



UNIVERSITY OF TWENTE.

Faculty of Behavioural,
Management and Social Sciences



Internal Risk Components Validation: Indicative Benchmarking of Discriminatory Power for LGD Models (Public Version)

Chris Sproates
MSc. Thesis
March 2017

Supervisors:

F. Bikker (Rabobank)
Dr. R.A.M.G. Joosten (UT)
Dr. B. Roorda (UT)

Dep. Industrial Engineering and
Business Information Systems
Faculty of Behavioural,
Management and Social Sciences
University of Twente
P.O. Box 217
7500 AE Enschede

Colophon

Title	Internal Risk Components Validation: Indicative Benchmarking of Discriminatory Power for LGD models.
Date	March 10, 2017
On behalf of	Rabobank - Team Quantitative Risk Analytics and University of Twente (UT)
Author	Chris Sproates
Project	Implied Gini
Rabobank Supervisor	F. Bikker
First Supervisor UT	R.A.M.G Joosten
Second Supervisor UT	B. Roorda
Contact	Chris.Sproates@rabobank.com or c.l.sproates@student.utwente.nl

"Data! Data! Data!" he cried impatiently. "I can't make bricks without clay!"

— **Sherlock Holmes**, *The Adventure of the Copper Beeches* written by Arthur
Conan Doyle

Management Summary

Table 1: Terminology

Risk Component	Term	Description
<i>PD</i>	Expected AR	Analytically determined AR, which describes the expected AR assuming that the ODR for each rating correspond with the PD ratings of the model.
	Implied AR	Simulation determined AR, which describes a distribution of AR values that result from simulated defaults according to the PD ratings of the model. The mean of this distribution is the same as the Expected AR.
	Realized AR	AR based on the actual model's performance.
<i>LGD</i>	Implied Gini	Approach developed by the bank to determine an implied AR based on the LGD estimations.
	Implied AR	Simulation determined AR, which unlike the implied Gini and the implied AR for PD models uses the actual observed loss distribution. The approach simulates a distribution of ARs taking into account portfolio characteristics.
	Expected AR	The mean of the Implied AR for LGD models.
	Realized AR	AR based on the actual model's performance.

In this study we attempt to find a systematic approach to set indicative benchmark values for the discriminatory power of an LGD model measured via the *Accuracy Ratio* (AR), which is a summary statistic for the discriminatory power of a classification model. LGD models are relatively new compared to PD models, but the last years have seen a significant number of papers discussing the LGD model. Unfortunately, there is not yet an approach which indicates whether a model's performance is sufficient in terms of discriminatory power. For these reasons the bank found it difficult determining what a sufficient level of discriminatory power for an LGD model is. Currently, they have set the threshold values fixed for each portfolio measured via the AR.

Various approaches to measure discriminatory power for LGD models exist and some are addressed in this thesis. Our investigation primarily focuses on the AR, which is also known as the Gini at the bank or Powerstat. The main reason to focus on the AR is that the bank uses this technique primarily to assess the discriminatory power of LGD models.

Implied Gini

As the bank felt that a threshold on discriminatory power that is the same for LGD models regardless of the underlying portfolio are not appropriate, it started a project which was aimed to find the AR that was implied by the LGD model. Initial research by the bank showed that the score of the AR is heavily dependent on the underlying product class for which the LGD model has been built. In order to find the potential discriminatory power of a model, the bank has developed the *implied Gini* coefficient. This approach was intended to indicate how well a model in potential can discriminate between the size of potential losses of debt instruments.

The idea of the implied Gini followed from a similar approach that has been developed for PD models, which indicated an expected AR if the PD model performed to its function. However, unlike the approach of the implied AR for the PD models the implied Gini for LGD models relies on a significant number of assumptions. We critically assessed these assumptions and found that the approach could not be generalized for four portfolios that were available for our investigation. Hence, based on our test results we conclude that the implied Gini is in its current form not valid. Adjusting the approach is not a possibility given the current available data and due to some critical assumptions that do not hold true.

Influencing Factors and Alternative Approach

Although the initial research question was to validate the implied Gini, the goal for which the approach was developed still stands. Therefore, we looked into factors that give more insight in the resulting AR. From our test results we conclude that factors such as probability of cure, variance, fraude and tolerance levels influence the AR heavily if they are highly present in a portfolio. Therefore, we are confident enough to conclude that setting a performance benchmark for LGD models that is fixed over all portfolios is unrealistic (the current situation). Due to portfolio characteristics it might be impossible for some LGD models to achieve this benchmark. Therefore, we argue that benchmarks should be set up dependent on the portfolio, while taking the characteristics of the portfolio into account.

In order to set these indicative benchmarks that depend on the portfolio characteristics we developed the *implied AR*, which is a simulation driven approach that uses actual loss distribution instead of the estimations.¹ For the development of an alternative approach we had to make choices due to the limit availability of data and cope with the current data quality. For the development of the alternative approach we limited ourselves by using the AR as a summary statistic. The main reasons for this choice are that for other summary statistics similar questions remain (e.g. "what are acceptable values?"), and that the usage of the AR is common practice at the bank (and industry).

Within the process of developing an alternative approach we had to cope with several issues that cannot be solved due to the lack of data availability. Only the estimated LGD and the corresponding losses were available with some overall information concerning the portfolio, such as the historical cure rate. Therefore, we had to make choices, which can be considered to be non-optimal. We argue, however, that regardless of this limitation, we can still prove that the current review approach of the bank can be improved by taking into account different portfolio characteristics. The alternative approach for the implied Gini we developed in this research still indicates that setting the same fixed threshold value for each portfolio is unnecessary penalizing particular portfolios. Our approach gives more insight in realistic AR values for specific portfolios compared to the current situation.

We acknowledge that the approach of the implied AR still relies on some basic assumptions that cannot be tested due to the lack of data or due to data quality. However, if more data becomes available or the data quality improves it could be the case that some assumptions are not necessary any more (e.g. such as the randomness of cure or fraud). The approach also relies primarily on quantitatively measurable factors that influence the AR. There might be, however, qualitative factors that indicate that it is more difficult for some portfolios to achieve higher levels of discriminatory power than others. In our research we found it difficult to objectively pinpoint these factors as they are in our opinion very context dependent. Nevertheless, we think that the implied AR helps with the understanding of what (quantitatively) drives the AR for LGD models and therefore helps to set up indicative benchmarks for discriminatory power measured via the AR.

¹The name 'implied' is strictly taken not correct, but due to the project implied Gini is currently taken on as a working title.

Preface

Every student Industrial Engineering and Management at the University of Twente completes his MSc. degree with a graduation project either carried out internally at the university or externally at a company or an institution. I have chosen to conduct the final stage of my study program externally and I would like to thank the Rabobank for granting me this opportunity.

The project started in the summer of 2016 and over the course of the project many people helped me in getting to know the Rabobank, exchanged thoughts on the project or provided me with other valuable resources, which helped to write this thesis. I would like to especially thank to my supervisor Floris Bikker for guiding me through this process. I would also like to thank the members of the Risk Management team at the Rabobank, who were always available to help me with data issues or provide me with feedback on the project. Finally, I would like to thank Reinoud Joosten and Berend Roorda for providing feedback and new ideas.

Chris Sproates

Contents

Colophon	iii
Management Summary	vii
Preface	xi
List of Figures	xv
List of Tables	xvi
List of Acronyms	xvii
1 Introduction	1
1.1 Credit Risk at Rabobank	2
1.2 Research Objective	4
1.3 Research Approach	4
1.4 Research Questions	5
1.5 Outline	6
2 LGD Models and Their Discriminatory Power	8
2.1 The Basel II Framework for the (advanced) IRB Approach	9
2.2 Factors Affecting the LGD Scores	11
2.3 LGD Modelling and Validation	13
2.4 Risk Model Structure	16
2.5 General Techniques to Get Insight in the Performance of LGD Models	16
2.6 Measuring Discriminatory Power	21
2.7 Current Approach	29
2.8 Conclusion	29
3 Implied Gini	31
3.1 Roots of the Implied Gini	31
3.2 Relation with the Expected AR for PD Models	31
3.3 Conclusion	37
4 Validity Tests	38
4.1 The Beta Distribution	38
4.2 Loss Distributions for LGD Estimation	45
4.3 Validity Tests	47
4.4 Test Results	47
4.5 Summary and Conclusion	47

5	An Alternative Approach	48
5.1	The Proposed Alternative	49
5.2	Explanatory Factors	51
5.3	Other Explanatory Factors for an LGD	60
5.4	Summary and Conclusion	61
6	Possible Approach to Set Indicative Benchmarks	63
6.1	Implied AR as an Indicative Benchmark	63
7	Conclusion and Further Research	65
	References	67
A	Example EAD Dataset	71
B	Computation of AR25, AR50 and AR75	72
C	Derivation of Gamma Function Properties	73
D	Determining the First and Second Moment of the Beta Distribution	76
E	Loss Distributions of the Available Portfolios	77
F	Estimated Beta Parameters	78
G	Anderson-Darling Test Results	79
H	Bootstrap for Sum of Individual Loss Distributions	80

List of Figures

1.1	Research Approach.	5
2.1	The Basel II framework as drafted by BCBS (2006).	9
2.2	Probability density function of losses (BCBS, 2005).	11
2.3	Validation methodology as drafted by BCBS (2005).	15
2.4	Examples of scatter plots for LGD data.	17
2.5	Example of a box-and-whisker plot.	17
2.6	Example of a CAP Curve.	23
2.7	Surfaces for AR.	23
2.8	Possible distributions for rating scores as found in BCBS (2005).	24
2.9	Example of an ROC Curve.	26
2.10	Surfaces for the AUC.	26
2.11	Example of an LC curve.	27
3.1	The histogram of resulting ARs for the development sample.	34
3.2	The histogram of resulting ARs for the validation sample.	34
4.1	Examples of different beta distributions (1).	40
4.2	Examples of different beta distributions (2).	41
5.1	Factors influencing the AR.	50
5.2	Impact of cures on the AR.	52
5.3	Distribution of all simulated ARs (example of Portfolio A).	54
5.4	Test results for experiment: 2 PC -2 of Portfolio A.	58

List of Tables

1	Terminology	vii
2.1	Factors that impact the LGD score.	13
2.2	Six-point scale for LGD buckets as found in Cantor <i>et al.</i> (2006).	18
2.3	Example of a confusion matrix (count-based)	19
2.4	Weights for the MAD as found in Li <i>et al.</i> (2009)	20
2.5	Example of a confusion matrix (count-based)	21
2.6	Example data for determining the Gini of a PD model.	22
2.7	Coordinates Perfect Model CAP curve.	23
2.8	Coordinates Rating Model CAP curve.	23
2.9	Example data for determining the ROC of a PD model.	26
2.10	Overview of techniques for measuring discriminatory power in this study.	30
3.1	Assigned debtors per rating and sample.	34
3.2	Assigned debtors per rating and sample.	34
3.3	Implied AR for different calibrations.	35
5.1	Implied AR of the portfolios.	55
5.2	Implied AR of the structural LGD model.	55
5.3	Experiments for effect of different cure rates on the AR.	56
5.4	Test results for multiple cure rates in an LGD model (extreme cases).	57
5.5	Test results for multiple cure rates in an LGD model (random assignment).	58
5.6	Example data for the influence of the variance of a portfolio on the AR.	59
5.7	Test results for various fraud rates.	60

List of Acronyms

AR	Accuracy Ratio
AUC	Area Under the Curve
AUROC	Area Under the Receiver Operating Characteristic
BCBS	Basel Committee on Banking Supervision
BIS	Bank of International Settlements
CEBS	Committee of European Banking Supervisors
CAP	Cumulative Accuracy Ratio
DNB	De Nederlandse Bank
EAD	Exposure at Default
EBA	European Banking Authority
EL	Expected Losses
FAR	False Alarm Rate
GBR	Generalized Beta Regression
GDP	Gross Domestic Product
Gini	Gini Coefficient
HR	Hit Rate
IRB	Internal Ratings-Based
LC	Loss Capture
LGD	Loss Given Default
LGL	Loss Given Loss
LR	Loss Rate
M	Effective Maturity
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error

MLE	Maximum Likelihood Estimation
MOM	Method of Moments
MSE	Mean Square Error
ODR	Observed Default Rate
PD	Probability of Default
RAROC	Risk-Adjusted Return on Capital
ROC	Receiver Operating Characteristic
RR	Recovery Rate
RWA	Risk-Weighted Asset
SA	Standardized Approach
SME	Small or Medium Enterprises
UL	Unexpected Losses

Introduction

In 2004 the Bank of International Settlements (BIS) published Basel II, which is the international standard for the amount of capital that banks need to hold in reserve to deal with current and potential financial and operational risks (Persaud & Saurina, 2008).² Part of this framework are the estimates of risk components for determining the amount of capital required for a given exposure. Subject to certain minimum conditions and disclosure requirements, some banks have received supervisory approval to use the Internal Ratings-Based (IRB) approach for determining their own internal risk components (BCBS, 2006). Following from the framework of Basel II the risk components include the following measures:

1. Probability of default (PD).
2. Loss given default (LGD).
3. Exposure at default (EAD).
4. Effective maturity (M).

In the credit risk literature significant attention has been devoted to the estimation of the PD measurement, while much less attention has been devoted to the LGD measurements (Caouette *et al.*, 2008). LGD is defined as the credit that is lost by a financial institution in the case a debtor defaults, expressed as a fraction of the exposure at default (Bastos, 2010). The accuracy of the LGD estimations are essential for computing economic capital and potential credit losses (Gupton, 2005). If a financial institution has accurate estimates of the risk components and models with adequate discriminatory power, it would mean in principal that a financial institution could gain a competitive advantage over its competitors as it is better in separating the 'bad' instruments from the 'good'. Therefore, it is necessary that a financial institution is capable to estimate and validate the risk models used for finding the risk measures properly.

In order to determine the discriminatory power of an LGD model Rabobank uses the Gini coefficient (Gini), which is also known as the Accuracy Ratio (AR) or PowerStat. The term AR will be used in the remainder of this thesis as it used more frequently in literature. Previous research by the bank has indicated that the AR is an adequate measure to determine the discriminatory power of an LGD model. In this thesis, however, some critical side notes on the AR are given. The AR is described in Chapter 2.

²Additional regulations were added in Basel III.

Measuring the discriminatory power of LGD models is relatively new compared to PD models, which means that there is less experience in modelling those type of models. In recent years more research and best practices on LGD models emerged, but compared to PD models it is still significantly less. The relative inexperience in modelling the LGD model component has made it difficult to determine what a sufficient level of discriminatory power for an LGD model is compared to PD models.

Hence, the bank started a project to find indicative benchmarks for discriminatory power measured via the AR. Initial research showed that the score of the AR is heavily dependent on the underlying product class for which the LGD model has been built. In order to find the maximum attainable discriminatory power of a model, the bank has developed the so-called *implied Gini* coefficient. This approach was intended to indicate how well a model in potential can discriminate between the size of potential losses of debt instruments. It was intended to derive an indicative benchmark for discriminatory power from the implied Gini. The realized AR (via historical data) indicates the actual performance of a model on discriminatory power. If the realized AR is above the threshold derived via the implied Gini, then the model is perceived to have performed to its abilities according to the bank. The formal definition of the implied Gini that has been developed by the bank is presented in Chapter 3.

In this study we review the approach developed by the bank and recommend whether it can be used in its current form. We review the underlying assumptions on the implied Gini made by the bank and research alternative methods for setting the target value for the discriminatory power of LGD models based on the AR. We provide recommendations on which method should be used for setting a target value of the discriminatory power. This chapter has the following outline:

Section 1.1: We describe the background and motivation of this study.

Section 1.2: We present the research objective.

Section 1.3: We describe the research approach.

Section 1.4: We cover the research questions, which are to be answered in this thesis.

Section 1.5: We present the outline of the thesis.

1.1 Credit Risk at Rabobank

The Rabobank is an international financial services provider operating on the basis of cooperative principles (Rabobank Group, 2015). It offers services such as retail banking, wholesale banking and private banking. Furthermore, in 2015 it was the second largest bank in the Netherlands measured in total assets (TheBanks.eu, 2016). One of the core activities of the Rabobank is providing savings and borrowing services, which leads to a private sector loan portfolio (outstanding credit) of EUR 426,157 million euro compared to the total assets of EUR 670,373 million (Rabobank Group, 2015).

According to Caouette *et al.* (2008) credit can be defined as "nothing but the expectation of a sum of money within limited time," which means that credit risk is "the chance that this

expectation will not be met". In order to cope with the risk that a debtor is not able to meet his financial obligations the bank has to hold capital. For this form of risk mitigation banks are required to hold mandatory regulatory capital and in addition they can hold economic capital, which is an internally computed measurement by banks to manage their risks. As a result of these regulations in 2015 the Rabobank reserved EUR 17.0 billion of which 86% is for credit and transfer risk. The bank computed their economic capital to be EUR 26.7 billion of which 54% attributes to credit and transfer risk.

In accordance with the regulation, the bank uses the advanced IRB approach to calculate its regulatory capital for credit risk for basically the whole loan portfolio. In accordance with the supervisor the Standard Approach (SA) is used for some portfolios with relatively limited exposure and a few small foreign portfolios as the advanced IRB is not suited (Rabobank Group, 2015). The difference between the two approaches is the way in which the risk-weighted assets (RWAs) are computed, which is the input variable for computing the first pillar capital requirements of Basel II. As Hull (2012) describes, the total required capital is computed via Eq. (1.1).

$$\text{Total Capital} = 0.08 \cdot (\text{credit risk RWA} + \text{market risk RWA} + \text{operational risk RWA}) \quad (1.1)$$

The current SA prescribes that external credit ratings are used as input in order to determine the (credit risk) RWA, while for the advanced IRB the banks supply their own estimates of the PD, LGD, EAD, and M to estimate the RWA (BCBS, 2015; Hull, 2012). In Chapter 2 a detailed overview of the Basel II framework and its risk components (especially the LGD component) is given. We describe how these risk components are used to compute the RWA via the (advanced) IRB approach.

As the bank is allowed to estimate its own risk components for a large part of its portfolio (under compliance of supervisory standards) it is important that the estimates are accurate. Underestimation of the risk components means that the mitigation of risk is not sufficient and extreme losses on the loan portfolio are not covered. Overestimation of risk is also undesired as the bank then holds additional capital, which does not yield a return.

It is important that the estimates of a risk model are accurate, but a model can be accurate in estimating the total portfolio losses without correctly estimating the risk components on an individual (observation) level. If a risk model is not able to differentiate between the 'good' and 'bad' loans, it should be considered as invalid because clients with an actual high credit risk will have to post relatively less collateral compared to clients that are in reality less risky. Therefore, it is not only important to validate the estimates of the risk models, but also the discriminatory power of the risk models. Kraft *et al.* (2002) indicates that there is no formal definition for the discriminatory power of risk models. For that reason, study follows the definition given by Prorokowski (2016), namely "the ability to differentiate between defaults and non-defaults, or high and low losses". For LGD models the discriminatory power would then be the ability to differentiate between the severity of losses.

In order to gauge the risk models for improving the quality and accuracy of the estimates, banks conduct a process called backtesting. This process is defined by BCBS (2005) as "using statistical methods to compare estimates of the three risk components to realised outcomes". The bank uses a graphical representation of the realized scores against estimated scores as well as the AR in order to assess the discriminatory power of its LGD models.

It has developed the implied Gini to create a target value for the assessment. We review the original approach that has been developed by the bank and we suggest an alternative approach to set indicative benchmarks for discriminatory power via the AR.

1.2 Research Objective

The objective of this study is to develop an approach to set indicative benchmarks for measuring discriminatory power via the AR. The general method should provide guidance in setting a target value for the discriminatory power of an LGD model. The indicative benchmarks should give the bank insight in the performance of the model on discriminatory power. It helps the bank in deciding, whether the model can be accepted or should be redeveloped. Part of this study is the review of the implied Gini developed by the bank. The method has been developed on the basis of assumptions that not have been validated or proven to be correct prior to this study. Before the implied Gini can be applied for model validation, it is required that these assumptions are investigated in-depth and tested on their validity. The first part of this study focusses on the validity of the current method.

XX
 XX
 XX
 XX
 XX
 XX
 XX

As the goal of this study is to develop a new general approach to set indicative benchmark values for LGD models this study can be described as theory oriented research following the design methodology for setting up a research project by Verschuren & Doorewaard (2007). They distinguish two types of theory oriented research, namely theory developing and theory testing. This study contains both types as we review the approach developed by the bank (implied Gini), which is theory testing oriented, while we also develop a new approach for setting indicative benchmarks for LGD models, which is theory developing oriented.

(Goal) *Develop a systematic approach to set indicative benchmark values for the discriminatory power of an LGD model measured via the AR.*

1.3 Research Approach

Before the actual study can start, it is important that all stakeholders agree on what the scope and content of the research is. A good approach to get an agreement and to manage expectations is making use of a research model, which allows every stakeholder to get a quick overview of the contents of the study. Resulting from the contextual framework from which the study subject originates (see Section 1.1) and the research objective (see Section 1.2) it is possible to derive a research approach. This model is visualised in Fig. 1.1.

