



TECHNISCHE
UNIVERSITÄT
WIEN

DIPLOMARBEIT

Modellierung von Ausfallwahrscheinlichkeiten für risikoarme Wertpapierportfolios

Zur Erlangung des akademischen Grades

Diplom-Ingenieur

Im Rahmen des Masterstudiums

Finanz- und Versicherungsmathematik

Ausgeführt am

Institut für

Stochastik und Wirtschaftsmathematik

TU Wien

unter der Anleitung von

Univ.-Prof. Dipl.-Math. Dr. rer. nat. habil. Uwe Schmock

durch

Katarína Juhásová, BSc

Matrikelnummer: 1527450

Wien, am 31.12.2022

(Unterschrift Verfasserin)

(Unterschrift Betreuer)



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.



TECHNISCHE
UNIVERSITÄT
WIEN

DIPLOMA THESIS

Modelling of Probability of Default for Low Risk Securities Portfolios

Submitted for the degree of

Master of Science

Within the master's program

Financial and Actuarial Mathematics

Written at the

Research Unit of

Financial and Actuarial Mathematics

TU Wien

under the supervision of

Univ.-Prof. Dipl.-Math. Dr. rer. nat. habil. Uwe Schmock

by

Katarína Juhásová, BSc

Matriculation number: 1527450

Vienna, 31.12.2022

(Signature Author)

(Signature Supervisor)



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Kurzfassung

Das Ziel dieser Diplomarbeit ist es, Methoden zur Schätzung der Ausfallwahrscheinlichkeit in risikoarmen Wertpapierportfolios zu untersuchen, einen theoretischen Hintergrund zum Kreditrisiko und seine Bedeutung zu geben, sowie einen Überblick über die aktuellen Vorschriften zur Modellierung für diese Portfolios zu spezifizieren. Es wird näher erläutert, warum die Analyse solcher Modelle für Banken von großer Bedeutung ist. Es werden drei verschiedene Ansätze für die Modellierung der Ausfallwahrscheinlichkeit definiert. Diese Ansätze sind das kumulative Genauigkeitsprofil, die Operationscharakteristik eines Beobachters und der Bayessche Ansatz. Um diese Ansätze zu testen und die Ergebnisse zu vergleichen, werden fiktive Datensätze verwendet.

Schlagnote: Ausfallwahrscheinlichkeit, Kreditrisiko, risikoarme Wertpapierportfolios, kumulatives Genauigkeitsprofil, Operationscharakteristik eines Beobachters, Bayessche Statistik, Ausfall



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Abstract

The aim of this thesis is to study methods for estimating probability of default in low risk securities portfolios, to give a theoretical background on credit risk, its importance as well as review of the current regulations on modeling for these portfolios. It will be explained in further detail, why analysis of these kind of models is of high importance for banks. Definition of three different approaches for modeling of probability of default will be given. These approaches are cumulative accuracy profile, receiver operating characteristic and Bayesian approach. Fictitious data set will be used to test these approaches and to compare results.

Key words: probability of default, credit risk, low default portfolio, cumulative accuracy profile, receiver operating characteristic, Bayesian statistics, default



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Acknowledgement

First and foremost, I would like to thank my supervisor Prof. Uwe Schmock who guided me in writing this thesis. He provided me with invaluable advice and helped me along the way. His motivation and help contributed tremendously to the successful completion of this thesis.

Besides, I would like to thank Dr. Gerald Wunderer, my supervisor at work, for letting me choose my working hours freely in order to finish my thesis on time.

At last but not in least, I would like to thank my family, my boyfriend Daniel Dishlievski and also friends for their support. Without their support I couldn't have succeeded in completing this thesis and my studies.

Thank you.

Katarína Juhásová



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Statement of originality

I hereby declare that I have authored the present master's thesis independently and did not use any sources other than those specified. I have not yet submitted the work to any other authority in the same or comparable form. It has not been published yet.

Vienna, 31.12.2022

Katarína Juhásová



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Contents

1	Introduction	1
1.1	Background	1
1.2	Chapter outline	1
2	Theoretical background	3
2.1	What is risk?	3
2.2	Credit risk	4
2.2.1	Default	4
2.2.2	Probability of default	4
2.2.3	Exposure at default	5
2.2.4	Loss given default	5
2.2.5	Expected loss	6
2.3	Bank regulations	6
2.3.1	Basel I Accord	7
2.3.2	Market Risk Amendment	8
2.3.3	Basel II Accord	8
2.3.4	Basel III Accord	10
2.4	Low risk securities portfolios	10
2.4.1	Different approaches	11
3	Cumulative accuracy profile and receiver operating characteristic approaches	13
3.1	Cumulative accuracy profile	13
3.2	Important property of the cumulative accuracy profile curve	15
3.3	Accuracy ratio	18
3.4	Receiver operating characteristic	19
3.5	The area under the receiver operating characteristic curve	20
3.6	Cumulative accuracy profile vs. receiver operating characteristic	21
3.7	The generation of probability of default with cumulative accuracy profile approach	22
3.8	The generation of probability of default with receiver operating characteristic approach	27
4	Bayesian approach	31
4.1	Bayesian statistics	31
4.2	The generation of probability of default with Bayesian approach	32
4.2.1	Statistical models for defaults	33
4.2.2	Expert opinion	35
4.2.3	Calculation of probability of default	37

5	Statistical modelling	41
5.1	Data	41
5.2	Cumulative accuracy profile approach	41
5.3	Receiver operating characteristic approach	46
5.4	Bayesian approach	50
5.5	Discussion of results	56
6	Summary and conclusion	59
7	Appendix: SAS Codes	61
	List of Figures	79
	Bibliography	81

1 Introduction

1.1 Background

Nowadays, the predictive ability of financial institutions about the behavior of clients who are unlikely to pay is one of the most important tools. Unfortunately, there exist portfolios which have historically experienced a low number of defaults, also called low default portfolios. From its name it can be understood that a low default portfolio is a portfolio of obligors where the observed amount of defaults is low. A problem with these portfolios in general, is how they are defined. The biggest issue is, that there is no exact definition for them and it therefore may vary from one bank to another. This is also the main reason why the concern about these portfolios has been increasing over the past years. Examples of such portfolios are portfolios with an overall good quality of borrowers, therefore often called low risk securities portfolios, like sovereigns, bank portfolios, insurance companies or project finance. Credit risk or also known as default risk, is the largest risk that most banks have to deal with. Therefore, credit risk management is crucial in every bank. Banks model various risk components, to be able to understand and predict the behavior of customers. The estimation of the most known risk component, probability of default, is the topic of this master's thesis. It defines a probability of a borrower defaulting on some obligation. The main problem with estimating probability of default (PD) in low default portfolios is its underestimation and volatility. Therefore, these estimates are unreliable in a statistical sense and may result in an inadequate assessment of the true risk. In practice, exposure of low default portfolios is very high and that is the reason why the problem of correct PD estimation in these portfolios is very important for the bank. The modeling of probability of default is addressed in the second of the Basel Accords, but as already mentioned above, for low default portfolios it has to be done differently than in high default portfolios. The topic of how to estimate PD in low default portfolios has been subject of much discussion and there exist many different approaches. In this master's thesis, three different approaches (cumulative accuracy profile, receiver operating characteristic and Bayesian approach) will be described and compared.

1.2 Chapter outline

The following is the chapter outline for the thesis.

Chapter 2: In this chapter, an overview of the subject of credit risk, its components and the regulations in this area is presented in order to give context to credit risk modeling problems in general.

Chapter 3: Cumulative accuracy profile, receiver operating characteristic and their associated quantitative measures are explained. Cumulative accuracy profile approach by Burgt

and receiver operating characteristic approach by Tasche are presented in this chapter.

Chapter 4: The Bayesian approach is presented by first reviewing Bayesian statistics in a general setting. Then, the theory covering the Bayesian approach is given.

Chapter 5: The data which is used, is presented in this chapter along with results for PD calculation of the cumulative accuracy profile approach, receiver operating characteristic approach and Bayesian approach. A comparison of the resulting PD for these approaches is performed.

Chapter 6: In this chapter, the summary and conclusion about the master's thesis is given.

Chapter 7: SAS codes used for the practical part are written in this last chapter.

2 Theoretical background

In this chapter, the definition of risk will be given. It will be analyzed, what the main risks in banking are and why risk management is so important. Credit risk will be discussed in detail, because estimation of one of its components is the main focus of this thesis. Further, the regulatory risk management control framework for banks will be mentioned, which is commonly known as the “Basel Accords”. The second of the accords (“Basel II”) will be explained in further detail. Lastly, a short introduction to low risk securities portfolios will be concluded.

2.1 What is risk?

Doing business in the field of financial institutions is undoubtedly associated with a whole group of specific risks. Therefore the definition of risk is needed for further understanding. There are many different versions of risk definition and there is not a single one that can be used in every context. [20] defines risk as ‘hazard, a chance of bad consequences, loss or exposure to mischance’. In [9] also the distinction between risk and uncertainty was made. According to the author of this book, Frank Knight, risk applies to situations where the outcome of a given situation is not known, but the odds can be accurately measured. Uncertainty, on the other hand, applies to situations where all the information needed is not known in order to set accurate odds. It is also called Knightian uncertainty, named after the author of the book. In the context of finance and insurance, two definitions can be found in [11] which are ‘any event or action that may adversely affect an organization’s ability to achieve its objectives and execute its strategies’¹, or alternatively ‘the quantifiable likelihood of loss or less than expected returns’¹.

As explained in [11] and many other books on this topic, there are three main risks in the banking industry.

- market risk (definition: ‘the risk of a change in the value of a financial position due to changes in the value of the underlying components on which that position depends, such as stock and bond prices, exchange rates, commodity prices, etc.’²)
- credit risk (definition: ‘the risk of not receiving promised repayments on outstanding investments such as loans and bonds, because of the default (the definition of default will be given in Subsection 2.2.1) of the borrower’²)
- operational risk (definition: ‘the risk of losses resulting from inadequate or failed internal processes, people and systems, or from external events’²)

¹[cf. 11, p. 1]

²[cf. 11, p. 2-3]

2.2 Credit risk

Credit risk or also known as default risk, is the largest risk that most banks have to deal with. Therefore, credit risk management is crucial in every bank. It is responsible for identification, quantification, monitoring, controlling and management of credit risk at both transaction and portfolio level as written in [10, p. 175–176]. Since it is the largest one, it is also the most difficult one to measure. The main purpose of credit risk measurement is to quantify the potential loss occurred during credit transactions of a bank or financial institution. On the basis of these findings, the financial entity decides if the transaction can be executed, how high the interest rate should be, if some reserves should be created and if close monitoring is necessary. Financial entities use several methods to quantify credit risk. They model risk components described below, to better understand and predict the behavior of customers. They often use historical data or rating systems as a basis for credit risk measuring.

2.2.1 Default

The elementary term that was already used few times and will be used quite often in this master's thesis is default. Firstly, it is therefore important to understand the meaning of this word. There are many different definitions of a default event but the most used one is a delay of payment for at least three months. The following definition of a default event is given in [13] 'a default is considered to have occurred with regard to a specific obligor when either or both of the two following events have taken place:

- The bank considers that the obligor is unlikely to pay its credit obligations to the banking group in full, without resource by the bank to actions such as realizing security, if held.
- The obligor is past due more than 90 days on any material credit obligation to the banking group. Overdrafts will be considered as being past due once the customer has breached an advised limit or been advised of a limit smaller than current outstanding.³

2.2.2 Probability of default

Probably the most known risk component is the probability of default (PD) and it is crucial for credit risk estimation. Learning the meaning of PD is very important for understanding the topic of this thesis, since it is its main focus. It defines a probability of a borrower defaulting on some obligation. It takes values in the interval $[0, 1]$ where the value 0 represents zero risk of default and 1 represents default happening. There are two primary modes of using the PD measure. The first one is in the measurement of credit risk associated with an individual borrower. In this case, the probability of default is defined as the likelihood that there is a credit event (this event is indicated by a random variable D , which takes the value $D = 0$ in case there is no event or the value $D = 1$ in case there is a credit event). PD is in this case $PD = E[\mathbb{1}_{\{D=1\}}]$. The second mode is in the

³[cf. 13, p. 100]

measurement of credit risk associated with a pool of similar borrowers, which are not modelled individually. In this case, the probability of default is defined as the expected default rate DR (= the ratio between the total number of defaults and total number of borrowers) or mathematically written $PD = E[DR]$. In both modes, is therefore PD a random variable. In most cases, the term PD is used to reflect the default risk in a time period of one year. Default risk is defined on a counterparty level, not on a product level. When a counterparty defaults on one loan or obligation, it is likely to default also on its other loans. Therefore it is understandable to define it rather on this level.

2.2.3 Exposure at default

Exposure at default (EAD) is the predicted amount of loss a bank may be exposed to when a debtor defaults on a loan. EAD is a random variable since its value is never known beforehand and the actual value is always known only after default. Since it is very important for banks to be aware of the estimated value, it has to be modeled. In [25, p. 28–29] it is explained, that the amount is fixed for some products like a bond or a straight loan. Unlike for credit cards or overdraft facilities, the amount varies with the liquidity needs of the borrower. For these products the calculation of EAD is split into drawn and undrawn part. The drawn part is the one that was already taken and is therefore known for the bank. Whereas the undrawn part needs to be estimated to come up with a value of EAD. The counterparty can withdraw amount till the negotiated credit limit (for example, the monthly limit for credit cards in Austria is around 3800 € for many clients). The credit limit bounds the commitment of the bank. Other products have no explicit limit, but each additional drawing needs approval of the bank. The uncertainty on the exact amount at risk at the very moment of a future default is exposure risk. Apart from product and contract properties, one can expect that the exposure risk depends on features of the borrower and on the general state of the economy. The exposure risk is typically expressed in the currency of the product or of the bank (euro, dollar...). The EAD calculation can be easily understood with the help of an example: If the client defaults with an outstanding debt of 500,000 € and the bank or insurance is able to sell the security (for example a flat) for a net price of 400,000 € (including costs related to the repurchase), then the EAD is 100,000 €.

2.2.4 Loss given default

As described in [25, p. 26–28], loss given default (LGD) determines the loss as a fraction of the exposure in the case of default. In the case of no loss, the LGD is equal to zero. Exactly like EAD, also LGD is a random variable since its value is always known only after default. When one loses the full exposure amount, the LGD is equal to 100%. A negative LGD indicates a profit (for example due to penalty fees and interest rate). In some cases, the LGD can be above 100%, for example, due to litigation costs and almost zero recovery from the defaulted counterpart. Unlike PD, LGD is measured on a product basis and has typically values between 0 and 100%. From the EAD example above, the LGD would be 20% ($= 100,000 \text{ €} / 500,000 \text{ €}$).

2.2.5 Expected loss

With help of the last three risk components, the very important one can be calculated. The expected loss (EL) is what is used by lending institutions to help them understand losses that may incur as a result of lending to a borrower that may default. EL is not time-invariant and should always be recalculated when circumstances change. For example currently in year 2022 because of Covid-19 pandemic and war in Ukraine the inflation is really high. Therefore it would now make sense to recalculate it. Mainly because change of default behavior for some borrowers is expected, since they have to spend more money on basics and have no money left for credit repayment. EL is always calculated on transaction level and EL of portfolio is simply the sum of ELs of the individual transactions. EL is also a random variable and it is calculated in the following way: $EL = PD \cdot EAD \cdot LGD$.

2.3 Bank regulations

This subchapter is based on [1, chapter 3]. It is explained there that, banks play a very important role in the economy by offering payment services and supplying credit. Their role is crucial because any disruption to the banking system could have huge effects on businesses and people. That is the main reason why banks have to be regulated. Ensuring continuity of a bank is not only a chief concern of a bank's stakeholders (including management, its shareholders, customers, and employees) but also of bank regulators and various agencies and authorities that are in charge of managing the economy. Efficient bank regulations try to reduce risks and those make sure that individual banks operate prudently and do not fail. The main role of regulations is to institute adequate supervision of individual banks to reduce the chance that runs on individual banks will happen and escalate to systemwide bank panics, contagion, and economic crises. To avert the devastating economic effects of bank failures and to ensure a stable banking industry in well-functioning financial markets, banking and financial regulators actively aim to meet the following objectives:

- Ensure that banks are operated prudently: Regulators set rules that give banks incentives to follow strict operational standards and to avoid risky loans. Regulators also impose capital requirements on the bank, requirements that take into account the risk inherent in the bank's activities.
- Reduce systemic risk: Regulators and bank supervisors monitor banks on an individual basis through examinations, inspections, and regulatory audits. In these examinations, the bank supervisors focus on identifying non-systemic sources of risks that can increase the possibility of a bank run. Usually banks with low equity and regulatory capital levels, risky loans, internal managerial problems, weak earnings, and limited funding sources receive additional regulatory monitoring.
- Implement systemwide support mechanisms: These mechanisms reduce the impact of a possible bank run by offering deposit insurance that insures each depositor's money. Other approaches include reserve requirements and access to various liquidity support systems in the banking system. Ultimately, the country's central bank or monetary



Figure 2.1: Timeline of bank regulations, source: [1, p. 67]

authority can act as the “lender of last resort” and may step in to offer temporarily emergency liquidity support to weak but otherwise viable banks.

For a long time, banking regulations were national. This meant that governments and their regulatory agencies developed rules and banking supervisory guidelines specific to the country’s needs. As a result of this, regulatory rules were significantly different in many countries. Much of the regulatory drive originated from the Basel Committee of Banking Supervision. This committee was established by the Central Bank Governors of the Group of Ten (G10, surprisingly consisting of following eleven countries: Belgium, Canada, France, Germany, Italy, Japan, the Netherlands, Sweden, Switzerland, the United Kingdom and the United States) at the end of 1974. This group was created to consult and cooperate on economic, monetary and financial matters. G10 was working mainly of two basic principles:

- Every international bank should be subject to supervision.
- The supervision should be substantial enough to ensure compliance.

To achieve this, the committee introduced two accords and one amendment, whose timeline can be seen in Figure 2.1.

2.3.1 Basel I Accord

With this accord, the international banking regulations were started. Its main emphasis was on credit risk and appropriate risk weighting of assets. The intention was to generate a risk identification system to make it possible to compare different types of banks and the different types of assets they held. Assets of banks were divided according to credit risk into four categories:

- Cash or loans to governments in OECD (the Organization of Economic Cooperation and Development consisting of 30 developed countries with a democratic government and market economy) countries were assigned with risk weight of 0%.
- Loans to banks in OECD countries were assigned with risk weight of 20%.
- Residential mortgages (credit to buy a home) were assigned with risk weight of 50%.
- Corporate loans or consumer loans were assigned with risk weight of 100%.

Each bank should calculate its total risk weighted assets (RWA) as the sum of the absolute value of each asset multiplied by its risk weight.

Secondly, Basel I Accord created the framework for the structure of bank capital. It states that the banks should hold their capital in two tiers:

- Tier 1 (also called core capital) is primarily the bank's equity.
- Tier 2 (also called supplementary capital) includes mainly reserves and provisions as well as hybrid capital instruments and subordinated debt. Tier 2 capital can be at most 50% of total regulatory capital.

Finally, the Basel I Accord also sets a minimum capital requirement of 8% for the ratio of the regulatory capital of the bank divided its risk weighted assets (RWA). This ratio is called the capital ratio or capital standard. For example, if the RWA of the bank is 2,000 million € then the bank's minimum regulatory capital is 160 million € (= 2,000 million · 8%).

2.3.2 Market Risk Amendment

Banking changed dramatically after the introduction of Basel I Accord. Banks became broad-based providers of financial services. Trading activities also started to become more significant. This was the reason, why the Basel Committee issued the Market Risk Amendment, formally titled Amendment to the Capital Accord to Incorporate Market Risks. The main focus of this Amendment was the effect of a bank's positions in various market-traded financial assets (like foreign exchange, debt securities, equities, commodities, and derivatives). The risks coming from trading positions in bonds, equities, foreign exchange, and commodities were separated from credit risk calculations and were assigned to a new risk category, the so called market risk (defined in Subsection 2.1). Banks were allowed to use their own systems for measuring market risk and the capital required to cover this risk was based on how effective their models had been at measuring the bank's market risk (if the model was more effective, than less capital was required). Another change that came with Market Risk Amendment, was incorporation of Tier 3 which consisted of short-term subordinated debt for the sole purpose of meeting a proportion of the capital requirements for market risks.

