

KTH ROYAL INSTITUTE OF TECHNOLOGY

DEBEKA BAUSPARKASSE AG

MASTER THESIS

---

# Assessment of a Credit Value at Risk for Corporate Credits

---

Author: Laura Kremer

Supervisors: Henrik Hult, Dorothe Schönberg

Stockholm/Koblenz, June 2013



## **Abstract**

In this thesis I describe the essential steps of developing a credit rating system. This comprises the credit scoring process that assigns a credit score to each credit, the forming of rating classes by the k-means algorithm and the assignment of a probability of default (PD) for the rating classes. The main focus is on the PD estimation for which two approaches are presented. The first and simple approach in form of a calibration curve assumes independence of the defaults of different corporate credits. The second approach with mixture models is more realistic as it takes default dependence into account. With these models we can use an estimate of a country's GDP to calculate an estimate for the Value-at-Risk of some credit portfolio.

*Keywords:* Bernoulli Mixture Models



## Acknowledgements

I want to express my thanks to the Risk Controlling department at Debeka Bausparkasse AG for enabling this thesis. My special thanks goes to my supervisor Dorothé Schönberg for her great support and guidance. Further I would like to thank my supervisor Henrik Hult from KTH for his most valuable comments and ideas for this thesis. To my parents I am truly grateful for enabling my studies at KTH and for their moral support. Finally I would like to thank my boyfriend for his encouragement, for proofreading this thesis and for just being with me.

Stockholm, June 2013

Laura Kremer



# Contents

<b>List of Abbreviations</b>	<b>1</b>
<b>Introduction</b>	<b>3</b>
<b>1 Basics of Credit Evaluation</b>	<b>5</b>
1.1 Regulatory Framework . . . . .	5
1.2 Credit Scoring . . . . .	10
1.3 Credit Rating . . . . .	12
1.3.1 Clustering Algorithm . . . . .	13
1.3.2 Example: Clustering . . . . .	14
1.4 Expected Loss and Value-at-Risk . . . . .	18
<b>2 Models to estimate a PD</b>	<b>21</b>
2.1 Calibration Curve . . . . .	21
2.2 Mixture Models . . . . .	22
2.2.1 Probit-Normal Mixing Distribution a Bernoulli Mixture Model without Economical Data . . . . .	23
2.2.2 Beta Mixing Distribution a Bernoulli Mixture Model with Economical Data . . . . .	26
<b>3 Data Base</b>	<b>29</b>
<b>4 Testing the Models on Data</b>	<b>33</b>
4.1 Calibration Curve . . . . .	33
4.2 Probit-Normal Mixing Distribution . . . . .	34
4.3 Beta Mixing Distribution . . . . .	35
4.4 Estimates for Expected Loss and Value-at-Risk . . . . .	38
4.5 Scenario Analysis . . . . .	39
<b>5 Conclusion</b>	<b>41</b>

## Appendices

A Conditional PD	43
B Beta Distribution	47
C Historical German GDP Data	51
D Simulation of Value-at-Risk in R	53



# List of Abbreviations

CRSA	Credit Risk Standardized Approach
DF	Default Frequency
EAD	Exposure At Default
EL	Expected Loss
IRBA	Internal Ratings Based Approach
LGD	Loss Given Default
LPB	Legal Persons' business
NPB	Natural Persons' business
PD	Probability of Default
VaR	Value-at-Risk



# Introduction

**The Company.** This thesis was written at one of Germany's biggest insurance groups: the Debeka-group. Together with the Debeka Bausparkasse AG it is a prestigious financial services business<sup>1</sup>. The thesis was written at the Risk Controlling department at the Debeka Bausparkasse AG.

A Bausparkasse is a bank that offers building loan agreements, so-called Bauspar-contracts. Entering a Bauspar-contract grants the customer the possibility to obtain a low-interest loan, which may be used for residential purposes only. A Bauspar-contract can be described in two phases. In the first phase the customer saves at least 40% of the Bauspar-sum. The Bauspar-sum is the amount agreed upon that can be paid to the customer and consists of the savings amount (at least 40%) and the loan (at most 60%). If the customer withdraws the Bauspar-sum, the second phase of loan redemption immediately follows the first. The customer's advantage of entering this kind of contract is that the loan's interest rate is quite low<sup>2</sup>.

The total credit portfolio of the Debeka Bausparkasse AG mainly consists of debt claims against natural persons but also of debt claims against legal persons. The granted loans have the form of Bauspar-contracts or common credits that are intended to be used for residential purposes (within the meaning of Section 1 *Bausparkassen Act*, Bausparkassengesetz) as well. Since all credits are at some risk of default, the borrowers' default risk has to be assessed.

For claims against natural persons there is a sufficiently large amount of data to use score cards based on statistical analysis. This enables an internal credit rating and the assessment of the default probability (PD) and the loss given default (LGD) rate. With these values, the Value-at-Risk (VaR) can be computed for this part of the credit portfolio. The big difference between testing a natural person's credit worthiness and testing a legal person's credit worthiness lies within the features asked for. A natural person has a job, children, etc. and a company has a balance sheet, a legal form, etc. so that the same score card with the same features cannot be used.

---

<sup>1</sup>Debeka-Group: Debeka eine Unternehmensgruppe mit Zukunft

<sup>2</sup>Debeka Bausparkasse AG: Ihr Bausparvertrag und was Sie darüber wissen sollten

**Problem Formulation.** There are score cards based on expert opinions for claims against companies. Due to an insufficient amount of data the statistics that work well within the Natural Persons' Business (NPB) do not give reliable results for the Legal Persons' Business (LPB). The aim of this thesis is to find estimates for PD and LGD on the basis of the available data. With these parameters, an estimate for the VaR shall be given.

Note: In this version of the thesis no real data will be used. The data used does not relate to the Debeka's data except for the total amount of data and the time period which the data stems from.

**Content.** Chapter 1 starts with a short presentation of the legal background formed by the Basel regulations. As a Bausparkasse is a kind of banking institution these regulations are relevant for its lending business. The Basel regulations give the outline for the credit scoring and credit rating systems that are introduced thereafter. The rating classes of the rating system are determined by a clustering method. Furthermore this first chapter briefly presents the commonly used VaR risk measure and shows how to compute the loan creditor's expected loss from a lending business.

In Chapter 2 models to find a PD are presented. There are two different kinds of models that will be studied. The first one is called calibration curve. The idea is that the default probabilities can be expressed as a function value of an exponential function. The second kind of model is the mixture model. In the mixture model a dependence structure among corporate credits is allowed, in contrast to the calibration curve where independence is assumed. However, the default dependence in the mixture models is eliminated by the assumption that there is a common factor that influences the corporate credits so much that we can assume independence of the defaults conditional on the common factor. This is the more advanced approach to PD modeling and it is the central topic of this thesis. The key question in the mixture models is how to model the common factor. This thesis presents two alternatives: the first one is that the common factor is Normal-distributed and the second one is that it is Beta-distributed.

Chapter 3 gives a short presentation of the fictional data. The historical default frequency is presented and the credits are grouped by their scores into rating classes. These are the rating classes along with their respective default frequencies that the models from Chapter 2 will use.

In Chapter 4 the estimates for the PD and VaR are presented for all kinds of models. In one of the mixture models the German GDP data is used as the Beta-distributed common factor. We will also do a scenario analysis to study the influence of the GDP growth rate to the VaR estimate.

In Chapter 5 the conclusions are drawn.

# Chapter 1

## Basics of Credit Evaluation

In this chapter we go through the Basel accords as the legal background of credit risk measuring. Then we will see how a banking institution values a credit. It starts with credit scoring, which takes the information on a possible obligor and enables the classification of the customer to a rating class. These rating classes can be formed by a clustering method, which is presented thereafter. The difference between a scoring and a rating system is that a scoring system does not provide a probability of default for any obligor. The rating system gives a PD estimate for each rating class. Together with the credit amount and a loss rate, both the expected loss and the loss distribution can be studied. From the loss distribution it is just a small step to give an estimate for VaR, which is presented in the concluding section.

### 1.1 Regulatory Framework

**The Basel Accords.** In July 1988 the Basel Committee on Banking supervision published the Basel I guidelines in the hope that these would "strengthen the soundness and stability of the international banking system"<sup>1</sup>. These guidelines aim at the banking industry of the G-10 states holding buffer capital in the height of the implemented minimum capital standards for credit risk<sup>2</sup>.

In June 2004 a new revised framework of the Basel Accord was published. This version, the 1988 accord and the additions made in between, was released in a 'Comprehensive Version' in June 2006. It is this comprehensive revised framework that is known as the Basel II Accord. It emphasizes the computation and management of risk. The key principles of Basel I basically remained unchanged, but the regulations on risk management got stronger and also more detailed. The major innovation is the "use of assessments of

---

<sup>1</sup>Basel Committee on Banking Supervision (Basel I), p. 3.

<sup>2</sup>Bank for International Settlements, <http://www.bis.org/publ/bcbs04a.htm>, accessed June 13th 2013.

risk provided by banks' internal systems as inputs to capital calculations"<sup>3</sup>, referring to the Internal Ratings-Based Approach that will be presented later on in this section. The Basel II regulations consist of the three pillars that are shown in Figure 1.1.

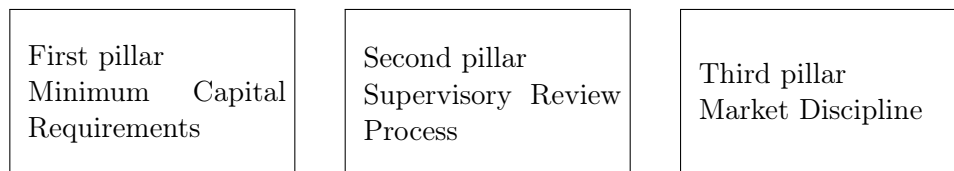


Figure 1.1: Basel II framework.<sup>4</sup>

This work lies within the area of the first pillar. The pillar describes how to calculate the capital requirement for three different kinds of risks: credit risk, operational risk and market risk. Here we will only study credit risk.

The Basel II guidelines have been included into German law. The corresponding laws are the *German Banking Act*<sup>5</sup> (Kreditwesengesetz), the *Regulation governing the capital adequacy of institutions, groups of institutions and financial holding groups*<sup>6</sup> (Solvabilitätsverordnung – SolvV, based on the first and the third pillar of Basel II<sup>7</sup>) and the *Minimum Requirements for Risk Management*<sup>5</sup> (Mindestanforderungen an das Risikomanagement – MaRisk, based on the second pillar of Basel II<sup>7</sup>).

In December 2010 the Basel III accord was published as an answer to the financial crisis that started in 2007. This crisis proved that the capital requirements and the risk management principles from Basel II were not sufficiently strong. The Basel III guidelines require "higher and better-quality capital, better risk coverage"<sup>8</sup> and "measures to promote the build up of capital that can be drawn down in periods of stress"<sup>8</sup>. So far the Basel III Accord has not been incorporated into German law and therefore we will present the Basel II framework (as in the SolvV) in the following.

<sup>3</sup>Basel Committee on Banking Supervision (Basel II), p. 2.

<sup>4</sup>Basel Committee on Banking Supervision (Basel II), p. 6.

<sup>5</sup>Translation: [http : //www.bafin.de/EN/Supervision/BanksFinancialServices Providers/RiskManagement/riskmanagement\\_artikel.html](http://www.bafin.de/EN/Supervision/BanksFinancialServicesProviders/RiskManagement/riskmanagement_artikel.html) (accessed on Feb. 28th, 2013).

<sup>6</sup>Translation: Deutsche Bundesbank. Note that this and the following translation(s) concerning the SolvV "is for information purposes only. The original German text is binding in all respects". Source: [http : //www.bafin.de/SharedDocs/Aufsichtsrecht/EN/Verordnung/solvv\\_en\\_ba.html](http://www.bafin.de/SharedDocs/Aufsichtsrecht/EN/Verordnung/solvv_en_ba.html) (accessed on Feb. 28th, 2013).

<sup>7</sup>Ettmann, Wolff, Wurm: Kompaktwissen Bankbetriebslehre.

<sup>8</sup>Bank for International Settlements, [http://www.bis.org/publ/bcbs189\\_dec2010.htm](http://www.bis.org/publ/bcbs189_dec2010.htm) , accessed on June 13th, 2013.