It is necessary to establish a theoretical framework (Phase A), before the process of theory testing and development can start. This framework forms the basis for establishing academic references and best practices from the industry, which provide validation techniques to test the implied Gini. Furthermore, the framework provides an overview of alternatives

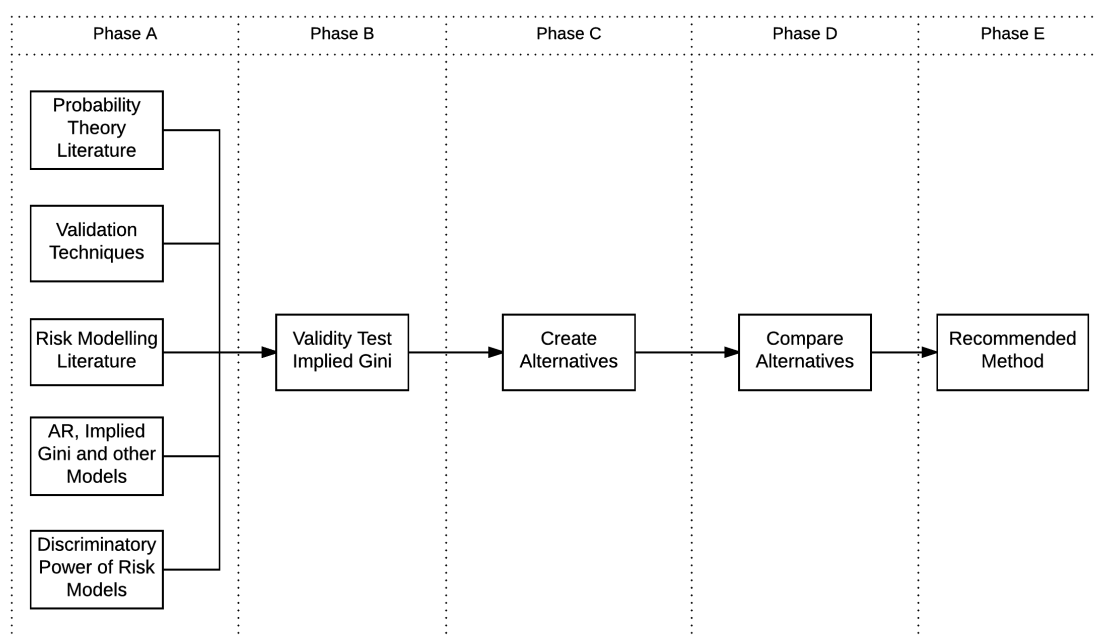


Figure 1.1: Research Approach.

and possible ideas to develop new approaches. Besides the theoretical framework, it is important to create a clear description of the method including all its assumptions, which is the subject of the validation process. Based on the test results following from the application of the validation techniques and the theoretical framework, it is possible to determine whether the current method of the implied Gini is valid for measuring the discriminatory power of a LGD model (Phase B). The results of Phase B determine, whether the implied Gini is taken into account as a valid alternative for setting indicative benchmarks for measuring the discriminatory power of an LGD model. In Phase C different approaches for setting indicative benchmarks are examined and developed. For all the valid methods the (dis-)advantages are compared (Phase D), which after consideration leads to a recommended method (Phase E).

1.4 Research Questions

From the research objective and model it is possible to derive the questions that this study has to answer. The main research question of this study is:

(Main) *Which factors should the bank take into account to determine the indicative benchmarks for the discriminatory power of an LGD model?*

To answer this question, the working of LGD models in general and the process of validation need to be described first in order to create a contextual framework and a general understanding. Once the framework has been established and the challenges concerning LGDs have been described, methods for assessing discriminatory power of an LGD model are discussed.

(1.1) *Which methods are available to assess the discriminatory power of an LGD model?*

(1.2) *How do the current LGD models differ from each other?*

- (1.3) *What are the differences between a PD model and an LGD model for measuring and benchmarking discriminatory power?*

If a general overview is created of possible methods to assess discriminatory power the current workings of the implied Gini and its assumptions can be explored in-depth. The current method of the implied Gini is based on the research done by the bank.

- (2.1) *How does the current method of the implied Gini coefficient work?*

- (2.2) *Which assumptions have been made for developing the implied Gini?*

After the explanation of the implied Gini method, it is required to validate the assumptions that have been made.

- (3) *Do the assumptions for the implied Gini hold?*

If the method is valid, it can be used to get more insight in establishing the benchmark for backtesting the discriminatory power of an LGD model via the AR. Regardless of whether the implied Gini is valid or not, it still does not provide a solution to setting an actual threshold value for the back-test of discriminatory power for an LGD model. The second part of the study focusses on what drives the AR score of an LGD model and when an LGD model is considered to be 'good'.

- (4) *Which factors impact the AR score of an LGD model?*

Based on the answers on all the research questions possible approaches for setting indicative benchmarks of an LGD model are developed. After the comparison of the possible methods a suggested systematic approach for establishing a benchmark value is presented.

1.5 Outline

The remaining part of the thesis has the following structure:

Chapter 2: We provide an in-depth analysis of LGD models, the discriminatory power and the methods to assess this attribute. The findings from literature are put into context of this study. We conclude this chapter with the answer to the first sub-questions.

Chapter 3: We explain the current method for determining the implied Gini coefficient. Furthermore, we describe the main differences between PD and LGD models that are relevant for measuring discriminatory power as there exist a similar concept of an implied AR value for PD models. The chapter concludes with the answers to sub-questions 2.1-2.2.

Chapter 4: We focus on the underlying assumptions of the implied Gini coefficient and investigates whether there is empirical and theoretical evidence in support of these claims. Furthermore, we discuss whether the implied Gini is possible as a method and if necessary adjustments are needed in order to make the method valid. We conclude with the answers to sub-question 3.

Chapter 5: We investigate LGD models more in-depth and determine which characteristics impact the AR score of an LGD model. Within this chapter we develop an alternative approach for establishing a threshold value based on the AR given the limitations on data in practice. We conclude with the answers to sub-question 4.

Chapter 6: We propose an approach to set indicative benchmarks for testing the discriminatory power of an LGD model based upon the approach we develop in Chapter 5.

Chapter 7: We provide our final conclusions on this study and discuss future research.

LGD Models and Their Discriminatory Power

Before the method of the implied Gini is discussed in-depth, it is necessary to establish a common understanding of risk models in general, and the challenges for modelling LGDs. Therefore, this chapter firstly elucidates the framework of Basel II for the (advanced) IRB approach. Secondly, an overview of processes and common practices in modelling and validating LGD models is given. The overview is based on findings from literature as well as internal documents and practices at the Rabobank. Finally, approaches for measuring the discriminatory power as well as the method for determining the AR of an LGD model is discussed.

The terminology, used in the internal documentation at the Rabobank, is adopted in this thesis as well for consistency reasons. The term LGD refers to the loss given default *estimate*, which is expressed as a percentage of the Exposure at Default (EAD). The actual *observed* loss has the term loss rate (LR) and is expressed as a percentage of the *observed* EAD. The LGD and LR term are sometimes also called estimated LGD and the realized LGD. This documentation will use the terms from the latest policy documentation, hence LGD and LR are used.

This chapter has the following outline:

- Section 2.1:** We discuss the Basel II Framework for the (advanced) IRB approach.
- Section 2.2:** We illustrate the complexity of modelling of LGD models as a lot of factors have to be taken into account.
- Section 2.3:** We describe a high level overview of the validation process and approaches for estimating LGDs.
- Section 2.4** We describe the general structure of risk models used at the bank.
- Section 2.5:** We give an overview of techniques for assessing the performance of LGD model.
- Section 2.6:** We give an overview of techniques for assessing the discriminatory power.
- Section 2.7:** We describe the internal guidelines and processes of LGD validation at the bank.
- Section 2.8:** We provide answers to the first sub-questions.

2.1 The Basel II Framework for the (advanced) IRB Approach

As discussed in the introduction, the BIS published in 2004 Basel II, which is i.a. a comprehensive framework of standards on how to measure various risk components and forms the basis on the amount of regulatory capital a financial institution has to reserve in order to cope with various risks. This section will focus on the regulatory and economic capital that is kept for dealing with credit risk. This is part of the first pillar of Basel II (which in total consists of three different pillars). The first pillar of Basel II describes the conditions and guidelines for determining the minimum capital requirements. It differentiates between three forms of risk, namely credit, operational and market risk (BCBS, 2006). An overview of the structure of Basel II based on BIS documentation can be found in Fig. 2.1.

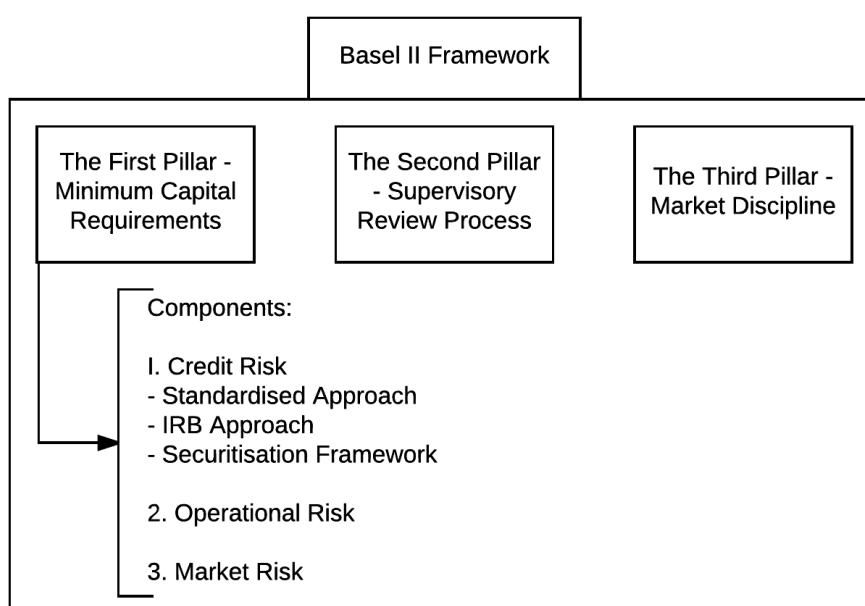


Figure 2.1: The Basel II framework as drafted by BCBS (2006).

Technically the IRB approach can be split up into two different approaches, namely the foundation IRB approach and the advanced IRB approach. The difference between the two approaches is that for the foundation approach the PD is estimated by internal models of a financial institution, but the LGD and EAD are determined on fixed values provided by the supervisor. In the advanced approach the bank uses internal models to estimate the PD, LGD and EAD. M is computed in the same way for both approaches, if M plays a role in the portfolio (DNB, 2007). If a bank has received supervisory approval, it may use the advanced IRB approach. Using the IRB approach has benefits for the regulators as well as the financial institutions. Financial institutions are incentivised to take on customers with low scores for PDs and LGDs as they result in lower risk weightings and therefore lower capital reserve requirements. It results in some form of self-surveillance, which also decreases the costs of regulation and potential legal battles with banks (Balin, 2008). As this study primarily focuses on LGD models we only the advanced IRB Approach is discussed in detail.

As is described in the BIS documentation³ the following equations are to be used to derive the RWA, which is used as input for calculating the minimum required capital via Eq. (1.1).

³International convergence of capital measurement and capital standards by BCBS (2006).

Please note that the PD and LGD are measured in percentages, and the EAD is measured as a currency in the BIS documentation. The PD value represents the probability that the loan will go into default within one year. The LGD is expressed as the fraction of the EAD that is lost if the loan goes into default (see Chapter 1). The \ln in Eq. 2.2 denotes a natural logarithm. $N(x)$ in Eq. 2.3 denotes the cumulative distribution function of a standard normal random variable. $G(z)$ denotes the inverse cumulative distribution function for a standard normal random variable (see BCBS, 2006).

$$R = 0.12 \cdot \frac{1 - e^{-50 \cdot PD}}{1 - e^{-50}} + 0.24 \cdot \left(1 - \frac{1 - e^{-50 \cdot PD}}{1 - e^{-50}}\right) \quad (2.1)$$

$$b = (0.11852 - 0.05478 \cdot \ln PD)^2 \quad (2.2)$$

$$K = \left(\text{LGD} \cdot N \left((1 - R)^{-50} \cdot G(PD) + \left(\frac{R}{1 - R} \right)^{0.5} \cdot G(0.9999) \right) - PD \cdot \text{LGD} \right) \cdot (1 - 1.5 \cdot b)^{-1} \cdot (1 + (M - 2.5) \cdot b) \quad (2.3)$$

$$\text{RWA} = K \cdot 12.5 \cdot \text{EAD} \quad (2.4)$$

Eq. 2.1 - 2.4 illustrate how the RWA for credit risk is determined under the advanced IRB approach. The correlation R and maturity adjustment b are input values among the LGD , PD and M estimates for the computation of the capital requirements K . The RWA is a function of K and the EAD estimate. There are some adjustments for specific asset classes, which are not treated in this study. For a complete overview see the documents published by the BIS (see References). The output of Eq. 2.4 is used as input for Eq. 1.1.

Economic capital is, as mentioned in Chapter 1, the internal estimate of the capital that is required to be held by a financial institution in order to cope with the risk it is taking. A more formal definition is that economic capital is the allocated capital a financial institution needs in order to absorb losses over one year with a certain confidence level (Hull, 2012). Regulatory capital is, more or less, computed via one-size-fits-all rules created by the BCBS. Hull (2012) states that "economic capital can be regarded as a 'currency' for risk-taking within a financial institution". He explains that a business unit is only allowed to take a certain risk when it has allocated the right amount of economic capital, and the profitability of the business unit is measured relative to the allocated economic capital. The latter is measured via the risk-adjusted return on capital (RAROC), which is not discussed further in this study as it lies beyond the scope of this research.

In Fig. 2.2⁴ a typical density function for credit losses can be found, which describes the likelihood of losses of a certain magnitude (BCBS, 2005). Capital reserved to cope with risk is used to cover unexpected losses (UL). The confidence level depends on the credit rating a financial institution wishes to pursue. If a bank for instance wishes to maintain an AA credit rating, then their probability of default in one-year would be about 0.03%. This suggests the confidence level for the determination of the amount of economic capital should be set at 99.97% (Hull, 2012). For financial institutions the expected losses (EL) of credit are a

⁴Figure from "An Explanatory Note on the Basel II IRB Risk Weight Functions" by BCBS (2005).

useful measure as well, as that indicates the amount it should hold in reserve from fees and interest revenues to absorb the losses that are likely to occur over the course of a year thus it is an important input factor for loan pricing decisions (Herring, 1999).

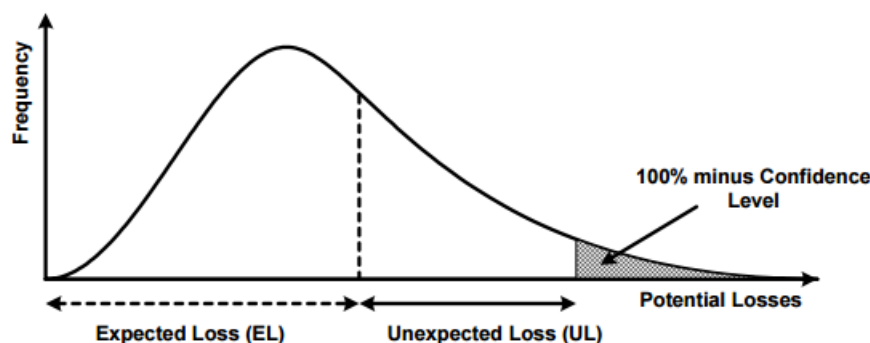


Figure 2.2: Probability density function of losses (BCBS, 2005).

The expected loss from defaults is determined via the risk components PD, LGD and EAD, which are estimated for each counterparty i . Hull (2012) shows that the total expected losses from defaults are then:

$$\sum_i EAD_i \cdot LGD_i \cdot PD_i \quad (2.5)$$

The LGD is related to the recovery rate (RR), which is the amount that is recovered from a defaulted credit. It has the following definition (expressed in percentages):

$$RR = 1 - LGD \quad (2.6)$$

From Basel II we conclude that it is crucial to have accurate estimations for the risk components to determine the necessary capital required for absorbing the potential risks. In the upcoming sections more attention will be paid to which factors affect the LGD and how it is validated.

2.2 Factors Affecting the LGD Scores

The LGD becomes relevant if a particular obligor has gone into default, which will start the process of trying to (partly) recover or salvage the outstanding credit. Default for a particular obligor has been defined by the BCBS (2006) if one or both of the following conditions has been met:

- The obligor is unlikely to pay its credit obligations to the financial institute in full from the creditors perspective, without recourse by the bank to actions.
- The obligor is past due more than 90 days on any material credit obligation to the banking group.

How the LR is measured by a financial institution depends on how it records *default* events on its debt instruments. There are examples of events in which, according to the definition by BCBS (2006), an obligor went into default (for instance it is more than 90 days past due

on credit obligations), but makes good on all their obligations in the next period. If a bank ignores such events and does not record this as recoveries in the bank's loss data, it might underestimate the RR on loans (Schuermann, 2004). This can be considered to be a cure, which is the event that a loan, which is declared to be in default, recovers and no loss is observed. It should be noted that an LGD of 0% does not explicitly mean that the default cured. Baesens & Van Gestel (2009) illustrate a similar issue of underestimating scores via technical defaults, which they define as the event that a counterparty fails to pay timely due to reasons that are not related to the financial position of the borrower. If a bank classifies the technical default as a default (from the definition of Basel II), it will typically result in a higher number of registered defaults, but in general a lower LR value.

In general, the LGD following from a default is a ratio of the losses to the EAD. One would think that it is not possible to have losses that are larger than the EAD, but it might occur in practice that the actual LGD is larger than 100%. There are namely various sources of a loss, which Schuermann (2004) distinguishes into three types:

- The loss of the principal, which is the original amount that has been lent (e.g., book value).
- The carrying cost of non-performing loans (e.g., interest income that cannot be retrieved).
- Workout expenses, which are e.g. costs made to collect the loan or collateral, or legal costs.

Due to for instance high workout costs, it might occur that the LR is larger than the actual EAD, resulting in an LR greater than 100%. Besides the workout expenses in case of a default, the seniority of the debt instrument and the posted collateral play important roles. The seniority determines the priority for all the creditors in the case value can be salvaged from a default. Collateral is a specific asset or property pledged to the creditor in the case of a default and used to secure debt instruments. Empirical evidence suggests that the seniority of bonds is one of the driving forces behind an RR, as the mean recovery increases for higher levels of seniority (Altman & Kishore, 1996). As loans are generally senior to bonds, it is expected that they will also provide higher RRs than bonds. Statistics resulting from Moody's database of 1970 till 2003 imply that RRs for loans are typically higher (Schuermann, 2004). A similar correlation can be found between collateral and RR (Grunert & Weber, 2009).

Another factor that has been described by Altman *et al.* (2005) is the negative correlation between the observed default rate (ODR, which is the actual number of defaults observed in a time period, and the RR (for corporate bonds): Higher aggregated levels of the ODR tend to coincide with lower RRs. Frye (2000) also indicates that in the US years that have a relatively high default rate, will result in a lower RR on average. The years that he indicated as the years with a high default rates (1990 and 1991) coincide with relatively low growth in gross domestic product (GDP), namely 1.9% and -0.1% on an annual basis (The World Bank, 2016). The other years in Frye's data set had a minimum growth of 2.7%. This relation could imply that in recessions or times of slow economic growth the RRs are lower than in times of expansions.

Also differences between industries are observed in the average realized RR (and thus LGD) by Altman & Kishore (1996). The industry condition is suggested to be an important determinant for the RR as industry-wide distress leads to lower recoveries (Acharya *et al.*, 2007).

However, Qi & Zhao (2013) suggest that the impact of the industry and macroeconomic variables on the LGD vary with the sample, model specification and modelling technique used. Their research suggested that the debt structure of a firm should be considered for modelling the RR.

A difference in RRs between countries is observed by Davydenko & Franks (2008). Their research suggested i.a. that the local bankruptcy code affects the RR, as in countries that are perceived to be debtor-friendly (e.g., France) the RR is considerably lower than in more creditor-friendly countries. They state that the influence of the bankruptcy code is the greatest in explaining the differences between the recoveries. Furthermore, they state that the bankruptcy codes also result in different lending and reorganization practices as e.g. the posted collateral in France is higher than in Germany and U.K., which are perceived to be more creditor-friendly.

The brief overview from the literature illustrates some of the difficulties to model LGDs. A lot of variables need to be taken into account in order to estimate a LGD correctly. Different portfolios have their own specific LGD models. Table 1 includes, but is not limited to, factors that influence the values of an LGD. Additional explanatory factors for LGD estimation can be found in Peters, C. (2011).

Table 2.1: Factors that impact the LGD score.

Factor	Literature
<i>Definition and Recording of Defaults</i>	Schuermann (2004), Baesens & van Gestel (2009)
<i>Debt Type, Seniority and Collateral</i>	Altman & Kishore (1996), Schuermann (2004), Grunert & Weber (2009)
<i>Debt Structure</i>	Qi & Zhao (2013)
<i>Default Rates and Macroeconomic Developments</i>	Frye (2000), Schuermann (2004), Altman <i>et al.</i> (2005)
<i>Type of Industry</i>	Altman & Kishore (1996), Schuermann (2004), Acharya <i>et al.</i> (2007)
<i>Bankruptcy Code (in Country)</i>	Davydenko & Franks (2008)

2.3 LGD Modelling and Validation

This section gives a brief overview of different approaches to estimate the LGD values, and standards and processes for validating of the LGD models.

Approaches for LGD Estimation

In order to estimate the LGD various techniques (or combinations) can be used. There are four approaches described by the Committee of European Banking Supervisors (CEBS)⁵ (2006) to estimate an LGD, which are the following:

Workout LGD	calculates the (discounted) cash flows resulting from a workout and/or collections process.
Market LGD	determines the LGD estimations on the basis of the market prices of defaulted obligations.
Implied Market LGD	is similar to the Market LGD, but estimates the LGD from non-defaulted loans, bonds or credit default instruments. The implied market LGD is derived via a theoretical asset pricing model (Schuermann, 2004).
Implied Historical LGD	is a technique that derives the LGD from the realised losses on exposures within a loan portfolio and PD estimations. CEBS (2006) allows this technique only for the Retail exposure class (e.g. loans made to individuals such as mortgages).

As CEBS (2006) points out, the supervisors do not require that a specific technique is used for the LGD estimations. Nevertheless, financial institutions will need to demonstrate that the assumptions underlying their models are justified and that the approach is appropriate for the specific portfolios to which it is applied as the results might be different per approach. For example, the realized workout LGD usually takes some time to be computed as it uses the discounted values of actual cash and assets after a default has been settled, while the market LGD is easy to observe from actual market prices. Renault & Scaillet (2004) state that the market LGD has its drawbacks.⁶ They state i.a. that the trading price recovery tends to have lower means compared to the ultimate recovery. An advantage, however, is that it is not required to choose a specific discount rate and because the market LGD is derived from actual market prices an investor can determine the RR if he liquidates his position immediately. For the estimation of an LGD one should carefully assess the drawbacks and advantages of the different approaches. The choice is also dependent on the type of portfolio, as not all approaches can be used for some portfolios. For example, not all types of portfolios have market data to derive RRs.