2.3.3 Basel II Accord

However, there were few issues with Basel I Accord that were not solved with introduction of the Market Risk Amendment. For example, banks that lent to companies with a very good ratings were obliged to have exactly the same amount of regulatory capital as banks that were lending to companies with bad ratings. Another problem with Basel I was, that the benefits of credit mitigation techniques weren't recognized. With help of credit mitigation techniques, banks could reduce the credit risk that is associated with loans through the use of collateral and loan guarantees. For instance, even some not so good rated company that is applying for credit, shouldn't have problems getting it, if the company already owns a

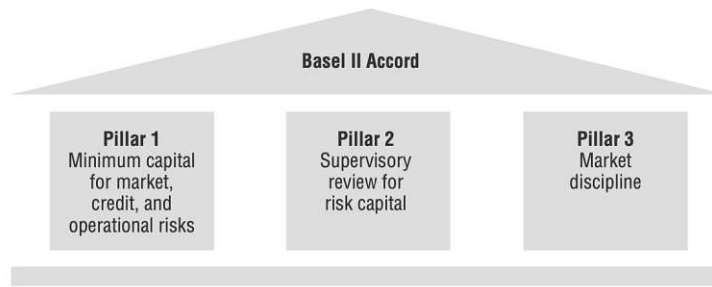


Figure 2.2: Three pillars of Basel II Accord, source: [1, p. 72]

land (this land could be used as collateral). In case that the company would default in the future, it could sell the land and repay the credit with it. One more issue of Basel I worth mentioning, is that the benefits of diversification were not recognized for credit risk reduction. Problem was, that a bank that lends to the same type of customer in the same region faces greater credit risk than a bank that lends to a diverse group of customers in the same or different regions of the world. As a result of these concerns, the committee issued a proposal for a new Basel Accord in 1999. After 5 years, a new capital framework also called Basel II was introduced. It consists of three pillars that are shown in Figure 2.2.

- Pillar 1 (main topic: minimum capital requirements): ‘It consists of the improved minimum capital requirements. These upgraded requirements should reflect the three major types of risk that a bank faces: credit risk, market risk, and operational risk. It should fix the extremism that was found Basel I. Now banks are able to choose from different alternatives of varying complexity to calculate their minimum regulatory capital requirements. It also represents the first attempt to assign a regulatory capital charge to the management of operational risk.’⁴
- Pillar 2 (main topic: supervisory review): ‘It complements and strengthens Pillar 1 by establishing a prudential supervision process. It covers all the risks in Pillar 1 and adds some other considerations:
 1. The calculation by the bank of the amount of economic capital, the amount of capital a firm will require to survive during times of distress to cover Pillar 1 risks.
 2. The creation of a governance structure within the bank to ensure internal supervision and oversight from the board of directors and senior management.
 3. The evaluation, by the banking supervisor of the bank’s own risk profile level and the processes the bank used to determine that level.’⁴
- Pillar 3 (main topic: market discipline): ‘This part outlines the effective use of market discipline as a lever to strengthen disclosure and encourage sound banking practices.

⁴[cf. 1, p. 72]

Market discipline is public disclosure of a bank's financial condition to depositors and other interested parties, allowing these to assess the condition of the bank. It relates to transparency of the bank and its activities. Disclosure, or transparency, is the degree to which a bank or any company reveals its assets, liabilities, and/or inner workings. Disclosure affords the market (other banks, depositors, and borrowers) a better picture of the bank's overall risk position and allows the bank's counterparties to price and deal appropriately.¹⁴

These three pillars are intended to support each other in an approach designed to strengthen the safety and soundness of the global financial system. The complex regulatory framework has two main objectives. Firstly, to improve how regulatory capital requirements reflect underlying risks and secondly to address the effects of financial innovation that has occurred. When a country wants to incorporate the Basel Accord into its banking regulatory and supervisory framework, it has to be done by adjusting it to its own laws and regulations. It is possible to do some amendments and adjustments to best reflect needs of each country. During development of this new Accord, the Basel Committee worked very closely with the European Union and as a result, the Basel II Capital Accord has become the basis for the EU regulatory framework and has been implemented across all EU countries.

2.3.4 Basel III Accord

This Accord is not yet mentioned in [1] on which this Subsection was based so far. Instead [14] together with [15] will be used to explain the Basel Committee's reforms to strengthen global capital and liquidity rules with the goal of promoting a more resilient banking sector. The goal of the third Accord is to improve the banking sector's ability to absorb shocks arising from financial and economic stress thus reducing the risk of spillover from the financial sector to the real economy. It was developed in response to the deficiencies in financial regulations revealed by the financial crisis of 2007–2008. These two documents were both published in 2010, with initial plan to introduce Basel III Accord between 2013–2015. It was postponed few times, with a final plan to introduce it in January 2022. This was again postponed because of Covid-19 pandemic and the current plan is, that the implementation will start in January 2023 and will be phased in over five years.

2.4 Low risk securities portfolios

The Basel Accords provide no formal definition of low risk securities portfolio also called low default portfolio (LDP). It was only mentioned in Basel Committee's newsletter [12]. From the name of this portfolio it can be understood that a low default portfolio is a portfolio of obligors where the observed number of defaults is low. These kind of portfolios are quite common in a financial world. For instance, when looking in the history, it can be seen that some portfolios have experienced low numbers of defaults. LDP are normally considered to be low risk (one example could be portfolios of exposures to sovereigns because only a very few countries defaulted in the past, another example could be banks, insurance companies or project finance, which is finance for large projects such as building highways or nuclear

reactors.). The main problem for low default portfolios is the lack of modeling data, to be exact default data, which makes the usage of models for estimation of PD, EAD or LGD not only difficult but also unreasonable. To calculate a historical average default rates and use those, is also not appropriate because they have been calculated only on a few observations. Because of this data problem, the credit risk could be underestimated or overestimated for this kind of portfolios. Underestimation would be a huge problem for banks, especially given the fact that a large portion of a bank's assets might consist of low default portfolios. Another problem could be the volatility of the observed defaults. There can be many years without observing any defaults, but that does not mean that there will be no defaults in the future (possible financial crisis or war. . .). Because of these problems that are feared not only by regulators but also by the financial industry, LDP problem is widely discussed in the literature and many different approaches are suggested by different authors. The focus of this thesis will be the estimation of probability of default in this problem setting.

2.4.1 Different approaches

1. Schuermann and Hanson [19] propose an approach to estimate PDs by using migration matrices. They suggested two different ways: cohort and duration approach. The difference is, that the cohort approach takes into account only the observations at the beginning of the period and at the end of the period, and the duration approach takes into account the observations during the whole period as well. Authors recommend the usage of the second one.
2. Most prudent estimation by Pluto and Tasche [16] is a proposition to estimate PDs by using upper confidence bounds while guaranteeing an ordering of PDs that respects the differences in credit quality as indicated by rating grades. Three ways were proposed for estimating of PDs, each time using different confidence levels: first was without correlation, second one was with correlation and the last one was a multi-period case. The choice of confidence levels is the biggest question that has to be answered in this approach, but authors suggest that using confidence levels of less than 95% appears to be intuitively appropriate.
3. Benjamin, Cathcart and Ryan [2] propose an approach with regulator having a role as well. In the paper it is written that the regulator would give a criteria for identifying a LDP and also a look-up table which gives a portfolio average PD as a function of numbers of company years in the historical sample and numbers of defaults observed in the historical samples. If the calculated weighted average PD by financial institutions is lower than the one calculated by the regulator, then it has to be adjusted upwards.
4. CAP approach by Burgt and ROC approach by Tasche will be further discussed in Chapter 3.
5. Bayesian approach by Kiefer will be further discussed in Chapter 4.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

3 Cumulative accuracy profile and receiver operating characteristic approaches

This chapter is based on [5, 18, 21, 22, 23, 24].

3.1 Cumulative accuracy profile

CAP (cumulative accuracy profile, in some literature also called power curve) is a graphical tool to evaluate the discriminatory power of a given rating model. A high discriminatory power indicates that the rating model can differentiate well between borrowers with high and low creditworthiness.

Let S be interpreted as the credit score for a solvent borrower. S takes on values on a continuous scale in some open interval $I \subset \mathbb{R}$ or on discrete scale in a finite set $I = \{1, 2, \dots, k\}$. Low values of S indicate low creditworthiness and high values indicate high creditworthiness. Let Z be interpreted as the borrower's state of solvency one observation period after the score was observed, this is usually 1 year (as described in Subsection 2.2.2). Z takes values in $\{0, 1\}$, where 0 means the survival of the borrower and 1 means default. Let N be defined for the survival event, that means $\{Z = 0\}$ and D for default event, which means $\{Z = 1\}$. The marginal distribution of the variable Z is characterized as the unconditional probability of default p which is defined as⁵

$$p = P[D] = P[Z = 1] \in [0, 1].$$

The joint distribution of (S, Z) can be specified by two conditional distributions of S given the state of Z so either the event D or N respectively. These conditional distributions are denoted by $F_D(s)$ and $F_N(s)$, they are called the hit rate and the false alarm and are defined as⁵

$$F_D(s) = P[S \leq s \mid D] = \frac{P[\{S \leq s\} \cap D]}{p}, \quad s \in I,$$

$$F_N(s) = P[S \leq s \mid N] = \frac{P[\{S \leq s\} \cap N]}{1 - p}, \quad s \in I.$$

The unconditional distribution of the credit score S , also called the alarm rate, can be defined with help of the law of total probability as⁶

$$F(s) = pF_D(s) + (1 - p)F_N(s), \quad s \in I. \quad (3.1)$$

⁵[cf. 23, p. 4]

⁶[cf. 23, p. 5]

The joint distribution of the pair (S, Z) can be specified by the unconditional distribution $P[S \leq \cdot]$ of S and combining it with the conditional probability of default $P[D | S] = 1 - P[N | S] = p_D(S)$. In general, $p_D(S)$ can be characterized by the property⁶

$$E[p_D(S)\mathbb{1}_{\{S \in A\}}] = P[D \cap \{S \in A\}], \quad (3.2)$$

for all Borel sets $A \subset \mathbb{R}$. Equation (3.2) implies the representation of $p_D(S)$ in two important cases:⁶

- S is a discrete variable. Then follows

$$P[D | S = j] = \frac{pP[S = j | D]}{pP[S = j | D] + (1 - p)P[S = j | N]}, \quad j \in I.$$

- S is a continuous variable such that there are Lebesgue densities f_N and f_D of the conditional distributions F_N and F_D . Then follows

$$P[D | S = s] = \frac{pf_D(s)}{pf_D(s) + (1 - p)f_N(s)}, \quad s \in I. \quad (3.3)$$

These basic definitions are needed for the rest of this chapter. With help of these, the definition of *CAP* can be stated.

Definition 3.1.1 (CAP). The cumulative accuracy profile function $CAP : [0, 1] \rightarrow [0, 1]$ is defined as⁷

$$CAP(u) = F_D(F^{-1}(u)), \quad (3.4)$$

where $F^{-1}(u) = \sup_{s \in \mathbb{R}} \{P[S \leq s] \leq u\}$ is the generalized inverse function of F .

Alternatively, *CAP* curve is obtained as the graph of all points $(F(s), F_D(s))$ for $s \in \mathbb{R}$, where the points are connected by straight lines (linear interpolation).

The most important examples of *CAP* are shown in Figure 3.1. In the perfect model, all defaults are observed only within borrowers with the lowest rating scores (for example in Figure 3.1 is the lowest close indicated by x). The *CAP* curve linearly increases at the beginning (for example in Figure 3.1, this is till the rating score x) and then stays at 1 for all other rating scores, since all defaults have already happened in the worst rating class. This is an ideal outcome. Other extreme is called random model. That is when the rating system has no discriminatory power and it randomly assigns ratings to borrowers. Then the cumulative percentage of defaults increases proportionally with the cumulative percentage of borrowers. The *CAP* curves of perfect and random model are used to evaluate real rating models. An average good model has a *CAP* curve that is in the middle between perfect and random *CAP* curves. The closer a model is to the perfect *CAP* curve, the better it is.

⁷[cf. 23, Theorem 3.2]

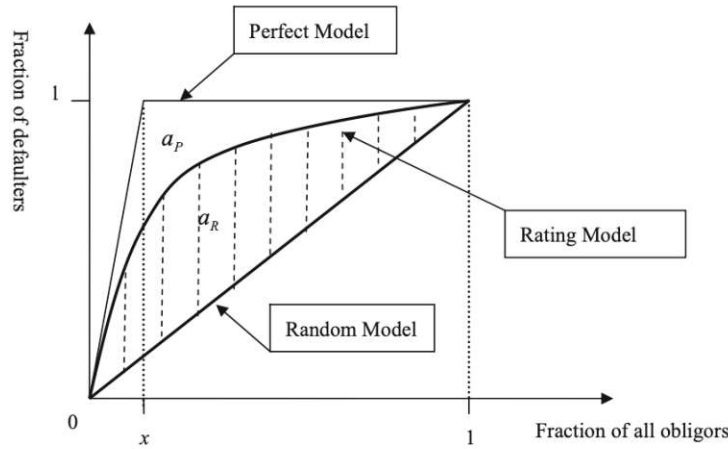


Figure 3.1: Cumulative accuracy profile, source: [5]

3.2 Important property of the cumulative accuracy profile curve

CAP is not only used for evaluation of discriminatory power of given rating model but also for PD estimation that can be used in the low default portfolios. The key property of this approach for PD estimation is the concavity. Therefore in this Subsection the concave shape of *CAP* curve is shown. Monotonicity of the rating system in the sense that better ratings should correspond to lower default probabilities is one of the most important requirements. Therefore first, the monotonicity of the probability of default function needs to be shown. Following proposition is based on [21].

Proposition 1. *The conditional probability of default is monotonically decreasing.*

Proof. It is important to note, that this condition is not always guaranteed. The aim is, to determine the conditions where the conditional probability of default is monotonically decreasing. First, an optimal test is determined. For this purpose, a debtor is randomly selected and the score is noted. Since no information about the future state is available, the score is used to infer the state variable X . Let assume that (s, x) is a realization of (S, X) . s has already been observed and x is not yet known. The question is, which value does x have, whether $x = 0$ (= no default) or $x = 1$ (= default). A set E is chosen as a set of all debtors who are in default, which means that

$$s \in E \Rightarrow x = 1,$$

$$s \notin E \Rightarrow x = 0.$$

To determine the set E , the conditional score distribution of the defaulter and non-defaulters needs to be considered. The aim is to choose E in a way that the conditional score distribution of the defaulter will be the maximum and the conditional score distribution of the non-defaulter will be the minimum. For this purpose, the hypothesis H_0 (=

the debtor is the future defaulter) or the alternative H_1 (= the debtor remains liquid) is formulated. Then these hypotheses are tested against each other. Two mistakes can occur in such test. Either a Type I error (= the hypothesis H_0 is rejected although it is true) or Type II error (= the hypothesis H_0 is maintained although the alternative H_1 is true). The probabilities of these two possible errors are considered, in order to make an optimal decision

$$P[\text{Type I error}] = P[S \notin E \mid X = 1],$$

$$P[\text{Type II error}] = P[S \in E \mid X = 0].$$

Since the Type I error is the more critical one, it will be bounded from above and called Type I error level α . To continue further, the so-called likelihood ratio needs to be defined.⁸

Definition 3.2.1 (Likelihood ratio). The following

$$L(s) = \frac{f_D(s)}{f_N(s)}, \quad s \in I, \quad (3.5)$$

is called the likelihood ratio at score s .

The main result will be based on Neymann–Pearson fundamental lemma which needs to be stated (without the proof) at this point.⁹

Theorem 3.2.1 (Neymann–Pearson fundamental lemma). *Fix a Type I error level $\alpha \in (0, 1)$. Assume that the densities f_D and f_N are positive continuous functions in some open interval $I \subset \mathbb{R}$ and it follows that for any $x > 0$ the set $U_x = \{s \in I : xf_N(s) = f_D(s)\}$ has Lebesgue measure 0. Then it follows that an S -based test ϕ for H_0 against H_1 is most powerful at level α if and only if for Lebesgue-almost $s \in I$*

$$\phi(s) = \begin{cases} 1, & \text{if } L(s) > L_\alpha \\ 0, & \text{otherwise} \end{cases}, \quad (3.6)$$

where L is the likelihood ratio defined by (3.5) and L_α is any constant such that $P[L \circ S > L_\alpha] = \alpha$.

Few more definitions are needed to understand Neymann–Pearson fundamental lemma and to be able to continue.¹⁰

Definition 3.2.2 (Randomized test). Any measurable function $\phi : I \rightarrow [0, 1]$ is called S -based test for H_0 against H_1 .

A S -based test ϕ is called test at level $\alpha \in (0, 1)$ if $E[\phi \circ S \mid H_1] \leq \alpha$.

A S -based test ϕ^* is called most powerful test at level $\alpha \in (0, 1)$ if it is a test at level α and for any S -based tests ϕ at level α it follows that $E[\phi^* \circ S \mid H_0] \geq E[\phi \circ S \mid H_0]$.

Definition 3.2.3 (α -quantile). α -quantile of the likelihood ratio is defined as

$$r_\alpha = \inf_{r \leq 0} \{P[L(s) \leq r \mid X = 1] \geq 1 - \alpha\}.$$

⁸[cf. 21, Definition 2.2]

⁹[cf. 21, Theorem 3.2]

¹⁰[cf. 21, Definition 3.1]

Due to the continuity, the α -quantile can be determined by solving the following equation

$$1 - \alpha = P[L(s) \leq r_\alpha \mid X = 1].$$

Using the Neymann–Pearson fundamental lemma, it follows that $\phi(s)$ defines the most powerful test at the α level. In this case, $\phi = 1$ means the rejection of the hypothesis H_0 . This result is applied on the above described problem.

If

$$S \notin E \Leftrightarrow \phi = 1,$$

$$P[S \notin E \mid X = 1] \leq \alpha,$$

hold, then follows

$$P[L(s) \leq r_\alpha \mid X = 0] \leq P[S \in E \mid X = 0]. \quad (3.7)$$

Using (3.7), the set E can be defined as

$$E = \{s : L(s) \leq r_\alpha\}. \quad (3.8)$$

Therefore it is shown that the set E can be used to distinguish between defaulters and non-defaulter. In order to investigate the monotonicity of the default probabilities, the Bayes' theorem (4.1) is used. With this theorem, it can be shown that the conditional default probability $P[X = 1 \mid S = s]$ is monotonic if and only if the likelihood ratio is monotonic. Combining Neymann–Pearson fundamental lemma and (3.8), it follows that the score variable S is linear with respect to the distribution $P[S \in \cdot \mid X = 1]$ and $P[S \in \cdot \mid X = 0]$ if the likelihood ratio is monotonic and thus also the conditional probability of default is monotonic. \square

At this point, it is possible to show the key property of the cumulative accuracy profile approach, which is the concavity of the CAP curve.

Theorem 3.2.2. *Let F_D and F_N be distribution functions on an open interval $I \subset \mathbb{R}$. Assume that F_D has a density f_D which is continuous on I and F_N has a positive density f_N that is continuous on I . Let $0 < p < 1$ be a fixed probability and define alarm rate F as in (3.1). Write f for a density of F . Define $CAP(u)$ for $u \in (0, 1)$ as in (3.4). Then $CAP(u)$ is concave for all u .¹¹*

Proof. Firstly, the derivative of the CAP function is calculated. Therefore, the derivative of the generalized inverse function of F is needed. The distribution function F is continuous and strictly monotone on the interval $[0, 1]$. Furthermore, let F be differentiable in s with $F'(s) \neq 0$. Then F^{-1} is differentiable in $F(s) = u$ and it holds

$$F^{-1'}(u) = \frac{1}{F'(s)} = \frac{1}{F'(F^{-1}(u))}.$$

With help of Bayes' theorem (4.1), it can be shown that

$$P[D \mid S = s] = \frac{pf_D(s)}{f(s)}.$$

¹¹[cf. 23, Theorem 5.1]

Now it is easy to calculate the derivative of CAP

$$\begin{aligned}
 CAP'(u) &= \frac{f_D(F^{-1}(u))}{f(F^{-1}(u))} \\
 &= \frac{\frac{1}{p}P[D | S = F^{-1}(u)]f(F^{-1}(u))}{f(F^{-1}(u))} \\
 &= \frac{P[D | S = F^{-1}(u)]}{p}.
 \end{aligned} \tag{3.9}$$

The concavity follows easily from (3.9). The conditional probability of default is monotonically decreasing (see Proposition 1), therefore $CAP^{-1}(u)$ is decreasing in u and thus the CAP function is concave. \square

3.3 Accuracy ratio

The visualization of model's performance can easily be done by CAP plots, but it is often convenient to have a single quantitative measure as well. The quality of a given rating model can be measured by the accuracy ratio (AR).

