make a decision if the population shift is expected and the current score model and scorecards still work for the shifted population. If that can't be confirmed, then a decision should be made to update the score model and scorecards.

We run the PSI test on the scores produced for the performance and non-performance populations by the binned score model (Fig. 7.6) and corresponding scorecards (Fig. 7.7). Since the model and scorecards are built on the combined population, so we expected the PSI test on these two sets of scores easily passes. Table 7.17 presents the PSI test results, and we see the total PSI based on ten bins is 0.009, which is much less than the lower threshold of 0.1 for "green" light.

Sensitivity analysis is used to assess how sensitive a model's outputs are to the change of model inputs, which could be data and assumptions. Models should have proper sensitivity to corresponding inputs. In general, score models should have proper sensitivity to score drivers and other inputs as designed. The sensitivity of score models to score drivers is represented by the magnitude of score changes along the bins. Low sensitivity to a score driver indicates flat scores among bins of this score driver and low information value provided by this score driver, and vice versa; high sensitivity to a score driver indicates steep scores among bins of this score driver and high information value. Either too high or too low sensitivity would result in further investigation and model adjustments or overlay. Under such cases, the initial step should be checking if models have fundamental issues or the modeling data are abnormal. Such issues could lead to model rejection.

Models could be sensitive to model assumptions. The assessment of model assumptions should be a part of the model concept soundness. Critical model assumptions should have been included in model limitations. Model outputs should be assessed for some minor deviations from critical model assumptions. For example, the monotonicity of scores on score drivers is the popular assumption for monotone scorecards, which are preferred and lead to the choice the WoE based score models given monotone WoE score drivers. The minor violation of this assumption shown by very few non-monotone granular bins in the binned score models can be tolerated. However, if a large portion of bins do not follow the monotone score trend, especially when scores fluctuate up and down among bins, then the monotonicity assumption on scorecards with the score driver may be risky and alternative score driver may be explored if a monotone scorecard is preferred.

7.6.5 Performance Monitoring

Model performance monitoring is essential for score models and scorecards due to the requirement of continuous assessments of their risk rating capability. Different from credit models related to loss projection, the performance of score models is more focused on their ability to discriminate the population based on creditworthiness commonly measured as the log odds of repayment risk. The monitoring of the score model performance is not only on the total creditworthiness estimate but also the contribution of individual score drivers. As predictive model, for score models, ongoing performance assessment (OPA), back-testing, and benchmarking are also common practices for model performance assessments.

Ongoing performance assessment (OPA) monitors whether production models continue to perform with the time advancing. With the changes in markets, products, activities, clients, or business practices, production models not updated promptly could deviate from the business trend and result in poor performance as measured by both statistical and business criteria. There is also the possibility that model assumptions could be broken and new model limitations are needed. Models on which business decisions depend requires close performance monitoring in a timely manner; otherwise flawed and costly business decisions could be made before deterioration in model performance becomes apparent.

Fundamentally, score models project the repayment risk event observed in fixed time window based on the score drivers observed at the application time. Over the time, new repayment risk events become realized in the performance population, and data are often observed by bins of the scorecards quarterly or monthly. So, the predicted odds from the score model and the realized odds observed for each bin of the scorecard can be obtained and compared for model performance. Large deviation of the projections from the actuals indicates the production model is not performing at least with the current observation, and investigation is required and potential model enhancements may be necessary. There is also the requirement that when new data and information inputs are added, the production model should be competitive in the model selection process as required by the model selection criteria. So, both the performance measured by those statistical criteria in the model selection process and the model ranking in the model selection process should not deteriorate significantly.

In addition to the performance of score models, population stability should also be checked as we have discussed in Sect. 7.6.4, since the change of population will certainly change the model performance on the new population.

Back-testing involves the comparison of actual outcomes with model projections during a sample time period not used in model development and at an observation frequency that matches the projection horizon or performance window of the model. The comparison is generally done using expected ranges or statistical confidence intervals around the model projections. When outcomes fall outside those intervals, one should analyze the discrepancies and investigate the causes that are significant in terms of magnitude or frequency. The objective of the analysis is to determine whether differences stem from the omission of material factors from the model, whether they arise from errors with regard to other aspects of model specification such as interaction terms or assumptions of linearity or monotonicity, or whether they are purely random and thus consistent with acceptable model performance. Analysis of in-sample fit and of model performance in holdout samples (data set aside and not used to estimate the original model) are important parts of model development but are not substitutes for back-testing.