LGD Validation Methodology

Once an approach for the LGD estimation has been chosen and a model has been developed the process of validation may begin. An LGD model should be reviewed periodically as it is important to see whether the LGD model is still accurate. This is done by validating the model by i.a. backtesting and benchmarking the risk components.

The BCBS (2005) notes that the use of statistical tests for backtesting may be difficult as the data from financial institutions may be constrained. The reason behind this might be

⁵Their tasks and responsibilities have been taken over by the European Banking Authority (EBA) as from 2011.

⁶Renault & Scaillet (2004) illustrate the differences via recovery rates, which they call ultimate recoveries and trading price recoveries. From their description it follows that the ultimate recovery rate is 1 minus the "workout LGD" and the trading price recovery rate is 1 minus the "market LGD". They state that the trading price recovery tends to lead to lower means compared to the ultimate recovery.

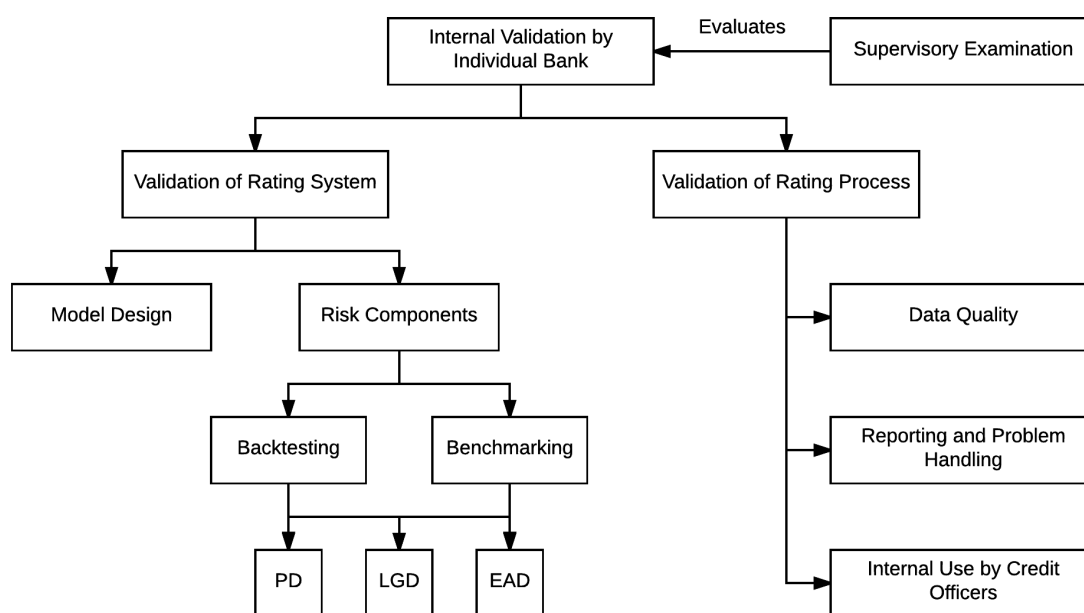


Figure 2.3: Validation methodology as drafted by BCBS (2005).

the low number of defaults in a portfolio or low internal data quality. Initiatives have been started according to the BCBS (2005) to build consistent data sets. There is an emphasis on the validation of LGDs in the studies conducted by the BCBS, as i.a. the capital charges are quite sensitive to the LGD (see Eq. 2.1-2.4). A high level overview of the validation methodology drafted by the BCBS (2005) can be found in Fig. 2.3.

A practical framework for the validation of an LGD model is described by Li *et al.* (2009). They define three specific performance goals, which are of interest for the credit risk manager. The goals are as follows:

- Good performance on the rank-order of different LGD estimations (discriminatory power).
- Accurate predictions of the LR (calibration).
- Accurate prediction of the total observed portfolios loss amounts, which assumes that the PD model correctly distinguishes between defaulters and non-defaulters.

Loterman *et al.* (2012) point out that a good ranking does not imply that the calibration is good, but on the other hand if the calibration is good it always implies that the discriminatory power is good as well. After the (re-)validation process Li *et al.* (2009) distinguishes three possible outcomes that may result from the assessment, namely:

- LGD model is a reasonable reflection of the current portfolio and performs well enough—no adjustments to its current form are required.
- LGD model has a (moderate) discrepancy from the original specification or a previous revalidation process—a refit is required.

- LGD model is significantly different from the original specification or a previous revalidation process, and is not likely to meet its expected performance level—a redevelopment is required.

In order to validate the risk models a variety of tools and metrics are available to recognize under-performing LGDs. As this study intends to find a valid method to establish a target value for the discriminatory power which can be used as a benchmark in the validation process an overview of frequently used techniques are presented in Section 2.4. The main focus is on the assessment of the discriminatory power of LGD models.

2.4 Risk Model Structure

XX
 XX
 XX
 XX
 XX
 XX
 XX

2.5 General Techniques to Get Insight in the Performance of LGD Models

Before techniques for assessing the discriminatory power of an LGD model is described, some general techniques to get insight in the performance of an LGD model are introduced. First possible techniques to visualise the performance of an LGD model are described. After the description of these visualisation techniques error measures are introduced, which indicates the performance of the model on accuracy and calibration. In Section 2.6 techniques for assessing discriminatory power are discussed, which is the main focus of this thesis.

Summary Plots

According to Li *et al.* (2009) one of the first plots that has to be examined when validating an LGD is the scatter plot of the LGD versus the LR. This helps to provide evidence whether the estimations of the model correspond with the actual LRs. A good model should be able to provide a scatter plot with points concentrated around the diagonal. A perfect model would be able to provide a line through the origin, while having a 45° angle. Fig. 2.4 provides a simplified example of the difference between the scatter plot of a ‘bad’ LGD model and a ‘good’ LGD model.

In order to see whether the assumed loss distribution corresponds with the realized loss distribution histograms allow a good visual comparison between the two distributions. To actually see whether the LGDs and LRs originate from the same distribution various statistical tests are at the credit risk managers’ disposal (e.g., two-sample Kolmogorov-Smirnov test). These techniques are discussed in-depth in Chapter 4.

Another approach suggested by Li *et al.* (2009) is to use box-and-whisker plots to get a sense of the magnitude and frequency of the outliers classified by the LGD buckets in order

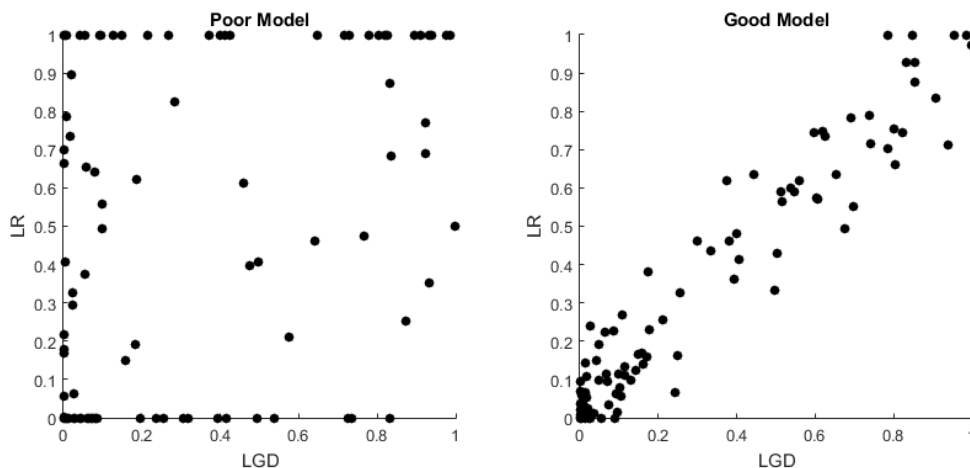


Figure 2.4: Examples of scatter plots for LGD data.

to grasp the characteristics of the EAD distribution as a histogram is often not enough. This graphical approach is simply a representation of the median and the quartiles (the box) and possible outliers (whiskers) for a particular data set. If this technique is used for getting insight in the performance of an LGD model then the box-and-whisker plot can show how the EADs of a portfolio are distributed against LGD estimations. Appendix A contains random sample data of the EAD and LGD for 50 debtors in a portfolio. In order to construct the box-and-whisker plot the data set is separated in 6 buckets based on the LGD estimation. The buckets ranges are based on the ranges suggested by Cantor *et al.* (2006) as found in Table 2.2. The sixth bucket is empty in our example (no estimations larger than 90%). For the remaining buckets a box-and-whisker plot is made on the basis of the EADs contained in a bucket. Fig. 2.5 shows the resulting box-and-whisker plot for the example data found in Appendix A. For this data set the plot shows that the EADs have a large range in most buckets and the spread also varies between the buckets, which is something to be taken into account for the validation process.

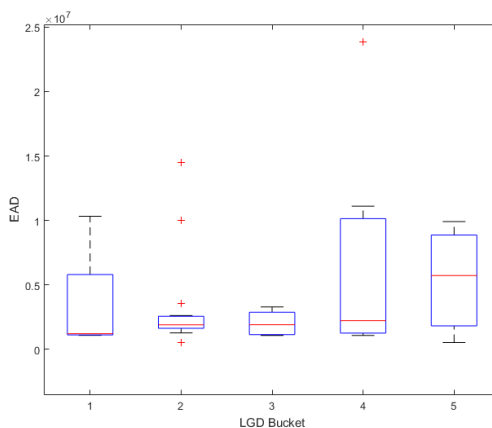


Figure 2.5: Example of a box-and-whisker plot.

Error Measures

Summary plots can be used to get a quick overview of the LGD models performance, but in order to quantify the LGD models performance error measures can be used. One approach suggested by Li *et al.* (2009) is the use of confusion matrices. These type of matrices are used to check how a classifier model performs. In order to construct the confusion matrix for an LGD model, it is required to have the LGD estimations and realizations (the LR). Each LGD estimation is classified to its corresponding bucket (e.g. via the six-point scale as found in 2.2), and the same process is followed for the realizations. We denote for each instance its corresponding estimated LGD as LGD_p^{EST} and its LR as LGD_p^{REA} for $p = 1, 2, \dots, n$ with n instances. We denote the total number of classes (e.g. the buckets as found in 2.2) as k and each possible assignment for LGD_p^{REA} as $i = 1, 2, \dots, k$ and for LGD_p^{EST} as $j = 1, 2, \dots, k$. Each bucket has a lower bound and upper bound, which determines whether an estimation or realization as assigned to the bucket. These bounds are denoted by LB and UB . Via Eq. (4.10) it is possible to determine the corresponding buckets for both the estimated and realized LGD.

$$c_{p,i,j} = \begin{cases} 1, & \text{if } LB_i \leq LGD_p^{REA} < UB_i \text{ and } LB_j \leq LGD_p^{EST} < UB_j \\ 0, & \text{otherwise} \end{cases} \quad (2.7)$$

If for each of the paired observations the corresponding classes or buckets are determined, then it is possible to compute the confusion matrix. The idea behind the confusion matrix is to count for each cell in the matrix the number of pairs within such a cell. The total number of pairs in a cell is denoted via $a_{i,j}$ and computed via Eq. (2.8).

$$a_{i,j} = \sum_{p=1}^n c_{p,i,j} \quad (2.8)$$

In Table 2.3 the confusion matrix for the sample data in Appendix A can be found (based on the buckets in Table 2.2). The diagonal of the matrix shows how many LGDs have been estimated in the correct bucket (i.e., they have been correctly classified). The other cells of the matrix show how many LGDs have been falsely classified. Take for example the cell (LGD 6, LGD 4), which has the value of 1. This particular instance has been estimated to be in Bucket 4 (50% till 70%). The realized LGD of this instance is, however, above 90% and therefore its actual 'class' is Bucket 6.

Table 2.2: Six-point scale for LGD buckets as found in Cantor *et al.* (2006).

LGD Assessment	Loss Range
LGD 1	$\geq 0\%$ and $< 10\%$
LGD 2	$\geq 10\%$ and $< 30\%$
LGD 3	$\geq 30\%$ and $< 50\%$
LGD 4	$\geq 50\%$ and $< 70\%$
LGD 5	$\geq 70\%$ and $< 90\%$
LGD 6	$\geq 90\%$ and $\leq 100\%$

Table 2.3: Example of a confusion matrix (count-based)

Realized LGD	Estimated LGD						Total
	LGD 1	LGD 2	LGD 3	LGD 4	LGD 5	LGD 6	
LGD 1	4	0	0	1	1	0	6
LGD 2	2	8	1	1	0	0	12
LGD 3	1	12	2	3	0	0	18
LGD 4	0	0	1	2	0	0	3
LGD 5	0	5	1	2	2	0	10
LGD 6	0	0	0	1	0	0	1
Total	7	25	5	10	3	0	50

This approach is called 'count basis' by Li *et al.* (2009). The other approaches they have described are intuitively the same, but might require more information such as EAD or observed losses values, which are defined respectively as the 'total EAD basis' and 'observed loss basis' approach. In the 'total EAD basis' approach one sums the total EAD for each cell. If the observation in the cell (LGD 6, LGD 4) is again taken as an example, the value for that cell in the 'total EAD basis' approach would be €11.100.000. This is 6.09 % of the total EAD for the whole portfolio (€182.400.000), while it only represents 2% of all the observations (50). If the EADs are more or less equal in size the 'count-basis' approach would be appropriate, but in our example (also indicated by the box-and-whisker plot in Fig. 2.5) the exposure at risk might be underestimated as some miss-classifications represent a large risk.

The 'observed loss basis' approach uses the realized losses instead of the EAD. This approach illustrates the impact of particular miss-classifications. The realized loss of the instance in the cell (LGD 6, LGD 4) is €10.989.000, which is 14.75% of the total losses (€74.488.650). The realized losses are obtained by multiplying the LR with the EAD. This shows that the impact of this miss-classification is quite severe in our example portfolio.

Li *et al.* (2009) argue that, besides the overview the confusion matrix gives of the model's performance, there still is a need to capture the information contained by the confusion matrix in a single metric. This metric allows a comparison between LGD models. They suggest two metrics as a measurement, namely the 'percent matched' and 'Mean Absolute Deviation (MAD)'. The metric percent matched indicates how many LGDs are correctly estimated to belong in the realized bucket. Looking to the confusion matrix in Table 2.3, this means that all the values on the diagonal are the correctly estimated values. The score on this metric is the sum of all the elements of the diagonal divided by the total number of elements in the data set. This is defined in Eq. 2.9 for which k is defined as the number of buckets (thus leading to a k -by- k confusion matrix), $a_{i,j}$ as a matrix cell (thus $a_{i,i}$ is the element on the diagonal), and n represents the total number of observations in the data set. In sample confusion matrix (see Table 2.3) 18 out of the 50 elements are on the diagonal, which leads to a performance score of 36%.

$$\text{Percent Matched} = \frac{\sum_{i=1}^k a_{i,i}}{n} \quad (2.9)$$

The other performance metric suggested by Li *et al.* (2009) is the 'Mean Absolute Error (MAE)', which takes into account how 'bad' the miss-classification was. They argue that a classification for a neighboring bucket can be considered as less severe than a classification of bucket that is further away from the actual realized bucket. For instance a classification in the cell (LGD1, LGD2) is 'less' wrong than a classification in the cell (LGD1, LGD6). The standard 'percent matched' approach does not take this into account. In order to take the 'severeness' of a miss-classification into account they compute the absolute deviation between the buckets. In order to do so they have determined weights for each bucket, which can be found in Table 2.4. These weights can of course be set differently.

Table 2.4: Weights for the MAD as found in Li *et al.* (2009)

Realized LGD	Estimated LGD					
	LGD 1	LGD 2	LGD 3	LGD 4	LGD 5	LGD 6
LGD 1	0	0.1	0.2	0.35	0.575	0.825
LGD 2	0.1	0	0.1	0.25	0.475	0.725
LGD 3	0.2	0.1	0	0.15	0.375	0.625
LGD 4	0.35	0.25	0.15	0	0.225	0.475
LGD 5	0.575	0.475	0.375	0.225	0	0.25
LGD 6	0.825	0.725	0.625	0.475	0.25	0

The absolute deviation is computed by multiplying the weight for each cell with the counted number in each cell as defined by Eq. 2.10. The $w_{i,j}$ is the weight for the combination (i,j) and $a_{i,j}$ is the number of elements in each combination (i,j).

$$AD_{i,j} = w_{i,j} \cdot a_{i,j} \quad (2.10)$$

If the same weights used by Li *et al.* (2009) are used for the sample confusion matrix in Table 2.3 then the result for all the absolute deviations can be found in Table 2.5. Li *et al.* (2009) determines the MAD by dividing the sum of all absolute deviations by the total number of estimations in the data set (see Eq. 2.11). For the sample data set this results in 7.15 divided by 50, which equals 14.3%. The lower the MAD is, the better the model is in correctly estimating the LGD of a loan.

$$MAD^{CM} = \frac{\sum_{i,j} AD_{i,j}}{n} \quad (2.11)$$

This error measurement is also used by Bellotti & Crook (2008), but they do not include weights and use the RR. In addition they use the 'Mean Square Error (MSE)'. They compute these error measures for each paired observation (thus on an individual level), while Li *et al.* (2009) computes the MAD via the confusion matrix, which gives different results. The mathematical definitions for the MSE and MAD (or MAE) can be found in Eq. 2.12-2.13 respectively. R is the realized RR and P is the estimated RR, while m denotes the number of observations (Bellotti & Crook, 2008).

$$MSE = \frac{1}{m} \sum_{i=1}^m (R_i - P_i)^2 \quad (2.12)$$

Table 2.5: Example of a confusion matrix (count-based)

Realized LGD	Estimated LGD						Total
	LGD 1	LGD 2	LGD 3	LGD 4	LGD 5	LGD 6	
LGD 1	0	0	0	0.35	0.575	0	0.925
LGD 2	0.2	0	0.1	0.25	0	0	0.55
LGD 3	0.2	1.2	0	0.45	0	0	1.85
LGD 4	0	0	0.15	0	0	0	0.15
LGD 5	0	2.375	0.375	0.45	0	0	3.2
LGD 6	0	0	0	0.475	0	0	0.475
Total	0.4	3.575	0.625	1.975	0.575	0	7.15

$$\text{MAD} = \frac{1}{m} \sum_{i=1}^m |R_i - P_i| \quad (2.13)$$

More error measurements are available for the calibration of LGD models, but these will not be covered within this study.⁷

2.6 Measuring Discriminatory Power

The approaches in Section 2.5 can be used to validate the accuracy of the LGD model or get a quick grasp of the LGDs performance, but give less information about the discriminatory power of a LGD model. A lot of statistical tools⁸ are available to assess the discriminatory power of PDs, but for LGDs these are not always applicable. This section introduces possible techniques to assess the discriminatory power of LGD models. As some techniques are mainly used for the validation of PD models, these models are explained first from an PD perspective, before the LGD approach is explained.

Cumulative Accuracy Profile

An approach that is used quite frequently is the Cumulative Accuracy Profile (CAP), which has been described by Sobehart *et al.* (2000). It is also known as the Gini curve, Power curve or Lorenz curve (BCBS, 2005). For explanatory purposes the CAP is first explained for PD models and Section 2.7 elaborates how it can be applied for LGD models, which is the current approach of the bank for measuring discriminatory power of LGD models. The CAP term used in Sobehart *et al.* (2000) represents the cumulative probability (of going into default) for the entire population. The method is explained via an example. Suppose that the following data from a PD model is available (see Table 2.6). The PD score is the expected probability that a debtor will go into default, while the binary score of 0 or 1 reflects whether the loan actually went into default (if the value is 1, then the loan defaulted).

⁷An overview can be found in the paper of Loterman *et al.* (2012), who tests various regression techniques for the modelling and predicting of LGDs.

⁸An extensive overview can be found in Section III in BCBS (2005) written by Tasche, D.

Table 2.6: Example data for determining the Gini of a PD model.

PD Score	Default
0.10	0
0.15	0
0.20	1
0.25	0
0.30	1
0.35	1
0.40	0
0.50	0
0.55	1
0.60	1

The CAP exists out of three curves, namely the perfect model curve, the rating model curve and the random model curve. The perfect model is a model that has perfect discriminatory power, which means that you can exactly separate the defaults from the non-defaults. The curve is constructed by plotting for each fraction of the population (%) the cumulative amount of defaults as a percentage relatively to the total amount of defaults in the population. As the perfect model perfectly separates defaults from non-defaults all the defaults are correctly 'predicted'. In order to construct the perfect model array is created, which first describes all the default cases and then the non-default cases. Please note, that the PD scores from the model are for the perfect model irrelevant. For the example, the following array is obtained:

$$[1, 1, 1, 1, 1, 0, 0, 0, 0, 0] \quad (2.14)$$

The curve is constructed by computing the cumulative defaults for each fraction of the population. Suppose that n is the size of the population and m is the number of defaults in the population. Let i be the position in the array in 2.14 and a_i the score. For the integers 1 till k with k equal to the population size each point of the curve is constructed via 2.15.

$$(X, Y) = \left(\frac{i}{n}, \frac{\sum_{i=1}^k a_i}{m} \right) \quad (2.15)$$

Thus, for the example the following points in Table 2.7 construct the perfect model CAP curve.

The curve of the rating model is quite similar only that the actual PD values for each observation are used for the ranking. Each observation is ranked based on the PD score from highest to lowest (thus from highest to lowest probability to go into default). The following array for the rating model can be obtained for the example:

$$[1, 1, 0, 0, 1, 1, 0, 1, 0, 0] \quad (2.16)$$

Table 2.7: Coordinates Perfect Model CAP curve.

Fraction of Population (%)	Cumulative Defaults (%)
0.10	0.20
0.20	0.40
0.30	0.60
0.40	0.80
0.50	1
0.60	1
0.70	1
0.80	1
0.90	1
1.00	1

Table 2.8: Coordinates Rating Model CAP curve.