Definition 3.3.1 (AR). Accuracy ratio is defined as the ratio of the area a_R which is between the CAP of the rating model being validated ($AUCAP$) and the CAP of the random model ($AUCAP_{random}$) and the area a_P which is between the CAP of the perfect rating model ($AUCAP_{perfect}$) and the CAP of the random model ($AUCAP_{random}$). Mathematical definition is^{12,13}

$$\begin{aligned}
 AR &= \frac{a_R}{a_P} \\
 &= \frac{AUCAP - AUCAP_{random}}{AUCAP_{perfect} - AUCAP_{random}} \\
 &= \frac{\int_0^1 CAP(u) du - \frac{1}{2}}{\frac{1}{2} - \frac{p}{2}} \\
 &= \frac{2 \int_0^1 CAP(u) du - 1}{1 - p}.
 \end{aligned}$$

These two areas are also indicated in Figure 3.1. AR has values between 0 and 1, where values near 0 indicate only slightly better model than the random one, whereas values near 1 indicate almost perfect model. Most of the models that are considered to be good, have values between 0.5 and 0.75. Alternatively, AR is obtained from the difference between the probability that the defaulter's score is smaller than the non-defaulter's score and the probability that the defaulter's score is larger than the non-defaulter's score or mathematically written as

$$AR = P[S_D < S_N] - P[S_D > S_N].$$

¹²[cf. 23, Definition 3.5]

¹³[cf. 5, p. 4]

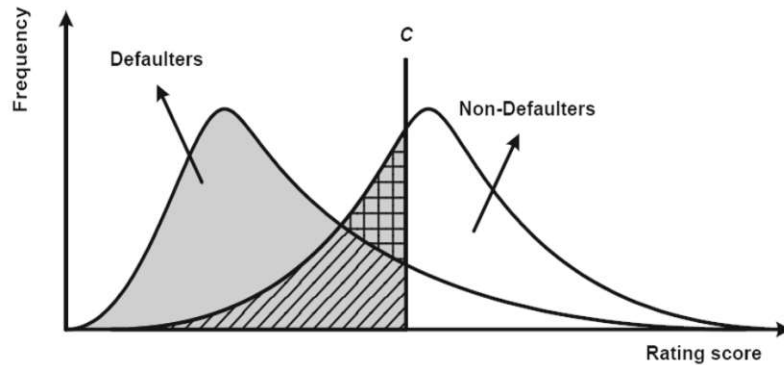


Figure 3.2: Distribution of rating scores for defaulting and non-defaulting borrowers, source: [5]

3.4 Receiver operating characteristic

A concept very similar to *CAP* is the receiver operating characteristic (ROC). Its construction is illustrated in Figure 3.2.

Definition 3.4.1 (ROC). The receiver operating characteristic function $ROC : [0, 1] \rightarrow [0, 1]$ is defined as⁷

$$ROC(u) = F_D(F_N^{-1}(u)), \quad (3.10)$$

where F_N^{-1} is the generalized inverse function of F_N .

It is easy to calculate the derivative of ROC , that will be needed later¹¹

$$ROC'(u) = \frac{f_D(F_N^{-1}(u))}{f_N(F_N^{-1}(u))}. \quad (3.11)$$

For better understanding, ROC represents the possible distributions of rating scores for defaulting and non-defaulting borrowers. In a perfect case, the left distribution (defaulters) and the right distribution (non-defaulters) would not overlap. In the real rating systems, that is not the case and as it is illustrated in Figure 3.2, there exists an overlap between those two distributions. The idea of ROC is to distinguish defaulting and non-defaulting borrowers based on their rating score. The easiest way would be to define a cut-off value C (as illustrated in Figure 3.2) and to classify each borrower with rating score lower than C as potential defaulter and each borrower with a rating score higher than C as non-defaulter. After this classification, four different outcomes would be possible. If the rating score is below the cut-off value C , the borrower was classified as a defaulter and the borrower will default, than the classification was correct. Otherwise the borrower was incorrectly classified. If the rating score is above the cut-off value C , the borrower was classified as a non-defaulter and the borrower will not default, than the classification was correct. Otherwise a defaulter was wrongly classified and put into the non-defaulters' group. After this distinction, it is possible to define the hit rate and the false alarm rate. These two

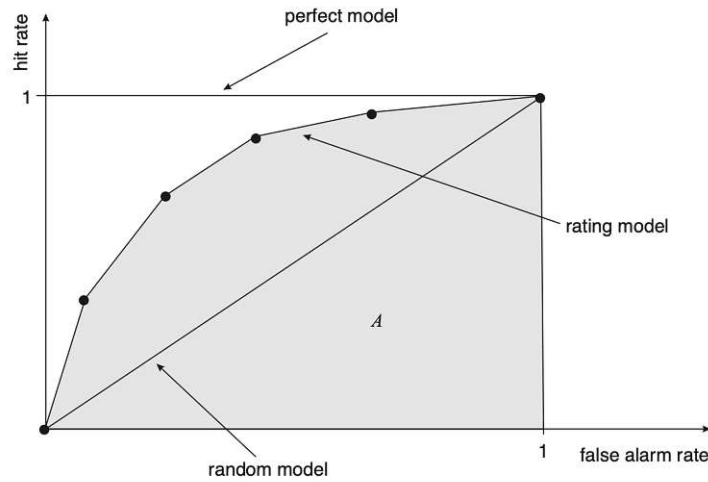


Figure 3.3: Receiver operating characteristic, source: [18]

rates are necessary for the understanding of *ROC*. The hit rate $HR(C)$ is defined as

$$HR(C) = \frac{H(C)}{N_D},$$

where $H(C)$ is the number of defaulters that were classified correctly with the cut-off value C and N_D is the total number of defaulters. The false alarm rate $FAR(C)$ is defined as

$$FAR(C) = \frac{F(C)}{N_{ND}},$$

where $F(C)$ is the number of non-defaulters that were predicted wrongly with the cut-off value C and N_{ND} is the total number of non-defaulters. In Figure 3.2, the area to the left of the cut-off value C under the score distribution of the defaulters (colored and hatched area) is the $HR(C)$ and the area to the left of C under the score distribution of the non-defaulters (chequered and hatched area) is the $FAR(C)$. To obtain the *ROC* curve, the quantities $HR(C)$ and $FAR(C)$ have to be computed for all possible cut-off values C that are contained in the range of the rating scores. The *ROC* curve is the plot of $HR(C)$ and $FAR(C)$. Figure 3.3 illustrates the concept of *ROC*. The performance of the rating model is better the steeper the *ROC* curve is at the left end and the closer the *ROC* curve's position is to the point $(0, 1)$.

3.5 The area under the receiver operating characteristic curve

As well as for *CAP*, also for *ROC* it is convenient to introduce a single quantitative measure. The area under the *ROC* curve is used for this purpose and is denoted by A (the gray area in Figure 3.3). The model is better the larger the area under the *ROC* curve is.

Definition 3.5.1 (AU ROC). The area under ROC can be calculated as¹²

$$A = \int_0^1 ROC(u) du = \int_0^1 F_D(F_N^{-1}(s)) ds. \quad (3.12)$$

Alternatively, A can also be interpreted as a probability in a way that the score of a randomly selected non-defaulter is greater than the score of a randomly selected defaulter. It can be shown that

$$A = P[S_D < S_N].$$

The area A is 0.5 for a random model and 1 for a perfect model. The value is therefore between 0.5 and 1 for any reasonable rating model.

3.6 Cumulative accuracy profile vs. receiver operating characteristic

There is a lot of literature about CAP curves and a lot of literature about ROC curve. It can be noticed that both curves look very similar. In fact, the CAP measure AR is just a linear transformation of the ROC measure A . That means that all analytical properties of CAP curves can simply be transferred to ROC curves and vice versa. Therefore, for example from the concave property of the CAP curve, the same property for the ROC curve can be derived.

Theorem 3.6.1. *There exists the following relation between the accuracy ratio and the area under ROC* ¹⁴

$$A = \frac{AR + 1}{2}.$$

¹⁴[cf. 23, Proposition 3.6]

Proof. Let Φ be the set of all possible score values s . Then it holds

$$\begin{aligned}
 A &\stackrel{(3.10)}{=} \int_0^1 F_D(F_N^{-1}(s)) \, ds \\
 &\stackrel{(\text{trafo})}{=} \int_{\Phi} F_D(s) \, d(F_N(s)) \\
 &\stackrel{(3.1)}{=} \int_{\Phi} F_D \, d\left(\frac{F(s) - pF_D(s)}{1-p}\right) \\
 &= \int_{\Phi} \frac{F_D(s)}{1-p} \, d(F(s)) - \int_{\Phi} \frac{pF_D(s)}{1-p} \, d(F_D(s)) \\
 &= \frac{\int_{\Phi} F_D(s) \, d(F(s))}{1-p} - \frac{p \int_{\Phi} F_D(s) \, d(F_D(s))}{1-p} \\
 &\stackrel{(\text{trafo})}{=} \frac{\int_0^1 F_D(F^{-1}(u)) \, du}{1-p} - \frac{p \int_0^1 F_D(F_D^{-1}(u)) \, du}{1-p} \\
 &= \frac{\int_0^1 CAP(u) \, du}{1-p} - \frac{\frac{1}{2}p}{1-p} \\
 &= \frac{1}{2} \left(\frac{2 \int_0^1 CAP(u) \, du - 1}{1-p} + \frac{1-p}{1-p} \right) \\
 &= \frac{1}{2} (AR + 1). \quad \square
 \end{aligned}$$

3.7 The generation of probability of default with cumulative accuracy profile approach

In the context of portfolios with little default data, Burgt [24] suggested estimating the conditional probability of default by fitting one-parameter curve to the observed CAP curve, taking the derivative of the fitted curve and then calculating the conditional probabilities of default based on the derivative. The probability of default conditional on a score value s is given in (3.3). It can be rewritten with help of CAP and ROC derivatives ((3.9) and (3.11)) as

$$\begin{aligned}
 P[D | S = s] &= \frac{pROC'(F_N(s))}{pROC'(F_N(s)) + 1 - p} \\
 &= pCAP'(pF_D(s) + (1-p)F_N(s)) \\
 &= pCAP'(F(s)).
 \end{aligned} \tag{3.13}$$

It should be noted, that by (3.13) the derivative of CAP curve for a borrower population with unconditional probability of default $p > 0$ is necessarily bounded from above by $1/p$. The following theorem shows, that this condition is not only necessary but it is also a sufficient condition for a distribution function on the unit interval to be a CAP curve.

Theorem 3.7.1. *Let $p \in (0, 1)$ be a fixed probability. Let f_D be a density on \mathbb{R} such that the set $I = \{f_D > 0\}$ is an open interval and f_D is continuous in I . Denote by*

3.7 The generation of probability of default with cumulative accuracy profile approach

$F_D(s) = \int_{-\infty}^s f_D(v) dv$ the distribution function associated with f_D . Then the following statements are equivalent:¹⁵

- (i) The function $u \mapsto C(u)$, $u \in (0, 1)$ is continuously differentiable in u with $\lim_{u \searrow 0} C(u) = 0$, $\lim_{u \nearrow 1} C(u) = 1$ and $0 < C'(u) \leq 1/p$, $u \in (0, 1)$.
- (ii) There is a density $f_N \geq 0$ such that $\{f_N > 0\} = I$, f_N is continuous I and

$$C(u) = F_D(F^{-1}(u)), u \in (0, 1),$$

where $F(s) = pF_D(s) + (1-p) \int_{-\infty}^s f_N(v) dv$.

Proof. (i) \Rightarrow (ii) : By assumption, C maps onto $(0, 1)$ and the inverse C^{-1} of C exists. Define $F(s) = C^{-1}(F_D(s))$. Then F is a distribution function with $\lim_{s \rightarrow \inf I} F(s) = 0$, $\lim_{s \rightarrow \sup I} F(s) = 1$ and density

$$f(s) = F'(s) = \frac{f_D(s)}{C'(F(s))}, \quad s \in I. \quad (3.14)$$

Observe that $f(s)$ is positive and continuous in I . Therefore the inverse F^{-1} of F exists. Let

$$F_N(s) = \frac{F(s) - pF_D(s)}{1-p}, \quad s \in \mathbb{R},$$

$$f_N(s) = f_D(s) \frac{1/C^{-1}(F(s)) - p}{1-p}, \quad s \in I.$$

By (3.14), is f_N continuous derivative of F_N and is positive in I by assumption on C^{-1} and f_D . This implies that F_N is a distribution function with $\lim_{s \rightarrow \inf I} F_N(s) = 0$, $\lim_{s \rightarrow \sup I} F_N = 1$ and density f_N . By construction of F and F_N , the functions C , F_D and F satisfy equation from (ii).

(ii) \Rightarrow (i) : By construction, $F_D(s)$ and $F(s)$ are distribution functions which converge to 0 for $s \rightarrow \inf I$ and to 1 for $s \rightarrow \sup I$. This implies the limit statements for C . Equation from (ii) implies that C is continuously differentiable with derivative

$$0 < C'(u) = \frac{f_D(F'(u))}{pf_D(F'(u)) + (1-p)f_N(F'(u))} \leq 1/p. \quad \square$$

This theorem is needed a little bit later in this Subsection. For the sake of completeness, also the same result (without proof) is given for the *ROC* curve.

Theorem 3.7.2. Let $f_D \geq 0$ be a density on \mathbb{R} such that the set $I = \{f_D > 0\}$ is an open interval and f_D is continuous in I . Denote by $F_D(s) = \int_{-\infty}^s f_D(v) dv$ the distribution function associated with f_D . Then the following statements are equivalent:¹⁶

¹⁵[cf. 23, Theorem 5.2]

¹⁶[cf. 23, Proposition 5.3]

(i) The function $u \mapsto R(u)$, $u \in (0, 1)$ is continuously differentiable in u with $\lim_{u \searrow 0} R(u) = 0$, $\lim_{u \nearrow 1} R(u) = 1$ and $0 < R'(u)$, $u \in (0, 1)$.

(ii) There is a density $f_N \geq 0$ such that $\{f_N > 0\} = I$, f_N is continuous I and

$$R(u) = F_D(F_N^{-1}(u)), \quad u \in (0, 1),$$

where $F_N(s) = \int_{-\infty}^s f_N(v) dv$.

As already mentioned and shown in Subsection 3.2, the key parameter in this approach is the concavity. The following function is introduced by Burgt [24] for estimating the CAP function¹⁷

$$y(x) = \frac{1 - e^{-kx}}{1 - e^{-k}}, \quad x \in [0, 1], \quad (3.15)$$

where $k \in \mathbb{R}$ is called the concavity parameter, y is the cumulative number of defaults and x is the cumulative number of borrowers. The concavity parameter can be interpreted as a measure of discriminatory power. When $k \rightarrow \infty$ than $y \approx 1$ and the CAP therefore corresponds to perfect model that was shown in Figure 3.1. When $k \rightarrow 0$ than $y \approx x$ and the CAP corresponds to random model that was shown in Figure 3.1. In practice, k is between 0 and ∞ . The CAP function has concave shape when $k > 0$. The CAP function is convex when k is negative. This is the case when the conditional probability of default is monotonically increasing, which means that the most of the defaults occur in the best rating classes and hardly any defaults in the worst rating classes. Therefore, negative k value corresponds to a rating system that assigns ratings in reverse order of default risk. This is called adverse selection (sometimes also referred to as anti-selection or negative selection). This was for example observed in the credit crisis of 2007: defaults of subprime mortgage loans had a negative impact on the hedge funds, investment banks and conduits by means of US mortgage securitized pools. Most of these securitized pools were very good rated, even though the underlying risk was high. It can be therefore concluded that the concavity k might be an indicator of the soundness of a rating system:

1. When k is high and positive, the rating system has good discriminatory power.
2. When $k = 0$, the rating system has no discriminatory power.
3. When k is negative, the rating system is subject to adverse selection.

The already mentioned accuracy ratio also measures discriminatory power, so it is not surprising that a relation between AR and k exists.

Theorem 3.7.3. *There exists the following relation between the accuracy ratio and the concavity parameter¹⁸*

$$k \approx \frac{2}{1 - AR}. \quad (3.16)$$

¹⁷[cf. 24, p. 4]

¹⁸[cf. 24, p. 6,14–16]

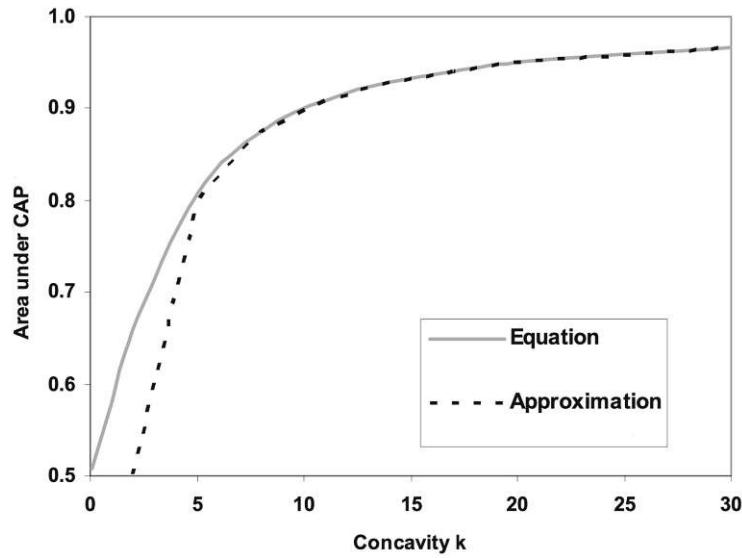


Figure 3.4: Comparison of AUCAP and its approximation, source: [24]

Proof. Firstly the area under the *CAP* curve (*AUCAP*) is calculated by integrating (3.15) over interval $[0, 1]$

$$AUCAP = \int_0^1 \frac{1 - e^{-kx}}{1 - e^{-k}} dx = \frac{1}{1 - e^{-k}} - \frac{1}{k}. \quad (3.17)$$

The *AUCAP* approaches 1 when $k \rightarrow \infty$, this is the case when the rating system has perfect discriminatory power (perfect model from Figure 3.1). When $k \rightarrow 0$, then the area *AUCAP* approaches 0.5. This is the case when rating system has no discriminatory power (random model from Figure 3.1). As e^{-k} goes to zero very fast, *AUCAP* can be approximated by the following expression

$$AUCAP \approx 1 - \frac{1}{k} \Rightarrow k \approx \frac{1}{1 - AUCAP}. \quad (3.18)$$

Figure 3.4 compares the *AUCAP* equation (3.17) with its approximation (3.18). It shows that the approximation is good for $AUCAP \geq 0.8$, which corresponds to $AR \geq 0.6$. This means, that for rating systems with good discriminatory power, (3.18) provides simpler method for calculation of the concavity k . Before the relation between k and the AR is derived, one more approximation needs to be done. It is assumed that $AUCAP_{perfect} \approx 1$. For a perfect discriminatory model, all defaults are observed in the worst rating class and therefore $AUCAP_{perfect}$ would be less than 1 and for this model the approximation would not be totally correct. But since the perfect model almost never occurs in real life, this approximation is considered to be correct

$$AR = \frac{AUCAP - AUCAP_{random}}{AUCAP_{perfect} - AUCAP_{random}} \approx 2AUCAP - 1. \quad (3.19)$$

A relation between k and the AR measure is obtained by combining (3.17) and (3.19)

$$AR = 2AUCAP - 1 = 2 \left(\frac{1}{1 - e^{-k}} - \frac{1}{k} \right) - 1.$$

Using the approximation from (3.18), the wanted relation between the AR and the concavity k is derived

$$AR \approx 2AUCAP - 1 = 1 - \frac{2}{k} \Rightarrow k \approx \frac{2}{1 - AR}.$$

□

The approach for PD estimation consists of few steps, where many of equations from above are used. The first step is to construct the CAP curve from the observations (it is also useful to construct the random and perfect CAP curve as well). The fitting of the CAP to the function in (3.15) is applied. Therefore, the concavity parameter needs to be calculated. There are different methods how this can be done. The easiest one, is to use the approximation (3.18), but since it is only an approximation, it is not so precise. Better way to calculate the concavity parameter is by minimizing root mean square error $RMSE$ ¹⁹

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \left\{ y_i - \frac{1 - e^{-kx_i}}{1 - e^{-k}} \right\}^2} \quad (3.20)$$

where N is the number of rating classes and x_i and y_i are the observed cumulative percentages of borrowers and defaults respectively in each rating class, so for all $i = 1, \dots, N$. It is always useful to check, if the calculation is correctly done. For that, firstly the derivative of (3.15) needs to be computed

$$y'(x) = \frac{ke^{-kx}}{1 - e^{-k}}. \quad (3.21)$$

It can be observed that for $k > 0$ following inequalities hold

$$y'(1) = \frac{k}{1 - e^{-k}} e^{-k} \leq y'(x) \leq \frac{k}{1 - e^{-k}} = y'(0), \quad x \in [0, 1],$$

and therefore Theorem 3.7.1 can be applied and implies the check for the concavity parameter

$$0 < \frac{k}{1 - e^{-k}} \leq \frac{1}{p},$$

where $p > 0$ denotes the unconditional probability of default.

After getting the concavity parameter, the PD can be derived from the CAP curve, using the following equation¹⁹

$$PD(s) = DR \frac{dy}{dx}, \quad (3.22)$$

where DR is the average observed default rate (= the total number of defaults divided by the total number of obligors for the whole portfolio) and s is the rating class. Combining (3.21) and (3.22) gives the following equation¹⁹

$$PD(s) = DR \left(\frac{ke^{-kx_s}}{1 - e^{-k}} \right), \quad (3.23)$$

¹⁹[cf. 24, p. 7]

where x_s represents the cumulative percentage of borrowers in rating class s . The value of x_s is calculated as the midpoint between the cumulative percentage of borrowers in rating class s and $s - 1$ ²⁰

$$x_s = \frac{z_N + z_{N-1} + \dots + z_{s-1} + (z_s/2)}{z},$$

where z is the total number of borrowers and z_i is the number of borrowers in the rating class i (whereas the rating class N represents the most risky rating class).