**Credit Risk.** The main focus of the Basel II Accord lies on credit risk, which is the most dangerous risk (in volume) a creditor faces<sup>9</sup>. The total capital charge with respect to SolvV can be computed as a product of the Exposure At Default (EAD) and some risk weight that depends on the credit or, more generally, on the financial position

$$\text{Total Capital Charge} = \sum_{P \in \text{Financial Position}} EAD(P) \times \text{Risk Weight}(P).$$

Here the set of all financial positions can but does not have to be divided into the set of positions valued with the Credit Risk Standardized Approach (CRSA) and the set of positions valued with the Internal Ratings Based Approach (IRBA). The difference lies within the risk weights; they differ within the IRBA and the CRSA. Additionally, the CRSA positions have to be multiplied by 0.08<sup>10</sup>. To put this formula in one sentence: The minimum buffer capital that should be held by a banking institution is the sum of its risk-weighted assets. It should be noted, however, that it is the credit receiver and his/her credit worthiness that determines the risk weight. The credit's features like the lending amount or the interest rate never stand alone but are always regarded together with the credit receiver when determining the risk weights.

If a financial institution uses the CRSA to assess the credit risk of standardized credits in the LPB, the risk weight is 75% (Section 34 SolvV). In the case that the obligor provides a security, the risk weight is lower, e.g. the risk weight for Bauspar-credits is 50% (Section 35, Subsection 4 SolvV).

In the IRBA, the risk weight<sup>11</sup> is the product of the (estimated) Loss Given Default (LGD) rate, the maturity  $M$  of the financial position and the difference of the conditional PD (from Section 87 and formula 1 SolvV, see appendix A) and the estimated PD that is to be estimated in this thesis. This is

$$\text{Risk Weight} = \text{LGD} \cdot M \cdot (PD^{cond} - PD^{est}). \quad (1.1)$$

The IRBA requires more work from the banking institution since it requires an internal rating system, but it allows the institution to hold a lower capital charge. There are two types of the IRBA. When using the basic IRBA, the Probability of Default (PD) must be calculated within the banking institution. In the advanced IRBA, estimates for the Exposure at Default (EAD) and for the Loss Given Default (LGD) have to be computed as well. Therefore the advanced approach requires every single credit to be looked at and to be examined under the aspects of risk.

<sup>9</sup>Hartmann-Wendels, Pfingsten, Weber: Bankbetriebslehre, p. 497.

<sup>10</sup>Hartmann-Wendels, Pfingsten, Weber: Bankbetriebslehre.

<sup>11</sup>According to Hartmann-Wendels, Pfingsten, Weber: Banklehre, p. 605.

**Internal Rating Systems.** It is the external rating systems that are known to all of us; Germany's Federal Financial Supervisory Agency (BaFin) has registered several credit rating agencies whose rating is accepted when using the CRSA. Among these are the German Creditreform Rating AG, S&P and Moody's<sup>12</sup>.

The internal rating systems are unknown to the public, as they are used within the banking institutions. If the institution wants to use an internal rating system to judge their obligors, then this rating system has to fulfill several criteria. These are given in Sections 107-123 SolvV. Some points will briefly be presented in this section.

A rating system for corporates must have at least seven rating classes for solvent debtors and one for defaulted obligors. This rating system must regard risk features of both the obligor and the credit (Section 110, Subsection 1 SolvV). But these criteria are not named. No matter which information enters the scoring process, they must be up to date and the credit rating itself must be reviewed at least once every year. If the available information is low, the obligor must be given a more conservative rating (Section 112, Subsection 2 SolvV). The risk features of the obligor are an indicator of the obligor's probability of default. So for example financial strength is an important criterion at this point. In contrast, the risk features of the credit such as e.g. the height of given securities is unimportant for the PD assessment (customer  $i$  defaults whether there are securities or not) but it is important for the calculation of the LGD rate.

Each of the at least eight rating classes is assigned a probability of default, such that all the credits in one rating class have the same PD. The PD may be estimated using the historical default data, but its minimum for any rating class is 0.03% (Section 88, Subsection 4 SolvV).

According to Section 132, Subsection 2 and 3 SolvV the estimate for the LGD shall be based on the historically realized LGD rates and should be conservative, i.e. the "LGD estimates [should be] [...] appropriate for an economic downturn if those are more conservative than the long-run average"<sup>13</sup>.

**Definition of Default.** The definition of the event 'default' can be found in the SolvV, Section 125<sup>14</sup> (the following definition is close to the original text but has been simplified):

The obligor has defaulted if at least one of the following events occurs:

---

<sup>12</sup>BaFin, Liste der für die Bankaufsichtliche Risikogewichtung anerkannten Ratingagenturen samt Mapping.

<sup>13</sup>Translation by Deutsche Bundesbank, [20].

<sup>14</sup>Compare also to Debeka Bausparkasse AG: Ausfall.



1. The obligor probably cannot meet his/her total payment obligations without making use of his/her securities.  
The indications of the improbability of payment are
  - (a) The institution adjusts the value of the debt
  - (b) The loan has to be restructured
  - (c) The institution applies for the insolvency of the obligor
  - (d) The obligor or a third person files for the insolvency of the obligor
2. The obligor is overdue for more than 90 succeeding calendar days with a significant part of his/her debt.  
Significant part of debt means at least 100 Euro and more than 2.5% of the total debt.

There is also a possibility of recovery; if none of these criteria above is fulfilled any longer, then the risk classification has to be done as if the obligor never had defaulted.

**This thesis and the SolvV.** The previous text presented a short and incomplete overview over the SolvV. It is impossible to get all these rules and regulations correctly (in the sense of complete, with all exceptions and special cases) and briefly. Therefore, in this thesis I do not even try to get all the estimates and calculations as postulated in the SolvV. Instead, the SolvV shall just give the basic framework to this thesis.

What is definitely used is the definition of the default event. With this definition, each credit is assigned a 1 or a 0, depending on whether the obligor defaulted or not. In the further use of the word default this means that one of the default criteria is fulfilled, but this does not necessarily mean that the obligor is definitely unable to pay back (which is often called 'defaulted' in the common sense). When referring to a total failure of paying back the loan and when the creditor starts to liquidate given securities we shall talk of 'insolvency' instead.

The Debeka Bausparkasse AG uses the CRSA to calculate risk in the corporate business. The aim is therefore to take a step into the IRBA direction. This includes looking for PD and LGD estimates.

There is one simple reason why this work can only be seen taking a step into the IRBA direction: the low amount of data. The original data set comprises of too few different companies and not a whole five years of data. This is definitely not enough to found proper statistical analysis which is the most important prerequisite for the estimation of parameters for the IRBA. So we have to leave the simple statistical methods that work nicely in the natural persons business and choose more complex models with some parameters that can be calibrated to the available data. At least this is what is done for the PD. For the LGD it turned out that there is too few data for

any kind of model. As the LGD rate can only be observed from insolvent companies, some insolvent companies are needed to give this estimate. But there is only one insolvent company and with that single one, no estimate for a loss rate given the company's default can be given.

## 1.2 Credit Scoring

To check a customer's degree of credit worthiness, credit scoring is a commonly used method. Here, several characteristics of the possible future obligor are aggregated to a credit score for this customer<sup>15</sup>. Every scoring system has an interval of integers as domain. Typically, a customer with a high degree of creditworthiness also achieves a high credit score.

In order to do any of the following, a data set is required. This data set must stem from several companies and for each company  $i$  there must be  $p$  characteristics with values  $x_i = (x_{i1}, \dots, x_{ip})$ . Typically this is information on the financial background, on the loan that is applied for, on the customer itself and on the securities. Some of this information must be transformed in order to enter the calculation of the credit score. As an example, the legal form as characteristic  $j$  of a company is partitioned into limited and unlimited liability. Then a company's characteristic 'limited by shares' is transformed by the mapping  $T_j$  from the set of characteristics into the set of numbers to the score  $-1$ , say. A company whose legal form gives unlimited liability would e.g. obtain the value  $+1$  in this characteristic. Some characteristics might be considered to be more relevant than others, so they are weighted with the factors  $w_j$  for  $j = 1, \dots, p$ . The total credit score  $S$  for customer  $i$  is then

$$S(x_i) = \sum_{j=1}^p w_j \cdot T_j(x_{ij})$$

if a simple linear model was chosen<sup>16</sup>.

The two key questions are which characteristics should be part of the scoring function and how to assess the weights  $w$ .

The first question can be answered by the power curve and the GINI coefficient that are presented in the following. The power curve gives some information about the discriminatory power of a rating system and the obligor's characteristics. The same holds for the GINI coefficient which is also known as accuracy ratio. The coefficient concentrates the information from the power curve into a single number.

The second question on how to assess the weights can be answered by implementing a logistic or linear regression, but we will not go into this subject.

---

<sup>15</sup>Hartmann-Wendels, Pfingsten, Weber: Banklehre, p. 498.

<sup>16</sup>This part follows Hartmann-Wendels, Pfingsten, Weber: Banklehre, p. 517.

**Power Curve.** To compute the power curve<sup>17</sup> take your historical data of company loans and make sure that each data point is given a value that tells us whether the customer defaulted or not. Then choose feature number  $j$  of your data for further analysis. Sort the data in increasing order by the feature you chose, i.e.  $T(x_{i_1j}) \leq \dots \leq T(x_{i_nj})$  where there are  $n$  credits and the set  $\{x_{i_1j}, \dots, x_{i_nj}\}$  is equal to the set  $\{x_{1j}, \dots, x_{nj}\}$ .

So far for the preparation. Now take the first  $x$  percent of the total amount of data and count the number of losses that occurred in this group. Divide this number of defaults by the total number of all defaults to obtain the value  $y$  of the defaulted credits in percentage. Do like this for  $x = 0, \dots, 100$ .

The result of this procedure can be seen in Figure 1.2, where a power curve for some fictional data is displayed. The thick line corresponds to the computed points  $(x, y)$ , the dashed line is the power curve of a perfect model and the diagonal would be a power curve of rating model based on randomness.

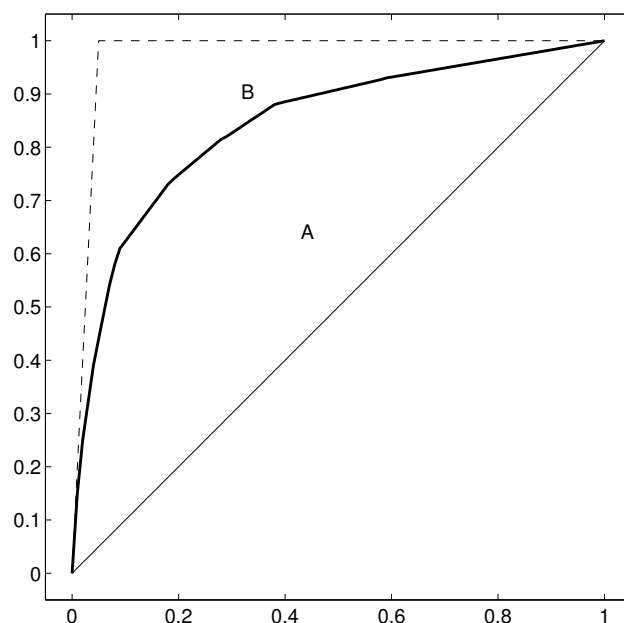


Figure 1.2: Power Curve. The thick line is the power curve of a fairly good scoring criterion. The dashed line represents the ideal scoring model, whereas the diagonal stands for a random model that does not give a relation between the resulting score and the observed default frequency.

<sup>17</sup>This part follows Ernst&Young, Validierung und Backtesting pp. 40 ff. [13].

In the case of Figure 1.2, the thick power curve would be a rather good example for a rating criterion. In the beginning the slope is high, then it is slowly decreasing. The explanation for this is rather simple; this happens if many of the defaulted credits had a low score value in the feature considered. The dashed line would picture the case of a feature with optimal discriminatory power. Here the list in increasing order has a block of all defaulted credits in the beginning. The diagonal is the power curve of an absolutely random model (with a big sample size) where the defaults are evenly spread among the sorted list.