Fraction of Population (%)	Cumulative Defaults (%)
0.10	0.20
0.20	0.40
0.30	0.40
0.40	0.40
0.50	0.60
0.60	0.80
0.70	0.80
0.80	1
0.90	1
1.00	1

Thus, the points in Table 2.8 construct the rating model CAP curve (similar to the points of the perfect model). The random model is simply a 45° line through the origin and the point (1,1). In Fig. 2.6 the CAP curves for the example are shown. If the PD model is accurate the CAP curve will be concave and will have a relatively high slope at the start of the curve, which will decline towards zero.

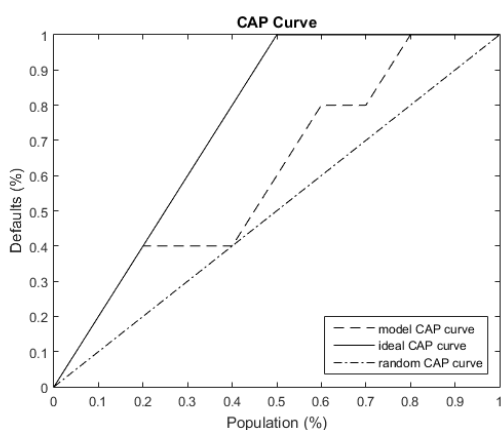


Figure 2.6: Example of a CAP Curve.

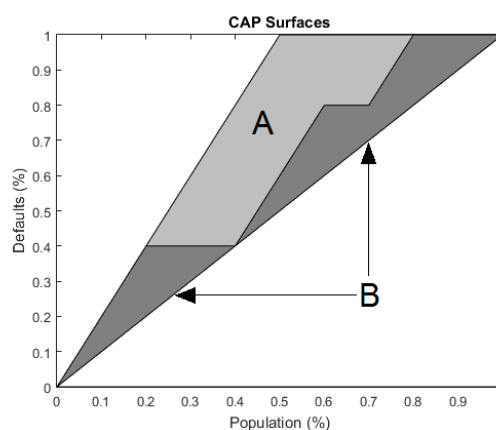


Figure 2.7: Surfaces for AR.

From these curves it is possible to derive the Accuracy Ratio (AR). The ratio is derived from the area enclosed by the random model and the rating model (model CAP curve) divided by the area enclosed by the random model and the perfect model (ideal CAP curve). The value of the *AR* will lie between 0 and 1, with a value near 0 means that the model has limited discriminatory power and the closer it is to 1 the more it represents the discriminatory

power of the perfect model. Suppose A is the area between the perfect model curve and the random model curve (light plus dark gray area in 2.7), while B is the area between the rating model curve and random model curve (dark grey area in 2.7), then the AR is computed via Eq 2.17. For the example used the AR equals 44%.

$$AR = \frac{B}{A + B} \quad (2.17)$$

Receiver Operating Characteristic

A concept that is closely related to the CAP curves is the Receiver Operating Characteristic (ROC). The ROC curve is constructed by using two distributions of rating scores for defaulting and non-defaulting debtors (BCBS, 2005).⁹ An example of possible distributions for rating scores can be found in Fig. 2.8.

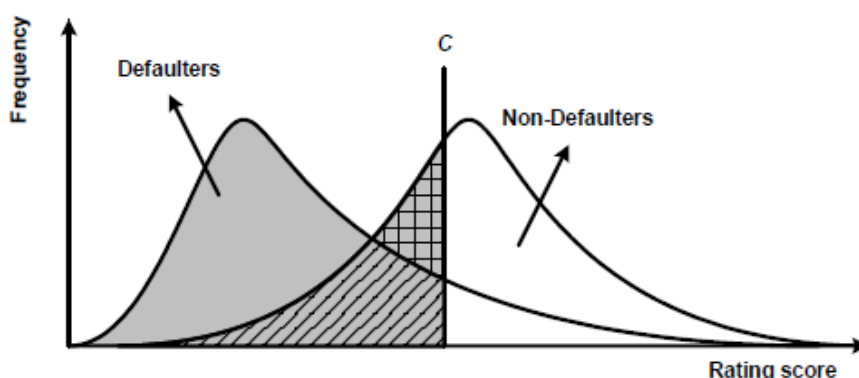


Figure 2.8: Possible distributions for rating scores as found in BCBS (2005).

A perfect model would be able to separate the two distributions completely, but in practice it is more likely that they will overlap as is shown in Fig. 2.8 (BCBS, 2005). If you have to find out which debtors will default in the next period, it is possible to introduce a cut-off point C , which results in the classification of potential defaulters (rating score below C) and non-defaulters, who have a rating score above C (BCBS, 2005). High PD scores express the probability that a loan is likely to default, which typically if one uses credit risk scorecards result in low scores. Loans with a low probability to go into default usually have high scores. Thus, if only the PD values for each instance are available, the cut-off point represents a PD score. If an instance has a PD value above the cut-off point it will be classified as a default and if it has a lower PD value it will be classified as a non-default. The ROC is constructed by computing the Hit Rates (HR) and False Alarm Rates (FAR), via Eq. 2.18-2.19. The HR for each C is the percentage that the model correctly classifies as default, while FAR is the percentage of non-default that has been wrongly classified as default given C .

$$HR(C) = \frac{H(C)}{N_D} \quad (2.18)$$

⁹Section III written by Tasche, D.

$$\text{FAR}(C) = \frac{F(C)}{N_{ND}} \quad (2.19)$$

In these equations $H(C)$ are the correctly predicted number of defaults with the cut-off point C , and N_D is the total number of defaults in the portfolio. The $F(C)$ is the number of non-defaults that have been predicted to go into default by applying the cut-off point C , and N_{ND} is the total of non-defaults in the sample (BCBS, 2005). The ROC curve is constructed by taking various values for C (within the range of the rating scores), and computing the corresponding $HR(C)$ and $FAR(C)$ values.

For the same data set used for the CAP curve it is possible to construct the ROC curve. First the HR and FAR for various cut-off points need to be computed. The PD values from Table 2.6 are taken as a cut-off point. The values 0 and 1 are taken as a cut-off point as well. If the PD value of an instance is lower or equal than the cut-off point it is classified as non-default, while higher than the cut-off point is classified as a default. The classifications are compared to the realizations. If the classification is a default and the realization is a default as well, then the instance is counted as a hit. If the classification is a default, but the realization shows that the instance did not go into default, then it is counted as false. If instance i is a hit then h_i is 1 (0 otherwise) and f_i is 1 if a non-default instance is classified as default (0 otherwise). If n is the total number of instances in the data set the total sum of hits and falsely classified instances is computed via Eq. 2.20-2.21. Table 2.9 contains the results for each cut-off point for the example data in Table 2.6.

$$H(C) = \sum_{i=1}^n h_i \quad (2.20)$$

$$F(C) = \sum_{i=1}^n f_i \quad (2.21)$$

The values of h_i and f_i depend on the cut-off point C , which is a value between 0 and 1. Suppose that we denote the outcome of whether a debtor i went into default or not with the term a_i for which the value 0 means that the debtor did not default, while the value 1 means that the debtor defaulted. The estimated PD for debtor i is denoted with the term PD_i . Then via Eq. 2.22-2.25 we can determine the hit rates h_i and false alarm rates f_i for all debtors.

$$h_i = 1 \text{ for } a_i = 1 \text{ and } PD_i > C \quad (2.22)$$

$$h_i = 0 \text{ for } a_i = 1 \text{ and } PD_i \leq C \quad (2.23)$$

$$f_i = 1 \text{ for } a_i = 0 \text{ and } PD_i > C \quad (2.24)$$

$$f_i = 0 \text{ for } a_i = 0 \text{ and } PD_i \leq C \quad (2.25)$$

The resulting ROC curve for this particular example can be found in Fig. 2.9. Due to the small data set used from the example this particular ROC looks blocky, but typically the ROC curve for larger data sets will look more smooth. In order to quantify the ROC curve the Area

Table 2.9: Example data for determining the ROC of a PD model.

C	H(C)	F(C)	HR(C)	FAR(C)
0.00	5	5	1.00	1.00
0.10	5	4	1.00	0.80
0.15	5	3	1.00	0.60
0.20	4	3	0.80	0.60
0.25	4	2	0.80	0.40
0.30	3	2	0.60	0.40
0.35	2	2	0.40	0.40
0.40	2	1	0.40	0.20
0.50	2	0	0.40	0.00
0.55	1	0	0.20	0.00
0.60	0	0	0.00	0.00
1.00	0	0	0.00	0.00

Under the Curve (AUC) is computed. Sometimes this is called the Area Under the Receiver Operating Characteristic (AUROC), which is the same. The size of the AUC determines how well the model performs (similar to the AR of the CAP) by the area under the rating model curve. For the used example it is the grey area in 2.10. The AUC for the example data set is 72%.

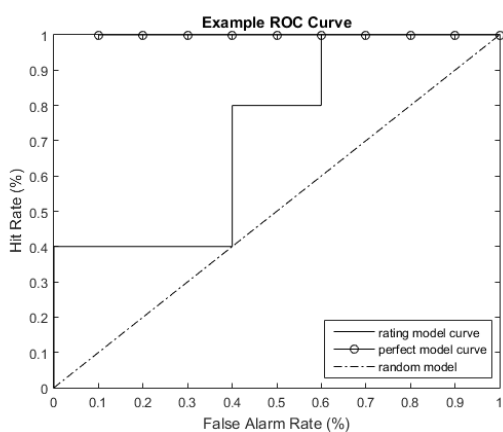


Figure 2.9: Example of an ROC Curve.

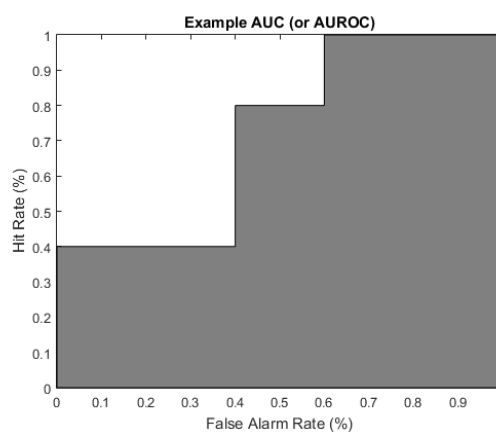


Figure 2.10: Surfaces for the AUC.

Engelmann & Tasche (2003) show that the *AR* is just linear transformation of the *AUC* as is shown in Eq. 2.26. This means that the summary statistics of both the ROC and CAP approach contain the same information (Engelmann & Tasche, 2003).

$$AR = 2 \cdot AUC - 1 \tag{2.26}$$

Generally, the AR or Gini is used in measuring the discriminatory power of LGD models. As is seen from the LC curve, using the approach for constructing the CAP can be applied to LGD models. This is not so obvious for the ROC curve as it uses binary values as an input for constructing the hit and false alarm rates. Usually, estimated LGD scores are continuous between 0 and 1 and not suitable to directly compute the hit and false alarm rates. A transformation of the data is therefore required. Hlawatsch & Reichling (2010) describe an approach to compute the ROC curve for a LGD model. For each observation (credit) they split the EAD up in n equal sized portions. They determine for each portion of a credit, whether it went into default or not and give it a binary value of 0 or 1 (similar to PD models). For example if a credit has a realized LGD of 50% then the first half of the portions would get a binary value of 1. This enables them to compute the hit and false alarm rates, which form (if they are cumulated) the ROC curve.

Loss Capture Ratio

An approach that has been designed to measure the discriminatory power on the LGDs ability to capture the portfolio's final observed loss amount is the loss capture (LC) ratio described by Li *et al.* (2009). As they point out it is very similar to the Cumulative Accuracy Profile (CAP). For this approach three plots are relevant, namely the model (rating) loss capture curve, ideal (perfect) loss capture curve and finally the random loss capture curve. These curves are constructed in the same way as the curves for the CAP. The main difference is the data, which is for LGDs and the LR a (continuous) percentage of the EAD, while for the CAP it is binary. An example of the LC curve can be found in Fig. 2.11. The LC ratio is constructed similar to the AR via Eq. (2.17). The LC can be percentage weighted, which simply uses the LGD and LR percentages as input, while it can also be EAD weighted, which uses the LGD and LR multiplied with the respective EAD as input. The results between the two approaches can differ especially if the portfolio is not-well balanced. In the remainder of this thesis the percentage weighted approach for the LC ratio will simply be referred to as AR as it essentially is the same. If the EAD weighted approach it is referred to as LC.

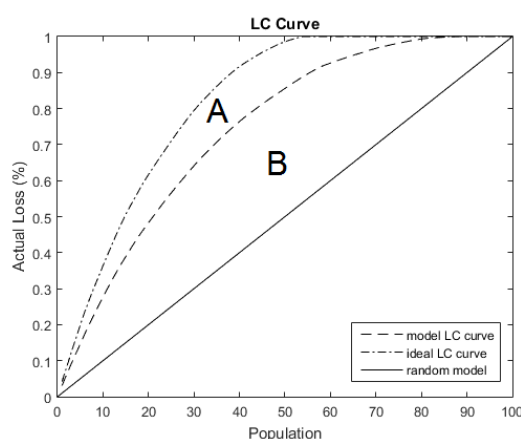


Figure 2.11: Example of an LC curve.

Other Summary Statistics

Aside from these graphical approaches with a summary statistic other metrics exist to measure discriminatory power. This study considers the following:

- Pearson's r
- Spearman's ρ
- Kendall's τ_b

Pearson's r is the index of a linear correlation between two variables, which is described extensively in Cohen *et al.* (2003). It is also called the Pearson Product Moment Correlation coefficient. The index measures the linear correlation between two variables, which can vary between -1 and 1. As both the estimated LGD and realized LGD are comparable scores no linear transformation is required and we can use the raw score formula to compute r , as described by Cohen *et al.* (2003). See Eq. 2.27 in which X can be for example the estimated LGD, while Y can represent the realized LGD. A perfect model would mean that the score for r is equal to 1 (perfect positive correlation). Pearson's r assumes a linear relationship between two variables, which are approximately normally distributed. Via a scatter plot the linearity between two variables can be indicated and the data can be tested on normality. LGDs and LR, however, generally do not have a normal distribution, which is shown in Chapter 4. Via Eq. (2.27) the statistic can be computed for n pairs of X and Y .

$$r_{XY} = \frac{n \sum XY - \sum X \cdot \sum Y}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}} \quad (2.27)$$

A simplification of Pearson's r is **Spearman's ρ** (or Spearman's rank correlation test), which only requires the ordinal position of each variable as input. It is a non-parametric test, which does not make any assumptions on the underlying distributions of the variables. The difference between the respective ranking of a paired observation X_i and realization Y_i (in the case of LGDs) is d_i . This is done for all n pairs. Eq. 2.28 shows how the Spearman rank correlation can be computed (as found in Cohen *et al.*, 2003). In the case of a perfect ranking for the observed and realized LGDs the index score equals 1. It should be noted that this does not necessarily mean that the estimated and realized LGDs are the same as the accuracy of the LGD model can be wrong, but still provide perfect discriminatory power (Loterman *et al.*, 2012).

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (2.28)$$

A different approach to measure discriminatory power is **Kendall's τ_b** . This summary statistic is a non-parametric measure of association, which can be computed by determining the number of concordances and discordances in paired observations (PSU, 2016). Suppose we have two paired observations, namely $(\text{LGD}_i^{\text{EST}}, \text{LGD}_i^{\text{REA}})$ and $(\text{LGD}_j^{\text{EST}}, \text{LGD}_j^{\text{REA}})$. $\text{LGD}_i^{\text{EST}}$ is the estimated LGD, while $\text{LGD}_i^{\text{REA}}$ is the realized LGD (or LR). If the statements (1) or (2) satisfy for two observations then a pair is concordant, if the statements (3) or (4) satisfy then the pair is discordant (PSU, 2016).

$$(1) \text{LGD}_i^{\text{EST}} < \text{LGD}_j^{\text{EST}} \text{ and } \text{LGD}_i^{\text{REA}} < \text{LGD}_j^{\text{REA}}$$

- (2) $LGD_i^{EST} > LGD_j^{EST}$ and $LGD_i^{REA} > LGD_j^{REA}$
- (3) $LGD_i^{EST} < LGD_j^{EST}$ and $LGD_i^{REA} > LGD_j^{REA}$
- (4) $LGD_i^{EST} > LGD_j^{EST}$ and $LGD_i^{REA} < LGD_j^{REA}$

It is also possible that a pair is tied, which leads to statement (5).

- (5) $LGD_i^{EST} = LGD_j^{EST}$ or $LGD_i^{REA} = LGD_j^{REA}$

Via Eq. 2.29 the total number of pairs that can be constructed for n observations is computed, which can be decomposed into five quantities (PSU, 2016). Eq. 2.30 shows this. For this equation P represents the number of concordant pairs, Q the number of discordant pairs, X_0 if the pair is tied for the estimated LGDs, Y_0 if the pair is tied for the realized LGDs and $(XY)_0$ if the pair is tied on both the estimated and realized LGDs.

$$N = \frac{1}{2}n(n - 1) \tag{2.29}$$

$$N = P + Q + X_0 + Y_0 + (XY)_0 \tag{2.30}$$

Once all variables are known, it is possible to compute Kendall's τ_b , which is shown in Eq. 2.31.

$$\tau_b = \frac{P - Q}{\sqrt{(P + Q + X_0)(P + Q + Y_0)}} \tag{2.31}$$

2.7 Current Approach

XX
 XX
 XX
 XX
 XX
 XX
 XX
 XX

2.8 Conclusion

We set out to establish a general understanding of risk models and the challenges associated with the LGDs in particular. The equations for determining the regulatory capital (2.1 - 2.4) demonstrate that the LGD estimations have a huge impact. Wrong estimates can lead to situations in which not all the allocated capital can absorb the potential risks or too much capital is held, which leads to opportunity costs (e.g. the capital can not be lent out or used

Implied Gini

XXX
XXX
XXX
XXX
XXX
XXX
XXX

This chapter has the following outline:

- Section 3.1:** We describe the motive behind the development of the implied Gini.

- Section 3.2:** We describe how the implied Gini works and is determined. Furthermore, it creates an overview of the underlying assumptions that have been made for the implied Gini.

- Section 3.3:** We describe the relation with an approach to assess the potential discriminatory power of PD models.

- Section 3.4:** We conclude with the answers to the sub-questions 2.1 and 2.2.

3.1 Roots of the Implied Gini

XXX
XXX
XXX
XXX
XXX
XXX
XXX

3.2 Relation with the Expected AR for PD Models

Before the implied Gini for an LGD model had been developed by the bank, a similar approach has been developed for PD models. This approach for PD models relies on fever

assumptions and it is generally accepted that the ODR can be approximated via a binomial distribution. Via a simulation approach the bank developed the so-called implied Gini for PD models, which serves the same purpose as the approach that is developed for LGD models. Research by Blochwitz *et al.* (2005) and Hamerle *et al.* (2003) show that similar results can be achieved analytically. From the implied Gini for PD models the idea arose that something similar could be done for LGD models.

XX
 XX
 XX
 XX
 XX
 XX
 XX

In Chapter 2 it is stated that the AR normally has a value between 0% and 100%. The implied AR shows that achieving an AR of 100% for your PD model is not realistic even for a model that correctly estimates the PD. The implied AR tells us the distribution of ARs we might expect given that our model correctly estimates the PD for each bucket or rating. In other words the ODR corresponds with the PD for each bucket or rating. Furthermore, it is shown that each portfolio has its own distribution of ARs, which indicates that the implied AR is portfolio dependent.

Let's assume that a 'perfect' PD model would be able to predict the right number of defaults for each rating (e.g. if 100 observations are assigned to the rating of 1% then it is expected that on average one default is observed). This also implies that the PD model holds 'perfect' discriminatory power for the possible ratings as you are able to separate debtors in terms of risk. In other words each PD corresponds with the ODR. The way in which the AR is determined, however, assumes that if a model holds perfect discriminatory power all defaults are given the 'worst' PD estimation. Thus, if you rank from 'worst' to 'best' based on the PD estimation the defaults and non-defaults are perfectly separated. It would mean that you know in advance, which debtor is going to default, which is not realistic (then you would not have given the loan). Thus, as it is highly unlikely that a perfect separation between defaults and non-defaults is possible, we assume that a 'perfect' model is a model where the PD corresponds with the ODR in each bucket or rating.

It is possible that the ODR deviates $x\%$ from the PD in a bucket. If each bucket or rating has the same deviation (thus ODR deviates $x\%$ from the PD for each bucket or rating) then the discriminatory power is expected to be the same compared to the 'perfect' model as the relative ranking remains intact.

Whether a debtor goes into default or not can be viewed as a process that follows a binomial distribution for which the PD of the debtor is the probability of 'success' (default) or 'failure' (non-default). Thus, we can describe the process of whether debtor X_i goes into default as a random variable, which follows a binomial distribution (3.1) for which PD_i is the respective PD of debtor i for $i = 1, 2, \dots, n$.

$$X_i \sim B(1, PD_i) \tag{3.1}$$

For the bucketing approach each X_i is assigned to a bucket j if $PD_i > PD_{j-1}$ and $PD_i \leq PD_j$. PD_j represents the probability of default for each bucket j , while PD_0 equals 0. For

the bucketing approach used at the bank $j = 1, \dots, 20$. Each debtor i assigned to bucket j gets as an estimate PD_j . To make this distinction clear we define this process as a random variable Y_i , which denotes that an individual debtor i assigned to bucket j goes into default or not as binomial distribution with PD_j (3.2).

$$Y_i \sim B(1, PD_j) \tag{3.2}$$

For each random variable Y_i we draw from a 'random experiment', which decides whether debtor i went into default or not. This results in a data set, which consists of paired observations with PD_i which is the original estimate of debtor i and a_i which represents whether debtor i defaulted ($a_i = 1$) or not ($a_i = 0$). For this data set it is possible to compute an AR.