One can extend the one-period back-test to a more comprehensive *walk-through test.* In a walk-though test, starting from a specified historical timeline with a specified forecast period (e.g., quarterly or yearly), the target model being tested (usually the production model) is refitted on the data before the selected historical timeline, and its projection on the next period is compared to that period's actual observations, and then these actual observations are added to the historical data to refit the target model, and its projection for the next period is used to compare with the actual observation in the corresponding period. Continue this process to measure the target model's projection performance while walk through all these periods to the most recent period. The walk-through test continuously measures the target model's performance over several periods to see its pertinent strength in performance instead of only the most recent period. It is also often used to detect whether the model catches the trend of the underlying changes.

Benchmarking is the comparison of a given model's inputs and outputs to estimates from alternative internal or external data or models. It can be incorporated in model development as well as in ongoing monitoring. For score models in credit scoring, examples of benchmarks include models from different methodologies, vendor firms or industry consortia, and data from credit bureaus. Whatever the source, benchmark models should be rigorous, and benchmark data should be accurate and complete to ensure a reasonable comparison.

Discrepancies between the model output and benchmarks should trigger investigation into the sources and degree of the differences and examination of whether they are within an expected or appropriate range given the nature of the comparison. The results of the analysis may suggest revisions to the model. However, differences do not necessarily indicate that the model is in error. The benchmark itself is an alternative prediction, and the differences may be due to the different data or methods used. If the model and the benchmark match well, that is evidence in favor of the model, but it should be interpreted with caution so we do not get a false degree of comfort.

As an example, if the score model based on WoE (Fig. 7.12) is taken as the production model, then the score model based on all bins (Fig. 7.6) can be the benchmark model. In this case, the benchmark model works as a check on the score monotonicity with score drivers.

7.6.6 Model Governance

On top of the previous components of model risk management is the model governance, which sets an effective framework with defined roles and responsibilities for clear communication of model limitations and assumptions, as well as the authority to restrict model usage.

The model risk management framework as shown in Fig. 4.18 is set up by the model governance through policies and procedures. The common practice is that an institute has an overall model risk policy, which covers all aspects of model risk management, including model and model risk definitions; assessment of model risk; acceptable practices for model development, implementation, and use; appropriate model validation activities; and governance and controls over the model risk management process. Then, within different legal entities, line of businesses, or modeling areas, there may be different model risk management policies and procedures. These policies and procedures cover specific model risk management activities, for example, specific policies for certain legal entities or specific procedures for model development, validation, and monitoring and reporting for a specified modeling area. Procedures usually provide more detailed guidance on the required activities.

Although score models are subject to less regulatory frameworks, the general model governance framework described above largely from the SR-11 model risk management guidance should be the base for score model risk governance. Business-specific model risk policies and procedures may be added to score models due to the periodic new regulatory requirements, business risk appetite, new models adopted, extended model use, and critical data risks.

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716–723.
Anderson, A. R. (2022). *Credit intelligence & modelling: Many paths through the forest of credit rating and scoring*. Oxford University Press.
BCBS. (2013). Fundamental review of the trading book: A revised market risk framework.
BIS. (2006). *Basel II International Convergence of Capital Measurement and Capital Standards: A Revised Framework*. BIS, Basel.
BIS. (2014). Revised Basel III leverage ratio framework and disclosure requirements, Bank for International Settlements, 2014b, Basel.
BOE. (2015). *Stress testing the UK banking system: Key elements of the 2015 stress test*. Bank of England Publications.
Borio, C., Drehmann, M., & Tsatsaronis, K. (2012). *Stress-testing macro stress testing: Does it live up to expectations?* BIS Working Paper 369, Bank for International Settlements, Basel, Switzerland.
Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.
Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Wadsworth.
C.F.R. (2019). Appendix A to Part 252 - policy statement on the scenario design framework for stress testing.
Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357.
Chen, C. (1993). *Mean loss of prediction and its asymptotic unbiased estimators*, MS Thesis, Institute of Systems Science, Academia, China.
Chen, C., & Zhang, Y. G. (2000). Variable selection of structural models. *Proceedings of World Multiconference on Systemics, Cybernetics, and Informatics, 8*, 769–773.
Chen, C., Chock, D., & Winkler, S. L. (1999). A simulation study of confounding in generalized linear models for air pollution epidemiology. *Environmental Health Perspectives, 107*, 217–222.
Consumer Financial Protection Bureau. (2013). CFPB consumer laws and regulations: Equal Credit Opportunity Act (ECOA).
Consumer Financial Protection Bureau. (2018). A summary of your rights under the Fair Credit Reporting Act.
David, J. H. (2020). *Dark data – Why we don't know maters*. Princeton University Press.
Deloitte. (2018). Capital and the allowance for credit losses - An opportunity exists to level the playing field among international banks.
Durand, D. (1941). Risk elements in consumer instalment financing. In *Studies in consumer instalment financing*. NBER.
EBA. (2020). *2020 EU-wide stress test, methodological note*. European Banking Authority.