With the help of the power curve it is easy to check whether a certain criterion could be of use in a credit scoring system. But the power curve is only applicable if this feature can (artificially) be arranged in increasing order by means of the mapping  $T(\cdot)$ .

**GINI Coefficient.** The GINI coefficient, also known as accuracy ratio, is the following quotient of the surface areas that can be seen in Figure 1.2:

$$GINI = \frac{A}{A + B}.$$

The higher the coefficient, the better the discriminatory power of a scoring system in the regarded feature or the credit scoring system in its whole.

If the accuracy ratio is greater than 60%, then the discriminatory power of the whole rating system is very good, with 50% it is good, with less than 40% it is rather poor, and with less than 20% it is not sufficient for practical purposes. For the discriminatory power of single characteristics, these percentages and their classifications do not hold. Ernst&Young takes the position that a GINI coefficient of 20% and higher for a single feature is a sign of a high discriminatory power.

### 1.3 Credit Rating

The scoring system is the basis of the rating system. Recall from Section 1.1 on legal requirements that an internal rating system for corporate credits must have at least eight rating classes. In this section it will be described how to form rating classes by a clustering method. The assessment of a PD to a rating class is then described in Chapter 2.

The main idea of clustering is to partition a set of  $I$  points or in this case score values  $p_1, \dots, p_I$  into  $K \leq I$  clusters  $C_1, \dots, C_K$  (rating classes). The optimization problem to be solved is

$$\min \sum_{k=1}^K \sum_{i=1: p_i \in C_k}^I \|p_i - \bar{C}_k\|^2$$

where  $\bar{C}_k$  is the mean value of all the points in the cluster  $C_k$  and  $\|\cdot\|$  denotes the Euclidean Norm.

The algorithm described in the following subsection was developed by Hartigan and Wong and is known as Algorithm AS 136. The formulation of the algorithm is partly kept close to the original text [15] but was changed when an easier reading was possible. Subsequently the operating mode is shown with a one-dimensional example that works for the clustering of score values into rating classes. Algorithm AS 136 is used by default when using the R function *kmeans*.

### 1.3.1 Clustering Algorithm

Input:  $I$  points in  $N$  dimensions and a matrix  $\bar{C}$  of size  $K \times N$  with the initial cluster centers.

**Step 1.** Introduce  $C_1$  and  $C_2$  as vectors of size  $I$ . For each point  $p_i$  find the closest center  $k_1$  (in terms of the Euclidean Norm) and the second closest center  $k_2$  and set  $C_1(i) \leftarrow k_1$  and  $C_2(i) \leftarrow k_2$ .  $C_1$  tells us for each point in which cluster this point is contained. If there is more than one cluster that could be assigned to a point, assign the cluster which comes first in your list of clusters<sup>18</sup>.

**Step 2.** Update the cluster centers  $\bar{C}$  with the cluster means.

**Step 3.** Introduce the  $K$ -dimensional indicator vector  $L$  with  $L[k] = 1$  if cluster  $k$  belongs to the so-called live set, 0 otherwise. Initially, all the clusters belong to the live set, i.e.  $L[k] \leftarrow 1$  for  $k = 1, \dots, K$ .

#### Step 4: Optimal Transfer (OPTRA) stage.

For all points  $p_i$  do:

Let  $p_i$  be in cluster  $k^*$ .

If this is the first time that the OPTRA stage is reached, continue with *Step 4a*.

ElseIf cluster  $k^*$  has been updated in the last QTRAN stage, then it belongs to the live set ( $L[k^*] \leftarrow 1$ ).

ElseIf cluster  $k^*$  has not been updated in the last  $I$  OPTRA steps assign  $L[k^*] \leftarrow 0$ .

If  $L[k^*] = 1$  continue with *Step 4a*,

Else continue with *Step 4b*.

---

<sup>18</sup>This is how it is done in the implementation given in [15], but it could also be done differently.

**Step 4a.**

If there are at least two points in cluster  $k^*$ , compute

$$\hat{k} \leftarrow \mathbf{argmin}_{k=1, \dots, K, k \neq k^*} \left\{ R_2(k) \leftarrow \frac{NC(k) \cdot \|p_i - \bar{C}[k]\|^2}{NC(k) + 1} \right\} \quad \text{and}$$

$$R_2(k^*) \leftarrow \frac{NC(k^*) \cdot \|p_i - \bar{C}[k^*]\|^2}{NC(k^*) - 1}$$

where  $NC(k)$  is the number of points in cluster  $k$ .

If  $R_2(\hat{k}) \geq R_2(k^*)$  the clusters stay as they are. Set  $C_2[i] \leftarrow \hat{k}$ .

Else assign the cluster  $\hat{k}$  to point  $i$ , i.e.  $C_1[i] \leftarrow \hat{k}$  and  $C_2[i] \leftarrow k^*$ .  
Update the cluster centers  $\bar{C}[\hat{k}]$  and  $\bar{C}[k^*]$ . Set  $L[k^*] \leftarrow 1$  and  $L[\hat{k}] \leftarrow 1$ .

**Step 4b.** Follow *Step 4a*, just minimize over the clusters in the live set only.

**Step 5.** Stop if  $L = (0, \dots, 0)$ . Otherwise do *Step 1* and *Step 2* and continue with *Step 6*.

**Step 6: Quick Transfer (QTRAN) stage.** Set  $\text{transfer} \leftarrow \text{TRUE}$ .

While (transfer)

transfer  $\leftarrow$  FALSE

For all points  $p_i$  compute

$$R_1 \leftarrow \frac{NC(C_1[i]) \cdot \|p_i - \bar{C}[C_1[i]]\|^2}{NC(C_1[i]) - 1} \quad \text{and}$$

$$R_2 \leftarrow \frac{NC(C_2[i]) \cdot \|p_i - \bar{C}[C_2[i]]\|^2}{NC(C_2[i]) + 1}.$$

If  $R_1 < R_2$  all clusters remain the same.

Else switch the clusters ( $C_1[i] \leftrightarrow C_2[i]$ ). Update the cluster centers. Set  $\text{transfer} \leftarrow \text{TRUE}$ .

Goto *Step 4*.

**1.3.2 Example: Clustering**

Let the points  $p = (p_1, \dots, p_6) = (0.5, 1.5, 2.5, 3.5, 6, 10)$  be given. Further, let  $\bar{C} = (\bar{C}_1, \dots, \bar{C}_3) = (0.75, 3, 7)$  be the initial cluster centers. This situation is displayed in Figure 1.3.

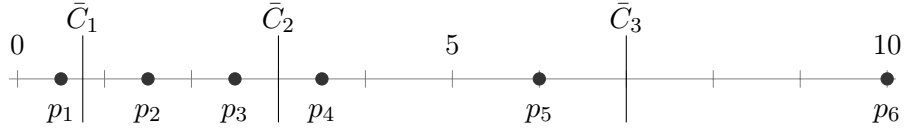


Figure 1.3: Points to be clustered and initial cluster centers

**Step 1.** For point  $p_1 = 0.5$  the closest center obviously is  $\bar{C}_1 = 0.75$  and the second closest center is  $\bar{C}_2 = 3$ . So  $C_1[1] \leftarrow 1$  and  $C_2[1] \leftarrow 2$ . Continuing like this for all six points leads to  $C_1 = (1, 1, 2, 2, 3, 3)$  and  $C_2 = (2, 2, 1, 1, 2, 2)$ .  $C_1$  can be seen in Figure 1.4

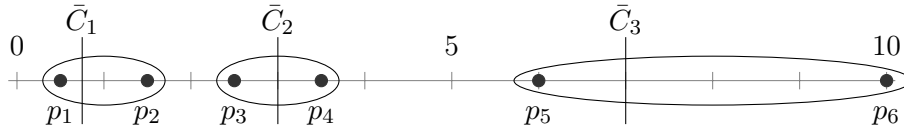


Figure 1.4: Initial cluster centers  $\bar{C}$  with cluster  $C_1$  (ellipses).

**Step 2.** The first cluster contains the points  $p_1$  and  $p_2$ . So the new cluster mean is  $(P_1 + P_2) \cdot 0.5 = (0.5 + 1.5) \cdot 0.5 = 1$ . For the other cluster means do the same. Then  $\bar{C} = (1, 3, 8)$  and the situation is as in Figure 1.5.

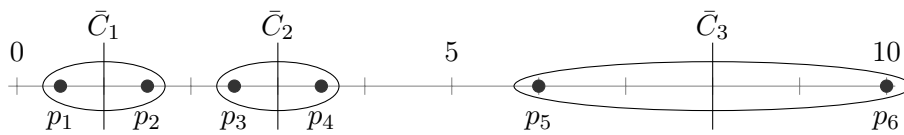


Figure 1.5: New cluster centers  $\bar{C}$  with cluster  $C_1$ .

**Step 3.** All clusters belong to the live set, therefore set  $L = (1, 1, 1)$ .

**Step 4.** Consider  $p_1$ , which is in cluster  $k^* = 1$ . As this is the first time we enter this step, we continue with *Step 4a*.

We compute

$$R_2(k^*) = \frac{NC(k^*) \cdot \|p_1 - \bar{C}[k^*]\|^2}{NC(k^*) - 1} = \frac{2 \cdot \|0.5 - 1\|^2}{2 - 1} = 0.5$$

$$R_2(2) = \frac{NC(2) \cdot \|p_1 - \bar{C}[2]\|^2}{NC(2) + 1} = \frac{2 \cdot \|0.75 - 3\|^2}{2 + 1} = 3.375$$

$$R_2(3) = \frac{NC(3) \cdot \|p_1 - \bar{C}[3]\|^2}{NC(3) + 1} = \frac{2 \cdot \|0.75 - 8\|^2}{2 + 1} \approx 35.04$$

since  $R_2(k^*) < R_2(2) < R_2(3)$ , the clusters stay as they are and we set  $C_2[1] \leftarrow 2$ .

For  $p_2, p_3$  and  $p_4$  nothing happens as well, not even  $C_2$  changes. But for  $p_5$  things are different:

$$R_2(k^*) = \frac{NC(k^*) \cdot \|p_5 - \bar{C}[k^*]\|^2}{NC(k^*) - 1} = \frac{2 \cdot \|6 - 8\|^2}{2 - 1} = 8$$

$$R_2(1) = \frac{NC(1) \cdot \|p_5 - \bar{C}[1]\|^2}{NC(1) + 1} = \frac{2 \cdot \|6 - 1\|^2}{2 + 1} \approx 16.67$$

$$R_2(2) = \frac{NC(2) \cdot \|p_5 - \bar{C}[2]\|^2}{NC(2) + 1} = \frac{2 \cdot \|6 - 3\|^2}{2 + 1} = 6$$

That is,  $R_2(2)$  is the minimum value. We set  $C_1[5] \leftarrow 2$  and  $C_2[5] \leftarrow 3$ , giving  $C_1 = (1, 1, 2, 2, 2, 3)$  and  $C_2 = (2, 2, 1, 1, 3, 2)$ . Updating the means of clusters 2 and 3 gives  $\bar{C} = (1, 4, 10)$  and a new Figure 1.6.

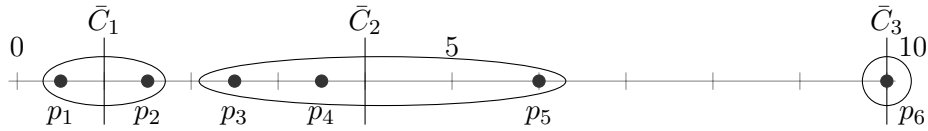


Figure 1.6: New clusters.

The point  $p_6$  can be skipped as it is the only point in cluster 3 and we are done with *Step 4*.

**Step 5.**  $L = (1, 1, 1)$  so *Step 1* and *Step 2* should be done; i.e. with the new cluster means assign to each point the closest and second closest cluster. The interesting point is  $p_3$  as this point could be assigned to both cluster 1 and cluster 2. However, as cluster 1 is the first cluster,  $C_1[3] \leftarrow 1$  and  $C_2[3] \leftarrow 2$ . So the new  $C_1 = (1, 1, 1, 2, 2, 3)$  and  $C_2 = (2, 2, 2, 1, 3, 2)$ . The new cluster means are  $\bar{C} = (1.5, 4.75, 10)$ . In Figure 1.7 the situation is shown.