This data set represents one experiment, and in order to construct the implied AR multiple experiments are conducted. For each experiment an AR is computed, which results in a distribution of ARs, which we denote as the implied AR. The average of the distribution is denoted as the sample mean AR, for which an analytic approach is addressed at the end of this section. Hence, each AR resulting from an experiment k is denoted as Z_k . The sample mean AR can be computed via (3.3), while the implied AR for the PD model can be made visible by plotting all k ARs in a histogram.

$$\bar{Z} = \frac{1}{k} \sum Z_k \tag{3.3}$$

A rule of thumb is that for the range $\pm 3\sigma$ of the mean almost all possible outcomes of a distribution (for a normal distribution 99,73%) are included. Therefore, the sample variance should be computed via (3.4).

$$s^2 = \frac{1}{k-1} \sum (Z_k - \bar{Z})^2 \tag{3.4}$$

XX
 XX
 XX
 XX
 XX
 XX
 XX

The implied AR is very dependent on the ratings and the number of ratings are used in the PD model. Furthermore, the distribution of the debtors over the ratings also heavily influence the resulting implied AR. This is illustrated via an example used by Blochwitz *et al.* (2005). In Table 3.1 we have illustrated an hypothetical PD model, which assign debtors to two different ratings, namely 1% and 5%. We have made a split between a development sample and a validation sample.

Suppose that the model performs accurately and is calibrated sufficiently. For that case the observed number of defaults corresponds with the estimated probabilities (see Table 3.2). For both the development and validation sample it is possible to determine the indicative benchmarks. For the development sample we would expect on average an AR of 37.12% for

Table 3.1: Assigned debtors per rating and sample.

	PD 1%	PD 5%
Development Sample	800	600
Validation Sample	200	400

the bounds (16.92%, 57.33%). The validation sample, however, has a different expectancy and indicative benchmarks. The implied AR for validation sample is 25.15% within the range (5.96%, 44.36%).

Table 3.2: Assigned debtors per rating and sample.

	Defaults PD 1%	Defaults PD 5%
Development Sample	8	30
Validation Sample	2	20

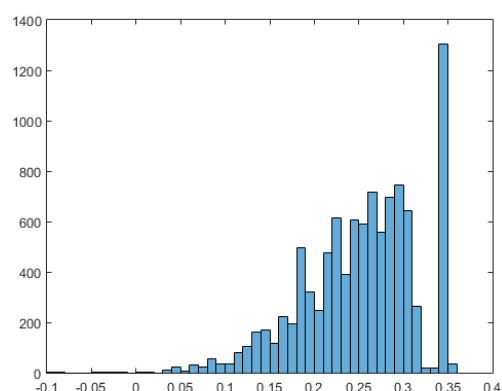
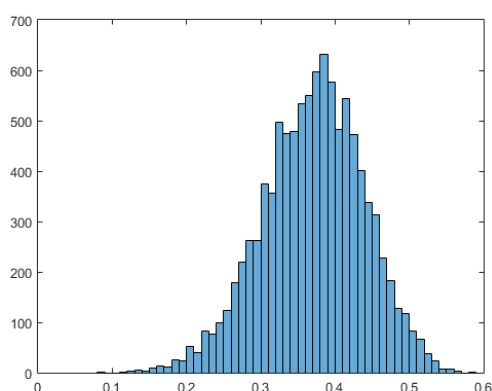


Figure 3.1: The histogram of resulting ARs for the development sample. **Figure 3.2:** The histogram of resulting ARs for the validation sample.

This example shows that the underlying estimation model as well as the proportion of assigned debtors to a rating determine the implied AR. Therefore, it is difficult to make comparisons on models based on the AR. This is also noticed and argued by Blochwitz *et al.* (2005). They argue that the AR is non-comparable in the following situations:

1. Different portfolios, same time period.
2. Same portfolio, different time periods.
3. Different portfolios, different time periods.

They state, however, that it is possible to make comparisons between rating models, which have the same underlying portfolio from the same time period. In practice this is meaningful when multiple rating models are developed for the same portfolio and the performances of

Table 3.3: Implied AR for different calibrations.

		ODR ₂									
		0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
ODR ₁	0.01	0.00	0.17	0.27	0.33	0.37	0.40	0.43	0.45	0.47	0.48
	0.02	-0.16	0.00	0.10	0.18	0.23	0.27	0.31	0.34	0.36	0.38
	0.03	-0.23	-0.10	0.00	0.07	0.13	0.18	0.22	0.25	0.28	0.31
	0.04	-0.28	-0.16	-0.07	0.00	0.06	0.11	0.14	0.18	0.21	0.24
	0.05	-0.31	-0.21	-0.13	-0.06	0.00	0.05	0.09	0.12	0.16	0.18
	0.06	-0.33	-0.24	-0.16	-0.10	-0.04	0.00	0.04	0.08	0.11	0.14
	0.07	-0.35	-0.26	-0.20	-0.14	-0.09	-0.04	0.00	0.03	0.07	0.10
	0.08	-0.36	-0.28	-0.22	-0.17	-0.12	-0.07	-0.03	0.00	0.03	0.06
	0.09	-0.37	-0.30	-0.24	-0.19	-0.15	-0.10	-0.07	-0.04	0.00	0.03
	0.10	-0.38	-0.32	-0.26	-0.22	-0.17	-0.13	-0.09	-0.06	-0.03	0.00

those ratings models need to be compared. Furthermore, it is argued by Blochwitz *et al.* (2005) that it is not meaningful to describe the quality of the rating system solely via its discriminatory power.

In our example from Table 3.2 the two populations cannot be compared as the two data sets result in non-comparable distributions over the ratings. In order to check whether the population distribution over the ratings are comparable it is possible to use the so-called Population Stability Index (PSI). If we denote each possible rating as i , the two populations we wish to compare as P_1 and P_2 with respective number of observations n and m , it is possible to compute the relative frequency of P_1 and P_2 denoted respectively as $F_{1,i}$ and $F_{2,i}$ for each rating i via Eq. (3.5). The observed frequencies of P_1 and P_2 in rating i is denoted respectively as n_i and m_i .

$$F_{1,i} = \frac{n_i}{n} \quad \text{and} \quad F_{2,i} = \frac{m_i}{m} \quad (3.5)$$

If the relative frequencies for each rating have been determined, it is possible to compute the PSI value via Eq. (3.6).

$$\text{PSI} = \sum \left((F_{1,i} - F_{2,i}) \cdot \ln \frac{F_{1,i}}{F_{2,i}} \right) \quad (3.6)$$

According to the internal guidelines of the bank the PSI is considered to be 'good' if it is less than 0.1. It is considered to be 'medium' if it lies between 0.1 and 0.25, while if it is larger than 0.25 it is considered to be 'bad'. For our example the PSI value is 0.2335, which is close to being classified as 'bad' according to internal guidelines. It illustrates that there is a major shift within the population and hence that the two populations are non-comparable.

Besides that one should carefully check whether the populations are comparable, one should also take into account that the ODR of a rating can be different compared to the PD. If the ODR is different it can affect the implied AR, and therefore discriminatory power cannot be

seen independently from calibration. This is illustrated with the example model described in Table 3.2. We denote the ODR of rating i as ODR_i . In Table 3.3 the different expected ARs can be found if the calibration is different than the estimated PDs. For example if the first rating has an ODR of 2% (PD was 1%) and the second rating has an ODR of 8% (PD was 5%) then the implied AR is 34%.

If it turns out that the PD model is not correctly calibrated, then the indicative benchmarks for measuring discriminatory power should also be re-calibrated.

Analytic Approach for the Expected AR of a PD Model

Previously, we discussed an simulation approach to determine the implied AR for PD models. In this subsection we describe the analytic approach to determine the expected AR of a perfectly calibrated model, which has been proven by Hamerle *et al.* (2003). This analytic approach to determine the expected AR corresponds with the mean of the implied AR.

Hamerle *et al.* (2003) show that the AR can be notated as the following:

$$AR = \frac{1}{1 - \bar{\lambda}} \cdot Gini \quad (3.7)$$

For which $\bar{\lambda}$ is described as the average default probability of the data set. The term *Gini* is defined as:

$$Gini = 1 + \frac{1}{N} - \frac{2}{N^2 \cdot \bar{\lambda}} \cdot (N \cdot \lambda_1 + (N - 1) \cdot \lambda_2 + \dots + \lambda_N) \quad (3.8)$$

where $\lambda_1, \dots, \lambda_N$ are the ordered default probabilities from the lowest to the highest and N is total number of debtors in the data set.

If we compare the outcomes with the results of the average value from the implied AR approach, we see that they are quite close. For the example from Table 3.1 and 3.2 we see that the analytic AR is respectively 37.10% and 25.17%, while the averages from the implied AR are respectively 37.12% and 25.15%.

Practical Use

Under the assumption that a rating model performs sufficiently if the PD estimates for each rating correspond with the ODR, then it is possible to determine the expected and implied AR. The suggestion is to check whether the realized AR of a model falls within the ranges of the implied AR, which are used as indicative benchmarks. However, it is required to check whether the model is calibrated correctly, which means that the ODR corresponds (statistically) with the PD for each rating. If the ODR does not correspond with the PD estimates for each rating the resulting AR is not comparable with the indicative benchmarks and the causes for the difference(s) need to be researched. If a model is re-calibrated, then the indicative benchmarks should be re-calibrated as well.

It is also indicated by Blochwitz *et al.* (2005) that it is only meaningful to compute the discriminatory power (via the AR) if the underlying portfolio and time period is the same. Hence, for the development of new PD models it is meaningful to compare different modeling approaches via the AR in order to see which one has more discriminatory power. Furthermore,

Validity Tests

This chapter tests the validity of all the assumptions that have been made for the construction of the implied Gini as it is proposed by the bank in order to assess the maximum achievable discriminatory power of an LGD model. As most of the assumptions are related to the assumption that a loss distribution follows a Beta distribution, we firstly discuss the characteristics of this distribution in detail and the role of the distribution within credit modelling. Secondly, the tests for fitting data to a specific distribution are formulated as well as tests to check whether data comes from a specific distribution. These tests are used to test the assumptions made by the bank. Finally, the test results are discussed and a verdict on the validity of the implied Gini is given. In addition, it is implicated that making adjustments to the banks approach for assessing the maximum achievable discriminatory power of an LGD model is difficult in its current form.

It should be noted that not each assumption can be tested in a straightforward matter. In that case a theoretical assessment is made, which demonstrates whether an assumption can be discarded beforehand or more research is required.

This chapter has the following outline:

Section 4.1: We discuss the characteristics of the Beta distribution.

Section 4.2: We discuss typical loss distributions used in the modelling of LGDs.

Section 4.3: We outline which tests are used in the validity test of the Implied Gini.

Section 4.4: We present the test results for each underlying assumption of the implied Gini.

Section 4.5: We discuss the implication of the test results and conclude with a verdict on the validity of the current approach of the implied Gini.

4.1 The Beta Distribution

As it is discussed in Chapter 3, the implied Gini mainly relies on the assumption that the loss distribution of a portfolio is beta distributed. The use of a beta distribution to estimate recoveries or losses (via LGDs) is not uncommon in the credit risk modelling domain. It is,

for example, used within commercial applications such as LossCalc, which is developed by Moody's (Gupton & Stein, 2002). The beta distribution is convenient in the estimation of LGDs as a realization typically is between zero and one and due to the two parameters of the distribution a variety of shapes can be modelled.

The two-parameter probability density function of the beta distribution with the shape parameters α and β is defined via Eq. (4.1) (Owen, 2008).

$$f(x|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1}, \text{ for } 0 \leq x \leq 1, \alpha > 0, \beta > 0 \quad (4.1)$$

From Eq. (4.1) it can be seen that the beta distribution is modelled via multiple gamma functions, which are defined as $\Gamma(\cdot)$. A Gamma function is defined via Eq. (C.1) (Freeden & Gutting, 2013).

$$\Gamma(x) = \int_0^{\infty} e^{-t} \cdot t^{x-1} \cdot dt, \text{ for } x > 0 \quad (4.2)$$

Johnson & Beverlin (1970) defined the beta distribution via Eq. (4.3), which includes a beta function defined via Eq. (4.4).

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1} \cdot (1-x)^{\beta-1}}{B(\alpha, \beta)}, \text{ for } 0 \leq x \leq 1, \alpha > 0, \beta > 0 \quad (4.3)$$

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} \cdot (1-t)^{\beta-1} \cdot dt, \text{ for } \alpha > 0, \beta > 0 \quad (4.4)$$

The beta function is equal to a ratio of gamma functions, which is described in Eq. (4.5).

$$B(\alpha, \beta) = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha + \beta)} \quad (4.5)$$

Johnson & Beverlin (1970) define the expected value and the variance of the beta distribution via the two parameters of the distribution. These are defined via Eq. (4.6) and Eq. (4.7).

$$E(x) = \frac{\alpha}{\alpha + \beta} \quad (4.6)$$

$$Var(x) = \frac{\alpha \cdot \beta}{(\alpha + \beta)^2 \cdot (\alpha + \beta + 1)} \quad (4.7)$$

According to Owen (2008) the shape of the beta distribution can change dramatically with the changes of the parameters. Owen (2008) lists a variety of possible shapes for the distribution for different values of the parameters:

- (1) $\alpha = \beta$: the distribution is unimodal and symmetric about 0.5. A special case is when $\alpha = \beta = 1$, which is equivalent to the Uniform (0,1) distribution.

- (2) $\alpha > 1$ and $\beta > 1$: the distribution is unimodal and skewed. The single mode of the pdf as $x = (\alpha - 1)/(\alpha + \beta - 2)$. The single mode defines the value at which the pdf attains its maximum value (highest probability). The distribution becomes positively skewed if β is greater than α , while it becomes negatively skewed if it is the opposite.
- (3) $\alpha = \beta < 1$: the distribution is *U*-shaped and symmetric about 0.5.
- (4) $\alpha < 1$ and $\beta < 1$: the distribution is *U*-shaped and skewed. The anti-mode (the lowest point of the pdf) is defined as $x = (\alpha - 1)/(\alpha + \beta - 2)$.
- (5) $\alpha > 1$ and $\beta \leq 1$ the distribution is strictly increasing and it is *J*-shaped without having a mode or anti-mode. If the values are the opposite ($\alpha \leq 1$ and $\beta > 1$) then the distribution is strictly decreasing and has a reverse *J*-shaped distribution.

In Fig. 4.1 and Fig. 4.2 some examples of the possible shapes as defined by Owen (2008) can be found.

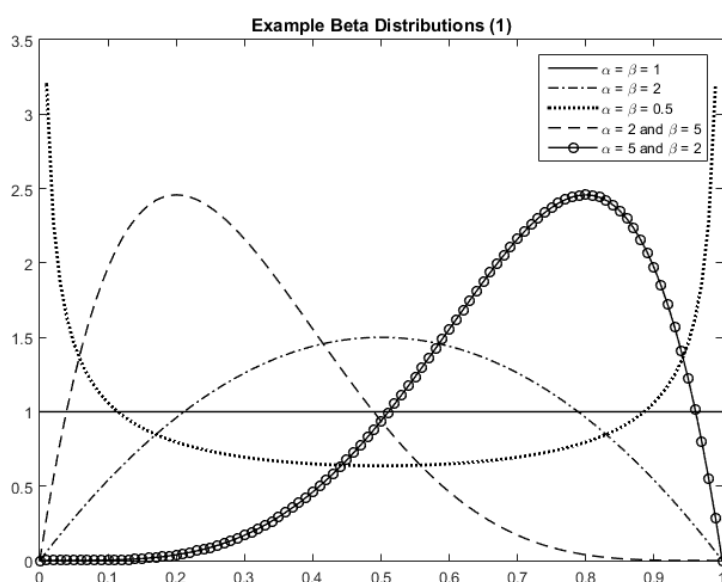


Figure 4.1: Examples of different beta distributions (1).

Fig. 4.1 and Fig. 4.2 illustrate that a lot of possible shapes for the beta distribution can be obtained. The data of the beta distribution is measured on the open interval $(0, 1)$. Ospina & Ferrari (2010) argue that if the data set contains a lot of zeroes and/or ones continuous distributions are not suitable for modelling the data. Unfortunately, the inclusion of zeroes and ones is frequently the case when assessing the LRs of a loss distribution. They propose to use a mixed continuous-discrete distribution to model the data that is observed on the intervals $[0,1)$, $(0,1]$ or $[0,1]$. If the data contains zeroes or ones (but not both) then they use a mixture of two distributions, namely a beta distribution and a deterministic distribution. The deterministic distribution is modelled in a known value c , where $c = 0$ or $c = 1$, which depends on the data set. If we define q as the mixture parameter, then it is possible to define the corresponding probability density function as follows:

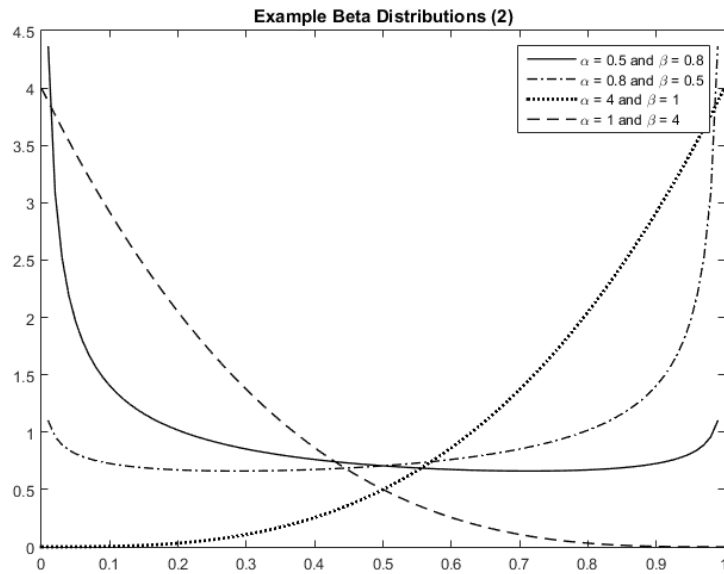


Figure 4.2: Examples of different beta distributions (2).

$$bi_c(x|q, \alpha, \beta) = \begin{cases} q, & \text{if } y = c \\ (1 - q) \cdot f(x|\alpha, \beta), & \text{if } x \in (0, 1) \end{cases} \quad (4.8)$$

A special case that Ospina & Ferrari (2010) describe is the case if the data contains zeroes as well as ones. For that particular situation they propose to use a mixture between the beta distribution and the Bernoulli distribution. The probability mass function of the Bernoulli distribution with probability p is defined via Eq. (4.9).

$$f(x|p) = p^x \cdot (1 - p)^{1-x}, \text{ for } x \in \{0, 1\} \quad (4.9)$$

Ospina & Ferrari (2010) describe this particular case as the zero-and-one-inflated beta distribution and the probability density function of this particular case is defined via the following equation:

$$beinf(x|p, q, \alpha, \beta) = \begin{cases} q(1 - p), & \text{if } x = 0 \\ pq, & \text{if } x = 1 \\ (1 - q) \cdot f(x|\alpha, \beta), & \text{if } x \in (0, 1) \end{cases} \quad (4.10)$$

Parameter Estimation

For the implied Gini it is necessary to determine whether the usage of the general structure of the beta distribution is correct. A mixture model such as the inflated beta distribution or perhaps a totally different probability distribution might describe the nature of current data sets better. Furthermore, it is necessary to determine an approach for the estimation of the parameters of the (inflated) beta distribution. Ospina & Ferrari (2010) recommend to use the maximum likelihood estimation (MLE), while Owen (2008) researched the impact of different estimation methods in which she among others concluded that the MLE and method of moments (MOM) perform sufficient in most situations. However, she states that MOM is more straightforward than MLE. A further complication is that MLE is quite sensitive. In her case the MLE was among others outperformed by the MOM in the case the sample size was small. She argues that it would have likely done fine if it was not required to use an iterative method to find estimators for α and β . She uses the Newton-Raphson method, which is an iterative approach for root finding. Iterative approaches are quite sensitive to the starting values for determining the root to which it converges.

Another approach is estimating the parameters via a chi-square function. The idea is to minimize the chi-square value for different combinations of parameters applied in the beta distribution. Berkson (1980) (re-)introduced this idea for different chi-square functions. Berkson (1980) suggests to use it as an approach to determine estimators for a probability distribution in some cases, but it is disputed whether it always provides better estimators than the MLE by other academics. By applying the chi-square function it is possible to retrieve estimators for the two parameters of the beta distribution. Via a simulation we can compute the chi-square value for each parameter combination under consideration. The minimum chi-square value can be considered to be estimators of a fitted beta distribution for a particular data set of losses.

As MOM is not outperformed by MLE in Owen (2008) and is more straightforward, we use MOM and the minimum chi-square value to estimate parameters for a beta distribution. The data sets for which we estimate parameters range have a minimum of 200 data points. The estimation of the parameters for a beta distribution do not tell whether the data actually follows a beta distribution. Therefore, we need to apply a goodness-of-fit test to check whether the data is actually beta distributed. These tests are discussed in Section 4.3.

Method of Moments

For the method of moments it is only required to have the sample mean and variance, which are easily obtained from the data. The moment generating function of order t for a beta distribution is defined as follows (Owen, 2008):

$$E(X^t) = \frac{\Gamma(\alpha + t)\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + t)\Gamma(\alpha)} \quad (4.11)$$

Then it is possible to derive the first and second moment, namely:

$$E(X) = \frac{\Gamma(\alpha + 1)\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + 1)\Gamma(\alpha)} \quad (4.12)$$

$$E(X^2) = \frac{\Gamma(\alpha + 2)\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + 2)\Gamma(\alpha)} \quad (4.13)$$

In order to solve these equations we have to make use of an useful property of the gamma function. In Gowers *et al.* (2008) it is shown that the gamma function has the following property:

$$\Gamma(x) = (x - 1) \cdot \Gamma(x - 1), \text{ for all } x > 1 \quad (4.14)$$

Evidently, we can derive $\Gamma(x+1)$ and $\Gamma(x+n)$ via integration by parts for which the complete derivation as shown in Appendix C. The statements for $\Gamma(x+1)$ and $\Gamma(x+n)$ are used to solve Eq. (4.12) and Eq. (4.13). The results are the first and second moment of the moment generating function:

$$E(X) = \frac{\alpha}{\alpha + \beta} \quad (4.15)$$

$$E(X^2) = \frac{\alpha \cdot (\alpha + 1)}{(\alpha + \beta + 1)(\alpha + \beta)} \quad (4.16)$$

In Appendix D the use of the statements for $\Gamma(x+1)$ and $\Gamma(x+n)$ for determining Eq. (4.15) and Eq. (4.16) is shown. From these moments it is possible to determine a statement for the variance in terms of the α and β parameter for the beta distribution. We know that:

$$\text{Var}(X) = E(X^2) - (E(X))^2 \quad (4.17)$$

This means that the variance of a beta distribution can be determined via the parameters as follows:

$$\text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \quad (4.18)$$

Owen (2008) shows that the parameters α and β can be rewritten into terms of the sample mean and the sample variance. We denote the sample mean as \bar{X} and the sample variance is determined via Eq. (4.19) for the population X_1, X_2, \dots, X_n with n instances.