3.8 The generation of probability of default with receiver operating characteristic approach

Tasche [23] on the other hand, suggested estimating the conditional probability of default by fitting two-parameter curve to the observed *ROC* curve. Here it is assumed that the score distributions conditional on default and survival are normal, meaning that

$$S_D \sim \mathcal{N}(\mu_D, \sigma_D^2), \quad S_N \sim \mathcal{N}(\mu_N, \sigma_N^2). \quad (3.24)$$

Therefore formulas for *ROC* and *A* can be easily derived from (3.10) and (3.12)²¹

$$\begin{aligned} ROC(u) &= \Phi\left(\frac{\sigma_N \Phi^{-1}(u) + \mu_N - \mu_D}{\sigma_D}\right), \quad u \in [0, 1] \\ A &= \Phi\left(\frac{\mu_N - \mu_D}{\sqrt{\sigma_N^2 + \sigma_D^2}}\right), \end{aligned} \quad (3.25)$$

where $\Phi()$ is a cumulative distribution function for a standard normal random variable.

The unconditional score distribution F can be derived from (3.1). However, F is not a normal distribution function under (3.24) for $p \notin \{0, 1\}$. Its inverse function F^{-1} can be evaluated numerically, but no closed form representation is known. Therefore, for the plots of *CAP* curve it is more efficient to obtain it from the graph of all points $(F(s), F_D(s))$ for $s \in \mathbb{R}$ where the points are connected by straight lines (linear interpolation). The value of *AR* can be derived from the value of *A* by Theorem 3.6.1. Figure 3.5 shows score densities vs. *ROC* and *CAP* curves in the case of normal conditional score densities with parameter values $\mu_D = 6.8$, $\sigma_D = 1.92$, $\mu_N = 8.5$, $\sigma_N = 2$ and unconditional PD 10%.

As already mentioned above, no closed form for *CAP* is available in case of normally distributed defaulter and survivor score distributions. That is the main reason why *ROC* curves and not *CAP* curves are used in this approach. Another reason is that fitting *ROC* curves is conceptually sounder than fitting *CAP* curves, because this way one can better separate the estimation of an unconditional probability of default from the estimation of parameters of the fitting function.

²⁰[cf. 24, p. 8]

²¹[cf. 23, p. 11]

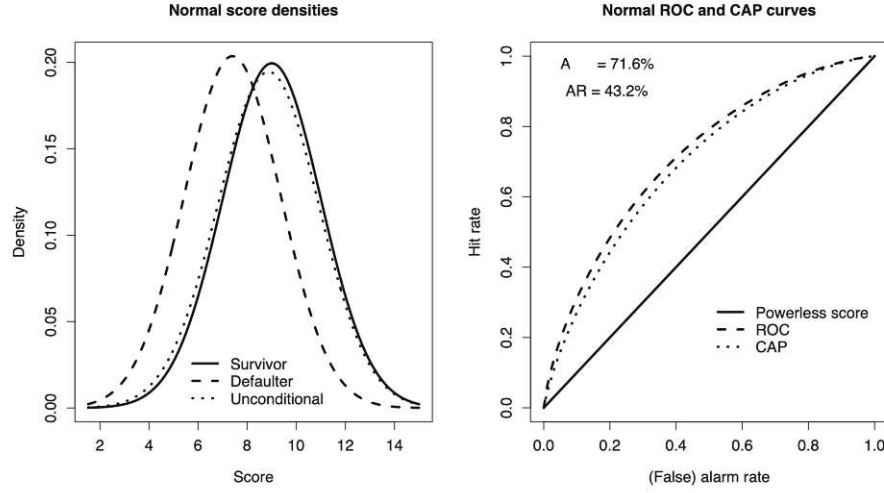


Figure 3.5: Score vs. alarm rate, source: [23]

By substitution of parameter $b > 0$ for $\frac{\sigma_N}{\sigma_D}$ and parameter $a \in \mathbb{R}$ for $\frac{\mu_N - \mu_D}{\sigma_D}$ into (3.25), the two-parametric family of *ROC* function is obtained as²²

$$R_{a,b}(u) = \Phi(a + b\Phi^{-1}(u)), \quad u \in (0, 1). \quad (3.26)$$

It is easy to calculate the derivative of $R_{a,b}$ and also to rewrite A with help of parameters a and b as

$$R'_{a,b}(u) = b \frac{\varphi(a + b\Phi^{-1}(u))}{\varphi(\Phi^{-1}(u))}, \quad (3.27)$$

$$A_{a,b} = 2\Phi\left(\frac{a}{\sqrt{b^2 + 1}}\right) - 1,$$

where $\varphi(\cdot)$ is a standard normal density.

This two-parameter family of functions would clearly give better fits than a one-parameter one but in order to compare it with the one-parameter *CAP* approach by Burgt [24], it is simplified into one-parameter family. That is defined by fixing b as $b = 1$ and instead of writing $R_{a,1}$, it is only written R_a . Therefore (3.26) and (3.27) are simplified into

$$R_a(u) = \Phi(a + \Phi^{-1}(u)),$$

$$R'_a(u) = e^{-a\Phi^{-1}(u) - a^2/2} \quad (3.28)$$

$$A_a = 2\Phi\left(\frac{a}{\sqrt{2}}\right) - 1.$$

When knowing the default rate DR which is easily calculated as the total number of defaults divided by the total number of obligors and $F_N(s)$ for each rating class s and using $R'_a(u)$

²²[cf. 23, p. 32]

3.8 *The generation of probability of default with receiver operating characteristic approach*

from (3.28) in (3.13), than the following representation of the probability of default for rating class s is derived

$$PD(s) = \frac{1}{1 + \frac{1-DR}{DR} e^{(a\Phi^{-1}(F_N(s))+a^2/2)}}. \quad (3.29)$$

This equation can be simplified into

$$PD(s) = \frac{1}{1 + e^{(\alpha+\beta\Phi^{-1}(F_N(s)))}}, \quad (3.30)$$

with $\alpha = \log(\frac{1-DR}{DR}) + \frac{a^2}{2}$ and $\beta = a$.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

4 Bayesian approach

This chapter is based on [3, 4, 6, 7, 8, 17, 27].

4.1 Bayesian statistics

This Subsection will introduce the theory that is necessary to describe the Bayesian methods used later in this chapter. The goal is to predict the PD, and prediction is very much concerned with probability. There are two main interpretations of probability: frequentism (or classical) and Bayesianism. In CAP and ROC approaches the frequentist (or classical) definition of probability is used, which is based on frequencies of events. Whereas the Bayesian definition of probability is based on the knowledge of events. Bayesian inference is a method of statistical inference in which Bayes' theorem is used to update the probability for a hypothesis as more evidence or information becomes available. Therefore, the Bayes' theorem needs to be introduced first with an explanation of its various components.

Theorem 4.1.1 (Bayes' theorem). *For events A and B , with $P[B] \neq 0$ the following equation holds, which is based on the conditional probability law²³*

$$P[A | B] = \frac{P[B | A]P[A]}{P[B]}. \quad (4.1)$$

where $P[A]$ is the probability of A before one knows the outcome of event B , $P[B | A]$ is the probability of B assuming what one knows about event A , and $P[A | B]$ is the probability of A knowing that event B has occurred. $P[A]$ is called „prior probability“ of A , while $P[A | B]$ is called „posterior probability“ of A .

Figure 4.1 provides a basic graphical illustration of how the Bayes' theorem combines the data with „prior distribution“ to get the „posterior distribution“.

For the Bayesian inference, firstly the formal definition of variables is needed²⁴

- x , a data point in general. This may be a vector of values.
- ϕ , the parameter of the data point's distribution, that is $x \sim p(x | \phi)$. This may be a vector of parameters.
- α , the hyperparameter of the parameter distribution, that is $\phi \sim p(\phi | \alpha)$. This may be a vector of hyperparameters.
- X is the sample, a set of n observed data points, that is x_1, \dots, x_n .

²³[cf. 3, p. 36]

²⁴[cf. 27]

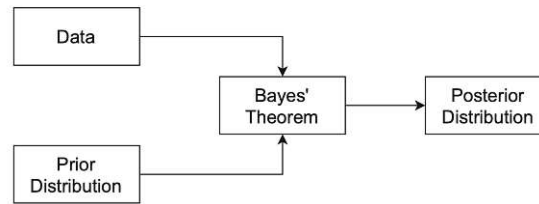


Figure 4.1: Bayes' theorem, source: compiled by author

In Bayesian inference, the „prior distribution“ is the distribution of the parameter(s) before any data is observed, that is $p(\phi | \alpha)$. The sampling distribution is the distribution of the observed data conditional on its parameters, that is $p(X | \phi)$. This distribution is also called „likelihood function“. The „marginal likelihood“ (also sometimes called the evidence) is the distribution of the observed data marginalized over the parameter(s), that is $p(X | \alpha) = \int p(X | \phi)p(\phi | \alpha) d\phi$. The „posterior distribution“ is the distribution of the parameter(s) after taking into account the observed data. This is determined by Bayes' theorem (4.1), which forms the heart of Bayesian inference²⁴

$$p(\phi | X, \alpha) = \frac{p(\phi, X, \alpha)}{p(X, \alpha)} = \frac{p(X | \phi, \alpha)p(\phi, \alpha)}{p(X | \alpha)p(\alpha)} = \frac{p(X | \phi, \alpha)p(\phi | \alpha)}{p(X | \alpha)}. \quad (4.2)$$

This outcome is called "posterior distribution" and it can be interpreted as what is known about the unknown parameter when the data is given. A simpler version of the „posterior“ can be written as

$$p(\phi | X, \alpha) \propto p(X | \phi, \alpha) \times p(\phi | \alpha),$$

„Posterior distribution“ \propto „Likelihood function“ \times „Prior distribution“.

4.2 The generation of probability of default with Bayesian approach

To use Bayesian statistics in the setting of estimating a probability of default was mentioned by Kiefer [7]. The steps in a Bayesian approach are:

1. Specify an economic model precisely defining the quantity of interest and generating a „likelihood function“ for well-defined data, refer to Subsection 4.2.1.
2. Identify a qualified expert and elicit information about the quantity of interest. Quantify this information in a probability distribution, which will be the „prior distribution“, refer to Subsection 4.2.2.
3. Use the rules of probability (Bayes' theorem) to combine coherently the „likelihood“ and the „prior distribution“, generating the „posterior distribution“, refer to Subsection 4.2.3.

4.2.1 Statistical models for defaults

Firstly, the „likelihood“ function needs to be defined. It is not known whether the default will occur or not, but it is possible to model this uncertain event with a probability distribution. Kiefer suggested that the simplest probability model for defaults of assets in a homogeneous segment of a portfolio is the binomial, in which the defaults are assumed independent across assets and over time, and occur with common probability $\phi \in [0, 1]$. Also expert judgement e is crucial at every step of a statistical analysis, this will be described in detail in Subsection 4.2.2. Further, Kiefer defined a random variable d_i , which takes the value $d_i = 0$ in case there is no default in the i -th observation or the value $d_i = 1$ in case there is a default in the i -th observation. It follows that the distribution of d_i is²⁵

$$P[d_i | \phi] = \phi^{d_i} (1 - \phi)^{1-d_i},$$

where ϕ is the binomial parameter, which is the already mentioned probability of default. This was the definition on observation level, but it also needs to be defined on the rating class level (this is the level on which the practical part of this thesis is based), and each rating class contains n observations. Let $D = \{d_i, i = 1, \dots, n\}$ be a multivariate random variable, which denotes whether each observation in the whole rating class data set was a default or not. Define $r = \sum_{i=1}^n d_i$ which counts the number of defaults in the whole rating class data set. Then the joint distribution of the data is²⁵

$$\begin{aligned} P[D | \phi, e] &= P[d_1, \dots, d_n | \phi, e] \\ &= \phi^{d_1} (1 - \phi)^{1-d_1} \dots \phi^{d_n} (1 - \phi)^{1-d_n} \\ &= \prod_{i=1}^n \phi^{d_i} (1 - \phi)^{1-d_i} \\ &= \phi^r (1 - \phi)^{n-r}. \end{aligned}$$

This is the searched „likelihood function“. Since this distribution depends on the data D only through r (n is regarded as fixed), the sufficiency principle implies that the attention can be concentrated only on the distribution of r ²⁵

$$P[r | \phi, e] = \binom{n}{r} \phi^r (1 - \phi)^{n-r}, \quad (4.3)$$

where $\binom{n}{r}$ is binomial coefficient that is defined as

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

Equation (4.3) defines the distribution of r defaults given that PD is ϕ and expert information e is available. Figure 4.2 illustrates an example of likelihood function for $n = 50$ and $r = 0, 1, 2$. It can be seen that functions move to the right with the increasing r .

²⁵[cf. 7, Chapter 2]

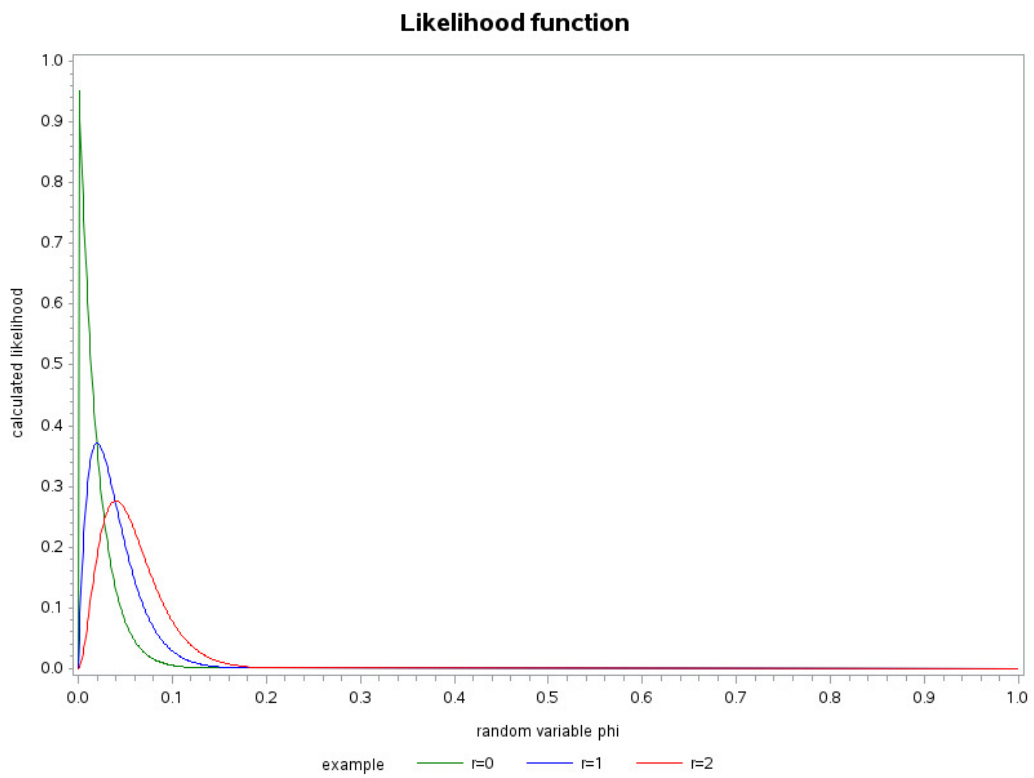


Figure 4.2: „Likelihood function“, source: compiled by author

4.2.2 Expert opinion

As already mentioned in previous Subsection, the expert judgement is very important at every step of statistical analysis. It is clear that there is some information available about ϕ in addition to the data information. Examples of this additional information are: it is expected that the portfolio in question is a low default portfolio, further it is assumed that no portfolio has default probability 0 or that the borrower is unwilling to make sure-loss bets. These beliefs can be combined according to the laws of probability (convexity, additivity and multiplication) and it can be concluded that the only one satisfactory way to represent this information available about ϕ is through probability distribution $P[\phi | e]$. It is clear that it is not easily possible to assess these beliefs in this level of detail and that some further objections can be easily raised. Few examples for these objections could be, if the events are really independent or if the observations are truly trials of the same experiment. The answer on both is no, but the expert judgement that is used, concludes that the statistical model captures the essential features of the given problem and that the remaining inaccuracies are minor. In scientific research, often a vigorous argument about the validity of the model can be found and there is also widespread use of statistical models that are not 100% correct but still accurate enough to be useful. Therefore the same should be applied in the assessment of the distribution $P[\phi | e]$.

As already mentioned above, this additional expert information should be represented in a probability distribution and this will be used as the prior distribution in the Bayes' theorem later in Subsection 4.2.3. The challenging step is to represent the expert's assessments with a statistical distribution. The conditioning argument e will be temporarily dropped and it will be substituted with other parameters. A distribution, that is commonly used for a parameter, that is constrained to lie in interval $[0, 1]$ is the beta distribution. The beta distribution for the random variable $\phi \in [0, 1]$ with parameters $\alpha > 0$ and $\beta > 0$ is²⁶

$$P[\phi | \alpha, \beta] = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \phi^{\alpha-1} (1 - \phi)^{\beta-1}. \quad (4.4)$$

This is a two-parameter beta distribution and it can be used as a „prior distribution“, instead of $P[\phi | e]$. An example of this distribution is shown in Figure 4.3 with parameters $(\{\alpha, \beta\} = \{2, 5\}, \{3, 1\})$ and an example of uniform distribution where the parameters are $(\{\alpha, \beta\} = \{1, 1\})$. It just needs to be decided how the parameters α and β are chosen. In [4, p. 4] it was suggested, that if the default events are completely independent (which is the assumption in this master's thesis), then $\alpha = n + 1$ and $\beta = n - r$, where n is the number of all borrowers in the whole rating class data set and r is the number of defaults in the whole rating class data set.

A more detailed specification to the two-parameter one, is the four-parameter beta distribution. Here the „prior distribution“ is defined as²⁶

$$P[\phi | \alpha, \beta, a, b] = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{(\phi - a)^{\alpha-1} (b - \phi)^{\beta-1}}{(b - a)^{\alpha+\beta-1}},$$

for the random variable $\phi \in [0, 1]$ and parameters $\alpha > 0$, $\beta > 0$, $a < \phi < b$. The four-parameter beta distribution allows flexibility within the range $[a, b]$, but in some situations

²⁶[cf. 7, Chapter 4]

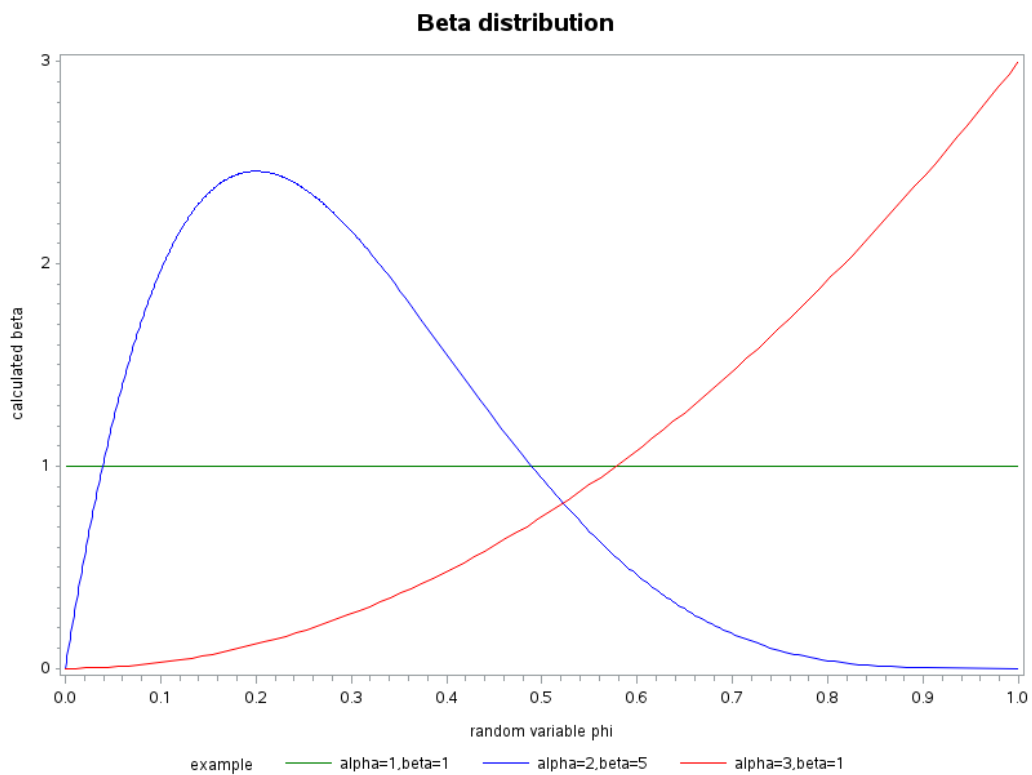


Figure 4.3: Two-parameter beta distribution, source: compiled by author

it may be too restrictive. This issue could be solved with usage of seven-parameter beta distribution that is defined as²⁶

$$P[\phi | \alpha_1, \beta_1, \alpha_2, \beta_2, \lambda, a, b] = \frac{\lambda \Gamma(\alpha_1 + \beta_1) (\phi - a)^{\alpha_1 - 1} (b - \phi)^{\beta_1 - 1}}{\Gamma(\alpha_1) \Gamma(\beta_1) (b - a)^{\alpha_1 + \beta_1 - 1}} + \frac{(1 - \lambda) \Gamma(\alpha_2 + \beta_2) (\phi - a)^{\alpha_2 - 1} (b - \phi)^{\beta_2 - 1}}{\Gamma(\alpha_2) \Gamma(\beta_2) (b - a)^{\alpha_2 + \beta_2 - 1}},$$

for the random variable $\phi \in [0, 1]$ and parameters $\alpha_1 > 0$, $\beta_1 > 0$, $\alpha_2 > 0$, $\beta_2 > 0$, $a < \phi < b$, $\lambda \in (0, 1)$. This is the mixture of two four-parameter beta distributions with common support. The approach would become more complicated and it would be difficult to derive this kind of unconditional distribution. Therefore, in the practical part of this master's thesis the two-parameter beta distribution will be used.