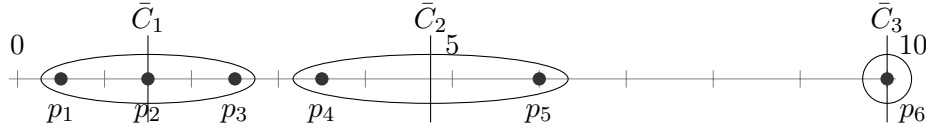


Figure 1.7: Reclustered clusters after the first OPTRA stage.

**Step 6.** We start with  $p_1$ .

$$R_1 = \frac{NC(C_1[1]) \cdot \|p_1 - \bar{C}[C_1[1]]\|^2}{NC(C_1[1]) - 1} = \frac{3 \cdot \|0.5 - 1.5\|^2}{2} = 1.5$$

$$R_2 = \frac{NC(C_2[1]) \cdot \|p_1 - \bar{C}[C_2[1]]\|^2}{NC(C_2[1]) + 1} = \frac{4 \cdot \|0.5 - 4.75\|^2}{5} = 14.45$$

Since  $R_1 < R_2$ , we go to the next point. For points  $p_2$  and  $p_3$  the same is the case. For  $p_4$  we obtain

$$R_1 = \frac{NC(C_1[4]) \cdot \|p_4 - \bar{C}[C_1[4]]\|^2}{NC(C_1[4]) - 1} = \frac{2 \cdot \|3.5 - 4.75\|^2}{1} = 3.125$$

$$R_2 = \frac{NC(C_2[4]) \cdot \|p_4 - \bar{C}[C_2[4]]\|^2}{NC(C_2[4]) + 1} = \frac{1 \cdot \|3.5 - 1.5\|^2}{2} = 2$$

Now  $R_1 > R_2$  and we switch the clusters  $C_1[4] \leftarrow 1$  and  $C_2[4] \leftarrow 2$ . The new situation is displayed in Figure 1.8 and the new cluster means are  $\bar{C} = (2, 6, 10)$ .

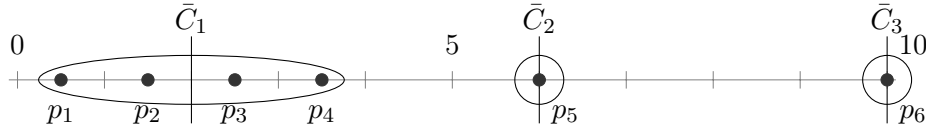


Figure 1.8: Reclustering within the first QTRAN stage.

We continue. Since the number of points in cluster 3 is one, we skip point  $p_6$  and start again with the while-loop because there was a transfer. However, there is no transfer the second time we enter the for-loop, so recalling that there was an update in clusters 1 and 2, we go to *step 4*.

**Step 4.** Since clusters 1 and 2 have been updated in the last QTRAN stage and since cluster 3 has been updated in the last 3 OPTRA steps, we still have  $L = (1, 1, 1)$ . However, there is no update necessary at this step.

**Step 5.** We do *step 1* and *step 2*. There we do not update anything at all.

**Step 6.** Since nothing has happened since the last time we were here, there will not happen anything here either.

**Step 4.** No cluster has been updated in the last QTRAN stage and not in the last  $I$  OPTRA steps, so the live set is empty, i.e.  $L = (0, 0, 0)$ .

Finally we enter *step 4b*, but since the live set is empty, there is nothing to do here.

**Step 5.** The algorithm terminates. The result is given by  $C_1$ , telling us for each point in which cluster it is contained. Also interesting is  $\bar{C}$ , the vector of cluster means. Figure 1.8 shows the final result.

## 1.4 Expected Loss and Value-at-Risk

Now we know how to form rating classes and in the next chapter we will see how to assess a PD to each rating class. But we still need a risk measure in order to calculate the risk a banking company faces. In this context the expected loss (EL) and the Value-at-Risk measure are quantities that often are used. But prior to the introduction of these risk quantities we specify the Exposure at Default (EAD) and the Loss Given Default (LGD) rate.

The EAD is the amount of money the obligor is legally obligated to pay at the time the calculations for the EL or VaR are done. In the easy case of assets from the balance sheet the EAD is the amount of the loan that still has to be payed back plus the interest on the loan (and eventually plus some other debt). In the more complicated case of off-balance sheet positions, an estimate of the EAD is requested<sup>19</sup>. These positions might be a credit that the banking institution has offered to a customer, but that is not used so far. In this case, the estimate of the EAD is 75% of the "undrawn part of a credit line"<sup>20</sup>.

The LGD is given in percentage and describes the part of the EAD that will be lost in the case of a company's default. Typically the LDG depends upon given securities. For example, the LGD might be 5% if the security is a savings account or 90% if there are no given securities.

Both EAD and LGD are often 'given'. These two factors should be random variables as neither the EAD nor the LGD are known prior to default. Often there are securities whose liquidation costs a different amount than expected or that vary in value themselves. Therefore the LGD might be different to what was expected in the first place. And the EAD can increase or decrease from one day to the other.

Anyway, in the real data set the estimates for the EAD and LGD have been computed. As there was just one insolvent company from that we

<sup>19</sup>Source: Ernst&Young, Fachkonzept LDG-Modell.

<sup>20</sup>Supervisory conversion factor; Section 101, Subsection 2.1.b SolvV [20].

could get an LGD rate, no statistical means could be used to model this rate. Nevertheless, there was a way such that an estimate could be assigned.

To have the same data structure for the real and the fictional data, both the values for the EAD and the LGD are given for this report. Therefore it remains just to estimate the PD for each obligor. This is all we need to calculate EL and VaR.

**Expected Loss.** The expected loss (EL) from a lending business for a single credit  $i$  is the product

$$EL_i = EAD_i \cdot LGD_i \cdot PD_i \quad (1.2)$$

where  $PD_i$  is the Probability of Default for customer  $i$ ,  $LGD_i$  and  $EAD_i$  correspondingly. For a portfolio of credits the EL is the sum of the  $EL_i$  of the single credits  $i$  with the assumption that the PDs are independent. There will not be any problem if one customer has more than one credit, as all the credits of a customer will be added together to a credit commitment.

**Value-at-Risk.** To obtain a value for this common risk measure, we have to simulate from the loss distribution. The loss from investing in commitment  $i$  is (today)

$$L_i = EAD_i \cdot LGD_i \cdot X_i$$

and the total portfolio loss is

$$L = \sum_{i=1}^n L_i = \sum_{i=1}^n EAD_i \cdot LGD_i \cdot X_i \quad (1.3)$$

where there are  $n$  commitments and  $X_i$  is the binary default indicator, i.e. if  $X_i = 1$  then company  $i$  defaults.  $X$  is an  $n$ -dimensional random vector following some distribution. As  $X$  is binary, a Bernoulli distribution is a good choice. The question is how to assess the probability  $p$  in the *Bernoulli*( $p$ ) distribution and this will be answered in Chapter 2. With equation (1.3) we can simulate from the loss distribution. The output will be viewed as a sample from the loss distribution  $L^{21}$ .

Based on a random sample  $\{L_1, \dots, L_m\}$  of size  $m$  the empirical distribution function  $F_{m,L}$  is given by

$$F_{m,L}(l) = \frac{1}{m} \sum_{k=1}^m \mathbb{1}_{\{L_k \leq l\}}$$

where  $\mathbb{1}$  is the indicator function. The empirical quantile function is then

$$F_{m,L}^{-1}(p) = \min\{x : F_{m,L}(l) \geq p\}.$$

---

<sup>21</sup>Source: Hult, Lindskog, Hammerlid, Rehn, Risk and Portfolio Analysis.

In Hult et al., Risk and Portfolio Analysis, it is shown in Chapter 7 that the empirical quantile is

$$F_{m,L}^{-1}(1-p) = L_{\lfloor mp \rfloor + 1, m}$$

where  $L_{1,m} \geq \dots \geq L_{m,m}$  is the ordered sample and  $\lfloor x \rfloor$  denotes the integer part of  $x$ .

So when calculating the VaR, this will be

$$VaR_p = L_{\lfloor mp \rfloor + 1, m},$$

the  $(\lfloor mp \rfloor + 1)$ th largest value of the simulated sample from the loss distribution.

Note that when calculating VaR, we use today's EAD and LGD and yearly default rates. So the VaR here is in principle the  $(\lfloor mp \rfloor + 1)$ th largest amount of money that could be lost if all commitments that are to default during the next year default tomorrow (without changed values for EAD and LGD).

The title of this thesis contains 'Credit' Value-at-risk. But there is no real difference between Credit VaR and VaR. When using Credit VaR, it is most often referred to the VaR being calculated with time horizon being one year. But the VaR is often calculated with a much shorter time horizon such as 10 days for example. In the following we will return to the shorter VaR notation but keep in mind that the time period for which it is computed is one year.

## Chapter 2

# Models to estimate a PD

Three different ways are presented in this chapter how to calculate a PD: a calibration curve and two mixture models.

### 2.1 Calibration Curve

There is an easy way to determine default probabilities if there is a possibility to score credits and if a historical data base is available. The historical data must contain both defaulted and non-defaulted credits along with their scores<sup>1</sup>.

To determine the calibration curve firstly sort the data by its scores. Then form  $R$  groups (buckets) and calculate the historical default frequency  $\hat{D}_r$  for  $r = 1, \dots, R$  within the buckets. The buckets may be rating classes resulting from a cluster analysis.

Here we use an exponential regression equation to determine the functional relation between the rating class  $r$  and the empirical default frequency  $DF_r$  for rating class  $r$ , i.e. it should hold that:

$$DF_r = b_0 \cdot e^{b_1 r} \quad r = 1, \dots, R. \quad (2.1)$$

Now the statistic software R can be used to solve the optimization problems

$$(P_1) \quad \min \sum_{r=1}^R |b_0 \cdot e^{b_1 r} - DF_r| \cdot w_r$$
$$(P_2) \quad \min \sum_{r=1}^R (b_0 \cdot e^{b_1 r} - DF_r)^2 \cdot w_r$$

where  $w_r$  is the fraction of commitments in rating class  $r$ .

---

<sup>1</sup>This section is based upon from Hartmann-Wendels, Pfingsten, Weber: Banklehre, pp. 528 ff.

With the estimates for the parameters  $\hat{b}_0$  and  $\hat{b}_1$ , the estimated probability of default  $\hat{PD}_r$  for each rating class  $r$  can be calculated analogously to (2.1):

$$\hat{PD}_r = \hat{b}_0 \cdot e^{\hat{b}_1 r}, \quad r = 1, \dots, R. \quad (2.2)$$

Note that Section 88, Subsection 4 SolvV determines a lower bound of 0.03% for the probability of default, hence it holds that

$$PD_r = \max \left\{ \hat{PD}_r, 0.0003 \right\} \quad r = 1, \dots, R.$$

Now we leave the simple PD model that did not regard any possible dependence structure of company defaults. But default dependence cannot be neglected as the default of some company influences the companies that depend in some way on the defaulted company. Just that this is very difficult to model. The mixture models find another way to express dependence.

## 2.2 Mixture Models

With these mixture models it is possible to model the default event without knowing the dependence structure among companies. If a stochastic 'common factor' is given, independence of the company defaults may be assumed conditional on this factor. The common factor typically is an indicator for the health of a country's economy<sup>2</sup>.

The mixture model belongs to the static credit risk models. In these models the loss of a portfolio in one year is estimated. Here it is not important when a default happens, it is only cared if a company defaults. This is modeled by a random variable indicating the default event<sup>3</sup>.

Consider  $N$  companies that are solvent today. Introduce the binary random variable  $X_i$  for each company  $i = 1, \dots, N$  as a default indicator. If  $X_i = 1$  then company  $i$  defaults within the next year, else  $X_i = 0$  and company  $i$  does not default. An assumption of the Bernoulli mixture model is that the probability of default is a function of the common factor  $Z$ , so for company  $i$  the probability of default is

$$P(X_i = 1 | Z = z) = \pi_i(z)$$

where  $\pi_i : \mathbb{R} \mapsto [0, 1]$ . The even stronger assumption is that the probabilities of default of the  $N$  companies are independent conditional on the common

<sup>2</sup>This chapter follows McNeil, Frey, Embrechts and Hult, Lindskog, Hammerlid, Rehn.