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (4.19)$$

Owen (2008) determined that the estimators for the beta distribution's parameters via the method of moments in terms of the sample mean and the sample variance are the following:

$$\hat{\alpha}_{\text{MOM}} = \bar{X} \cdot \left(\frac{\bar{X} \cdot (1 - \bar{X})}{S^2} - 1 \right) \quad (4.20)$$

$$\hat{\beta}_{\text{MOM}} = (1 - \bar{X}) \cdot \left(\frac{\bar{X} \cdot (1 - \bar{X})}{S^2} - 1 \right) \quad (4.21)$$

Pearson's Chi-Squared Test

Law (2007) states that the chi-square test is one of the oldest goodness-of-fit tests. He states that the chi-square test can be thought of as a more formal comparison of a histogram with the fitted density or mass function. In order to estimate parameters for the beta distribution, we compute the test statistic of Pearson's chi-squared test for a variety of parameter α and β combinations. The minimum test statistic achieved for these tests can be considered as the parameter combination that is the best fit with the data (on the assumption that the data is beta distributed). Solely, the determination of the minimum test statistic does not tell whether the data is actually beta distributed. In order to test whether the estimated parameters that lead to the minimum test statistic are actually beta distributed, we can compare these values against the critical value from the chi-squared distribution, which depends on the df degrees of freedom and confidence interval. The minimum chi-squared statistic and critical value are, however, very dependent on the number of bins and outliers in the data set. Alternative approaches to test the estimated parameter combinations on which we elaborate in Section 4.3.

Suppose we would like to find the minimum test statistic for a sample data set (S). S consists of n random variables, which we define as X_1, X_2, \dots, X_n . We hypothesize (our null hypothesis) that these random variables follow a beta distribution function \hat{F} with parameters $\hat{\alpha}$ and $\hat{\beta}$. Hence, our null hypothesis is:

$$H_0: \text{The } X_i\text{'s are independent and identically distributed random variables distributed via } \hat{F}(x|\hat{\alpha}, \hat{\beta}).$$

For the purpose of minimizing the test statistic we vary the parameters $\hat{\alpha}$ and $\hat{\beta}$ in predefined ranges and step sizes. It should be noted that for the purpose of the parameter estimation we do not determine whether the null hypothesis should be rejected.

In order to compute the chi-square test statistic it is necessary to divide the entire range of the fitted distribution into k adjacent intervals, namely $[a_0, a_1), [a_1, a_2), \dots, [a_{k-1}, a_k)$ (Law, 2007). Each observation is then tallied for each interval (Law, 2007):

$$N_j = \text{number of } X_j\text{'s in the } j\text{th interval } [a_{j-1}, a_j), \text{ for } j = 1, 2, \dots, k \quad (4.22)$$

The next step is to determine the expected proportion p_j of the X_i 's that would fall in the j th interval if a sample from the fitted distribution would have been used (Law, 2007):

$$p_j = \int_{a_{j-1}}^{a_j} \hat{f}(x|\hat{\alpha}, \hat{\beta}) \cdot dx \quad (4.23)$$

In Eq. (4.23) \hat{f} is the probability density function of the fitted distribution. Law (2007) describes that the test statistic can be determined via:

$$\chi^2 = \sum_{j=1}^k \frac{(N_j - np_j)^2}{np_j} \quad (4.24)$$

As described in Berkson (1980) there are multiple approaches for determining a chi-square test statistic, but for the estimation of the parameters via a chi-square function we limit ourselves to the approach developed by Pearson. In Berkson (1980) the resulting estimators via different approaches varied little. The main difficulty is to find appropriate interval lengths with sufficient observations. A rule of thumb is that each interval should at least contain five observations. If insufficient observations are present within an interval then intervals should be combined.

4.2 Loss Distributions for LGD Estimation

The choice to use the beta distribution for estimating a typical loss distribution is not uncommon. We addressed that normally the beta distribution is bounded on the interval $(0, 1)$, which makes it convenient to simulate an LGD or a RR. The basic assumption is of course that you can either lose nothing ($\text{LGD} = 0$) or lose the total value of the outstanding credit ($\text{LGD} = 1$). However, from practice we have seen that this is not always the case as values outside the interval of the beta distribution can occur. Nevertheless, the use of the beta distribution is generally accepted for estimating LGDs as the beta distribution can adapt to a variety of shapes.

The use of the beta random variables for loss estimation is reflected in practice. *LossCalc*, which is Moody's model for predicting LGDs for bonds, loans, and preferred stock, assumes that the actual distribution of recoveries is best reflected by a beta distribution (Gupton & Stein, 2002). However, a variety of distribution and techniques are used to model losses and recoveries, which not always rely on the beta distribution. Frye (2000) for example assumes that the recoveries are dependent on the state of the economy and normally distributed.¹⁰ As the normal distribution can take on negative values it might not always be convenient for the estimation of LGDs. Huang & Oosterlee (2011) reflect on extensions that can cope with the (in principal) nonnegativity of LGDs found in the literature. They conclude, however, that those models do not have the convenient economic interpretation of the parameters as in Frye (2000) and the models require that the transformed LGD is symmetric

¹⁰The inclusion of systematic risk and the state of the business cycle allows the computation of the 'downturn' LGD, which is a value that should reflect the occurring losses during stress scenarios or a downturn. Calabrese (2014) assumes that an LGD is a mixture of an expansion and recession (loss) distribution and that both distributions are distributed via a mixture of a Bernoulli random variable and a beta random variable (inflated beta distribution). We do not cover downturn LGD estimation as it is outside the research's scope.

and homoskedastic, which sometimes is contradicted by with empirical studies. Therefore, Huang & Oosterlee (2011) propose to use generalized beta regression (GBR) models for modelling LGDs (with the inclusion of systematic risk) for credit portfolio losses, for which the assumption is made that an LGD always is (conditionally) beta distributed.

Although models found in literature frequently rely on the beta assumption for LGD estimation, there are implications that the recovery distributions (and therefore loss distributions) are not beta distributed. Renault & Scaillet (2004) researched the nonparametric estimation of the recovery or loss distribution of defaulted bonds via the beta kernel approach. They conclude that (against common practice) the recovery distribution is not beta distributed.

If X_1, \dots, X_n is a random sample with unknown probability density function f on the unit interval $[0,1]$, then the beta kernel estimator of f at point x is defined as follows (Renault & Scaillet, 2004):

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K\left(X_i, \frac{x}{b} + 1, \frac{(1-x)}{b} + 1\right) \quad (4.25)$$

The asymmetric kernel $K(\cdot)$ the probability density function of the beta distribution, which is defined in Eq. (4.1). The variable b is called the bandwidth and functions as a smoothing parameter (Renault & Scaillet, 2004). Chen (1999) has determined that the optimal bandwidths are $O(n^{-2/5})$ for beta kernel estimators. Renault & Scaillet (2004) apply a rule of thumb to determine the bandwidth. They multiply the standard deviation of the empirical distribution of the observed data with $n^{-2/5}$.

Calabrese & Zenga (2010) propose to consider the recovery rate as a mixed random variable and to estimate the density function by a mixture of beta kernels estimators. A mixed random variable is simply a random variable that is a mixture between a Bernoulli random variable and a continuous random variable X on the unit interval (Calabrese & Zenga, 2010). An example of a mixed random variable is a draw from the inflated beta distribution as described by Osipina & Ferrari (2010), which has beta distributed continuous variable. Calabrese & Zenga (2010) define the mixture of beta kernels estimator as:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K\left(x, \frac{X_i}{b} + 1, \frac{(1-X_i)}{b} + 1\right) \quad (4.26)$$

Similar to Renault & Scaillet (2004), Calabrese & Zenga (2010) find that, although the end-points of the data set are removed (the zeros and ones), the estimated probability function of their data set cannot be replicated by the beta density function.

Sometimes an LGD is not estimated directly, but the actual loss amount is estimated. In this case the distribution is not necessarily required to return estimates on the interval $[0,1]$, but it returns a value, namely the estimated losses of a loan. It is possible to compute the LGD expressed as a percentage by dividing the estimated losses by the total value of the loan. Tong *et al.* (2013) used such an approach to model residential mortgage losses. They fit a zero adjusted gamma model to the loss distribution, which is similar to the inflated

An Alternative Approach

The results from our validity test suggest that the current approach of the implied Gini is not valid. Also, due to structure of the approach, the implied Gini is difficult to modify such that it can be implemented as a valid approach. Hence, we need to develop an approach that is able with the current available data to support the assessment of an LGD model's potential discriminatory power.

For the development of an alternative approach we have to make choices due to the availability of data and cope with the current data quality. For the development of the alternative approach we limit ourselves by using the AR as a summary statistic. The main reasons for this choice are that for other summary statistics similar questions remain (e.g. "what are acceptable values?"), and that the usage of the AR is common practice at the bank (and industry).

Within the process of developing an alternative approach we have to cope with several issues that cannot be solved due to the lack of data availability. Therefore, we have to make choices, which can be considered to be non-optimal. As we have described in Chapter 3 the implied Gini developed by the bank only uses LGD estimates to determine a model's potential discriminatory power. Therefore, we only have these LGDs available with the corresponding realizations, the LRs. We acknowledge that in principle an LGD is an expected value of a probability distribution based on various drivers or variables of a debtor. Unlike PD models, we do not know the exact relationship between an estimated LGD and a realized LR. In the case of PD models we know that this relationship is always binomial. For LGDs frequently the assumption is made that this relationship is beta on a portfolio level, but from our empirical data we concluded that this does not always hold true for each portfolio.

If we would like to assess this relationship more in-depth, we would need additional information concerning the LGDs (aside from the fact that the number of debtors in some cases is limited). The individual variables or drivers that led to an estimated LGD for an instance are, however, difficult to obtain or not available anymore. If the original drivers were available, we could have chosen to create within a portfolio different classes of debtors, which all have their own specific characteristics. Then we would be able to determine a specific probability distribution for the LRs of a class of debtor. The current situation concerning our data sets is that everything is aggregated within one portfolio (based on product class or type e.g. mortgages) without knowing the individual drivers behind the LGDs. Therefore, we cannot truly assess in an ideal situation world situation what a 'perfect' model estimation of an LR would be. Given our situation we can only state that in the case of a 'perfect' model you would like to have estimated the loss perfectly, hence the LGD equals the LR. Strictly taken

this is of course not ideal. We argue, however, that regardless of this limitation, we can still prove that the current review approach of the bank can be improved by taking into account different portfolio characteristics. The alternative approach for the implied Gini we developed in this research still indicates that setting the same fixed threshold value for each portfolio is unnecessary penalizing particular portfolios. Our approach gives more insight in realistic AR values for a specific portfolio compared to the current situation. We discuss our developed approach in this chapter and provide additional insights in explanatory factors that influence the performance of an LGD model. This chapter has the following outline:

Section 5.1: We describe our proposed alternative approach, which we have developed in order to cope with the current data availability. We explain various choices we have made and why our approach improves the current situation. Furthermore, we discuss the trade-off between the practicality of the approach and the 'true' perfect model.

Section 5.2: We discuss the mathematical formulation of our approach and how we can simulate an *implied AR* for a specific model. We only do this for LGD models, which are of a structural type as the portfolios that are available for this research are of this type (see Chapter 2).

Section 5.3: We address explanatory factors for estimating LGDs, which can influence the accuracy of an LGD model and suggest an approach to incorporate the uncertainty within our proposed alternative approach, which determines the implied AR of an LGD model.

Section 5.4: We summarize and conclude with the answer to the Sub-Research Question 4. We also argue why, in our opinion, our developed approach helps with understanding the differences between different realized ARs between portfolios despite the limitations of the approach due to data issues.

5.1 The Proposed Alternative

As we have addressed before our data only contains an LGD and an LR for a debtor in a portfolio without any additional data. It is clear that portfolios have quite different characteristics regarding the distribution of losses. However, for practical reasons we would like to create a general approach that can provide insight in the performance of an LGD model on discriminatory power, does not rely heavily on probability distributions which require data that is currently not available and is easy to compute on the basis of the characteristics of a portfolio.

From the available data we can determine multiple aspects of a portfolio. It is possible to compute the variance of the realized loss distribution, we can determine the number of zero observations and by using internal bank documents the historical cure probability for a specific portfolio. These are the factors we are able to determine for the portfolios. Due to this limited amount of information the variety approaches we can develop are limited. Without additional information it is, for example, impossible to determine which debtor cured or was a default with no losses for the bank. If that were possible we would be able to split the data set in *cure cases* and *loss given loss (LGL) cases*.

With the information we have available we can, however, still provide valuable insights concerning an LGD model performance on discriminatory power. As we know the historical cure probability of a portfolio and the ratio of zeros of a data set, we can determine via a simulation whether a zero observation either cured or actually defaulted. In order to determine the 'maximum' attainable AR score we simply set the LGD equal to the observed losses for the defaulted cases. For the cure cases we draw a number from the empirical loss distribution as we would not have known in advance whether this case would cure. The approach is mathematically defined in the next section. We acknowledge that setting the LGD equal to the LR for LGL cases in practice not realistic. However, we do not know the 'true' relationship and cannot determine it based on the current situation. In Section 5.3 we assess the impact of 'extreme' relations between the LGD and LR to determine the impact on the AR.

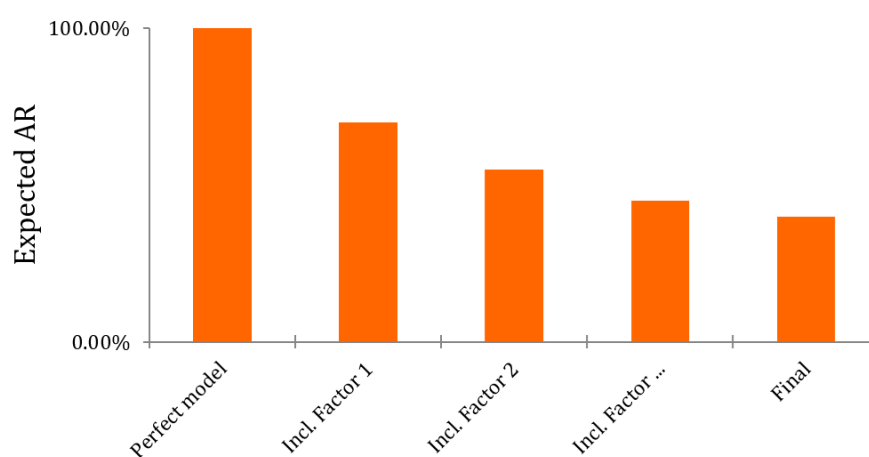


Figure 5.1: Factors influencing the AR.

If only the LGD and LR values are available and some overall portfolio statistics (e.g. historical cure or fraud rate), then we propose to use this alternative approach to get more insights in the 'maximum' attainable AR. In general it works as follows:

1. **Determine the different scenarios that can occur within a portfolio.**
2. **For each scenario determine how it impacts the realized losses.**
3. **Determine the applicable scenarios for a debtor dependent on the actual LR (e.g. if there are observed losses then a cure scenario is not applicable).**
4. **Simulate the assignment of scenarios to debtors.**
5. **Based on the assigned scenario determine the corresponding LGD.**
6. **Compute the AR.**
7. **Repeat step 4-6 a sufficient number of times and determine the average value, which we denote as the *expected AR* of a portfolio. The complete range of values for the simulated ARs for a portfolio is denoted as the *implied AR* for a portfolio.**

Currently it is only possible to assess to the impact of cures on the AR as we only have additional data on the historical cure rate of a portfolio. However, it is possible that other factors influence to AR, such as fraud cases. If we add these factors within our simulation then we can determine an expected AR based on portfolio characteristics and set more realistic thresholds for assessing the performance of an LGD model on discriminatory power. This process is visualised in Fig. 5.1.

In the following sections we discuss possible explanatory factors that can be included in our approach and the mathematical definition for including the probability of cure within our simulation model.

5.2 Explanatory Factors

This section explains the effect of the probability of cure on the AR for an LGD model. It suggests an approach to compute the implied AR for LGD models given that there are cures within the data set. The approach is first generally described for the LEA variant of LGD models used at the bank, and then for the scenario-structured LGD model. Furthermore, it addresses the influence of the variance of the loss distribution on the AR for an LGD model via logic reasoning.

XXXXXXXX

XX
 XX
 XX
 XX
 XX
 XX
 XX

Unfortunately, the data quality does not allow to filter out the cures from the actual losses as an LR of zero might also imply that simply no losses were observed (and the debtor defaulted nevertheless). These cures affect the maximum achievable score on the AR as this method assumes that in the case of a perfect model one should exactly predict the LR. Hence, in the case of a cure scenario the approach assumes that one should have predicted zero losses, while actually due to the cure scenario this is probably not true. This might lead to a wrong idea of how the model performs as the observed AR for the model will be lower due to the inclusion of cures. Thus, if we would like to measure the discriminatory power of an LGD model via the AR, it should not be measured against the 100% benchmark if the data set consists of cures. The portfolio might have a low observed AR score due to the inclusion of a lot of cures.

This is illustrated via an example. Suppose we have a data set solely consisting of non-zero LRs, which is denoted as S with n instances. In the case of the perfect model we would accurately estimate each LR in S , which means that for each instance i the statement $LGD_i = LR_i$ holds. In this particular setting the AR is equal to 100%. Suppose we would add cures to the data set, which are zero losses (hence for the cured instance j $LR_j = 0$), but normally the LGD will not be zero (there is always a floor value i.e. a minimum LGD). This is based on the assumption that you cannot predict with 100% accuracy, which facility is going to cure and becomes performing again in the case of a default. Thus, for each

XX
 XX
 XX
 XX

As we wish to determine the implied AR of a portfolio we need an approach to determine which zero loss is considered to be a cure. As we know the cure rate it is possible to make the split between cures and zero losses. The cure rate is the fraction of the portfolio that on average cures, thus the remainder of the zeroes is assumed to be the fraction of zero losses. We denote the fraction of cures for a portfolio t as frac_t^c and the fraction of a zero losses as frac_t^z . The total number of zero observations is denoted as $\text{frac}_t^{\text{TOT}}$. frac_t^z is determined via Eq. (5.2).

$$\text{frac}_t^z = \max\{\text{frac}_t^{\text{TOT}} - \text{frac}_t^c, 0\} \tag{5.2}$$

This allows us to determine the probability that a zero loss is a cure in portfolio t via Eq. (5.3).

$$p_t^c = \frac{\text{frac}_t^c}{(\text{frac}_t^c + \text{frac}_t^z)} \tag{5.3}$$

It is then possible for each observed zero j in the data set consisting of m zeroes to simulate whether it is a cure or a zero loss via the binomial distribution. Thus, we can describe the process of determining that a observed zero j is either a cure or zero loss via the random variable $X_{t,j}$ with p_t^c being the cure probability of portfolio t .

$$X_{t,j} \sim B(1, p_t^c) \tag{5.4}$$

If an observed zero j is considered to be simply a zero loss (hence $x_{t,j} = 0$) then we wish that our estimated LGD is 100% accurate (for a perfect model). However, if a zero j is considered to be a cure ($x_{t,j} = 1$) then we assume that we would have estimated an LGD (other than zero). To simulate the process that cures receive a number we draw a number from the empirical distribution of non-zero LRs (as we simulate the perfect model given the presence of cures). We denote a draw from the empirical distribution of non-zero LRs from portfolio t as $F_t(1)$. Thus, we can describe the LGD and LR of an observed zero j in portfolio t via Eq. (5.5) and (5.6).

$$\text{LGD}_{t,j} \sim F_t(1) \quad \text{and} \quad \text{LR}_{t,j} = 0 \quad \text{for} \quad x_{t,j} = 1 \tag{5.5}$$

$$\text{LGD}_{t,j} = 0 \quad \text{and} \quad \text{LR}_{t,j} = 0 \quad \text{for} \quad x_{t,j} = 0 \tag{5.6}$$

The results of this process can be described as data set, which we denote as C_t for portfolio t . As we assume that the non-zero LR should be estimated 100% accurately, we can denote the LGD of non-zero loss i in portfolio t for n non-zero losses via Eq. (5.7). The resulting data set is denoted as S_t .

$$\text{LGD}_{t,i} = \text{LR}_{t,i} \quad (5.7)$$

By combining the two data sets C_t and S_t it is possible to compute the AR. As the data set C_t is a result of random process, we should repeat this process a sufficient number of times in order to construct a distribution of ARs from which we can derive the expected AR. We denote each experiment as k and hence the resulting cure / zero loss data set as $C_{t,k}$. S_t remains constant for each experiment. For each experiment an AR is computed, which results in a distribution of ARs, which we denote as the implied AR. The average of the distribution is denoted as the sample mean AR. Hence, each AR resulting from an experiment k is denoted as $Z_{t,k}$. The sample mean AR can be computed via (5.8), while the implied AR for the LGD model can be made visible by plotting all k ARs in an histogram. An example of such an histogram can be found in Fig. 5.3.

$$\bar{Z}_t = \frac{1}{k} \sum_{\text{all } k} Z_{t,k} \quad (5.8)$$

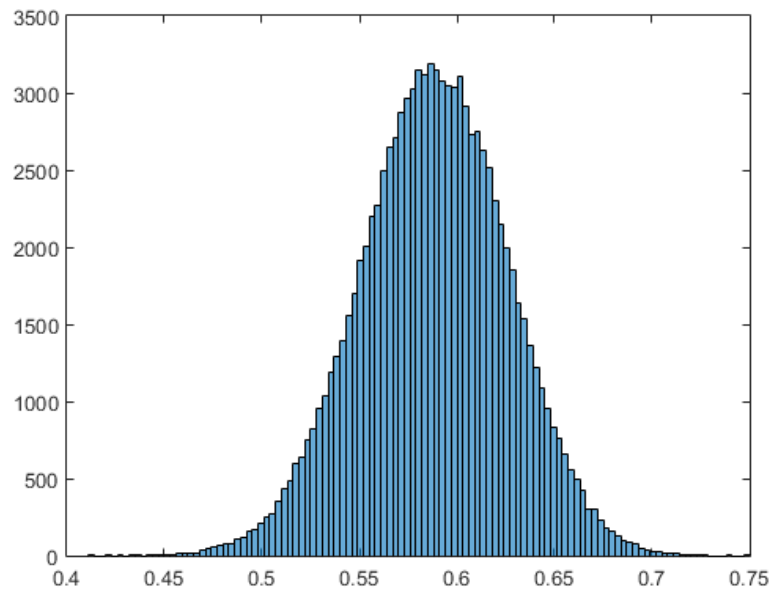


Figure 5.3: Distribution of all simulated ARs (example of Portfolio A).