C. Chen, *Practical Credit Risk and Capital Modeling, and Validation*, Management for Professionals, https://doi.org/10.1007/978-3-031-52542-1

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics, 7*, 1–26.
FASB ASU2016-13. (2016). Measurement of credit losses on financial instruments, an amendment of the FASB accounting standards codification, FASB ASU2016-13.
Federal Register. (2007). Final rule: Risk-based capital standards: Advanced capital adequacy framework — Basel II - Vol. 72, No. 235.
Federal Register. (2020). Regulatory capital rule: Revised transition of the current expected credit losses methodology for allowances, Vol. 85, No. 190.
Ferrari, S. L. P., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics, 31*(7), 799–815.
FHFA. (2019). PART 1254 – validation and approval of credit score models, final rule by the Federal Housing Finance Agency.
FRB. (2020a). *Comprehensive capital analysis and review 2020 summary instructions*. Board of Governors of the Federal Reserve System.
FRB. (2020b). *Dodd-Frank Act stress test 2020: Supervisory stress test methodology*. Board of Governors of the Federal Reserve System.
Geisser, S. (1974). A predictive approach to the random effect model. *Biometrika, 61*, 101–107.
Geisser, S. (1975). The predictive sample reuse method with applications. *JASA, 70*, 320–328.
Gordy, M. (2003). A risk-factor foundation for risk-based capital rules. *Journal of Financial Intermediation, 12*, 199–232.
Gouriéroux, C., Monfort, A., & Trognon, A. (1984). Pseudo maximum likelihood methods: Applications to Poisson models. *Econometrica, 52*(3), 701–720.
Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
Horst, P. (1941). *Prediction of personal adjustment*. Social Science Research Council (Bulletin 48).
IASB. (2014). *IFRS 9 Financial Instruments. July 2014. Technical report.* International Accounting Standards Board.
Kalbfleisch, J. D., & Prentice, R. L. (1980). *The statistical analysis of failure time data*. Wiley.
Kilian, L., & Lütkepohl, H. (2017). Chapter 4. In *Structural vector autoregressive analysis*. Cambridge University Press.
Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics, 54*, 159–178.
Larson, S. C. (1931). The shrinkage of the coefficient of multiple correlation. *Journal of Education & Psychology, 22*, 25–55.
Ljung, G. M., & Box, G. E. P. (1978). On a measure of a lack of fit in time series models. *Biometrika, 65*(2), 297–303.
McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics*. Academic Press.
Meeker, W. Q., & Escobar, L. A. (1998). *Statistical methods for reliability data*. Wiley.
Merton, R. (1974). On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance, 29*(2), 449–470.
Moral, G. (2011). EAD estimates for facilities with explicit limits. Chapter 11. In E. Hayden, D. Porath, B. Engelmann, & R. Rauhmeier (Eds.), *The Basel II risk parameters: Estimation, validation, stress testing - with applications to loan risk management*. Springer-Verlag.
Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics, 11*(6), 619–632.
Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T., & Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics, 34*(4), 541–554.
Rebonato, R. (2010). *Coherent stress testing: A Bayesian approach*. Wiley.
Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.
Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3–4), 591–611.

SIFMA/BMA. (1999). Standard formulas for the analysis of mortgage-Backed securities and other related securities, Uniform Practices/Standard Formulas.

Singer, J. E., & Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics, 18*(2), 155–195.

Stephens, M. A. (1986). Tests based on EDF statistics, Ch. 4. In R. B. D'Agostino & M. A. Stephens (Eds.), *Goodness-of-fit techniques*. Marcel Dekker.

Stone, M. (1974a). Cross-Validatory Choice and Assessment of Statistical Predictions (With discussion). *Journal of the Royal Statistical Society Series B, 36*, 111–147.

Stone, M. (1974b). Cross-validation and multinomial prediction. *Biometrika, 61*, 509–515.

Stone, M. (1977a). Asymptotics for and against cross-validation. *Biometrika, 64*, 29–35.

Stone, M. (1977b). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society Series B, 39*, 44–47.

Tobin, J. (1958). Estimation of relationship for limited dependent variables. *Econometrica, 26*, 24–36.

Vasicek, O. (2002). Loan portfolio value. *Risk, 15*(2), 160–162.

White, H. (1980). A Heteroskedasticity-consistent Covariance Matrix estimator and a direct test for Heteroskedasticity. *Econometrica, 48*(4), 817–838.