<sup>3</sup>The opposite models are dynamic models often used to evaluate credit derivatives. In these models a stochastic process is used to model the portfolio value as it changes over time.

factor  $Z$ :

$$P(X = x|Z = z) = \prod_{i=1}^N P(X_i = x_i|Z = z).$$

As  $X \in \{0, 1\}$ , it holds that

$$P(X = x|Z = z) = \prod_{i=1}^N \pi_i(z)^{x_i} (1 - \pi_i(z))^{1-x_i}. \quad (2.3)$$

The formal definition is as follows:

**Definition 2.1** *Given the common factor (and random variable)  $Z$ , the  $N$ -dimensional random vector  $X$  follows a **Bernoulli mixture model with factor  $Z$**  if there are functions  $\pi_i : \mathbb{R} \mapsto [0, 1]$ ,  $1 \leq i \leq N$ , such that the components of  $X$  are independent Bernoulli random variables satisfying  $P(X_i = 1|Z = z) = \pi_i(z)$ .<sup>4</sup>*

It is to emphasize that the PD is not  $\pi_i(Z)$ , as this is the probability of default conditional on the economical indicator  $Z$ . Instead, the probability of default  $PD_i$  for company  $i$  has to be the same for all possible scenarios for  $Z$ , and is thus given by

$$PD_i = \mathbb{E}[\pi_i(Z)] = \int_{-\infty}^{\infty} \pi_i(z) f_Z(z) dz$$

where  $f_Z(z)$  is the probability density function of the common factor  $Z$ .

So it remains to specify the so-called mixing variable  $\pi_i(z)$ . Two ways will be presented in the following: The first approach uses a Probit-link and the common factor  $Z$  is Standard Normal-distributed. Here a score value or a rating class can be used to influence  $\pi_i(z)$ . In the second approach a Beta distribution is fitted to historical economical data, i.e.  $Z$  has a Beta distribution.

### 2.2.1 Bernoulli Mixture Model without economical data: Probit-Normal mixing distribution

If there is no reliable economical data available, the mixing variable  $\pi_i(Z)$  might be modeled with a Probit-Normal mixing distribution.

The mixing variable is defined as

$$\pi_i(z) = P(X_i = 1|Z = z) = \Phi(\mu + cs_i + \sigma z) \quad (2.4)$$

---

<sup>4</sup>Very close to in McNeil, Frey, Embrechts: Quantitative Risk Management, Definition 8.10.

where  $\Phi$  denotes the cumulative distribution function of the standard Normal distribution,  $\mu$  is the intercept,  $s_i$  is a score value,  $c$  and  $\sigma$  are weights and  $Z \sim N(0,1)$ . Note that the score of a customer with a high credit worthiness must be a small value (and the score of a customer with a low credit worthiness must be a high value) as  $\Phi$  is strictly increasing. Inserting equation (2.4) into equation 2.3 gives

$$P(X = x|Z = z) = \prod_{i=1}^N \Phi(\mu + cs_i + \sigma z)^{x_i} (1 - \Phi(\mu + cs_i + \sigma z))^{1-x_i}$$

as conditional probability. The unconditional probability of default for company  $i$  is then again computed via integration over the common factor  $Z$ :

$$PD_i = \mathbb{E}[\Phi(\mu + cs_i + \sigma Z)] = \int_{-\infty}^{\infty} \Phi(\mu + cs_i + \sigma z) \phi(z) dz \quad (2.5)$$

where  $\phi$  denotes the probability density of the Standard Normal distribution.

Now we know how to compute the PD of a single company. But what about the PD of a rating class? In the remaining part of this section, we disregard the single companies and study rating classes instead. Therefore the companies are grouped by their score value into rating classes.

Assume that there are  $k$  different rating classes. Define the mixing variable for each rating class as

$$\pi_r(z) = \Phi(\mu_r + \sigma z) \quad \text{for } r = 1, \dots, k \quad (2.6)$$

where  $\mu_r$  is the intercept for rating class  $r$ . Further,  $\Phi$  is the Standard Normal probability distribution function and  $\sigma > 0$  is again a scaling parameter.

The unconditional probability of default for a company in rating class  $r$  is then

$$PD_r = \mathbb{E}[\Phi(\mu_r + \sigma Z)] = \int_{-\infty}^{\infty} \Phi(\mu_r + \sigma z) \phi(z) dz \quad (2.7)$$

Now we want to know the probability that  $M$  companies default. Write  $n_r$  for the number of companies in rating class  $r$  and let  $M_r$  be a random variable for the number of defaulted companies in rating class  $r$ . By inserting equation (2.6) into equation (2.3), we obtain for the conditional probability on  $Z$  that  $m = m_1, \dots, m_k$  firms in the rating classes  $r = 1, \dots, k$  default:

$$P(M = m|Z = z) = \prod_{r=1}^k \binom{n_r}{m_r} \Phi(\mu_r + \sigma z)^{m_r} (1 - \Phi(\mu_r + \sigma z))^{n_r - m_r}. \quad (2.8)$$

The binomial coefficient in (2.8) arises because  $M_r$  is the sum of  $n_r$  Bernoulli( $\Phi(\mu_r + \sigma z)$ ) random variables. The point here for the appearance of



the Binomial( $n_r, \Phi(\mu_r + \sigma z)$ ) distribution is that the conditional probability of default is the same for the companies within one group. Also the unconditional probability can be computed; for a rating class  $r$  the unconditional probability that exactly  $m = m_r$  out of  $n = n_r$  credits default is

$$P(M = m) = \int_{-\infty}^{\infty} \binom{n}{m} \Phi(\mu_r + \sigma z)^m (1 - \Phi(\mu_r + \sigma z))^{n-m} \cdot \phi(z) dz.$$

Now we know how to compute both the PD of single companies and rating classes. What remains is to find the parameters  $\mu$ ,  $c$  or  $\mu_r$  and  $\sigma$ . These can be computed by Maximum Likelihood (ML) Estimation. Therefore recall that the ML estimator  $\hat{\Theta}$  is given by

$$\hat{\Theta} = \operatorname{argmax}_{\Theta} \prod_{r=1}^k f_{Y;\Theta}(Y_r)$$

where  $f_{Y;\Theta}$  is the density function of the random variable  $Y$  as a function of  $\Theta$ . Back to the the actual problem with  $\mu_r$  and  $\sigma$  in (2.8). Since the number of defaults  $M_r$  within a rating class is dependent on the outcome of the random variable  $Z$ , we are dealing with the joint density  $f_{M,Z;\mu,\sigma}$  of the random vector  $M$  and the random variable  $Z$ . The bivariate density can be factored as

$$f_{M,Z;\mu,\sigma} = f_{M|Z;\mu,\sigma} \cdot f_Z$$

where  $f_Z$  is the standard Normal density. If there is data on the number of companies within a rating class and the number of defaulted companies within the rating class from the years  $t = 1, \dots, T$  available, the density  $f_{M|Z;\mu,\sigma}$  of  $M$  conditional on  $Z$  is

$$\begin{aligned} f_{M|Z;\mu,\sigma}(n_{t,r}, m_{t,r}, z_t) &= P(M_{t,r} = m_{t,r} | Z_t = z_t) \\ &= \binom{n_{t,r}}{m_{t,r}} \Phi(\mu_r + \sigma z_t)^{m_{t,r}} (1 - \Phi(\mu_r + \sigma z_t))^{n_{t,r} - m_{t,r}}. \end{aligned}$$

Here the number of companies  $n = (n_{t,r})$  and the number of defaulted companies  $m = (m_{t,r})$  in rating class  $r$  at time  $t$  is given and in matrix form.  $M$  is a random matrix.

Then the preliminary Likelihood function for the rating classes becomes

$$\begin{aligned} L^{prelim.} &= \operatorname{argmax}_{\mu,\sigma} \prod_{t=1}^T \prod_{r=1}^k \binom{n_{t,r}}{m_{t,r}} \Phi(\mu_r + \sigma z_t)^{m_{t,r}} \\ &\quad (1 - \Phi(\mu_r + \sigma z_t))^{n_{t,r} - m_{t,r}} \cdot \phi(z_t). \end{aligned}$$

Preliminary because this function still depends on the outcome of  $Z$ . Getting rid of the dependence on the outcome of  $Z$  leads us to the final Likelihood function:

$$L = \operatorname{argmax}_{\mu, \sigma} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{t=1}^T \prod_{r=1}^k \binom{n_{t,r}}{m_{t,r}} \Phi(\mu_r + \sigma z_t)^{m_{t,r}} (1 - \Phi(\mu_r + \sigma z_t))^{n_{t,r} - m_{t,r}} \cdot \phi(z_t) dz_1 \dots dz_T.$$

Assuming that  $Z_1, \dots, Z_T$  is not only Standard Normal but also independent gives the nicer expression

$$L = \operatorname{argmax}_{\mu, \sigma} \prod_{t=1}^T \int_{-\infty}^{\infty} \prod_{r=1}^k \binom{n_{t,r}}{m_{t,r}} \Phi(\mu_r + \sigma z_t)^{m_{t,r}} (1 - \Phi(\mu_r + \sigma z_t))^{n_{t,r} - m_{t,r}} \cdot \phi(z_t) dz_t$$

for the Likelihood function. To simplify further, the log-Likelihood function is

$$LL = \operatorname{argmax}_{\mu, \sigma} \sum_{t=1}^T \log \left( \int_{-\infty}^{\infty} \prod_{r=1}^k \binom{n_{t,r}}{m_{t,r}} \Phi(\mu_r + \sigma z_t)^{m_{t,r}} (1 - \Phi(\mu_r + \sigma z_t))^{n_{t,r} - m_{t,r}} \cdot \phi(z_t) dz_t \right) \quad (2.9)$$

and it can easily be solved using R or some other software. Finally, we have the parameters needed to compute the PD from equation (2.7) (for the PD from (2.5) the LL function is computed analogously).

## 2.2.2 Bernoulli Mixture Model with Economical Data: Beta Mixing Distribution

The common factor  $Z$  is of great importance in the mixture models as conditional on just this factor, independence of the defaults is assumed. McNeil, Frey and Embrechts refer to the book by Duffie and Singleton who found a negative correlation between GDP growth rates and default rates. As the regarded portfolio consists of credits given to German customers, the German GDP growth rates<sup>5</sup> will be studied in Section 4.3.

In general, the common factor is some economical indicator which is denoted by the random vector  $Z$ . In the Bernoulli mixture model with a Beta mixing distribution  $Z$  is assumed to be Beta( $a, b$ ) distributed. Then the density of  $Z$  is

$$g(z) = \frac{1}{\beta(a, b)} z^{a-1} (1-z)^{b-1} \quad \text{for } z \in (0, 1), a, b > 0$$

<sup>5</sup>Published by Eurostat, see appendix C.

where

$$\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 z^{a-1}(1-z)^{b-1}dz, \quad \text{for } a, b > 0$$

and  $\Gamma(\cdot)$  denotes the Gamma function. Further, the expected value and the variance are given by

$$\mathbb{E}[Z] = \frac{a}{a+b} \quad \text{and} \quad \text{Var}(Z) = \mathbb{E}[Z] \frac{b}{(a+b)(a+b+1)}. \quad (2.10)$$

If, in practice, there is historical data on the economical indicator available, both the expected value and the variance can be computed. From the equations in (2.10) we get

$$a = \frac{b\mathbb{E}[Z]}{1 - \mathbb{E}[Z]} \quad \text{and} \quad b = \frac{\mathbb{E}[Z]}{\text{Var}(Z)} + \mathbb{E}[Z] - 1 \quad (2.11)$$

and with the condition that

$$0 < \frac{\text{Var}(Z)}{1 + \text{Var}(Z)} < \mathbb{E}[Z] < 1 \quad (2.12)$$

it is fulfilled that  $a > 0$  and  $b > 0$ . For more details see appendix B.

Now we have a distribution for  $Z$ . The mixing variable  $\pi_i(Z)$  denoting the probability of default conditional on  $Z$  for company  $i$  is now

$$P(X_i = 1|Z = z) = \pi_i(z) = z$$

where  $Z \sim \text{Beta}(a, b)$ .