A rule of thumb is that for the range $\pm 3\sigma$ of the mean almost all possible outcomes of a distribution (for a normal distribution 99,7%) are described. Therefore, the sample variance

should be computed via (5.9).

$$s_t^2 = \frac{1}{k-1} \sum_{\text{all } k} (Z_{t,k} - \bar{Z}_t)^2 \quad (5.9)$$

This process is applied to the portfolio's A, B and C for which the results can be found in Table 5.1.

Table 5.1: Implied AR of the portfolios.

Portfolio	Expected AR	Lower Bound	Upper Bound
Portfolio A	58.80%	47.56%	70.04%
Portfolio B	41.84%	35.37%	48.30%
Portfolio C	65.04%	54.29%	75.79%

Scenario-Structured LGD Models

Different than the XXXXXXX, for the scenario-structured LGD models we can independently check each component on discriminatory power. The cure scenario component is similar to that of an PD model, and therefore it can be checked via the approach described in Section 3.3. The loss given loss (LGL) scenario can be checked regularly via the AR (against the 100%) if one thinks that 100% accuracy is a solid benchmark. Following from an expert discussion, it is, however, urged to also check the discriminatory power for the model as a whole. For that purpose, it is possible to use the approach described for the XXXXXXX and to assign zero losses according to the probability of certain scenario's. The disadvantage, however, is that for scenario-structured LGD models the probability of e.g. cure is not considered to be random. A small adjustment therefore can be considered to make an adjustment for that error. If the individual probabilities of cure are known for a zero loss j , then we can describe the determination process whether it is a cure via the random variable $X_{t,j}$ with $p_{t,j}^c$ being the cure probability zero loss j in portfolio t .

$$X_{t,j} \sim B(1, p_{t,j}^c) \quad (5.10)$$

Unfortunately, this data is not available. Therefore, the assumption is made for the structural LGD model under consideration that each zero observation has the same probability of cure. The results for this portfolio can be found in Table 5.2.

Table 5.2: Implied AR of the structural LGD model.

Portfolio	Expected AR	Lower Bound	Upper Bound
Portfolio D	65.88%	64.88%	66.88%

Although we cannot directly use the data from the portfolio, we can illustrate the effect of having individual cure rates present in an LGD model. Suppose that we perfectly know in advance what the losses will be if a facility defaults, hence we can perfectly predict the LGL. However, all the facilities have a probability of cure. We assume that we can predict the total number of cures for the whole portfolio (if our LGD model works perfectly). Thus, by Eq. (5.11) we can determine our estimated LGD.

$$\text{LGD} = p_{\text{cure}} \cdot \text{LGD}_{\text{cure}} + (1 - p_{\text{cure}}) \cdot \text{LGD}_{\text{default}} \quad (5.11)$$

Recall that for the XXXXXXs we assumed that we cannot discriminate on the basis of the cure probability, but only know the historical cure rate of a portfolio e.g. 50%. Hence, in order to establish a baseline for the AR we simply assume that in Eq. (5.11) p_{cure} equals 50% for each instance. In the case of a cure we assume that the losses are negligible, thus LGD_{cure} equals 0%. For determining the losses in the case of a liquidation scenario we simply use the available LR distributions of the four portfolio's. If the observed zero losses in a LR distribution are less than 50% we simply add a number of zero's such that we have a loss distribution in which 50% of the instances cured. Similar to the approach for the XXXXXXs for each observed zero we determine whether it was a cure or not. In the case of cure we draw a number from the corresponding LGL distribution. Once we have determined for each instance the losses in the case of a liquidation, it is possible to assign compute the corresponding LGD by (5.11).

Table 5.3: Experiments for effect of different cure rates on the AR.

Experiment	P_1^C	P_2^C	P_3^C	P_4^C
Baseline	0.5	-	-	-
2 P^C - 1	0.4	0.6	-	-
2 P^C - 2	0.3	0.7	-	-
2 P^C - 3	0.2	0.8	-	-
2 P^C - 4	0.1	0.9	-	-
4 P^C - 1	0.3	0.4	0.6	0.7
4 P^C - 2	0.2	0.4	0.6	0.8
4 P^C - 3	0.1	0.3	0.7	0.9
4 P^C - 4	0.2	0.3	0.7	0.8
4 P^C - 5	0.1	0.4	0.6	0.9

As we have simulated the LGDs and the LRs it is possible to compute the AR. We simulate this process a sufficient number of times in order to construct the implied AR for these settings. In order to make results comparable we made sure that the expected losses for each portfolio remain the same as well as the distribution.

In order to test the effect of having different cure rates present within the LGD model, we practically need more information on how the estimated cure rates are correlated with the

facilities. As this information is not present, we will simulate the most extreme settings and a random setting. The most extreme settings are either that the lowest LGLs have the highest cure probability (denoted as *setting 1*), or that the lowest estimated LGLs have the lowest cure rates (denoted as *setting 2*). Hence, in setting 1 the LGLs are negatively correlated with the probabilities of cure, and in setting 2 the correlation is positive. For the random setting we randomly assign probabilities of cure to the instances. We assign the probabilities of cure on the basis of the simulated LGL distribution in which we take into account that the correct proportions have cured. For example this means that if we have two probabilities of cure, say 30% and 70%, and we assume that these are negatively correlated with the LGL, then 30% of the cures originate from the upper half of the LGL distribution and visa versa. Zero losses are a special case as they strictly taken change the LGL distribution, although their losses are always zero regardless of their scenario. Therefore, we neglected these instances in the LGL distribution for assigning the probabilities of cure.

Table 5.4: Test results for multiple cure rates in an LGD model (extreme cases).

Experiment	Port. A		Port. B		Port. C		Port. D	
Setting	1	2	1	2	1	2	1	2
Baseline	54.62%		61.19%		59.11%		72.07%	
2 P^C - 1	57.06%	63.64%	62.65%	71.55%	61.36%	66.49%	73.52%	74.81%
2 P^C - 2	63.67%	72.50%	67.55%	83.47%	67.29%	77.52%	77.17%	81.92%
2 P^C - 3	74.56%	78.94%	75.92%	86.38%	76.23%	83.94%	82.85%	88.79%
2 P^C - 4	86.24%	86.71%	87.09%	93.96%	87.52%	89.95%	90.38%	93.76%
4 P^C - 1	63.22%	78.35%	69.01%	86.29%	66.91%	78.75%	76.43%	81.98%
4 P^C - 2	70.12%	75.24%	75.25%	84.12%	73.07%	84.37%	79.83%	90.78%
4 P^C - 3	70.47%	80.21%	79.43%	83.22%	80.87%	81.74%	84.44%	94.49%
4 P^C - 4	71.88%	76.50%	76.28%	84.87%	74.27%	84.77%	80.79%	90.51%
4 P^C - 5	70.44%	77.18%	79.07%	81.15%	78.34%	80.58%	83.27%	94.43%

As mentioned before we set the initial cure rate at 50%, which lead to an expected number of cures (namely $0.5 \cdot n$ for n instances). For our experiments we keep the expected number of cures the same in order to be able to make some comparisons. We acknowledge that the sum of binomial distributions leads to a different variance compared to the baseline binomial distribution, while the expected value is the same. In Table 5.4 an overview of the experiments we conduct to research the impact of different cure rates on the AR can be found.

We have simulated the experiments for similar population sizes found in the portfolio. The results for the simulation can be found in Table 5.4, which denotes the expected AR for each setting. In Fig. 5.4 an example of how the underlying distributions of ARs for an experiment look like.

When we simply assign different cure rates randomly to the instances (while keeping the expected value the same as the baseline), we see that if the difference between the individual cure rates increases, the difference in the expected AR increases as well. We see that the discriminatory power increases if the difference in cure rates increases. This is simply

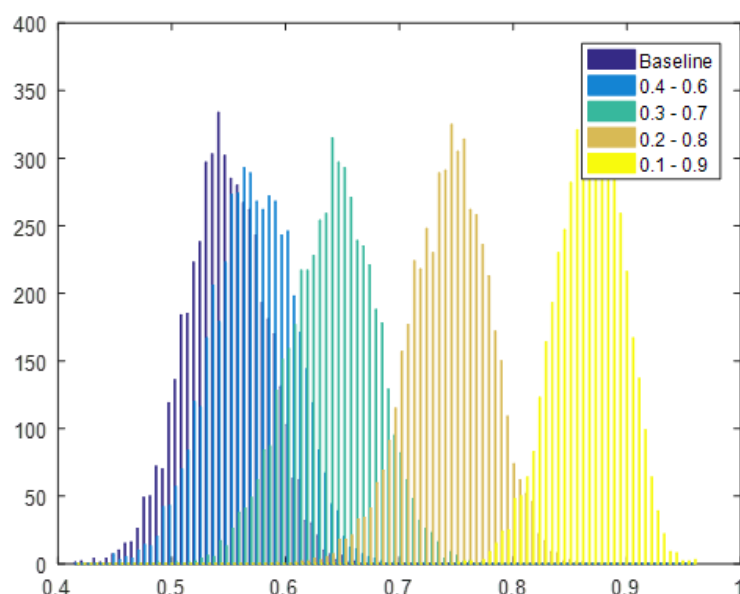


Figure 5.4: Test results for experiment: $2 P^C$ -2 of Portfolio A.

Table 5.5: Test results for multiple cure rates in an LGD model (random assignment).

Experiment	Port. A	Port. B	Port. C	Port. D
Baseline	54.62%	61.19%	59.11%	72.07%
$2 P^C$ - 1	56.89%	63.64%	61.43%	73.80%
$2 P^C$ - 2	64.38%	71.56%	68.57%	78.47%
$2 P^C$ - 3	74.42%	81.45%	78.33%	84.96%
$2 P^C$ - 4	87.01%	91.03%	89.67%	92.32%
$4 P^C$ - 1	60.26%	68.10%	65.12%	76.26%
$4 P^C$ - 2	66.23%	74.04%	70.75%	79.99%
$4 P^C$ - 3	77.13%	83.34%	80.13%	86.45%
$4 P^C$ - 4	69.54%	76.94%	73.53%	81.94%
$4 P^C$ - 5	74.25%	80.72%	77.63%	84.73%

due to the proportions of cure cases as the higher the cure probability is, the more cure cases with this probability actually cured (as we assume that the cure model is correct). The discriminatory power increases as the cure cases have a lower estimate LGD as explained by Eq. (5.11). The results are presented in Table 5.5.

Variance of the Loss Distribution

Not only the cure rate plays a role in the determination of the AR for LGD models. The distribution of the loss rates also influences the outcome of the AR. In the current approach, unlike e.g. Spearman's rank correlation, not only the rank of the instances, but also the size

of the losses influences the resulting AR. Hence, if a portfolio is more unbalanced (e.g. the variance is higher) than it is possible to achieve high values on the AR if the highest losses are classified as the 'worst' losses, while the ranking of the small losses does not really matter for the AR. This is illustrated via a small example.

Table 5.6: Example data for the influence of the variance of a portfolio on the AR.

LGD	LR ^A	LR ^B
0.10	0.10	0.30
0.20	0.20	0.40
0.30	0.50	0.50
0.40	0.90	0.70
0.50	0.80	0.60

Suppose we have the following data available from 5.6. We have an range of LGDs and corresponding LRs. Both portfolio's score the same on Spearman's rank correlation and have the same percentage-weighted losses. The LGDs are the same for both portfolio's, but the magnitude of the LGDs does not affect the AR (as the AR is measured via the total losses and not the total loss estimations). Hence, the only difference between the two portfolio's is the variance, which we can compute via (5.12). For portfolio *A* the variance is 0.10, while for portfolio *B* the variance is 0.02.

$$\text{Var}(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \quad (5.12)$$

The respective AR scores are for portfolio *A* 95.45% and for portfolio *B* 90%. Hence, we conclude that the variance of the loss rates in the portfolio's affects the resulting AR. This is an aspect that should be taken into account when reviewing LGDs on discriminatory power. Depending on the variance one portfolio can be penalized more severely for having ranked two instances incorrectly.

Fraud

One aspect that also influences the outcome of the losses is the case of fraud, which of course is unknown at the time credit is given (otherwise no credit would have been given). This also affects the outcome of the AR as it is not uncommon that fraud results in high losses. This factor is difficult to capture in an LGD model and affects the performance on discriminatory power. After establishing baselines for different portfolios we simulate fraud cases. This is done by randomly assigning a non-zero LR to be a case of fraud and changing the LR to one. The results can be found in Table 5.7. Initially, the increased losses randomly assigned in the portfolio impact the resulting AR, but the effect on the AR by increasing the number of fraud cases decreases. This can be simply explained that initially the impact of a high loss randomly assigned fraud case is a larger portion of the total losses and weights more heavily on the resulting AR.

Table 5.7: Test results for various fraud rates.

Experiment	Port. A	Port. B	Port. C	Port. D
Baseline	48.86%	37.47%	48.55%	64.75%
Fraud - 5%	43.83%	33.10%	43.08%	51.11%
Fraud - 10%	39.42%	29.58%	38.68%	41.77%
Fraud - 15%	35.43%	26.62%	34.83%	34.98%

5.3 Other Explanatory Factors for an LGD

For the approach described in Section 5.1 we assumed that each non-zero LR should be estimated 100% accurately. However, there might be reason to deviate from this 100% estimation. If the outcome of LR is expected to be very volatile (perhaps due to market circumstances or regulation), one might be satisfied if the LR lies within a range of the estimated LGD. If that is the case, then the AR will also change, because the actual ranking might shift. Suppose that we consider it is sufficient for an $LGD_{t,i}$ to lie within the range $R_{t,i}$ of $LR_{t,i}$ of observation i in portfolio t . Then we can redefine the data set S_t by drawing random numbers for $LGD_{t,i}$, which lie within the range. One approach to do that is via the Uniform distribution (see (5.13)).

$$LGD_{t,i} \sim U[LR_{t,i} - R_{t,i}, LR_{t,i} + R_{t,i}] \tag{5.13}$$

To illustrate the effect for various ranges on the AR a simulation is conducted. In this simulation the ranges are determined to be a certain percentage under and above of the respective LR. If we denote this percentage as r_t , then the estimate for each non-zero observation i in portfolio t is determined via Eq (5.14).

$$LGD_{t,i} \sim U[LR_{t,i} \cdot (1 - r_t), LR_{t,i} \cdot (1 + r_t)] \tag{5.14}$$

It should be noted that in this approach the values of an LGD might exceed 1. It is possible to use other approaches to derive ranges. If we vary r_t and compute the average AR of k experiments we see that for each portfolio an impact on the AR is observed.

XX
 XX
 XX
 XX
 XX
 XX
 XX

If one finds it reasonable to allow estimates to deviate from the LR, then the difficulty is in how to determine an appropriate range in which a deviation is allowed. One needs qualitative arguments to state whether a deviation is allowed and one needs to determine, whether

the portfolio characteristics are general enough to use a deviation for all observations (as we have done in our example) or that it is required to determine the deviation for each individual observation. The latter has of course a high workload. One might consider to include this process in the determination of the LGD as many explanatory factors for determining the LGD will also be a driving factor of how variable the LR can be.

There are many factors that could determine variability in the outcome of a loss process. It is not possible to say, whether e.g. Loan-to-Value (LTV) is a driving factor in solely determining the recoveries as also jurisdiction and the easiness to collect collateral plays a role. For instance in particular states in the US you can simply transfer the collateral (e.g. your house) of a mortgage to the bank, which relieves you of your obligations regarding that mortgage. Hence, if your portfolio consists of e.g. multiple jurisdictions then it is impossible to state something about acceptable ranges for the whole portfolio. If that is the case you should individually review each observation and determine acceptable ranges per observation. An overview of possible explanatory factors is given in Chapter 2.

5.4 Summary and Conclusion

As the implied Gini approach turned out to be difficult to maintain given the results in Chapter 4, we set out to try to indicate portfolio characteristics, which impact and partly explain the performance on discriminatory power. This chapter mainly focused on quantitative aspects of a portfolio as qualitative aspects (described in Chapter 2) are very context (or opinion) dependent on how these drivers impact the AR.

XX
 XX
 XX
 XX
 XX
 XX
 XX

The outcome of the simulation results in a distribution of ARs, which is dubbed as the implied AR (for consistency reasons with the implied Gini)¹¹. The expected AR is simply the mean of this distribution.

For assessing the impact of the cure rate we made a distinction between the XXXXXXXs and the scenario-structured LGD model. Due to their structure, the XXXXXXXs do not allow to explicitly derive an individual cure rate for an observation. Therefore, we have assumed that zero observations have the same probability of cure for the whole portfolio. In the scenario-structured LGD model an explicit probability of cure is estimated. For our current data sets these probabilities are, however, not available. In order to still assess the impact of having different cure rates in one portfolio we computed LGDs from a simulated LGL distribution with various settings of cure rates. The results indicate that the AR is affected if the LGD allows different amounts of ratings. Hence, these aspects should be taken into account for the review an LGD model. More research is, however, needed on this matter as we are unable to determine with the current data set how the probability of cure is correlated to the

¹¹Although the term is perhaps not suitable as 'implied' suggests that the AR follows from the LGD model, while in our approach we reason the AR from the loss distribution, which is independent from the LGD model.

drivers that lead to an LGL estimation.

We also indicated that the variance of a portfolio (or the balance) plays a determining role in the computation of the AR. Due to the nature of the AR some instances (e.g. due to higher losses) impact the resulting AR more heavily than others. Similar to the variance, fraud cases also impact the AR heavily as fraud is difficult to predict, but can attribute to a large proportion of the losses that are suffered in a portfolio.

Another aspect that we looked into is how different tolerance levels influence the AR. Deviation from the LR negatively influences the resulting AR for a portfolio. However, in practice the estimated LGD always deviates from the observed LR and in order to cope with that it is possible to use certain tolerance levels in which an LGD is acceptable although it deviates from the corresponding LR. Setting these levels are arbitrarily and should be discussed internally.

We conclude that the described factors, such as probability of cure, variance, fraude and tolerance levels influence the AR heavily if they are highly present in a data set. Therefore, setting a performance benchmark for LGD models that is fixed for each portfolio is unrealistic (the current case). Due to portfolio characteristics it might be impossible for some LGD models to achieve this benchmark. Therefore, we argue that benchmarks should be set up dependent on the portfolio, while taking the characteristics of the portfolio into account. Therefore, we have developed an approach to cope with the current data availability and the different explanatory factors.

If only the LGD and LR values are available and some overall portfolio statistics (e.g. historical cure or fraud rate), then we propose to use this alternative approach to get more insights in the 'maximum' attainable AR. In general it works as follows:

- 1. Determine the different scenarios that can occur within a portfolio.**
- 2. For each scenario determine how it impacts the realized losses.**
- 3. Determine the applicable scenarios for a debtor dependent on the actual LR (e.g. if there are observed losses then a cure scenario is not applicable).**
- 4. Simulate the assignment of scenarios to debtors.**
- 5. Based on the assigned scenario determine the corresponding LGD.**
- 6. Compute the AR.**
- 7. Repeat step 4-6 a sufficient number of times and determine the average value, which we denote as the *expected AR* of a portfolio. The complete range of values for the simulated ARs for a portfolio is denoted as the *implied AR* for a portfolio.**

Possible Approach to Set Indicative Benchmarks

In this chapter we discuss an approach to set an indicative benchmark for testing the discriminatory power of an LGD model. The approach of setting indicative benchmarks for assessing performance of an LGD model is more a subjective approach of the level of discriminatory power for which the bank (and regulator) is satisfied. It is difficult to find a universal score that given the portfolio characteristics determines a benchmark that that particular LGD model should achieve minimally. It is also difficult to derive them from other banks or academic papers as they sometimes vary in the way discriminatory power (via the AR) is measured. Hence, the approach outlined in this chapter is based upon the determination of the so-called implied AR given the current data availability (see Chapter 5). This means that it includes the historical cure probabilities and, as the empirical loss distribution of a portfolio is used, also the variance of the losses. This method can be extended by adding a fraud component, which is described in Chapter 5 as well. Compared to the original implied Gini, it relies less on critical assumptions on the loss distribution and it does not depend on the estimations of the LGD model. It still has made assumptions on the cure probability, which are difficult to validate based on the current data, but the approach relies on internal procedure documents of the bank. Furthermore, the current approach can only cooperate with two scenarios, namely an LGL scenario and a cure scenario, but if relevant it can be extended with additional scenarios. This chapter has the following outline:

Section 6.1: We describe an alternative approach to the implied Gini for setting indicative benchmarks for testing the discriminatory power of an LGD model.