4.2.3 Calculation of probability of default

Now it possible to combine the statistical model for the data information calculated in (4.3) and expert's opinion used as „prior distribution“ calculated in (4.4) with equation from Bayes' theorem (4.1) to get the „posterior distribution“. This distribution describes the uncertainty about ϕ given observation of r actual defaults in rating class data set with n obligors and having expert information e . It is denoted as $P[\phi | r, e]$ and calculated as²⁷

$$P[\phi | r, e] = \frac{P[r | \phi, e] P[\phi | e]}{P[r | e]}. \quad (4.5)$$

The one part of (4.5) that still needs to be calculated is $P[r | e]$, which is the unconditional distribution of the number of defaults. It is called the predictive distribution of the statistic r and is defined as²⁷

$$P[r | e] = \int P[r | \phi, e] P[\phi | e] d\phi.$$

For the two-parameter beta family as in (4.4), an exact functional form can be calculated. It is²⁷

$$P[r | e] = \frac{\Gamma(r + \alpha) \Gamma(n - r + \beta) \Gamma(\alpha + \beta) \Gamma(n + 1)}{\Gamma(r + 1) \Gamma(n - r + 1) \Gamma(\alpha) \Gamma(\beta) \Gamma(n + \alpha + \beta)}.$$

This predictive distribution for the two-parameter beta family has a simple form in a special case of uniform distribution. That is

$$P[r | e] = \frac{1}{n + 1}. \quad (4.6)$$

Figure 4.4 shows the predictive distribution corresponding to two-parameter beta family with $\alpha = 2$, $\beta = 5$ and also $\alpha = \beta = 1$, which is the uniform distribution (It can be used to double check with the equation (4.6). When this equation is used, the result is $P[r | e] = \frac{1}{n+1} = \frac{1}{51} = 0.01961$ which corresponds with the value from Figure 4.4.). For the four-parameter beta family and seven-parameter beta family, the predictive distributions

²⁷[cf. 7, Chapter 5]

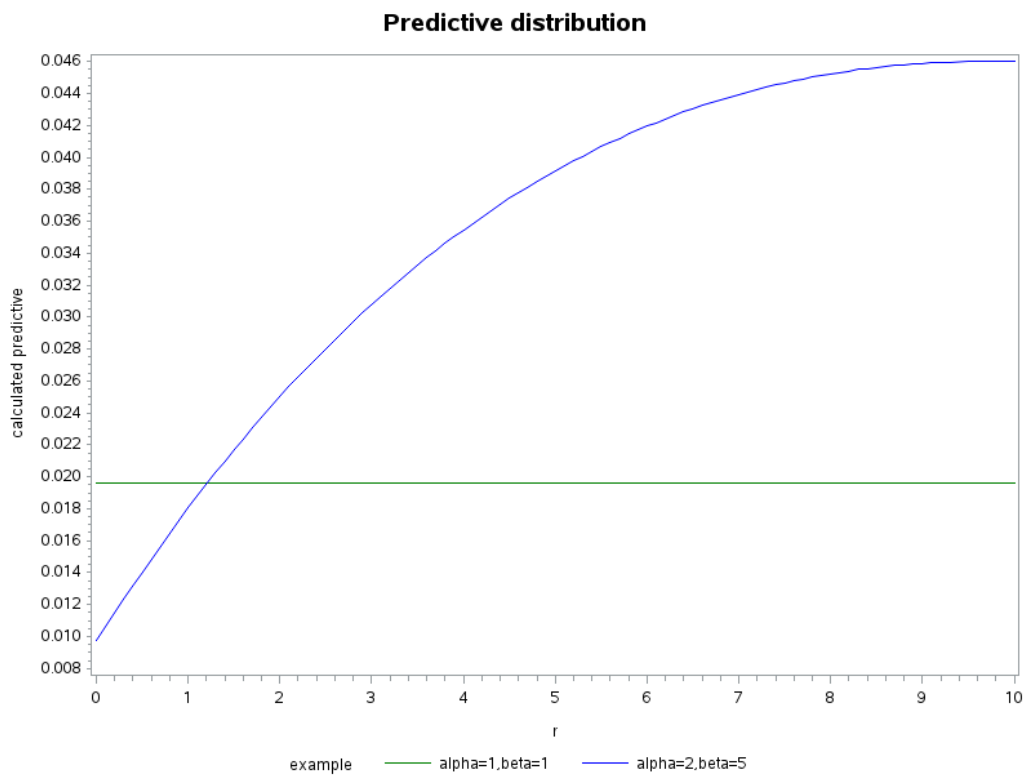


Figure 4.4: Predictive distribution, source: compiled by author

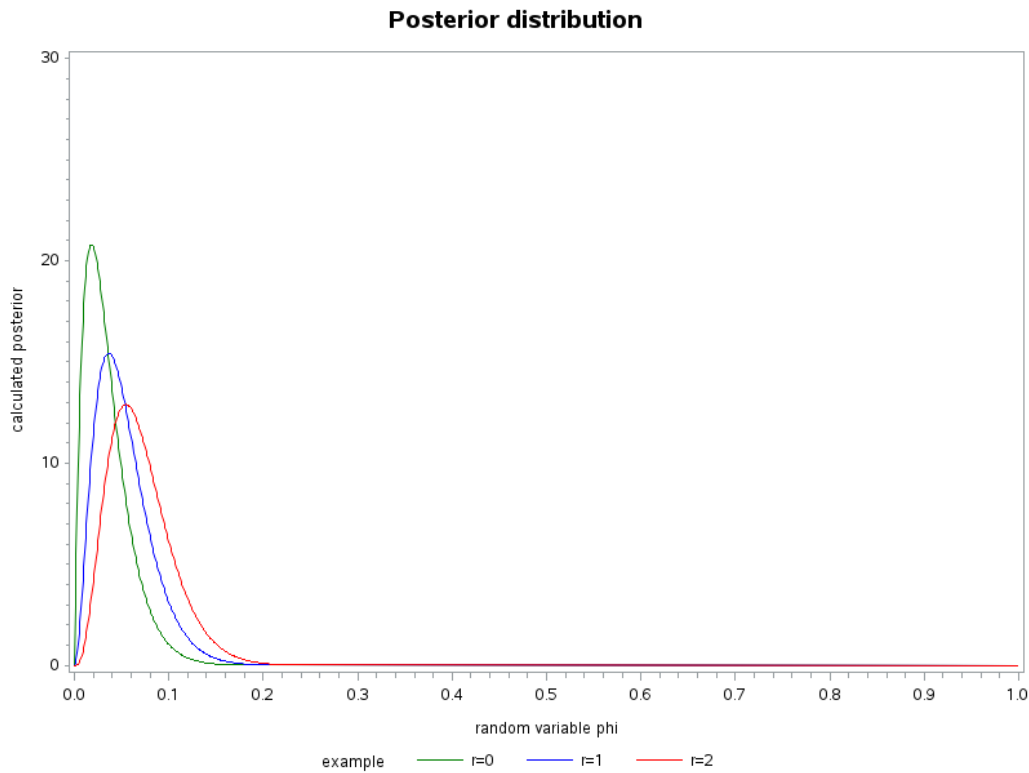


Figure 4.5: „Posterior distribution“, source: compiled by author

are not so simple. Therefore, in the practical part of this master’s thesis, the two-parameter beta family version of predictive distribution will be used.

At this point, everything is known that is needed to define „posterior distribution“. Therefore an example of it can be shown. Figure 4.5 illustrates this for $n = 50$ and $r = 0, 1, 2$. It can be seen that functions move to the right with the increasing r . Probability of default for rating class s is derived searching for the maximum value of the „posterior distribution“ (4.5) or mathematically written as

$$PD(s) = \max_{\phi} P[\phi | r, e]. \tag{4.7}$$



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

5 Statistical modelling

This chapter is based on fictitious data set. The author of this master's thesis has been working in the banking industry for the last four years and therefore the distribution of both borrowers and defaults does represent the real world situation. Almost always are the most defaults found in the worst four rating classes and the most borrowers are distributed in the middle four rating classes. In this chapter the data will be used to calculate the probability of default using different approaches, that were mentioned in Chapters 3 and 4. The results will be compared and discussed. Everything calculated here is based on SAS codes that can be found in Chapter 7.

5.1 Data

Figure 5.1 shows the data set that was used for testing. As already mentioned above, even though this data set is fictitious, it can be seen that almost 63% of all borrowers have one of the middle four rating classes and 7 out of 8 defaults are in the worst four rating classes. Both of these observations correspond with reality and therefore the results obtained in the further analysis can be applied on the real data sets. Figure 5.1 has the rating classes sorted in descending order from the worst (= rating class number 8) to the best (= rating class number 1), it also consists of total number of borrowers, total number of defaults and finally the default rate, which is the ratio between the total number of defaults and total number of borrowers in each grade. In high default portfolios is default rate of each rating class often used to model the probability of default. That is not the case in the low default portfolios and with help of this example, it can easily be seen why. In the rating grades 1, 2 and 4 there were no defaults observed. Sometimes this can even be the case for more than three rating classes and if the default rate of these rating classes would be used to model the probability of default, the outcome would be 0%. That is not allowed, PD can be small but it has to be greater than zero and therefore some other way to calculate it needs to be found. And that is the main intention of this master's thesis.

5.2 Cumulative accuracy profile approach

Firstly, the PD will be calculated with help of *CAP* approach, that was explained in Subsection 3.7. The columns cumulative number of borrowers, cumulative number of defaults need to be added into Figure 5.1 and therefore new Figure 5.2 will originate. Also the ratio between the cumulative number of borrowers and total number of borrowers in each grade needs to be calculated. This will be denoted with x . Finally, the ratio between the cumulative number of defaults and total number of defaults needs to be determined and also added into Figure 5.2. This second ratio will be indicated with y .

rating class	# borrowers	# defaults	default rate
8	21	2	9.52%
7	34	2	5.88%
6	51	2	3.92%
5	52	1	1.92%
4	53	0	0.00%
3	45	1	2.22%
2	30	0	0.00%
1	34	0	0.00%

Figure 5.1: Data set (actual model), source: compiled by author

rating class	# borrowers	# defaults	default rate	cumulative borrowers	cumulative defaults	x	y
8	21	2	9.52%	21	2	0.066	0.250
7	34	2	5.88%	55	4	0.172	0.500
6	51	2	3.92%	106	6	0.331	0.750
5	52	1	1.92%	158	7	0.494	0.875
4	53	0	0.00%	211	7	0.659	0.875
3	45	1	2.22%	256	8	0.800	1.000
2	30	0	0.00%	286	8	0.894	1.000
1	34	0	0.00%	320	8	1.000	1.000

Figure 5.2: Data set (actual model) for *CAP* approach, source: compiled by author

This is the actual model. As described in Subsection 3.1, the actual model is always compared with perfect and random model, in order to evaluate how good the actual model really is. Therefore, the perfect and random model need to be identified. The perfect model is a model where all defaults emerge from the worst rating grade. Figure 5.3 shows a profile of the perfect model, in which can be seen that all 8 defaults happen in rating class 8.

Whereas, the random model is a model where borrowers and defaults are equally distributed across all grades. Figure 5.4 shows a profile of the random model, in which can be seen that there are the same number of borrowers and defaults in each rating class.

The next step is the construction of a scatter plot. For that, the variables x and y

rating class	# borrowers	# defaults	default rate	cumulative borrowers	cumulative defaults	x	y
8	21	8	38.10%	21	8	0.066	1.000
7	34	0	0.00%	55	8	0.172	1.000
6	51	0	0.00%	106	8	0.331	1.000
5	52	0	0.00%	158	8	0.494	1.000
4	53	0	0.00%	211	8	0.659	1.000
3	45	0	0.00%	256	8	0.800	1.000
2	30	0	0.00%	286	8	0.894	1.000
1	34	0	0.00%	320	8	1.000	1.000

Figure 5.3: Data set (perfect model) for *CAP* approach, source: compiled by author

rating class	# borrowers	# defaults	default rate	cumulative borrowers	cumulative defaults	x	y
8	40	1	2.50%	40	1	0.125	0.125
7	40	1	2.50%	80	2	0.250	0.250
6	40	1	2.50%	120	3	0.375	0.375
5	40	1	2.50%	160	4	0.500	0.500
4	40	1	2.50%	200	5	0.625	0.625
3	40	1	2.50%	240	6	0.750	0.750
2	40	1	2.50%	280	7	0.875	0.875
1	40	1	2.50%	320	8	1.000	1.000

Figure 5.4: Data set (random model) for *CAP* approach, source: compiled by author

from Figures 5.2, 5.3 and 5.4 will be used. The x-axis contains the cumulative number of borrowers and the y-axis contains the corresponding cumulative number of defaults. The result of this plot is the *CAP* curve for actual, perfect and random model.

When Figure 5.5 is compared with Figure 3.1, it can be observed that the perfect and random models match each other and the actual model lies in between those two, exactly as it was assumed in Subsection 3.1. The goal is to generate the probability of default with help of *CAP* curve. As already shown in Subsection 3.7, the key parameter in this approach is the concavity. At this point, it makes sense to repeat the function that was introduced by Burgt. This function can be used in order to estimate the concavity parameter $k \in \mathbb{R}$

$$y(x) = \frac{1 - e^{-kx}}{1 - e^{-k}}, \quad x \in [0, 1], \quad (5.1)$$

since x and y are known. One way to calculate the concavity parameter is by minimizing root mean square error. The results are shown in Figure 5.6. Here x_{measured} and y_{measured} are the values from the actual model, that were shown in Figure 5.2 (the variables x and y). For the error variable, the function from Burgt is used, and it is calculated as

$$\begin{aligned} \text{error} &= y_{\text{measured}} - y_{\text{predicted}} \\ (5.1) \quad &= y_{\text{measured}} - \frac{1 - e^{-kx_{\text{measured}}}}{1 - e^{-k}}. \end{aligned}$$

Squared error is self-explanatory. Root mean square error is calculated as

$$\text{RMSE} = \sqrt{\frac{1}{8} \sum_{i=1}^8 (\text{squared error})}.$$

Concavity parameter k was calculated by minimizing RMSE across all eight rating grades. This was done in Excel by using the solver function. This calculation also matches the one from (3.20). The plot of y_{measured} (this is y variable from Figure 5.2) and $y_{\text{predicted}}$ (calculated with (5.1) as implied in calculation of error variable) is shown in Figure 5.7.

Another way to calculate the concavity parameter would be to use the approximation rule (3.16) from Theorem 3.7.3. AR can be easily calculated with the trapezoidal rule of integration. Figure 5.8 illustrates the trapezoidal rule. For a given a set of points

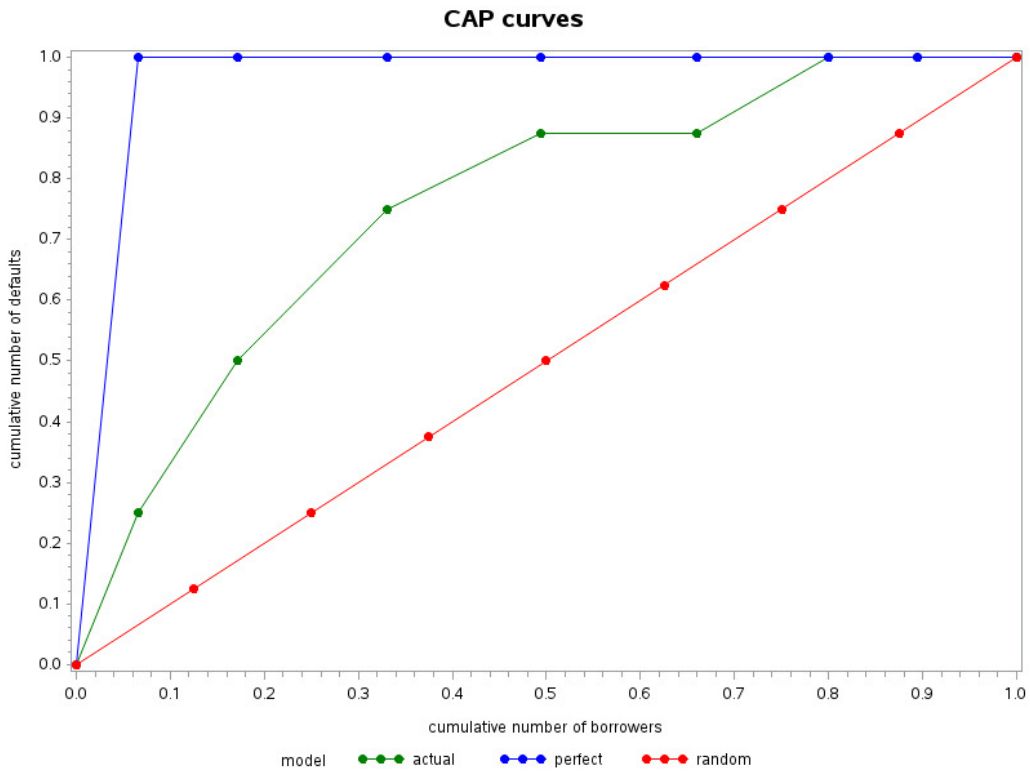


Figure 5.5: CAP curves, source: compiled by author

rating class	y measured	x measured	y predicted	error	squared error
8	0.250	0.066	0.232	0.018	0.000
7	0.500	0.172	0.500	0.000	0.000
6	0.750	0.331	0.741	0.009	0.000
5	0.875	0.494	0.873	0.002	0.000
4	0.875	0.659	0.943	-0.068	0.005
3	1.000	0.800	0.976	0.024	0.001
2	1.000	0.894	0.990	0.010	0.000
1	1.000	1.000	1.000	0.000	0.000
k=					3.912
RMSE =					0

Figure 5.6: Calculation of concavity parameter, source: compiled by author

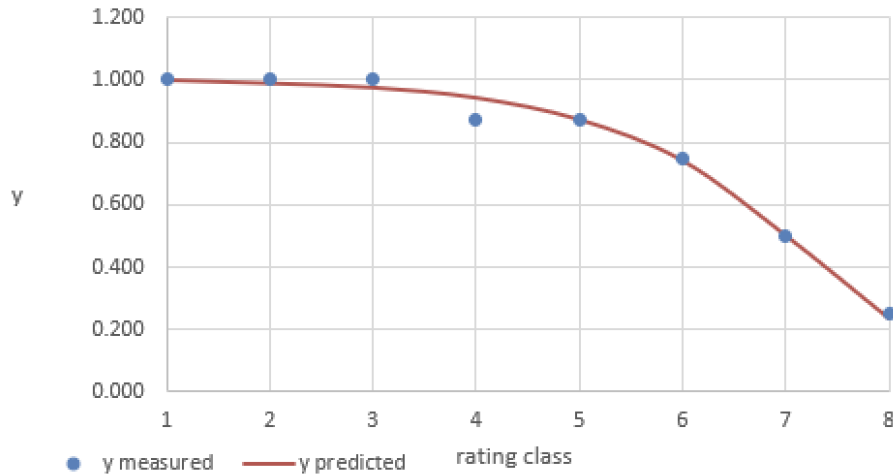


Figure 5.7: y_{measured} vs. $y_{\text{predicted}}$, source: compiled by author

$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, with $x_1 \leq x_2 \leq \dots \leq x_n$, the trapezoidal rule computes the area of the piecewise linear curve that passes through the points. The area under the piecewise linear segments can be computed by summing the area of the trapezoids A_1, A_2, A_3 , and A_4 . This rule was used to calculate $AUCAP$, $AUCAP_{\text{random}}$ and $AUCAP_{\text{perfect}}$. At this point, the definition of AR calculation is repeated and completed with further computation that was done with the above mentioned trapezoidal rule of integration

$$\begin{aligned}
 AR &= \frac{a_R}{a_P} \\
 &= \frac{AUCAP - AUCAP_{\text{random}}}{AUCAP_{\text{perfect}} - AUCAP_{\text{random}}} \\
 &= \frac{0.789 - 0.5}{1 - 0.5} \\
 &= 0.579.
 \end{aligned}$$