For the unconditional probability of default  $p$  we obtain

$$p = \mathbb{E}[\pi_i(Z)] = \mathbb{E}[Z] = \frac{a}{a+b}.$$

Let now  $M = X_1 + \dots + X_n$ . Then the probability of the default of  $m$  out of  $n$  companies given the information  $z$  is

$$P(M = m|Z = z) = \binom{n}{m} z^m (1-z)^{n-m}$$

and the unconditional probability of the default is

$$\begin{aligned} P(M = m) &= \int_0^1 \binom{n}{m} z^m (1-z)^{n-m} \frac{1}{\beta(a, b)} z^{a-1} (1-z)^{b-1} dz \\ &= \binom{n}{m} \frac{1}{\beta(a, b)} \int_0^1 z^{m+a-1} (1-z)^{n-m+b-1} dz \\ &= \binom{n}{m} \frac{\beta(m+a, n+b-m)}{\beta(a, b)}. \end{aligned}$$



## Chapter 3

# Data Base

Since one customer often has more than one credit, it makes more sense to regard loan commitments instead of single credits. Such a commitment groups several credits of one customer or it groups several credits of several customers that are related in some financial way. The advantage of using commitments instead of several credits for each customer is that if an obligor defaults, just the commitment has to be marked as defaulted.

The data set that will be used has been copy/pasted, changed, shifted and multiplied by my favorite (random) numbers. So the data does no longer represent the Debeka Bausparkasse AG's data. By using the changed data set the results from the three different models can be compared. The only thing that was kept the same is the low number of data points and the time period from 2008/03 to 2013/01 (but the number of commitments in a time step has been changed sometimes).

**Historical Default frequency.** To compute the historical default frequency, the historical data of the commitments is studied. The aim is to estimate the one-year PD. First all the historical one-year default frequencies are calculated, that is, from January to January, from February to February, ... evaluate the quotient

$$\frac{\# \text{ defaulted commitments between time } t - 12 \text{ and } t}{\# \text{ commitments}},$$

where  $t = 2008/03, \dots, 2013/01$ . The number of commitments and the number of defaulted commitments relates to the number of commitments that have already been valid and were non-defaulted loan contracts at time  $t - 12$  [11]. The historical default rate based on commitments can be seen in Figure 3.1.

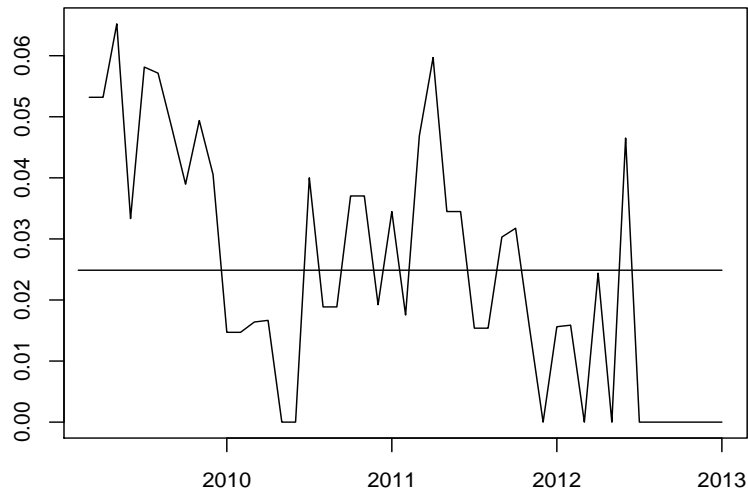


Figure 3.1: Historical default rate based on commitments. The straight line is the average PD of 0.0249.

**Scores and Rating classes.** In Figure 3.2 the distribution of the credit commitments' score values is shown.

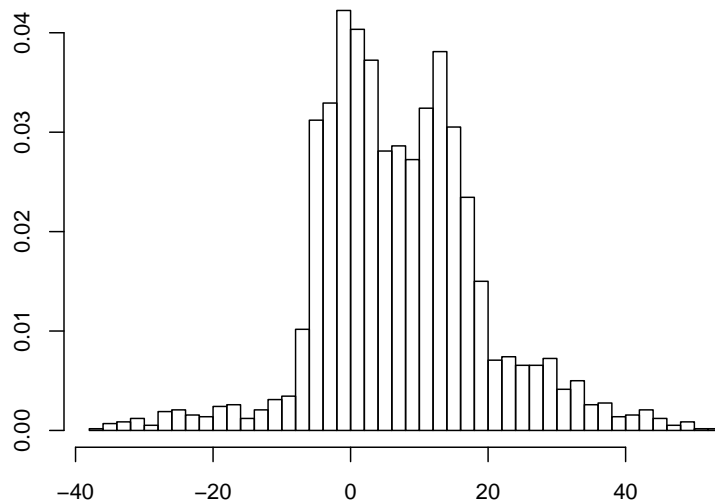


Figure 3.2: Score values.

The scores are now grouped into the rating classes by R's *kmeans* algorithm. At first we have to determine how many rating classes we want to obtain. Say we want to have the legal minimum of 8 rating classes. Here we have one rating class reserved for the defaulted credits that must not be part of the clustering. Observe that we want to find the historical default frequencies for the seven other rating classes and therefore we cannot separate the defaulted credits from the healthy ones. Second, we must assume that the credit scores that were obtained cover all the seven rating classes for solvent commitments.

So we run the *kmeans* algorithm in R on seven clusters. We run the algorithm for a couple of times as it only gives local minima and we choose the lowest local minimum.

The distribution of the commitments into the rating classes is as in Table 3.1. We observe a nicely shaped distribution of the commitments into the clusters. The calculated default frequencies tend to increase with the rating class. This coincides with other empirical studies that have already shown that the default rate increases when credit worthiness decreases<sup>1</sup>.

Table 3.1: Credit commitments in 7 clusters. *Second row*: Cluster mean. *Third row*: Distribution into the clusters. *Fourth row*: Default frequency.

Rating class	1	2	3	4	5	6	7
Cluster mean	41.65	28.82	17.71	10.75	2.16	-4.49	-22.77
Fraction (in %)	2.38	6.90	15.97	25.28	29.59	16.59	3.31
Default freq. (in %)	0.00	1.00	0.43	1.91	4.31	3.74	8.33

Based on the values in Table 3.1, the model fitting can begin.

---

<sup>1</sup>Frankfurt School of Finance and Management, Working Paper No. 179: Ratingverfahren: Diskriminanzanalyse versus Logistische Regression





## Chapter 4

# Testing the Models on Data

Using the fictional data presented in the previous chapter we can start testing the introduced models. First the calibration curve is calculated giving default rates for all rating classes. Then the mixing models are tested as well.

### 4.1 Calibration Curve

The calibration curve is a simple way of obtaining default probabilities from historical default frequencies. It is intended to be used if there are score values or rating classes available. Now rating classes are used.

We use the empirical default frequencies and the weights from the previous chapter to solve the optimization problems ( $P1$ ) and ( $P2$ ). In Table 4.1 and in Figure 4.1 the input values and results are displayed. The calibration curve fitted with the linear optimization problem gives the more extreme fit.

Table 4.1: Estimated default probabilities for the calibration curve and the empirical default frequency (DF) that was used to calibrate the curve.

Rating class	1	2	3	4	5	6	7
Emp. DF(%)	0.00	1.00	1.00	1.91	4.31	4.31	8.33
(P1)-PD(%)	0.42	0.69	1.15	1.91	3.17	5.26	8.72
(P2)-PD(%)	0.59	0.90	1.37	2.10	3.21	4.90	7.49

For the VaR calculations we will only use the result from (P1) as the estimated default probability is higher than the empirical default frequency in rating classes with a large expected number of defaults.

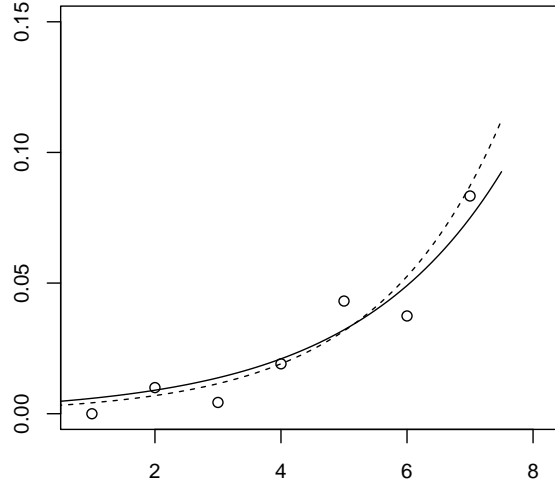


Figure 4.1: Default probabilities in dependence of rating classes. The historical default frequencies are displayed as circles. The dashed calibration curve corresponds to the optimization problem (P1) and the continuous line corresponds to the quadratic optimization problem (P2).

## 4.2 Probit-Normal Mixing Distribution

When using the normal mixture model, the input data is two matrices:

$n_{tr}$  is the number of commitments at time  $t$  in rating class  $r$ ,

$m_{tr}$  is the number of defaulted commitments at time  $t$  in rating class  $r$ ,

where  $t = 2009/03, \dots, 2013/01$  and  $r = 2, \dots, 7$  (as no defaults occur in rating class 1). These matrices are plugged into the log-likelihood function (2.9) and solved in R.

The estimated PD as in equation (2.7) in the Bernoulli mixture model with a normal mixing distribution has been calculated and can be seen in Table 4.2 as  $PD^{output}$ . As the minimum PD for any rating class is 0.03%, and since the PDs should be increasing with the rating class, the PDs that result from the model are adjusted to the  $PD$  in Table 4.2.

Table 4.2: The normal mixture model.  $PD^{output}$ : the PDs for the rating classes estimated by the model.  $PD$ : the adjusted PDs s.t. the minimum PD is 0.03% and s.t. the PDs are nondecreasing with increasing rating class.

Rating class	1	2	3	4	5	6	7
$PD^{output}$ (%)	-	0.00	0.90	2.07	6.33	4.62	0.01
$PD$ (%)	0.03	0.03	0.90	2.07	6.33	6.33	6.33

### 4.3 Beta Mixing Distribution

As there seems to be a connection between a country's GDP and the default rate of its companies (as suggested by Duffie and Singleton), a function of the German GDP will serve as common factor. The historical German GDP per capita in Euro is displayed in the upper plot of Figure 4.2. The data is now transformed:

Let  $GDP_t$  be the value of the GDP in year/month  $t = 2008/03, \dots, 2013/01$ . Then introduce

$$LGDP_t = \log\left(\frac{GDP_t}{GDP_{t-12}}\right) \quad \text{for } t = 2009/03, \dots, 2013/01.$$

The result is displayed in the lower plot of Figure 4.2.

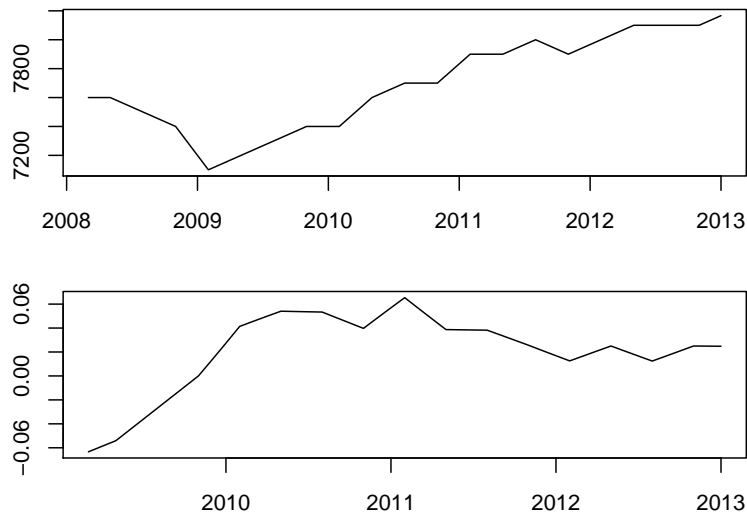


Figure 4.2: *Upper plot*: historical German GDP per capita in Euro from 2008/03 to 2013/01. *Lower plot*: German log-GDP-growth rates.