6.1 Implied AR as an Indicative Benchmark

In Section 5.1 an approach for determining the implied AR for LGD models has been proposed. For XXXXXXs it assumes that each zero observation (the LR is 0) within a data set can be either classified as a cure or a zero loss. If a zero observation is classified as a zero loss then it assumes that also zero losses should have been predicted. If it is classified as a cure, however, then the estimated LGD is determined via a random draw from the empirical distribution of non-zero losses is done. After the classification process it is possible to compute the AR. The AR is computed multiple times as we deal with randomness. The distribution of the ARs resulting from the simulation run can be considered as the implied AR. The sample mean of that distribution is the expected AR, while via the 3σ rule of thumb it is possible to determine the boundaries in which a perfectly accurate model discriminatory

power will be (given the presence of cures).

It is not suggested to use these boundaries as indicative benchmarks for XXXXXXs, but a minimum threshold of the lower bound (of the implied AR). It is, namely, unlikely that LGDs perfectly correspond with the LR for each observation. If we denote LB_t as the lower bound of the implied AR for portfolio t and think that the percentage x_t for portfolio t is sufficient as a threshold, then the indicative benchmark is determined via (6.1).

$$BM_t = x_t \cdot LB_t \quad (6.1)$$

For scenario-structured LGD models it is recommended that each individual component is tested for discriminatory power. A cure component follows a binomial process similar to the PD model and therefore, it is recommended that the approach for implied AR for PD models is followed. An LGL component can simply follow the regular approach computing the AR (as no filtering of cures is necessary). Following from an expert discussion, it is, however, urged to also check the discriminatory power for the model as a whole. For that purpose, it is possible to use the approach described for the XXXXXXs and to assign zero losses according to the probability of certain scenarios. The disadvantage, however, is that for scenario-structured LGD models the probability of e.g., cure is not considered to be completely random unlike the assumption for XXXXXXs. A small adjustment therefore can be considered to make an adjustment for that error (see Section 5.1). An indicative benchmark can be derived in a similar fashion as for the XXXXXXs.

This approach assumes that non-zero losses have to be estimated 100% accurately. However, there might be reason to deviate from this 100% estimation. If the outcome of LR is expected to be very volatile (perhaps due to market circumstances or regulation), one might be satisfied if the LR lies within a range of the estimated LGD. If that is the case, then the AR will also change, because the actual ranking might shift. It remains difficult to find unique characteristics for entire portfolio to support a particular deviation. However, in the estimation process it is possible to find drivers (the same used in a scorecard approach) to determine the allowed deviation e.g. jurisdiction or collateral. Another aspect that could be taken into account is the heterogeneity or homogeneity of a portfolio. If instances are quite comparable, we expect it should be easier to make accurate estimations compared to a portfolio that is diverse.

Conclusion and Further Research

In this research we attempt to find a systematic approach to set indicative benchmark values for the discriminatory power of an LGD model measured via the AR. As LGD models are relatively new compared to PD models, but the last years have seen a significant number of papers discussing the LGD model. Unfortunately, there is not yet an approach which indicates whether a model's performance is sufficient in terms of discriminatory power. For these reasons the bank found difficulty in determining what a sufficient level of discriminatory power for an LGD model is. Currently, they have set the threshold values fixed for each portfolio measured via the AR.

There are various approaches to measure discriminatory power for LGD models as some are addressed in our research. Our research primary focuses on the AR, which is also known as the Gini at the bank or Powerstat. The main reason to focus on the AR is that the bank uses this technique primary to assess the discriminatory power of LGD model.

Implied Gini

XX
XX
XX
XX
XX
XX
XX

Influencing Factors and Alternative Approach

Although the initial research question was to validate the implied Gini, the goal for which the approach was developed still stands. Therefore, we looked into factors that give more insight in the resulting AR. From our test results we conclude that factors, such as probability of cure, variance, fraude and tolerance levels influence the AR heavily if they are highly present in a portfolio. Therefore, we are confident enough to conclude that setting a performance benchmark for LGD models that is fixed for each portfolio is unrealistic (as it is the current case). Due to portfolio characteristics it might be impossible for some LGD models to achieve this benchmark. Therefore, we argue that benchmarks should be set up dependent on the portfolio, while taking the characteristics of the portfolio into account.

In order to set these indicative benchmarks that depend on the portfolio characteristics we developed the implied AR, which is a simulation driven approach that uses actual loss distribution instead of the estimations.¹² This approach computes a distribution of ARs, which results from randomly assigning zero observations to be either a cure or an actual zero loss per simulation run. For the XXXXXXs it is then possible to setup an indicative benchmark, which is a fraction of the lower bound of the implied AR. Scenario-structured LGD models can be tested per scenario, but in an expert discussion it was argued that the model as whole should also be tested. In the implied AR is possible to account for different scenario's such as assigning individual probabilities of cures and LGL to instances. Furthermore, it is possible to add extensions to the implied AR such as the possibility of fraud.

We acknowledge that the approach of the implied AR still relies on some basic assumptions that cannot be tested due to the lack of data or due to data quality. However, if more data becomes available or the data quality improves it could be the case that some assumptions are not necessary any more (e.g. such as the randomness of cure or fraud). The approach also relies primarily on quantitatively measurable factors that influence the AR. There might be, however, qualitative factors that indicate that it is more difficult for some portfolios to achieve higher levels of discriminatory power than others. In our research we found it difficult to objectively pinpoint these factors as they are in our opinion very context dependent. Nevertheless, we think that the implied AR helps with the understanding of what (quantitatively) drives the AR for LGD models and therefore helps to set up indicative benchmarks for discriminatory power measured via the AR.

Further Research

As we indicated, the implied AR still has some open ends. The implied AR currently only accounts for two scenarios either a cure or a loss given loss scenario (including fraud). However, there might be other scenarios worth and relevant to research their impact on the AR. Furthermore, we limited ourselves to the AR for measuring discriminatory power as it is the bank's primary tool to measure the performance of a model on this aspect. The AR has, in our opinion, some serious deficits in the case portfolio's losses are not balanced. In that case a small fraction of instances can determine the AR and the ranking of the remainder does not matter anymore. Therefore, we would recommend in further research to address more attention to alternative approaches. We have described some alternatives in Chapter 2, but there are more approaches or alternatives available to measure discriminatory power (e.g. the H-measure or the MAUC, which is variant of the AUC). Nevertheless, also for these methods the question still remains what levels of discriminatory power are sufficient.

Benchmarking discriminatory power is one thing, but improving models with different techniques can also be helpful. Academic papers such as Loterman *et al.* (2012) suggest that different techniques from those currently employed at the bank might lead to higher discriminatory power. Further research in these methods is considered to be helpful to advance LGD models, but we acknowledge that some complex techniques (although promising) are not practical as they are difficult to review as they are sometimes viewed as a 'black box' approach by some practitioners.

¹²The name 'implied' is strictly taken not correct, but due to the project implied Gini is currently taken on as a working title.

References

- Acharya, V. V., Bharath, S. T., & Srinivasan, A. (2007). Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries. *Journal of Financial Economics*, 85(3), 787-821.
- Altman, E. I., & Kishore, V. M. (1996). Almost everything you wanted to know about recoveries on defaulted bonds. *Financial Analysts Journal*, 52(6), 57-64.
- Altman, E. I., Brady, B., Rest, A., & Sironi, A. (2005). The link between default and recovery rates: theory, empirical evidence, and implications. *Journal of Business*, 78(6), 2203-2227.
- Baesens, B., & Van Gestel, T. (2009). *Credit Risk Management: Basic Concepts: Financial Risk Components, Rating Analysis, Models, Economic and Regulatory Capital*. New York: Oxford University Press Inc.
- Balin, B. J. (2008). *Basel I, Basel II, and Emerging Markets: A Non-Technical Analysis*.
- Bastos, J. A. (2010). Forecasting bank loans loss-given-default. *Journal of Banking & Finance*, 34(10), 2510-2517.
- BCBS. (2005). *An Explanatory note on the Basel II IRB risk weight functions*. Basel: Bank for International Settlements.
- BCBS. (2005). *Studies on the validation of internal rating systems*. Basel: Bank for International Settlements.
- BCBS. (2006). *International convergence of capital measurement and capital standards*. Basel: Bank for International Settlements.
- BCBS. (2015). *Revisions to the standardised approach for credit risk*. Basel: Bank for International Settlements.
- Bellotti, T., & Crook, J. (2008). Modelling and estimating Loss Given Default for credit cards. *Credit Research Centre*, University of Edinburgh Business School.
- Berkson, J. (1980). Minimum chi-square, not maximum likelihood!. *The Annals of Statistics*, 8(3), 457-487.
- Blochwitz, S., Hamerle, A., Hohl, S., Rauhmeier, R. & Rsch, D. (2005). Myth and reality of discriminatory power for rating systems. *Wilmott Magazine*, 2-6.

- Calabrese, R. (2014). Downturn loss given default: Mixture distribution estimation. *European Journal of Operational Research*, 237(1), 271-277.
- Calabrese, R., & Zenga, M. (2010). Bank loan recovery rates: Measuring and nonparametric density estimation. *Journal of Banking & Finance*, 34(5), 903-911.
- Cantor, R., Emery, K., & Stumpp, P. (2006). Probability of default ratings and loss given default assessments for non-financial speculative-grade corporate obligors in the United States and Canada. *Moody's Investors Service-Global Credit Research*, New York.
- Caouette, J. B., Altman, E. I., Narayanan, P., & Nimmo, R. (2008). *Managing Credit Risk*. Hoboken: John Wiley & Sons, Inc.
- CEBS. (2006). *Guidelines on the implementation, validation and assessment of advanced measurement (AMA) and internal ratings based (IRB) approaches*.
- Chalupka, R., & Kopecsni, J. (2008). Modelling Bank Loan LGD of Corporate and SME Segments: A Case Study. *IES Working Paper 27/2008*, IES FSV, Prague: Charles University.
- Chen, S. X. (1999). Beta kernel estimators for density functions. *Computational Statistics & Data Analysis*, 31(2), 131-145.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah: Lawrence Erlbaum Associates, Inc.
- Davydenko, S. A., & Franks, J. R. (2008). Do bankruptcy codes matter? A study of defaults in France, Germany, and the U.K. *The Journal of Finance*, 63(2), 565-608.
- DNB. (2007, January 1). *Credit risk: capital requirements based on internal models*. Retrieved from De Nederlandse Bank (DNB) Website: <http://www.toezicht.dnb.nl/en/2/51-202068.jsp#>
- Engelmann, B., Hayden, E., & Tasche, D. (2003). Testing Rating Accuracy. *Risk* 16(1), 8286.
- Freeden, W., & Gutting, M. (2013). *Special functions of mathematical (geo-) physics*. Basel: Springer Science & Business Media.
- Frye, J. (2000). *Depressing recoveries*. Chicago: Federal Reserve Bank of Chicago.
- Gowers, T., Barrow-Green, J., & Leader, I. (Eds.). (2010). *The Princeton companion to mathematics*. Princeton University Press.
- Grunert, J., & Weber, M. (2009). Recovery rates of commercial lending: Empirical evidence for German companies. *Journal of Banking & Finance*, 33(3), 505-513.
- Gupton, G. M. (2005). Advancing loss given default prediction models: how the quiet have quickened. *Economic Notes*, 34(2), 185-230.
- Gupton, G. M., & Stein, R. M. (2002). LOSSCALCTM: Model for predicting loss given default (LGD). New York: *Moody's KMV*.

- Hamerle, A., Rauhmeier, R., & Rsch, D. (2003). Uses and misuses of measures for credit rating accuracy.
- Herring, R. J. (1999). Credit risk and financial instability. *Oxford Review of Economic Policy*, 15(3), 63-79.
- Hlawatsch, S., & Reichling, P. (2010). A framework for loss given default validation of retail portfolios. *The Journal of Risk Model Validation*, 4(1), 23-48.
- Huang, X., & Oosterlee, C. W. (2011). Generalized beta regression models for random loss given default. *The Journal of Credit Risk*, 7(4), 45.
- Hull, J. C. (2012). *Risk Management and Financial Institutions*. Hoboken: John Wiley & Sons, Inc.
- Johnson, P., & Beverlin, M. (1970). Beta Distribution. *Continuous Univariate Distributions, Vol. 2*. New York: John Wiley and Sons, pp. 37-56.
- Kraft, H., Kroisandt, G., & Müller, M. (2002). Assessing the discriminatory power of credit scores.
- Larsen, R. J., & Marx, M. L. (2012). *An introduction to mathematical statistics and its applications*. Englewood Cliffs, NJ: Prentice Hall.
- Law, A. M. (2007). *Simulation Modeling and Analysis, Fourth edition*. Arizona: McGraw-Hill.
- Li, D., Bhariok, R., Keenan, S., & Santilli, S. (2009). Validation techniques and performance metrics for loss given default models. *The Journal of Risk Model Validation*, 33, 3-26.
- Loterman, G., Brown, I., Martens, D., Mues, C., & Baesens, B. (2012). Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting*, 28(1), 161-170.
- Ospina, R., & Ferrari, S. L. (2010). Inflated beta distributions. *Statistical Papers*, 51(1), 111-126.
- Owen, C. E. B. (2008). Parameter estimation for the beta distribution. Provo: *Brigham Young University*.
- Penn State University: Eberly College of Science. (2016, August 22). 18.3 - Kendall Tau-b Correlation Coefficient. Retrieved from STAT 509: Design and Analysis of Clinical Trials: <https://onlinecourses.science.psu.edu/stat509/node/158>
- Peter, C. (2011). Estimating loss given default: Experience from banking practice. In *The Basel II Risk Parameters* (pp. 151-183). Springer Berlin Heidelberg.
- Persaud, J., & Saurina, J. (2008). Will Basel II help prevent crises or worsen them. *Finance and Development*, 45(2), 29-33.
- Prorokowski, L. (2016). Rank-order statistics for validating discriminative power of credit risk

models. *Bank i Kredyt*, 47(3), 227-250.

Qi, M., & Zhao, X. (2013). Debt structure, market value of firm and recovery rate. *Journal of Credit Risk*, 9(1), 3-37.

Rabobank Group. (2015). *Annual Report 2015*. Utrecht: Rabobank Communications & Corporate Affairs.

Rabobank Group. (2016). *Model Development Procedure: Loss Given Default* Utrecht: RM ALM & Analytics.

Rabobank Group. (2016). *Model Development Procedure: Probability of Default* Utrecht: RM ALM & Analytics.

Raschke, M. (2009). The biased transformation and its application in goodness-of-fit tests for the beta and gamma distribution. *Communications in Statistics-Simulation and Computation*, 38(9), 1870-1890.

Raschke, M. (2011). Empirical behaviour of tests for the beta distribution and their application in environmental research. *Stochastic Environmental Research and Risk Assessment*, 25(1), 79-89.

Renault, O., & Scaillet, O. (2004). On the way to recovery: a nonparametric bias free estimation of recovery rate densities. *Journal of Banking & Finance*, 28(12), 2915-2931.

Schuermann, T. (2004). What do we know about loss given default?

Sobehart, J. R., Keenan, S. C., & Stein, R. (2000). Benchmarking quantitative default risk models: a validation methodology. *Moody's Investors Service*.

The World Bank. (2016, July 26). *GDP Growth (Annual %)*. Retrieved from data.worldbank.org: <http://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=US>

TheBanks.eu. (2016, July 7). *Major Banks in the Netherlands*. Retrieved from thebanks.eu: <https://thebanks.eu/articles/major-banks-in-the-Netherlands>

Tong, E. N., Mues, C., & Thomas, L. (2013). A zero-adjusted gamma model for mortgage loan loss given default. *International Journal of Forecasting*, 29(4), 548-562.

Verschuren, P. J., & Doorewaard, H. A. (2007). *Het Ontwerpen van Onderzoek*. Amsterdam: Boom Lemma Uitgevers.

Example EAD Dataset

XX
XX
XX
XX
XX
XX

Computation of AR25, AR50 and AR75

XX
XX
XX
XX
XX
XX

Derivation of Gamma Function Properties

This appendix shows how the necessary properties of the gamma function for determining the MOM estimators for the beta function are determined. The gamma function is defined as the following integral:

$$\Gamma(t) = \int_0^{\infty} y^{t-1} \cdot e^{-y} \cdot dy, \text{ for } t > 0 \quad (\text{C.1})$$

Via integration by parts it is possible to derive an equation for $\Gamma(t + 1)$ and $\Gamma(t + n)$, which we need to determine the estimators for the MOM approach. Integration by parts states that:

$$\int u(x)v'(x)dx = u(x)v(x) - \int v(x)u'(x)dx \quad (\text{C.2})$$

We define $\Gamma(t + 1)$ as the following integral:

$$\Gamma(t + 1) = \int_0^{\infty} y^t \cdot e^{-y} \cdot dy, \text{ for } t > 0 \quad (\text{C.3})$$

Hence, for the integration by parts we define the following:

$$u(t) = y^t \quad (\text{C.4})$$

$$du(t) = t \cdot y^{t-1} \quad (\text{C.5})$$

$$dv(t) = e^{-y}dy \quad (\text{C.6})$$

$$v(t) = -e^{-y} \quad (\text{C.7})$$

Thus, we can make the following statement:

$$\Gamma(t + 1) = \lim_{a \rightarrow \infty} [-y^t \cdot e^{-y}]_{y=0}^{y=b} + t \cdot \int_0^{\infty} y^{t-1} \cdot e^{-y} \cdot dt \quad (\text{C.8})$$

Provided that $t > 0$, we can state:

$$\Gamma(t + 1) = - \lim_{a \rightarrow \infty} \left[\frac{a^t}{e^a} \right] + t \cdot \Gamma(t) \quad (\text{C.9})$$

Via the L'Hôpital's Rule we can rewrite the first term (the limit), but before we apply L'Hôpital's Rule we need to rewrite the statement. First, we take the natural logarithm of the numerator and denominator. Second, we take the exponent of the numerator and the denominator.

$$- \lim_{a \rightarrow \infty} \left[\frac{b^t}{e^b} \right] = - \lim_{b \rightarrow \infty} \left[\frac{e^{t \cdot \ln(b)}}{e^b} \right] \quad (\text{C.10})$$

The next step is to combine the denominator with the numerator:

$$- \lim_{b \rightarrow \infty} \left[\frac{b^t}{e^b} \right] = - \lim_{b \rightarrow \infty} \left[e^{t \cdot \ln(b) - b} \right] \quad (\text{C.11})$$

We can rewrite the statement such that we can apply L'Hôpital's Rule and determine the limit:

$$- \lim_{b \rightarrow \infty} \left[\frac{b^t}{e^b} \right] = - \lim_{b \rightarrow \infty} \left[e^{t \cdot b \cdot \left(\frac{\ln(b)}{b} - 1 \right)} \right] \quad (\text{C.12})$$

Within the exponent we can clearly indicate that the first b goes to infinity, but for the fraction within the exponent we get infinity over infinity. Hence, we can apply L'Hôpital's Rule. Thus:

$$\lim_{b \rightarrow \infty} \left[\frac{\ln(b)}{b} \right] = \lim_{b \rightarrow \infty} \left[\frac{1}{b} \right] = 0 \quad (\text{C.13})$$

This means that we can write the statement as follows:

$$- \lim_{b \rightarrow \infty} \left[\frac{b^t}{e^b} \right] = - \lim_{b \rightarrow \infty} \left[e^{t \cdot b \cdot (0-1)} \right] \quad (\text{C.14})$$

Hence, as b is multiplied with -1 , we get an exponent of minus infinity. This means that the limit will go to 0. Thus, we can write:

$$-\lim_{b \rightarrow \infty} \left[\frac{b^t}{e^b} \right] = 0 \quad (\text{C.15})$$

This means that we can rewrite the gamma function $\Gamma(t + 1)$ as:

$$\Gamma(t + 1) = t \cdot \Gamma(t) \quad (\text{C.16})$$

The second gamma function we would like to rewrite is $\Gamma(t + n)$. We can reiterate the statement from (C.16) to find an expression for $\Gamma(t + n)$. (C.16) shows that:

$$\Gamma(t + n) = (t + n - 1) \cdot \Gamma(t + n - 1) \quad (\text{C.17})$$

Reiterating this step leads to a statement for $\Gamma(t + n)$ for $t + n > 0$. Hence:

$$\Gamma(t + n) = t \cdot (t + 1) \dots (t + n - 1) \cdot \Gamma(t) \quad (\text{C.18})$$

Determining the First and Second Moment of the Beta Distribution

This appendix shows how Eq. (4.15) and Eq. (4.16) are determined by applying the resulting statements from Appendix C. We have the following statements for respectively the first and second moment of the moment generating function for a beta distribution:

$$E(X) = \frac{\Gamma(\alpha + 1)\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + 1)\Gamma(\alpha)} \quad (\text{D.1})$$

$$E(X^2) = \frac{\Gamma(\alpha + 2)\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + 2)\Gamma(\alpha)} \quad (\text{D.2})$$

We can rewrite Eq. (D.1) as follows:

$$E(X) = \frac{\alpha \cdot \Gamma(\alpha) \cdot \Gamma(\alpha + \beta)}{(\alpha + \beta) \cdot \Gamma(\alpha + \beta) \cdot \Gamma(\alpha)} \quad (\text{D.3})$$

Rearranging this statement leads to the first moment for the beta distribution as found in Eq. (4.15):

$$E(X) = \frac{\alpha}{\alpha + \beta} \quad (\text{D.4})$$

Similarly, for Eq. (D.2):

$$E(X^2) = \frac{\alpha \cdot (\alpha + 1) \cdot \Gamma(\alpha) \cdot \Gamma(\alpha + \beta)}{(\alpha + \beta) \cdot (\alpha + \beta + 1) \Gamma(\alpha + \beta) \cdot \Gamma(\alpha)} \quad (\text{D.5})$$

Rearranging this statement leads to the first moment for the beta distribution as found in Eq. (4.16):

$$E(X^2) = \frac{\alpha \cdot (\alpha + 1)}{(\alpha + \beta) \cdot (\alpha + \beta + 1)} \quad (\text{D.6})$$

Loss Distributions of the Available Portfolios

XX
XX
XX
XX
XX
XX

Estimated Beta Parameters

XX
XX
XX
XX
XX
XX

Anderson-Darling Test Results

XX
XX
XX
XX
XX
XX

Bootstrap for Sum of Individual Loss Distributions

XX
XX
XX
XX
XX
XX