The next step is the already mentioned approximation rule (3.16) which can be applied now

$$k \approx \frac{2}{1 - AR} = \frac{2}{1 - 0.579} = 4.751.$$

Since this is only an approximation of the concavity parameter, it is not as exact as the one calculated by minimizing the root mean square error. But it is better than nothing and sometimes it can be useful, since it is easier and therefore quicker way of calculating the concavity parameter. The probability of default will be calculated with both k values. The final step is therefore the calculation of PD , which can be done with help of following equation

$$PD(s) = DR \left(\frac{ke^{-kx_s}}{1 - e^{-k}} \right),$$

where x_s represents the cumulative percentage of borrowers in rating class s and DR is the average observed default rate (= the total number of defaults divided by the total number

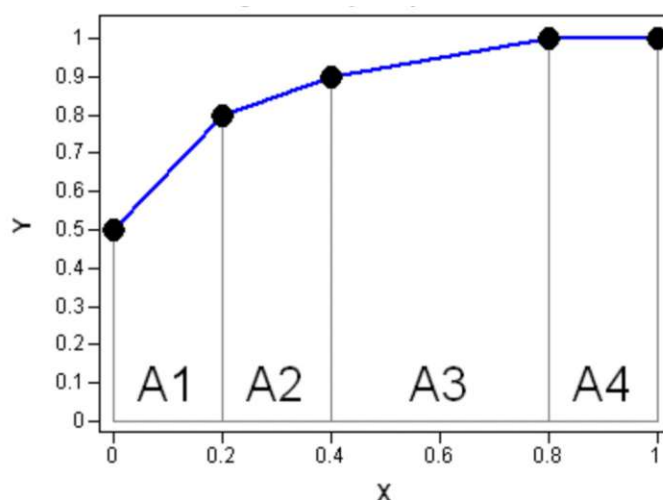


Figure 5.8: Trapezoidal rule, source: [26]

rating class	# borrowers	# defaults	x_s	PD for $k= 3.91$	PD for $k= 4.75$
8	21	2	0.066	7.720%	8.772%
7	34	2	0.119	6.271%	6.815%
6	51	2	0.252	3.730%	3.626%
5	52	1	0.413	1.988%	1.688%
4	53	0	0.577	1.046%	0.774%
3	45	1	0.730	0.575%	0.374%
2	30	0	0.847	0.363%	0.214%
1	34	0	0.947	0.246%	0.133%

Figure 5.9: Resulting PD for *CAP* approach, source: compiled by author

of obligors for the whole portfolio). The value of x_s is calculated as the midpoint between the cumulative percentage of borrowers in rating class s and $s - 1$

$$x_s = \frac{z_N + z_{N-1} + \dots + z_{s-1} + (z_s/2)}{z},$$

where z is the total number of borrowers and z_i is the number of borrowers in the rating class i (whereas the rating class N represents the most risky rating class, so that means rating class 8). Figure 5.9 shows the variable x_s which was used for PD calculation and also the resulting PD for each rating grade for both calculated concavity parameters. Figures 5.10 and 5.11 visualize the behavior of the PD curve.

5.3 Receiver operating characteristic approach

In this approach, the PD will be calculated in a way that was described in Subsection 3.8. Also here, Figure 5.1 is the basis for the further analysis. Total number of not defaults,

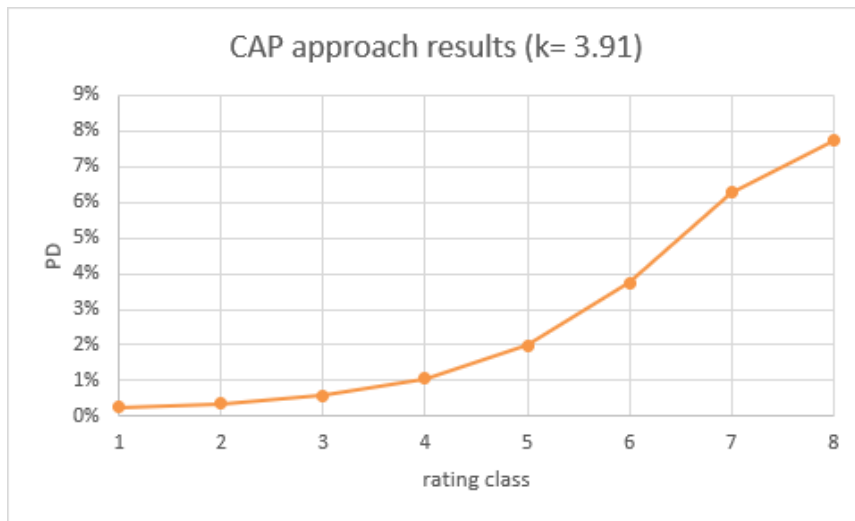


Figure 5.10: Behavior curve of the PD by *CAP* approach for $k=3.91$, source: compiled by author

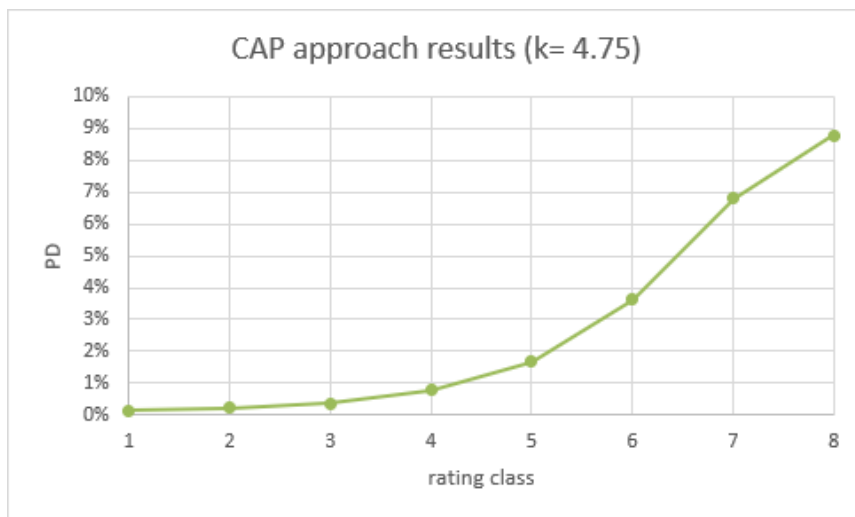


Figure 5.11: Behavior curve of the PD by *CAP* approach for $k=4.75$, source: compiled by author

rating class	# borrowers	# defaults	# not defaults	cumulative defaults	cumulative not defaults	x	y
8	21	2	19	2	19	0.061	0.250
7	34	2	32	4	51	0.163	0.500
6	51	2	49	6	100	0.321	0.750
5	52	1	51	7	151	0.484	0.875
4	53	0	53	7	204	0.654	0.875
3	45	1	44	8	248	0.795	1.000
2	30	0	30	8	278	0.891	1.000
1	34	0	34	8	312	1.000	1.000

Figure 5.12: Data set (actual model) for *ROC* approach, source: compiled by author

rating class	# borrowers	# defaults	# not defaults	cumulative defaults	cumulative not defaults	x	y
8	21	8	13	8	13	0.042	1.000
7	34	0	34	8	47	0.151	1.000
6	51	0	51	8	98	0.314	1.000
5	52	0	52	8	150	0.481	1.000
4	53	0	53	8	203	0.651	1.000
3	45	0	45	8	248	0.795	1.000
2	30	0	30	8	278	0.891	1.000
1	34	0	34	8	312	1.000	1.000

Figure 5.13: Data set (perfect model) for *ROC* approach, source: compiled by author

which is the subtraction of total number of defaults from total number of borrowers in each grade, needs to be added into the data set. The columns cumulative number of borrowers, cumulative number of not defaults need to be added as well and therefore new Figure 5.12 will to be created. The ratio between the cumulative number of not defaults and total number of not defaults in each grade need to be calculated. This will be denoted with x . Also the ratio between the cumulative number of defaults and total number of defaults needs to be determined and also added into Figure 5.12. This second ratio will be indicated with y , which is the same as in *CAP* approach.

Also here the actual model needs to be compared with the perfect and random model. Therefore, the creation of these two models in a same way as in *CAP* approach with the same new variables defined above for the actual model is necessary and is shown in Figures 5.13 and 5.14.

The next step is the construction of a scatter plot. For that, the variables x and y from Figures 5.12, 5.13 and 5.14 will be used. The x-axis contains the false alarm rate, which is the cumulative number of not defaults and the y-axis contains the hit rate, which is the cumulative number of defaults. The result of this plot is the *ROC* curve for actual, perfect and random model.

Also here, Figure 5.15 can be compared with Figure 3.3. It can be observed that the perfect and random models match each other and the actual model lies in between those two, exactly as it was assumed in Subsection 3.4. The goal is to generate the probability of default with help of *ROC* curve. At this point, it makes sense to repeat the equations

rating class	# borrowers	# defaults	# not defaults	cumulative defaults	cumulative not defaults	x	y
8	40	1	39	1	39	0.125	0.125
7	40	1	39	2	78	0.250	0.250
6	40	1	39	3	117	0.375	0.375
5	40	1	39	4	156	0.500	0.500
4	40	1	39	5	195	0.625	0.625
3	40	1	39	6	234	0.750	0.750
2	40	1	39	7	273	0.875	0.875
1	40	1	39	8	312	1.000	1.000

Figure 5.14: Data set (random model) for *ROC* approach, source: compiled by author

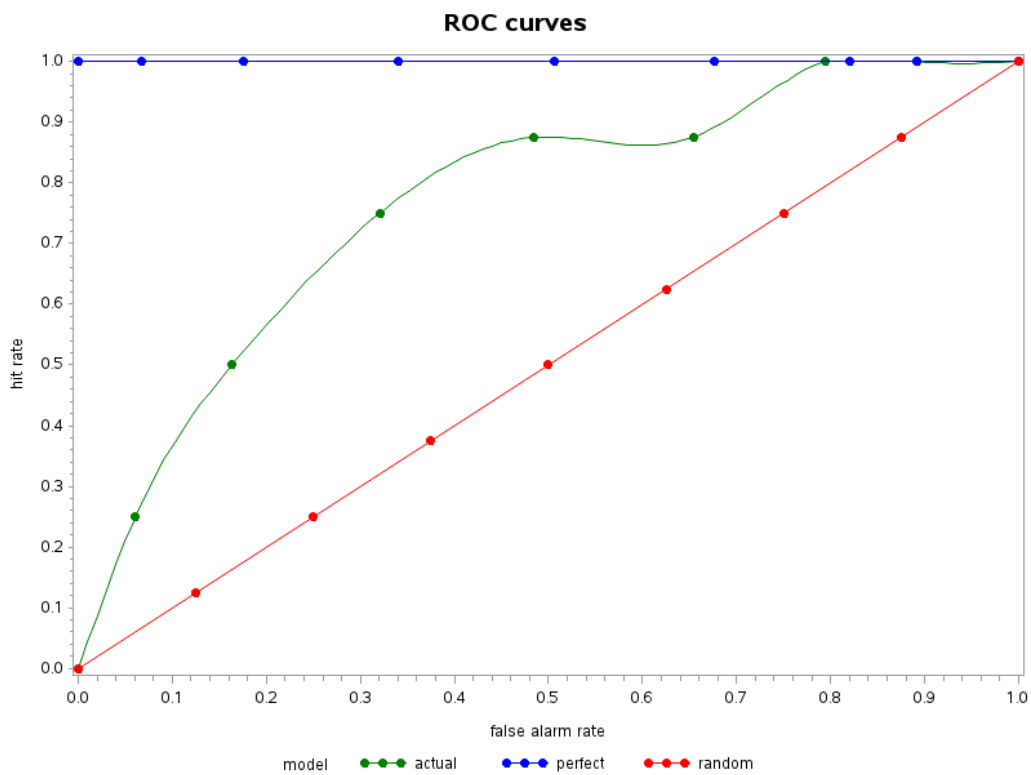


Figure 5.15: *ROC* curves, source: compiled by author

rating class	# borrowers	# defaults	# not defaults	cumulative defaults	cumulative not defaults	x	y	PD	
8	21	2	19	2	19	0.061	0.250	7.74%	
7	34	2	32	4	51	0.163	0.500	3.11%	
6	51	2	49	6	100	0.321	0.750	1.32%	
5	52	1	51	7	151	0.484	0.875	0.65%	
4	53	0	53	7	204	0.654	0.875	0.31%	
3	45	1	44	8	248	0.795	1.000	0.15%	
2	30	0	30	8	278	0.891	1.000	0.08%	
1	34	0	34	8	312	1.000	1.000	0.00%	
								A:	0.770
								a:	1.697

Figure 5.16: Resulting PD for *ROC* approach, source: compiled by author

used in this approach

$$A_a = 2\Phi\left(\frac{a}{\sqrt{2}}\right) - 1, \quad (5.2)$$

$$PD(s) = \frac{1}{1 + \frac{1-DR}{DR} e^{(a\Phi^{-1}(x(s))+a^2/2)}}, \quad (5.3)$$

where $\Phi()$ is a cumulative distribution function for a standard normal random variable, DR is the average observed default rate (= the total number of defaults divided by the total number of obligors for the whole portfolio, so it is the same value for all rating classes) and $x(s)$ is the ratio between the cumulative number of not defaults and total number of not defaults for the rating class s (from Figure 5.16).

With help of these equations, the probability of default can be calculated in the following three steps:

1. Calculate the area under the receiver operating characteristic curve A . This can be easily done with help of trapezoidal rule of integration.
2. Calculate a , with help of (5.2), since A_a is already known. $\Phi()$ can be easily calculated in Excel with „NORMSDIST“ function.
3. Use (5.3) to calculate PD for each rating class. $\Phi^{-1}()$ can be easily calculated in Excel with „NORMSINV“ function.

The results for A , a and also PD for each rating class are shown in Figure 5.16. Figure 5.17 visualizes the behavior of the PD curve for *ROC* approach.

5.4 Bayesian approach

Lastly, the PD will be calculated in a way it was described in Chapter 4. Also here, Figure 5.1 is the basis for the further analysis. At this point, a quick repetition of everything that will be needed for PD calculation is necessary. Probability of default is derived as maximum value of the „posterior distribution“ which is defined as

$$P[\phi | r, e] = \frac{P[r | \phi, e]P[\phi | e]}{P[r | e]},$$

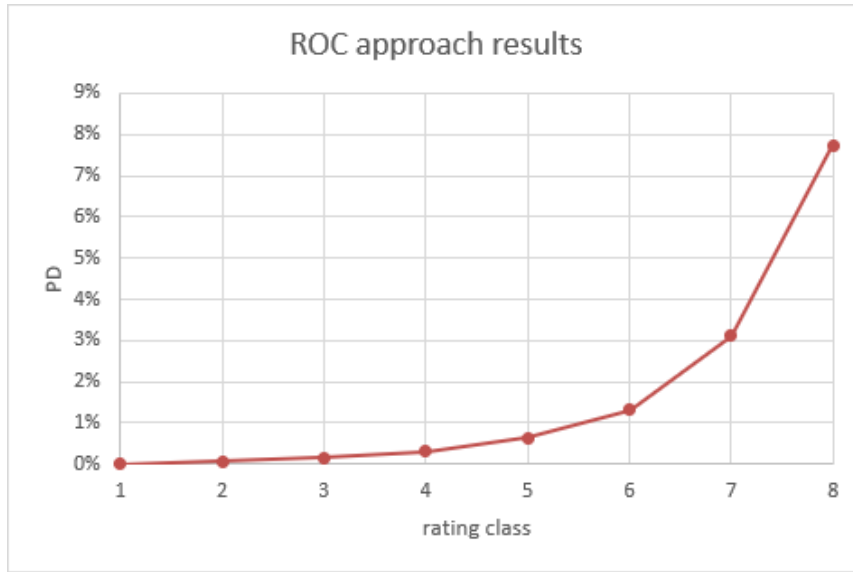


Figure 5.17: Behavior curve of the PD by *ROC* Approach, source: compiled by author

with „likelihood function“

$$P[r | \phi, e] = \binom{n}{r} \phi^r (1 - \phi)^{n-r},$$

„prior distribution“

$$P[\phi | e] = P[\phi | \alpha, \beta] = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \phi^{\alpha-1} (1 - \phi)^{\beta-1},$$

and predictive distribution

$$P[r | e] = \frac{\Gamma(r + \alpha)\Gamma(n - r + \beta)\Gamma(\alpha + \beta)\Gamma(n + 1)}{\Gamma(r + 1)\Gamma(n - r + 1)\Gamma(\alpha)\Gamma(\beta)\Gamma(n + \alpha + \beta)},$$

where ϕ is the probability of default, n is number of borrowers in the whole rating class, r is number of defaults in the whole rating class, $\alpha = n + 1$ and $\beta = n - r$. Therefore, the columns α and β need to be added into Figure 5.1, which will result in Figure 5.18. Firstly, the „likelihood function“ is calculated for all eight rating classes and is shown in Figure 5.19. Secondly, the „prior distribution“ which represents the expert’s assessments is created for all eight rating classes and shown in Figure 5.20. Last step that is needed before the „posterior distribution“ can be constructed, is the predictive function for all eight rating classes. This is shown in Figure 5.21. At this point, the „posterior distribution“ can be produced for all eight rating classes with help of information from Figure 5.19, 5.20 and 5.21. The result is shown in Figure 5.22. The final step is the calculation of PD, which is the value of maximum of the „posterior distribution“ and is shown in Figure 5.23 for each rating class. Figure 5.24 visualizes the behavior of the PD curve.

rating class	# borrowers	# defaults	default rate	alpha	beta
8	21	2	9.52%	3	19
7	34	2	5.88%	3	32
6	51	2	3.92%	3	49
5	52	1	1.92%	2	51
4	53	0	0.00%	1	53
3	45	1	2.22%	2	44
2	30	0	0.00%	1	30
1	34	0	0.00%	1	34