Because the Beta distribution does only take positive values, we have to transform the  $LGDP$  values, i.e. we have to find a function  $\tau$  such that  $\tau(LGDP) \mapsto [0, 1]$ . A general function is

$$\tau(LGDP) = \Phi(\sigma(LGDP - \mu))$$

where  $\Phi$  denotes the cumulative distribution function of the Standard Normal distribution. Then the parameters  $\sigma$  and  $\mu$  have to be found.

Here we use the historical frequencies within the rating classes. We want to determine the parameters  $\sigma_r$  and  $\mu_r$  for each rating class such that  $\Phi(\sigma_r(LGDP_t - \mu_r))$  is the same as the empirical default frequency  $DF_{r,t}$  in rating class  $r$  at time  $t$ . Therefore solve for each rating class  $r$

$$\min \sum_{t=2009/03}^{2013/01} (DF_{rt} - \Phi(\sigma_r(LGDP_t - \mu_r)))^2.$$

In Figure 4.3 a histogram of the common factor  $Z_r = \Phi(\sigma_r(LGDP - \mu_r))$  is shown along with the fitted beta distribution for the rating classes 5 to 7. For rating class 1 this method can not yield any proper results as the empirical default frequency is zero. In the following we therefore assume a deterministic default probability of 0.03% for rating class 1.

Table 4.3 contains the parameter estimates  $a_r$  and  $b_r$  for the beta distribution along with the estimate of  $PD^{Output} = \mathbb{E}[Z_r] = \frac{a_r}{a_r + b_r}$ . With these parameters it is possible to simulate the mixing variable  $\pi_r(Z) = Z$  where  $Z$  is Beta-distributed and  $r$  is the rating class. Alternatively, if we have an estimate of the future GDP, we can also use the representation of  $Z_r = \Phi(\sigma_r(LGDP - \mu_r))$  along with the parameters  $\mu_r$  and  $\sigma_r$  to obtain a deterministic PD for each rating class. The parameters  $\mu_r$  and  $\sigma_r$  are also given in Table 4.3.

Table 4.3: The beta mixture model.  $PD$  is the adjusted PD s.t. the minimum PD is 0.03% and s.t. the PDs are nondecreasing with increasing rating class.

Rating class	1	2	3	4	5	6	7
Parameter a	-	291.24	0.38	29.29	3.02	3.78	54717
Parameter b	-	56056	84.82	1895	81.41	114.71	1274054
Parameter $\mu$	-	-4.14	-0.23	-1.01	-0.25	-0.30	-27.44
Parameter $\sigma$	-	-0.62	-11.45	-2.11	-6.72	-5.97	-0.06
$PD^{Output}$ (%)	-	0.52	0.45	1.52	3.58	3.19	4.12
$PD$ (%)	0.03	0.52	0.52	1.52	3.58	3.58	4.12

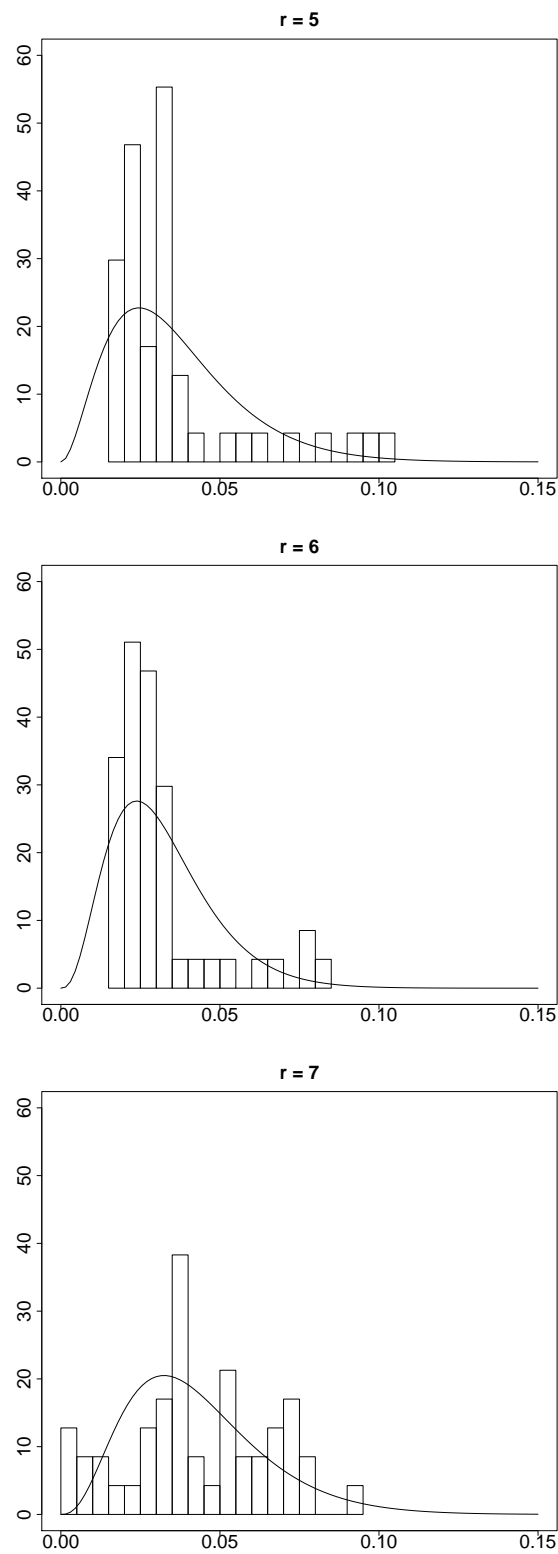


Figure 4.3: Histograms of the common factor along with the fitted Beta-distribution.

## 4.4 Estimates for Expected Loss and Value-at-Risk

Now that we have estimates for the PD in all models we can compute the Expected Loss and we can simulate the Value-at-Risk.

In order to compare the PDs resulting from the various models, see Table 4.4. In Table 4.5 the PDs that are used for the EL and VaR calculations are shown. These PDs respect the minimum PD of 0.03% according to the SolvV and that the PD should be nondecreasing with increasing rating class. Further, the latter table shows the

$$EL = \sum_{i=1}^{n_0} EAD_i \cdot LGD_i \cdot PD_i$$

estimate, where  $i = 1, \dots, n_0$  are the credits in the current time period.  $PD_i$  is given by the estimates of the PD values in Table 4.5 and depends on the rating class.

Table 4.4: Historical default frequencies and estimated unconditional default probabilities for the three models given in percentage.

Rating class	1	2	3	4	5	6	7
emp. DF	0.00	1.00	0.43	1.91	4.31	3.74	8.33
Calibration	0.42	0.69	1.15	1.91	3.17	5.26	8.72
Normal mixture	-	0.00	0.90	2.07	6.33	4.62	0.01
Beta mixture	-	0.52	0.45	1.52	3.58	3.19	4.12

Table 4.5: Estimated unconditional default probabilities for the three models given in percentage and EL in EUR.

Rating class	1	2	3	4	5	6	7	EL
Calibration	0.42	0.69	1.15	1.91	3.17	5.26	8.72	107086
Normal mixture	0.03	0.03	0.90	2.07	6.33	6.33	6.33	86878
Beta mixture	0.03	0.52	0.52	1.52	3.58	3.58	4.12	56334

For the VaR estimate we have to run several simulations. We calculate 5000 values for

$$L = \sum_{i=1}^{n_0} EAD_i \cdot LGD_i \cdot X_i,$$

where  $X_i$  is Bernoulli-distributed with probability given by the model we want to use and depending both on the random factor  $Z$  and on the rating

class of the credit  $i$ . Then we can evaluate

$$VaR_p = L_{\lfloor 5000p \rfloor + 1, 5000}$$

for  $p = 0.01$  and  $p = 0.05$ . The VaR estimates from the three models are shown in Table 4.6. For the code, see appendix D.

Table 4.6: VaR estimates.<sup>1</sup>

	$VaR_{0.01}$	$VaR_{0.05}$
Calibration	765211	643799
Normal mixture	740537	643799
Beta mixture	650416	446559

The VaR estimates are not the way we would expect. If we take a look at Tables 3.1 and 4.5, about 70% of the data is in the rating classes 4, 5 and 6, and for all these rating classes the estimated PD from the calibration curve is lower than the estimated unconditional PD from the Normal model. But the observation  $VaR_{0.01}^{Calibration} > VaR_{0.01}^{Normal}$  can be explained by the data set: most of the credits that are large in volume are in the upper rating classes 1 to 3 and here, the estimated PD from the calibration curve is higher than the estimated unconditional PD from the Normal model.

The Beta model yields reasonable VaR estimates given the estimated unconditional PDs.

## 4.5 Scenario Analysis

The Beta mixture model can easily be used for a scenario analysis on GDP growth rates. We want to use the representation  $Z_r = \Phi(\sigma_r(LGDP - \mu_r))$  which, given a fixed value for  $LGDP$ , then gives a deterministic default probability for rating class  $r$ .

Let therefore  $G$  be the expected one-year GDP growth rate, e.g.  $G = 0.03$  which corresponds to an expected GDP growth rate of 3%. First we have to calculate the  $LGDP$  value that is used. The relation is  $LGDP = \log(1 + G)$ . Together with the parameters  $\mu_r$  and  $\sigma_r$  for each rating class (see Table 4.3) the one-year VaR estimates can be calculated. For the scenarios that the GDP growth rate is  $-3\%$ ,  $0\%$  and  $+3\%$ , the VaR estimate is displayed in Table 4.7.

<sup>1</sup>There is no error in the values for  $VaR_{0.05}$ .

Table 4.7: VaR estimates from the Beta mixture model with different scenarios.

GDP growth	-3%	0%	+3%
$VaR_{0.01}$	666642	657109	648892
$VaR_{0.05}$	505454	465335	440913

We observe that the VaR estimates for the different scenarios are as we would expect them to be. If the GDP growth rate is negative the VaR estimates increase and if the GDP growth rate is positive, the VaR estimates decrease. Also the VaR estimate obtained by the Beta model from Table 4.6 suits this picture: The historical GDP growth rate is most often positive (see Figure 4.2) and therefore it is reasonable that the previous estimate is somewhere close to the scenario where the GDP growth rate is positive.

Considering Figure 4.4, it can be said that an increase of +3% of the GDP results for this data set in at least a 1%-decrease for  $VaR_{0.01}$ .

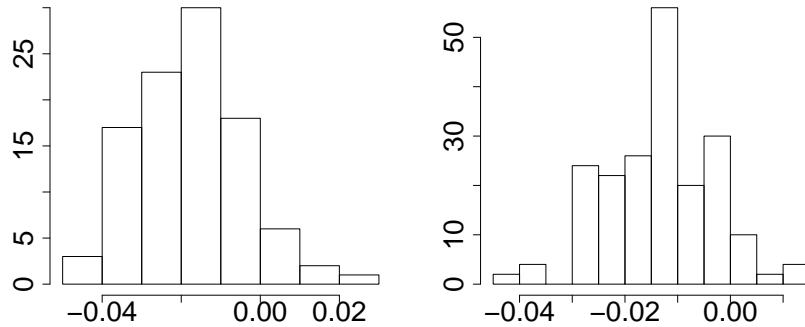


Figure 4.4: Histograms of 100 values for the quotients: *left:*  $VaR_{0.01}^{0\%}/VaR_{0.01}^{-3\%} - 1$ , *right:*  $VaR_{0.01}^{+3\%}/VaR_{0.01}^{0\%} - 1$ .



## Chapter 5

# Conclusion

We have seen that there are various ways to calculate default probabilities. The method of the calibration curve is a simple model that assumes independence among defaults. The more sophisticated the mixture models are closer to reality as they allow default dependence. However, this dependence could easily be modeled by the common factor.

All the models work on rating classes and they could theoretically be used on score values instead (this would correspond to more rating classes). However, in practice there will be limitations on the amount of data such that the models work better with rating classes.

The calibration curve is the most simple model of the presented ones. But it is also the model with the lowest amount of input data: Just the rating classes and the corresponding historical default frequencies were used.

The Bernoulli models both take the total amount of companies and the amount of defaulted ones at different times into account. The Beta mixture model even takes GDP growth rates as input Data. This makes the Beta mixture model somewhat superior to the Normal mixture model, if we accept the assumption that the GDP growth rate and the default rate are correlated<sup>1</sup>.