Figure 5.18: Data set for Bayesian approach, source: compiled by author

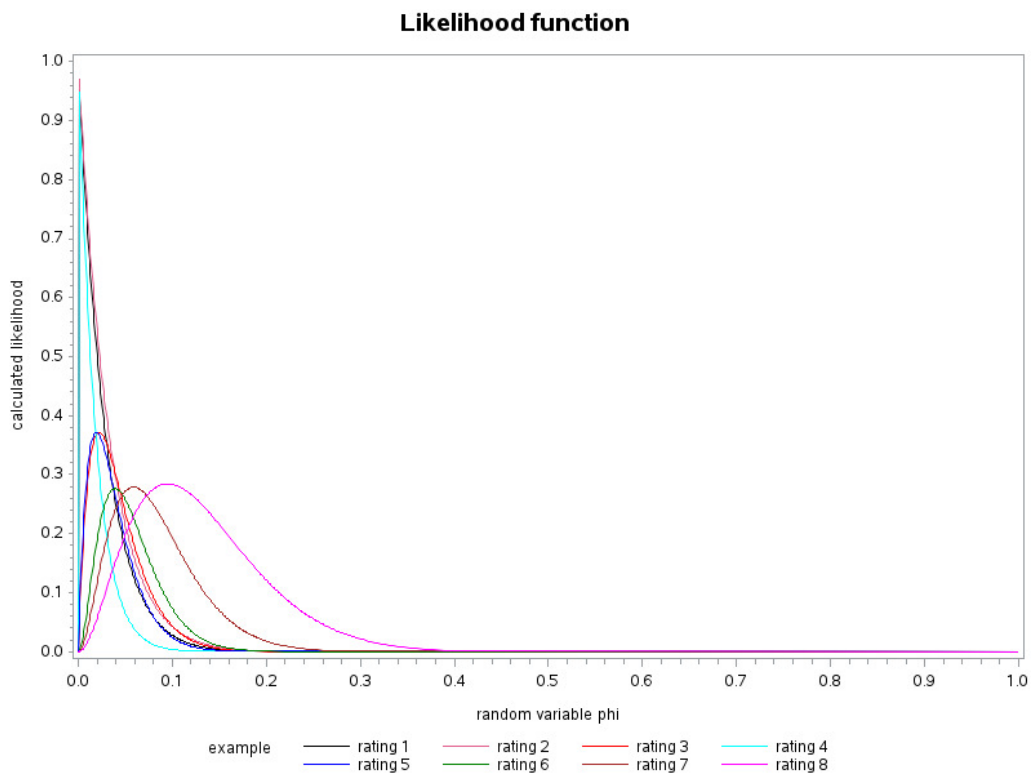


Figure 5.19: „Likelihood function“, source: compiled by author

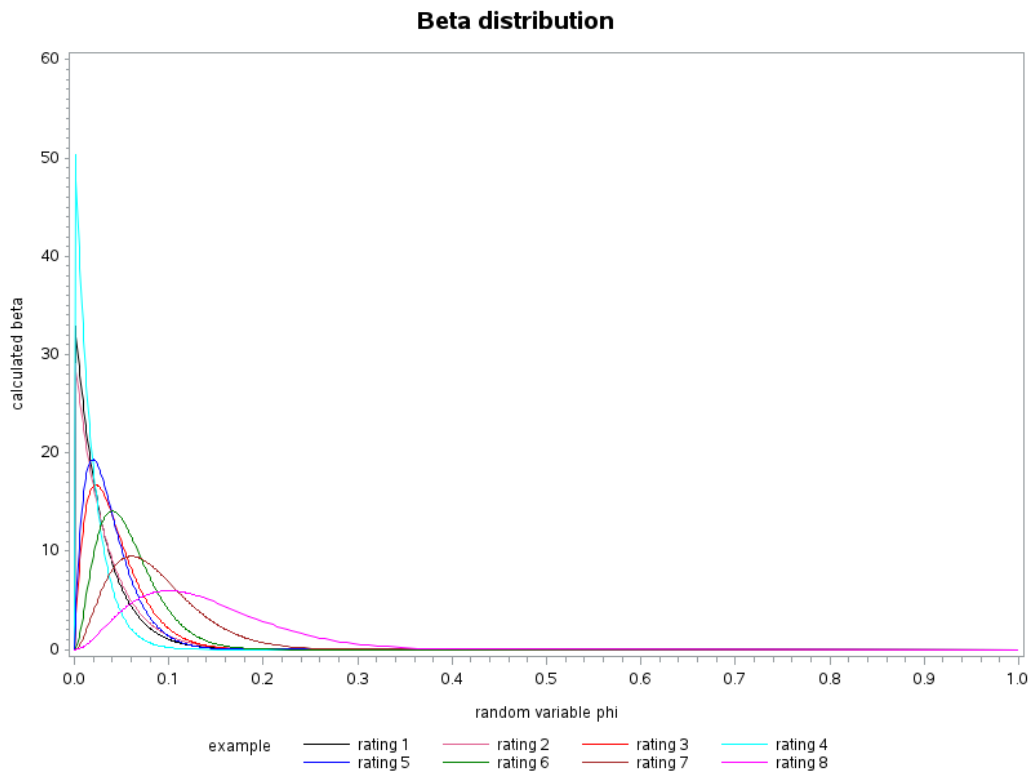


Figure 5.20: „Prior distribution“, source: compiled by author

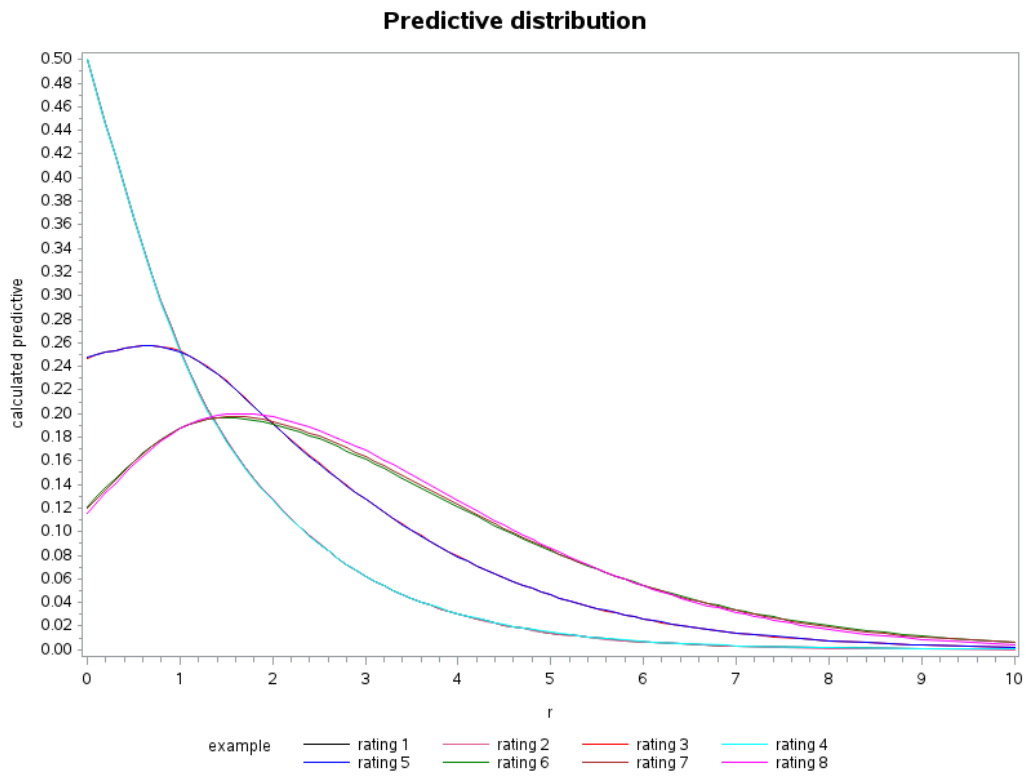


Figure 5.21: Predictive function, source: compiled by author

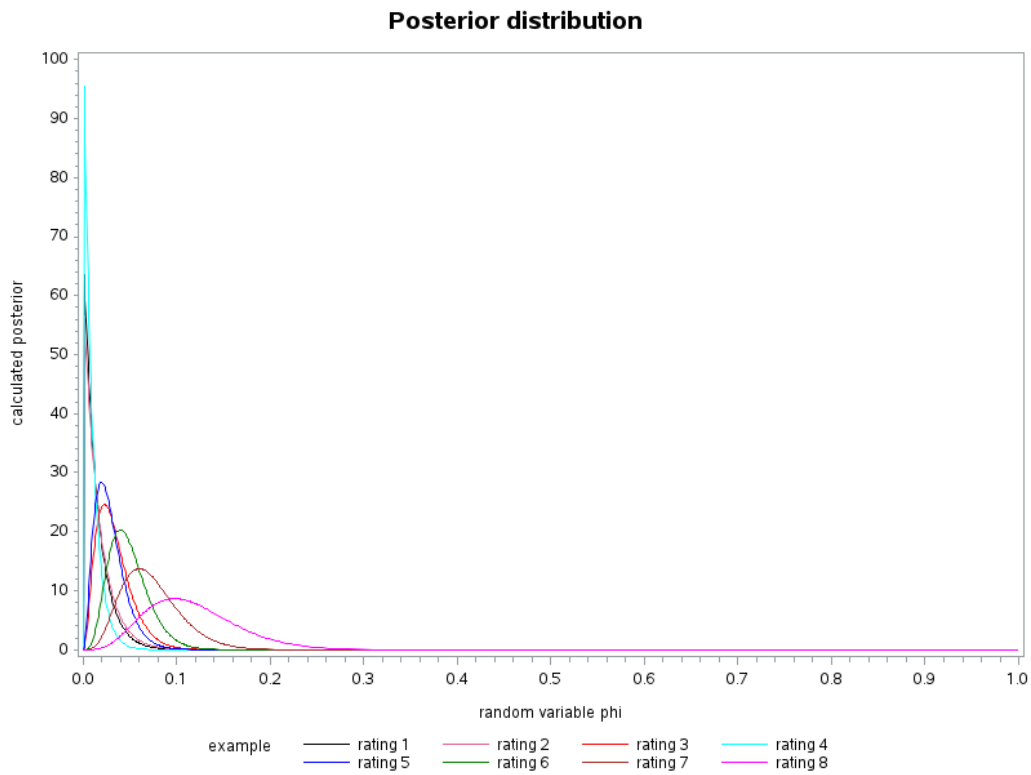


Figure 5.22: „Posterior distribution“, source: compiled by author

rating class	# borrowers	# defaults	max posterior	PD
8	21	2	8.635	9.80%
7	34	2	13.698	6.00%
6	51	2	20.329	4.00%
5	52	1	28.413	1.90%
4	53	0	95.429	0.10%
3	45	1	24.626	2.20%
2	30	0	56.560	0.10%
1	34	0	53.591	0.10%

Figure 5.23: Resulting PD for Bayesian approach, source: compiled by author

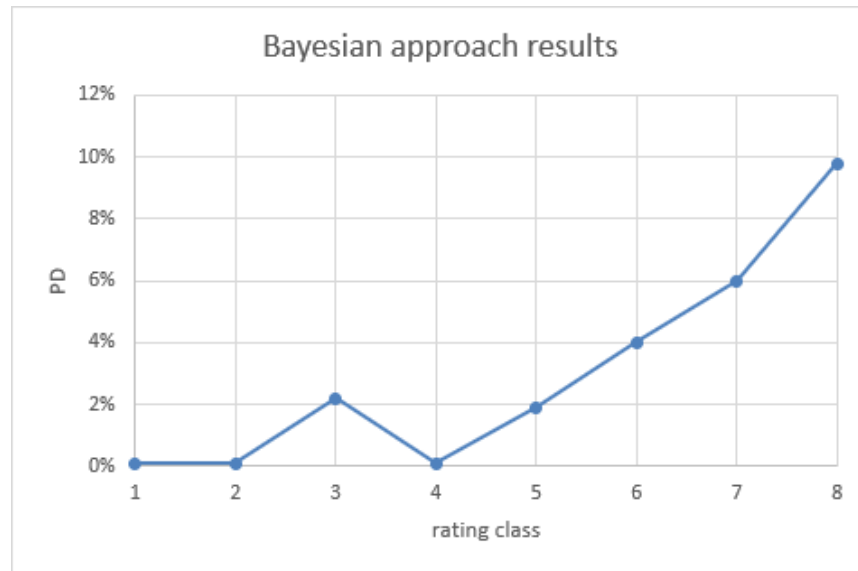


Figure 5.24: Behavior curve of the PD by Bayesian approach, source: compiled by author

rating class	# borrowers	# defaults	default rate	PD CAP	PD ROC	PD Bayes
8	21	2	9.52%	7.72%	7.74%	9.80%
7	34	2	5.88%	6.27%	3.11%	6.00%
6	51	2	3.92%	3.73%	1.32%	4.00%
5	52	1	1.92%	1.99%	0.65%	1.90%
4	53	0	0.00%	1.05%	0.31%	0.10%
3	45	1	2.22%	0.58%	0.15%	2.20%
2	30	0	0.00%	0.36%	0.08%	0.10%
1	34	0	0.00%	0.25%	0.00%	0.10%

Figure 5.25: Summary of PD results, source: compiled by author

5.5 Discussion of results

Figure 5.25 shows the summary of calculated probabilities of default for each rating grade and for each approach. It should be noted that the PD results for the cumulative accuracy profile approach in Figure 5.25, are the ones calculated by minimizing RMSE and not the approximation ones, since these are more accurate.

It can be noticed that the results for the receiver operating characteristic approach are much lower than the results for other two approaches. In this approach, the calculated PD is even lower than the actual default rate (also for rating classes where at least few defaults were observed). This definitely should not be the case and therefore this approach is not recommended by author of this master's thesis for further use.

When looking at the results for Bayesian approach, it can be seen that the calculated PD for the rating classes where at least one default occurred seems appropriate. Unfortunately, the PDs for rating classes where zero default occurred are too low. Another problem in this

approach is, that it does not comply with the monotony requirement, as in this example the PD for rating class number 4 is lower than for the rating class number 3. This can also be seen in Figure 5.24. In a real data sets, the zero default issue could also be the case in the higher rating classes. Therefore, the author does not recommend using this approach to estimate PD for portfolios with no default rating classes.

At first look, the resulting PDs for the cumulative accuracy profile approach comply with the monotony requirement and seem to be very reasonable. The biggest advantage of this approach, is the fact that it does not only use the data information about each rating class but also the data information about the whole data set. This results in reasonable PDs even for rating classes where no default was observed and solves the issue from Bayesian approach. Another advantage of this approach is the existence of the approximation rule. This could be used for a moments when the PD should be calculated a little bit faster. In Figure 5.9, it can be seen that the approximation rule underestimates the PD for the first six rating classes and overestimates the PD for the two worst rating classes. Specifically the overestimation in the worst two rating classes could sometimes even be the desired outcome. The disadvantage of this approach is the fact, that it does not work for zero default portfolios since the default rate of the whole portfolio is used to estimate the PD for each rating class. If the default rate of the whole portfolio would be zero, the resulting estimated PD for each rating class would also be zero. In this case, for example the most prudent estimator approach by Tasche that is described in [16] could be used. For low default (not zero default) portfolios, author of this thesis recommends the usage of *CAP* approach for PD estimation.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

6 Summary and conclusion

The purpose of this thesis was to assess approaches for estimating probability of default in low default portfolios. To achieve this aim, data set was created, PD was calculated for each rating class with each approach and results were compared. The approaches for PD estimation studied in this thesis were the cumulative accuracy profile approach, receiver operating characteristic approach and Bayesian approach. The conclusion of this thesis is that estimation of PD in low default portfolios is a complex problem with many issues, but if one of these three approaches discussed in this master's thesis, should be used to estimate probability of default for low default portfolios, the cumulative accuracy profile approach works the best. An advantage of this approach is the fact, that it does not only consider data information of a specific rating class to calculate the probability of default, but it uses the data information from whole portfolio. A disadvantage of this approach is, that because it is based on modeling of the cumulative accuracy profile curve, it does not work when there are zero defaults observed in the whole portfolio. The second best approach is the Bayesian approach, which works comparably good to the cumulative accuracy profile one, for the rating classes with at least one default. Since it doesn't use the data information from whole portfolio, the calculated probability of default for rating classes without any single default is too low. The third approach, which is the receiver operating characteristic approach, is not recommended by author for further usage. The results from this approach were too low for almost every rating class.

There are several avenues how this work could be extended. The first of which, would be the inclusion of one important topic that was left out completely in this thesis. And that is the dependence of defaults. This was for example discussed in [4, 8]. The second extension would be to compare the results of these three models to other models, for example the ones mentioned in Subsection 2.4.1. Another extension would be to include some models that work in zero defaults case, which is not the case in the three mentioned approaches of this thesis. In this case, the principle of the most prudent estimator described in [16] seems appropriate.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

7 Appendix: SAS Codes

In this chapter, the SAS codes that were used to get results for the Chapter 5 can be found. All codes were written and run under SAS version 9.4. It is always commented what the code does.

```
/* create low default table, perfect model table and  
random model table on which the analysis will be based */
```

```
proc sql;  
create table low_default  
(rating num,  
borrowers num,  
defaults num);  
insert into low_default  
values(1,34,0)  
values(2,30,0)  
values(3,45,1)  
values(4,53,0)  
values(5,52,1)  
values(6,51,2)  
values(7,34,2)  
values(8,21,2);  
values(9,0,0);  
quit;
```

```
proc sql;  
create table perfect_model  
(rating num,  
borrowers num,  
defaults num);  
insert into perfect_model  
values(1,34,0)  
values(2,30,0)  
values(3,45,0)  
values(4,53,0)  
values(5,52,0)  
values(6,51,0)  
values(7,34,0)  
values(8,21,8);
```

```
values(9,0,0) ; /* for CAP curve */
/*values(9,8,8); this is for ROC curve */
quit;

proc sql;
create table random_model
(rating num,
borrowers num,
defaults num);
insert into random_model
values(1,40,1)
values(2,40,1)
values(3,40,1)
values(4,40,1)
values(5,40,1)
values(6,40,1)
values(7,40,1)
values(8,40,1);
values(9,0,0);
quit;

/* create all_defaults and all_borrowers variable */

proc sql;
create table low_default as
select *, sum(borrowers) as all_borrowers, sum(defaults)
as all_defaults from low_default;
quit;

proc sql;
create table perfect_model as
select *, sum(borrowers) as all_borrowers, sum(defaults)
as all_defaults from perfect_model;
quit;

proc sql;
create table random_model as
select *, sum(borrowers) as all_borrowers, sum(defaults)
as all_defaults from random_model;
quit;

/* sort the tables from worst rating class to best */

proc sort data=low_default;
by descending rating;
```

```

run;

proc sort data=perfect_model;
by descending rating;
run;

proc sort data=random_model;
by descending rating;
run;

/* create default rates and cumulative sum */

data low_default;
set low_default;
DR = defaults/borrowers;
all_DR= all_defaults/all_borrowers;
retain cum_borrowers;
if _n_ = 1 then cum_borrowers=borrowers;
else cum_borrowers= cum_borrowers+borrowers;
retain cum_defaults;
if _n_ = 1 then cum_defaults=defaults;
else cum_defaults= cum_defaults+defaults;
run;

data perfect_model;
set perfect_model;
DR = defaults/borrowers;
all_DR= all_defaults/all_borrowers;
retain cum_borrowers;
if _n_ = 1 then cum_borrowers=borrowers;
else cum_borrowers= cum_borrowers+borrowers;
retain cum_defaults;
if _n_ = 1 then cum_defaults=defaults;
else cum_defaults= cum_defaults+defaults;
run;

data random_model;
set random_model;
DR = defaults/borrowers;
all_DR= all_defaults/all_borrowers;
retain cum_borrowers;
if _n_ = 1 then cum_borrowers=borrowers;
else cum_borrowers= cum_borrowers+borrowers;
retain cum_defaults;
if _n_ = 1 then cum_defaults=defaults;

```

```
else cum_defaults= cum_defaults+defaults;
run;

/* create x and y variables for the plot */

proc sql;
create table low_default as
select *, cum_borrowers/all_borrowers as x, cum_defaults/all_defaults
as y, 'actual ' as model from low_default;
quit;

proc sql;
create table perfect_model as
select *, cum_borrowers/all_borrowers as x, cum_defaults/all_defaults
as y, 'perfect' as model from perfect_model;
quit;

proc sql;
create table random_model as
select *, cum_borrowers/all_borrowers as x, cum_defaults/all_defaults
as y, 'random ' as model from random_model;
quit;

data all;
set low_default perfect_model random_model;
run;

/* create CAP plot */

proc gplot data=all;
title1 'CAP curves';
plot y*x=model /vaxis=axis1 haxis=axis2;
symbol1 color=green interpol=join value=dot height=1;
symbol2 color= blue interpol=join value=dot height=1;
symbol3 color=red interpol=join value=dot height=1;
axis1 label=(angle=90 rotate = 0'cumulative number of defaults'
justify=center);
axis2 label=('cumulative number of borrowers' justify=center);
run;
quit;

/* calculate area under CAP curve */

data Trapezoidal_c1;
set low_default end=last;
```

```
retain integral;
lag_x=lag(x); lag_y = lag(y);
if _N_ eq 1 then integral = 0;
else integral = integral + (x - lag_x) * (y + lag_y) / 2;
if last;
drop lag_;;
run;

data Trapezoidal_c2;
set perfect_model end=last;
retain integral;
lag_x=lag(x); lag_y = lag(y);
if _N_ eq 1 then integral = 0;
else integral = integral + (x - lag_x) * (y + lag_y) / 2;
if last;
drop lag_;;
run;

data Trapezoidal_c3;
set random_model end=last;
retain integral;
lag_x=lag(x); lag_y = lag(y);
if _N_ eq 1 then integral = 0;
else integral = integral + (x - lag_x) * (y + lag_y) / 2;
if last;
drop lag_;;
run;

/* ROC approach, create non_defaults variable */

data low_default;
set low_default;
not_defaults= borrowers-defaults;
run;

data perfect_model;
set perfect_model;
not_defaults= borrowers-defaults;
run;

data random_model;
set random_model;
not_defaults= borrowers-defaults;
run;
```

```
/* create cumulative sum for non_defaults */

data low_default;
set low_default;
retain cum_non_d;
if _n_ = 1 then cum_non_d=not_defaults;
else cum_non_d= cum_non_d+not_defaults;
run;

data perfect_model;
set perfect_model;
retain cum_non_d;
if _n_ = 1 then cum_non_d=not_defaults;
else cum_non_d= cum_non_d+not_defaults;
run;

data random_model;
set random_model;
retain cum_non_d;
if _n_ = 1 then cum_non_d=not_defaults;
else cum_non_d= cum_non_d+not_defaults;
run;

/* create all_non_d variable */

proc sql;
create table low_default as
select *, sum(not_defaults) as all_non_d from low_default;
quit;

proc sql;
create table perfect_model as
select *, sum(not_defaults) as all_non_d from perfect_model;
quit;

proc sql;
create table random_model as
select *, sum(not_defaults) as all_non_d from random_model;
quit;

/* create x variable for roc plot, y stays the same as in cap */

proc sql;
create table low_default as
select *, cum_non_d/all_non_d as x_2 from low_default;
```

```

quit;

proc sql;
create table perfect_model as
select *, cum_non_d/all_non_d as x_2 from perfect_model;
quit;

proc sql;
create table random_model as
select *, cum_non_d/all_non_d as x_2 from random_model;
quit;

data all;
set low_default perfect_model random_model;
run;

/* create ROC plot */

proc gplot data=all;
title1 'ROC curves';
plot y*x_2=model /vaxis=axis1 haxis=axis2;
symbol1 color=green interpol=spline height=1;
symbol2 color= blue interpol=join height=1;
symbol3 color=red interpol=spline height=1;
axis1 label=(angle=90 rotate = 0'hit rate' justify=center);
axis2 label=('false alarm rate' justify=center);
run;
quit;

/* calculate area under ROC curve */

data Trapezoidal_r1;
    set low_default end=last;
    retain integral;
    lag_x=lag(x_2); lag_y = lag(y);
    if _N_ eq 1 then integral = 0;
    else integral = integral + (x_2 - lag_x) * (y + lag_y) / 2;
    if last;
    drop lag_;;
run;

data Trapezoidal_r2;
    set perfect_model end=last;
    retain integral;

```

```
lag_x=lag(x_2); lag_y = lag(y);
if _N_ eq 1 then integral = 0;
else integral = integral + (x_2 - lag_x) * (y + lag_y) / 2;
if last;
drop lag_;;
run;

data Trapezoidal_r3;
set random_model end=last;
retain integral;
lag_x=lag(x_2); lag_y = lag(y);
if _N_ eq 1 then integral = 0;
else integral = integral + (x_2 - lag_x) * (y + lag_y) / 2;
if last;
drop lag_;;
run;

/* another way to calculate A and create ROC plot */

ods graphics on;
proc logistic data=low_default plots(only)=roc;
ods select ROCcurve;
model defaults/not_defaults=rating /outroc=roc;
run;

/* creating figures 4.2, 4.3, 4.4 and 4.5 for chapter 4 */

data start (drop = i);
x = -1;
do i = 1 to 1000;
x + 1;
output;
end;
run;

/* preparation likelihood macro that can be used here but also
in chapter 5 */

%macro likelihood(sample,r,n,example);
data &sample.;
set start;
x_2 = x/1000;
example = &example.;
r= &r.;
n= &n.;
```

```
run;

proc sql;
create table &sample. as
select *, (fact(n)/(fact(r)*fact(n-r)))*(x_2**r)*((1-x_2)**(n-r))
as y_2 from &sample.;
quit;

data &sample.;
set &sample.;
if y_2 eq . then y = 0;
else y=y_2;
run;

data &sample.;
set &sample.;
drop y_2;
rename y = y_2;
run;
%mend;

/* preparation figure 4.2 */

%likelihood(table_4_2_1,0,50,'r=0');
%likelihood(table_4_2_2,1,50,'r=1');
%likelihood(table_4_2_3,2,50,'r=2');

data table_4_2;
set table_4_2_1 table_4_2_2 table_4_2_3;
run;

/* create figure 4.2 */

proc gplot data=table_4_2;
title1 'Likelihood function';
plot y_2*x_2=example /vaxis=axis1 haxis=axis2;
symbol1 color=green interpol=spline height=1;
symbol2 color= blue interpol=spline height=1;
symbol3 color=red interpol=spline height=1;
axis1 label=(angle=90 rotate = 0'calculated likelihood'
justify=center);
axis2 label=('random variable phi' justify=center);
run;
quit;
```

```
/* preparation beta macro that can be used here but also
in chapter 5 */

%macro beta(sample,a,b,example);
data &sample.;
set start;
x_2 = x/1000;
example = &example.;
a=&a.;
b=&b.;
run;

proc sql;
create table &sample. as
select *,
(GAMMA(a+b)/(GAMMA(a)*GAMMA(b)))*((x_2)**(a-1))*((1-(x_2))**(b-1))
as y_3 from &sample.;
quit;

data &sample.;
set &sample.;
if y_3 eq . then y = 0;
else y=y_3;
run;

data &sample.;
set &sample.;
drop y_3;
rename y = y_3;
run;
%mend;

/* preparation figure 4.3 */

%beta(table_4_3_1,2,5,'alpha=2,beta=5');
%beta(table_4_3_2,3,1,'alpha=3,beta=1');
%beta(table_4_3_3,1,1,'alpha=1,beta=1');

data table_4_3;
set table_4_3_1 table_4_3_2 table_4_3_3;
run;

/* create figure 4.3 */

proc gplot data=table_4_3;
```

```

title1 'Beta distribution';
plot y_3*x_2=example /vaxis=axis1 haxis=axis2;
symbol1 color=green interpol=spline height=1;
symbol2 color= blue interpol=spline height=1;
symbol3 color=red interpol=spline height=1;
axis1 label=(angle=90 rotate = 0'calculated beta' justify=center);
axis2 label=('random variable phi' justify=center);
run;
quit;

```

```

/* preparation predictive macro that can be used here but also
in chapter 5 */

```

```

%macro predictive(sample,a,b,n,example);
data &sample. (drop = i);
r = -1;
do i = 1 to 11;
r + 1;
output;
end;
run;

```

```

data &sample.;
set &sample.;
a=&a.;
b=&b.;
n=&n.;
example=&example.;
run;

```

```

proc sql;
create table &sample. as
select *, (GAMMA(r+a)*GAMMA(n-r+b)*GAMMA(a+b)*GAMMA(n+1))/
(GAMMA(r+1)*GAMMA(n-r+1)*GAMMA(a)*GAMMA(b)*GAMMA(n+a+b))
as y_5 from &sample.;
quit;
%mend;

```

```

/* preparation figure 4.5 */

```

```

%predictive(table_4_5_1,2,5,50,'alpha=2,beta=5');
%predictive(table_4_5_2,1,1,50,'alpha=1,beta=1');

```

```

data table_4_5;
set table_4_5_1 table_4_5_2;

```

```
run;

/* create figure 4.5 */

proc gplot data=table_4_5;
title1 'Predictive distribution';
plot y_5*r=example /vaxis=axis1 haxis=axis2;
symbol1 color=green interpol=spline height=1;
symbol2 color= blue interpol=spline height=1;
axis1 label=(angle=90 rotate = 0'calculated predictive' justify=center);
axis2 label=('r' justify=center);
run;
quit;

/* preparation posterior macro that can be used here but also
in chapter 5 */

%macro posterior(in,out,like,beta,pred,a,b,r,n,example);
data &in.;
set start;
x_2 = x/1000;
example = &example.;
a=&a.;
b=&b.;
r= &r.;
n= &n.;
run;

proc sql;
create table &out. as
select a.*, b.y_2
from &in. a
left join &like. b
on a.x_2=b.x_2;
quit;

proc sql;
create table &out. as
select a.*, b.y_3
from &out. a
left join &beta. b
on a.x_2=b.x_2;
quit;

proc sql;
```

```
create table &out. as
select a.*, b.y_5
from &out. a
left join &pred. b
on b.r=&r.;
quit;

proc sql;
create table &out. as
select *, (y_2*y_3)/y_5 as y_4 from &out.;
quit;
%mend;

/* preparation figure 4.4 */

%posterior(table_4_4_1,table_4_4_1_r0,table_4_2_1,table_4_3_1,
table_4_5_1,2,5,0,50,'r=0');
%posterior(table_4_4_2,table_4_4_2_r1,table_4_2_2,table_4_3_1,
table_4_5_1,2,5,1,50,'r=1');
%posterior(table_4_4_3,table_4_4_3_r2,table_4_2_3,table_4_3_1,
table_4_5_1,2,5,2,50,'r=2');

data table_4_4;
set table_4_4_1_r0 table_4_4_2_r1 table_4_4_3_r2;
run;

/* create figure 4.4 */

proc gplot data=table_4_4;
title1 'Posterior distribution';
plot y_4*x_2=example /vaxis=axis1 haxis=axis2;
symbol1 color=green interpol=spline height=1;
symbol2 color= blue interpol=spline height=1;
axis1 label=(angle=90 rotate = 0'calculated posterior' justify=center);
axis2 label=('random variable phi' justify=center);
run;
quit;

/* bayesian approach chapter 5 */

/* using likelihood macro for all 8 rating grades */

%likelihood(likelihood_rating1,0,34,'rating 1');
%likelihood(likelihood_rating2,0,30,'rating 2');
%likelihood(likelihood_rating3,1,45,'rating 3');
```

```

%likelihood(likelihood_rating4,0,53,'rating 4');
%likelihood(likelihood_rating5,1,52,'rating 5');
%likelihood(likelihood_rating6,2,51,'rating 6');
%likelihood(likelihood_rating7,2,34,'rating 7');
%likelihood(likelihood_rating8,2,21,'rating 8');

data likelihood;
set likelihood_rating1 likelihood_rating2 likelihood_rating3
likelihood_rating4 likelihood_rating5 likelihood_rating6
likelihood_rating7 likelihood_rating8;
run;

/* create likelihood plot */

proc gplot data=likelihood;
title1 'Likelihood function';
plot y_2*x_2=example /vaxis=axis1 haxis=axis2;
symbol1 color= black interpol=spline height=1;
symbol2 color= lilac interpol=spline height=1;
symbol3 color=red interpol=spline height=1;
symbol4 color= cyan interpol=spline height=1;
symbol5 color= blue interpol=spline height=1;
symbol6 color= green interpol=spline height=1;
symbol7 color= brown interpol=spline height=1;
symbol8 color= magenta interpol=spline height=1;
axis1 label=(angle=90 rotate = 0'calculated likelihood'
justify=center);
axis2 label=('random variable phi' justify=center);
run;
quit;

/* using beta macro for all 8 rating grades */

%beta(beta_rating1,1,34,'rating 1');
%beta(beta_rating2,1,30,'rating 2');
%beta(beta_rating3,2,44,'rating 3');
%beta(beta_rating4,1,53,'rating 4');
%beta(beta_rating5,2,51,'rating 5');
%beta(beta_rating6,3,49,'rating 6');
%beta(beta_rating7,3,32,'rating 7');
%beta(beta_rating8,3,19,'rating 8');

data beta;
set beta_rating1 beta_rating2 beta_rating3 beta_rating4
beta_rating5 beta_rating6 beta_rating7 beta_rating8;

```

```
run;

/* create beta plot */

proc gplot data=beta;
title1 'Beta distribution';
plot y_3*x_2=example /vaxis=axis1 haxis=axis2;
symbol1 color= black interpol=spline height=1;
symbol2 color= lilac interpol=spline height=1;
symbol3 color=red interpol=spline height=1;
symbol4 color= cyan interpol=spline height=1;
symbol5 color= blue interpol=spline height=1;
symbol6 color= green interpol=spline height=1;
symbol7 color= brown interpol=spline height=1;
symbol8 color= magenta interpol=spline height=1;
axis1 label=(angle=90 rotate = 0'calculated beta' justify=center);
axis2 label=('random variable phi' justify=center);
run;
quit;

/* using predictive macro for all 8 rating grades */

%predictive(predictive_rating1,1,34,34,'rating 1');
%predictive(predictive_rating2,1,30,30,'rating 2');
%predictive(predictive_rating3,2,44,45,'rating 3');
%predictive(predictive_rating4,1,53,53,'rating 4');
%predictive(predictive_rating5,2,51,52,'rating 5');
%predictive(predictive_rating6,3,49,51,'rating 6');
%predictive(predictive_rating7,3,32,34,'rating 7');
%predictive(predictive_rating8,3,19,21,'rating 8');

data predictive;
set predictive_rating1 predictive_rating2 predictive_rating3
predictive_rating4 predictive_rating5 predictive_rating6
predictive_rating7 predictive_rating8;
run;

/* create predictive plot */

proc gplot data=predictive;
title1 'Predictive distribution';
plot y_5*r=example /vaxis=axis1 haxis=axis2;
symbol1 color= black interpol=spline height=1;
symbol2 color= lilac interpol=spline height=1;
```

```

symbol3 color=red interpol=spline height=1;
symbol4 color= cyan interpol=spline height=1;
symbol5 color= blue interpol=spline height=1;
symbol6 color= green interpol=spline height=1;
symbol7 color= brown interpol=spline height=1;
symbol8 color= magenta interpol=spline height=1;
axis1 label=(angle=90 rotate = 0'calculated predictive' justify=center);
axis2 label=('r' justify=center);
run;
quit;

```

```
/* using posterior macro for all 8 rating grades */
```

```

%posterior(posterior_r1,posterior_rating1,Likelihood_rating1,
Beta_rating1,Predictive_rating1,1,34,0,34,'rating 1');
%posterior(posterior_r2,posterior_rating2,Likelihood_rating2,
Beta_rating2,Predictive_rating2,1,30,0,30,'rating 2');
%posterior(posterior_r3,posterior_rating3,Likelihood_rating3,
Beta_rating3,Predictive_rating3,2,44,1,45,'rating 3');
%posterior(posterior_r4,posterior_rating4,Likelihood_rating4,
Beta_rating4,Predictive_rating4,1,53,0,53,'rating 4');
%posterior(posterior_r5,posterior_rating5,Likelihood_rating5,
Beta_rating5,Predictive_rating5,2,51,1,52,'rating 5');
%posterior(posterior_r6,posterior_rating6,Likelihood_rating6,
Beta_rating6,Predictive_rating6,3,49,2,51,'rating 6');
%posterior(posterior_r7,posterior_rating7,Likelihood_rating7,
Beta_rating7,Predictive_rating7,3,32,2,34,'rating 7');
%posterior(posterior_r8,posterior_rating8,Likelihood_rating8,
Beta_rating8,Predictive_rating8,3,19,2,21,'rating 8');

```

```

data posterior;
set posterior_rating1 posterior_rating2 posterior_rating3
posterior_rating4 posterior_rating5 posterior_rating6
posterior_rating7 posterior_rating8;
run;

```

```
/* create posterior plot */
```

```

proc gplot data=posterior;
title1 'Posterior distribution';
plot y_4*x_2=example /vaxis=axis1 haxis=axis2;
symbol1 color= black interpol=spline height=1;
symbol2 color= lilac interpol=spline height=1;
symbol3 color=red interpol=spline height=1;
symbol4 color= cyan interpol=spline height=1;

```

```
symbol5 color= blue interpol=spline height=1;
symbol6 color= green interpol=spline height=1;
symbol7 color= brown interpol=spline height=1;
symbol8 color= magenta interpol=spline height=1;
axis1 label=(angle=90 rotate = 0'calculated posterior' justify=center);
axis2 label=('random variable phi' justify=center);
run;
quit;
```

```
/* create PD as a maximum */
```

```
proc sql;
create table PD as
select *, max(y_4) as max_y4 from posterior
group by example;
quit;
```

```
data PD;
set PD;
where y_4 eq max_y4;
PD = x_2*100;
run;
```



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

List of Figures

2.1	Timeline of bank regulations, source: [1, p. 67]	7
2.2	Three pillars of Basel II Accord, source: [1, p. 72]	9
3.1	Cumulative accuracy profile, source: [5]	15
3.2	Distribution of rating scores for defaulting and non-defaulting borrowers, source: [5]	19
3.3	Receiver operating characteristic, source: [18]	20
3.4	Comparison of AUCAP and its approximation, source: [24]	25
3.5	Score vs. alarm rate, source: [23]	28
4.1	Bayes' theorem, source: compiled by author	32
4.2	„Likelihood function“, source: compiled by author	34
4.3	Two-parameter beta distribution, source: compiled by author	36
4.4	Predictive distribution, source: compiled by author	38
4.5	„Posterior distribution“, source: compiled by author	39
5.1	Data set (actual model), source: compiled by author	42
5.2	Data set (actual model) for <i>CAP</i> approach, source: compiled by author	42
5.3	Data set (perfect model) for <i>CAP</i> approach, source: compiled by author	42
5.4	Data set (random model) for <i>CAP</i> approach, source: compiled by author	43
5.5	<i>CAP</i> curves, source: compiled by author	44
5.6	Calculation of concavity parameter, source: compiled by author	44
5.7	y_{measured} vs. $y_{\text{predicted}}$, source: compiled by author	45
5.8	Trapezoidal rule, source: [26]	46
5.9	Resulting PD for <i>CAP</i> approach, source: compiled by author	46
5.10	Behavior curve of the PD by <i>CAP</i> approach for $k=3.91$, source: compiled by author	47
5.11	Behavior curve of the PD by <i>CAP</i> approach for $k=4.75$, source: compiled by author	47
5.12	Data set (actual model) for <i>ROC</i> approach, source: compiled by author	48
5.13	Data set (perfect model) for <i>ROC</i> approach, source: compiled by author	48
5.14	Data set (random model) for <i>ROC</i> approach, source: compiled by author	49
5.15	<i>ROC</i> curves, source: compiled by author	49
5.16	Resulting PD for <i>ROC</i> approach, source: compiled by author	50
5.17	Behavior curve of the PD by <i>ROC</i> Approach, source: compiled by author	51
5.18	Data set for Bayesian approach, source: compiled by author	52
5.19	„Likelihood function“, source: compiled by author	52
5.20	„Prior distribution“, source: compiled by author	53
5.21	Predictive function, source: compiled by author	54

List of Figures

5.22 „Posterior distribution“, source: compiled by author	55
5.23 Resulting PD for Bayesian approach, source: compiled by author	55
5.24 Behavior curve of the PD by Bayesian approach, source: compiled by author	56
5.25 Summary of PD results, source: compiled by author	56

Bibliography

- [1] R. Apostolik, Ch. Donohue, and P. Went. *Foundations of Banking Risk (An Overview of Banking, Banking Risks, and Risk-Based Banking Regulation)*. John Wiley and Sons, 2009.
- [2] N. Benjamin, A. Cathcart, and K. Ryan. Low default portfolios: A proposal for conservative estimation of default probabilities. *Financial Services Authority*, 2006.
- [3] L. D. Broemeling. *Bayesian Inference for Stochastic Processes*. Taylor and Francis Group, 2018.
- [4] D. J. C. Dinis. Low default portfolios—a proposed rule to identify differences between imprudence, conservatism, and exaggeration. *Journal of Financial Risk Management*, 11, 2022.
- [5] B. Engelmann, E. Hayden, and D. Tasche. Measuring the discriminatory power of rating systems. *Series 2: banking and financial supervision*, 2003.
- [6] M. Jacobs and N. M Kiefer. Bayesian approach to default risk: A guide. *CAE Working Paper*, 2010.
- [7] N. M Kiefer. Default estimation for low-default portfolios. *OCC Economics Working Paper*, 2006.
- [8] N. M Kiefer. Default estimation, correlated defaults, and expert information. *Cornell University*, 2008.
- [9] F. H. Knight. *Risk, Uncertainty, and Profit*. Houghton Mifflin Company, 1921.
- [10] J. Lam. *Enterprise Risk Management: From Incentives to Controls*. John Wiley and Sons, 2014.
- [11] A. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, 2005.
- [12] Basel Committee on Banking Supervision. Validation of low-default portfolios in the basel 2 framework. *Bank for International Settlements*, 2005.
- [13] Basel Committee on Banking Supervision. International convergence of capital measurement and capital standards. *Bank for International Settlements*, 2006.
- [14] Basel Committee on Banking Supervision. Basel 3: A global regulatory framework for more resilient banks and banking systems. *Bank for International Settlements*, 2010.

- [15] Basel Committee on Banking Supervision. Basel 3: International framework for liquidity risk measurement, standards and monitoring. *Bank for International Settlements*, 2010.
- [16] K. Pluto and D. Tasche. Estimating probabilities of default for low default portfolios. *The Basel 2 Risk Parameters*, 2006.
- [17] J. A. Rice. *Mathematical Statistics and Data Analysis*. The Thomson Corporation, 2007.
- [18] S. Satchell and W. Xia. Analytic models of roc curve: Application to credit rating model validation. *Quantitative finance research centre*, 2006.
- [19] T. Schuermann and S. Hanson. Estimating probabilities of default. *Federal Reserve Bank of New York Staff Reports*, 2004.
- [20] C. Soanes and A. Stevenson. *Oxford English Dictionary*. Oxford University Press, 2005.
- [21] D. Tasche. Remarks on the monotonicity of default probabilities. *Deutsche Bundesbank*, 2002.
- [22] D. Tasche. *The Analytics of Risk Model Validation*. Academic Press, 2008.
- [23] D. Tasche. Estimating discriminatory power and pd curves when the number of defaults is small. *Lloyds Banking Group*, 2010.
- [24] M. J. Van der Burgt. Calibrating low-default portfolios, using the cumulative accuracy profile. *Journal of Risk Model Validation*, 2008.
- [25] T. Van Gestel and B. Baesens. *Credit Risk Management, Basic Concepts: Financial Risk Components, Rating Analysis, Models, Economic and Regulatory Capital*. Oxford University Press, 2009.
- [26] R. Wicklin. The trapezoidal rule of integration. <https://blogs.sas.com/content/iml/2011/06/01/the-trapezoidal-rule-of-integration.html>, 2011.
- [27] Wikipedia. Bayesian inference. https://en.wikipedia.org/wiki/Bayesian_inference.