The problem is, however, that we now have three different estimates for the Value-at-Risk. On the one hand they are justified by the way the PDs emerged. On the other hand we do not know which PDs are the best estimates. This problem can only be solved by the permanent collection of data. This way we can update the estimates for the PDs on the basis of more data and we can also check the quality of today's PD and VaR estimates by backtesting.

We can conclude that we have to wait for the future to find out which model describes reality best.

---

<sup>1</sup>The historical default frequencies of the fictional data set used in this version is just slightly positive correlated with the GDP growth rate. But for these theoretical purposes we can be ok with the assumption.



## Appendix A

# Conditional PD according to Section 87 SolvV

The conditional PD is given by

$$\begin{aligned} PD^{cond} &= \Phi \left( (1 - \rho)^{-0.5} \cdot \Phi^{-1}(PD^{est}) + \left( \frac{\rho}{1 - \rho} \right)^{0.5} \cdot \Phi^{-1}(0.999) \right) \\ &= \Phi \left( \frac{\Phi^{-1}(PD^{est}) + \sqrt{\rho} \Phi^{-1}(0.999)}{\sqrt{1 - \rho}} \right) \end{aligned} \quad (A.1)$$

where  $PD^{est}$  is the PD that is estimated in this thesis and  $\rho$  is a value for the correlation with the economical factor (see Section 89 SolvV).  $\Phi(\cdot)$  is the cumulative Standard Normal distribution function and  $\Phi^{-1}$  is its inverse. In the following it will very briefly be described how this formula arises. For more (complete) information on the following the book Quantitative Risk Management by McNeil et al. [18] is recommended.

First, let  $Y \in \{0, 1\}^n$  be the default indicator (i.e.  $Y_i = 1$  if company  $i$  defaults).

Let the  $n$ -dimensional random vector  $X$  follow a factor model<sup>1</sup>, that is  $X = \sqrt{\rho}F + \sqrt{1 - \rho}\varepsilon$  where  $\rho$  is a (correlation) parameter,  $F \sim N(0, 1)$  is a random vector of common factors and  $\varepsilon$  is a standard normally distributed random vector with independent components and it is independent from  $F$ .

Let then  $(X, d)$  be a threshold model<sup>2</sup>. We only explain what is needed: That is, if  $X$  is any  $n$ -dimensional random vector of asset returns or asset values and if  $d$  is an  $n$ -dimensional vector of deterministic values that are the threshold values, then this simple threshold model indicates the default of company  $i$  if  $X_i < d_i$ . Hence we have

$$X_i < d_i \iff Y_i = 1.$$

---

<sup>1</sup>See Definition 3.33 in [18].

<sup>2</sup>See Definition 8.4 in [18].

Now the random vectors  $X$  and  $Y$  follow a Bernoulli mixture model (see Chapter 2)<sup>3</sup> if conditional on the one-dimensional common factor  $\Psi$ , their components are independent Bernoulli and satisfy

$$P(Y_i = 1 | \Psi = \psi) = P(X_i < d_i | \Psi = \psi) = p_i(\psi)$$

Set  $\Psi = -F$  and we obtain for the conditional probability

$$X |_{\Psi=\psi} = -\sqrt{\rho}\psi + \sqrt{1-\rho}\epsilon \sim \Phi_n(-\sqrt{\rho}\psi, \sqrt{1-\rho}^2 I).$$

Where  $I$  is the  $n \times n$  identity matrix. The covariance matrix happens to be a diagonal matrix as the components of  $\epsilon$  are independent and standard normally distributed.

Therefore for each company  $i$ ,

$$p_i(\psi) = P(X_i < d_i | \Psi = \psi) = \Phi\left(\frac{d_i + \sqrt{\rho}\psi}{\sqrt{1-\rho}}\right).$$

Now we still need a value for the threshold  $d_i$ . Here comes the estimated PD for company  $i$  into the equation. The threshold will be the quantile value of the estimated PD, so  $d_i = \Phi^{-1}(PD_i^{est})$ . Using the same technique for the common factor  $\psi$ , a large value for the common factor is used:  $\psi = \Phi^{-1}(0.999) \approx 3.09$ .

So the conditional PD as in (A.1) could be explained. In the SolvV, Section 90, the minimum and maximum value of the correlation is given,  $\rho \in [0.12, 0.24]$  for corporate credits. In the left plot of Figure A.1 the conditional PD as a function of the estimated PD is shown. In the right plot of the same figure, the difference of conditional minus estimated PD, that goes into the calculation of the risk weight (1.1), is displayed. Let us call that difference  $PD^D$ . We observe that the function  $PD^D(x) = PD^{cond}(x) - PD^{est}(x)$  on the right hand side attains its maximum value at 27%. Further we can see that the estimated default probabilities that are less than 45% are adjusted to the larger  $PD^D$  whereas estimated default probabilities that are bigger than 45% are diminished. This observation should be kept in mind whenever a rating system is implemented and its default probabilities are assessed.

---

<sup>3</sup>Alternatively see Definition 8.10 in [18].

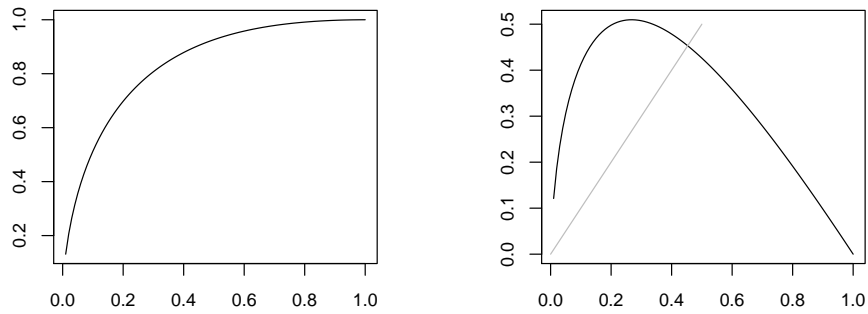


Figure A.1: *Left plot:* conditional PD in dependence of the estimated PD (x-axis). The correlation parameter was chosen to be 0.18. *Right plot:* Difference of conditional PD and estimated PD with identity function.



## Appendix B

# Beta Distribution

The Beta-distribution depends on two parameters and it takes values on the interval  $[0, 1]$ . It is a continuous distribution. Let  $Z \sim \text{Beta}(a, b)$ .

**Density.** The density of  $Z$  is given by<sup>1</sup>

$$f(z) = \frac{1}{\beta(a, b)} z^{a-1} (1-z)^{b-1} \quad \text{for } z \in (0, 1), a, b > 0,$$

where the gamma function appears in  $\beta(a, b)$ :

$$\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 z^{a-1} (1-z)^{b-1} dz, \quad \text{for } a, b > 0.$$

Recall that for the Gamma function it holds that  $\Gamma(a+1) = a \cdot \Gamma(a)$ . The density for different parameters  $(a, b)$  can be viewed in Figure B.1.

**Expected Value.**

$$\begin{aligned} \mathbb{E}[Z] &= \int_0^1 z \cdot \frac{1}{\beta(a, b)} z^{a-1} (1-z)^{b-1} dz = \frac{1}{\beta(a, b)} \int_0^1 z^a (1-z)^{b-1} dz \\ &= \frac{\beta(a+1, b)}{\beta(a, b)} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+1+b)} \\ &= \frac{\Gamma(a+b)}{\Gamma(a)} \cdot \frac{a\Gamma(a)}{(a+b)\Gamma(a+b)} = \frac{a}{a+b} \end{aligned} \tag{B.1}$$

**Variance**

$$\text{Var}(Z) = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2$$

---

<sup>1</sup>See also: Hult, Lindskog, Hammerlid, Rehn, p. 322.

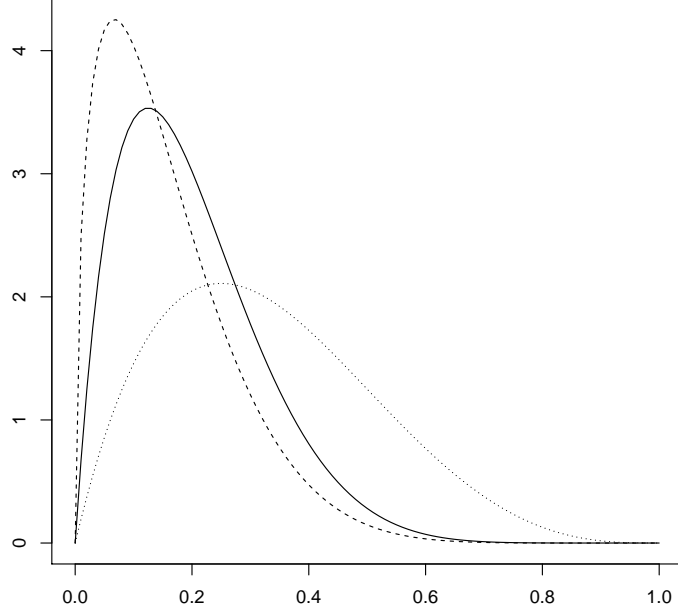


Figure B.1: The density of the  $Beta(a, b)$  distribution for parameters  $(2, 8)$  (solid line),  $(1.5, 8)$  (dashed line) and  $(2, 4)$  (dotted line).

With

$$\begin{aligned}
 \mathbb{E}[Z^2] &= \int_0^1 z^2 \cdot \frac{1}{\beta(a, b)} z^{a-1} (1-z)^{b-1} dz = \frac{1}{\beta(a, b)} \int_0^1 z^{a+1} (1-z)^{b-1} dz \\
 &= \frac{\beta(a+2, b)}{\beta(a, b)} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{(a+1)a\Gamma(a)\Gamma(b)}{(a+b+1)(a+b)\Gamma(a+b)} \\
 &= \frac{a(a+1)}{(a+b+1)(a+b)},
 \end{aligned}$$

this gives

$$\begin{aligned}
 \text{Var}(Z) &= \frac{a(a+1)}{(a+b+1)(a+b)} - \frac{a^2}{(a+b)^2} \\
 &= \frac{ab}{(a+b)^2(a+b+1)} = \mathbb{E}[Z] \frac{b}{(a+b)(a+b+1)}. \tag{B.2}
 \end{aligned}$$



**Calculating  $a$  and  $b$  from  $\mathbb{E}[Z]$  and  $\text{Var}(Z)$ .** From the equation for the expected value (B.1) we get

$$\mathbb{E}[Z] = \frac{a}{a+b} \iff a = \frac{b\mathbb{E}[Z]}{1-\mathbb{E}[Z]}.$$

Assume for now that  $b > 0$ . To fulfill  $a > 0$ , it is essential that  $0 < \mathbb{E}[Z] < 1$ . Then check that indeed  $b > 0$ ; equation (B.2) gives:

$$\begin{aligned} \text{Var}(Z) &= \mathbb{E}[Z] \frac{b}{(a+b)(a+b+1)} = \mathbb{E}[Z] \frac{b}{\left(\frac{b\mathbb{E}[Z]}{1-\mathbb{E}[Z]} + b\right)\left(\frac{b\mathbb{E}[Z]}{1-\mathbb{E}[Z]} + b + 1\right)} \\ &= \frac{b\mathbb{E}[Z]}{b(b+1-\mathbb{E}[Z])} = \frac{\mathbb{E}[Z]}{b-\mathbb{E}[Z]+1}, \end{aligned}$$

and solving for  $b$  yields

$$b = \frac{\mathbb{E}[Z]}{\text{Var}(Z)} + \mathbb{E}[Z] - 1.$$

$b$  is positive if

$$\mathbb{E}[Z] > \underbrace{\frac{\text{Var}(Z)}{1+\text{Var}(Z)}}_{>0}.$$

As we know, the sample mean  $\bar{Z} = \frac{1}{n}(Z_1 + \dots + Z_n) \rightarrow \mathbb{E}[Z]$  as  $n \rightarrow \infty$  for some sample  $\{Z_1, \dots, Z_n\}$ . So given the sample mean  $\bar{z}$  and the variance  $s^2$  of some empirical data, the parameters  $a > 0$  and  $b > 0$  for the beta distribution can be computed if

$$0 < \frac{s^2}{1+s^2} < \bar{z} < 1.$$



## Appendix C

# Historical German GDP Data

The Data on Germany's GDP can be found on the homepage of Eurostat ([epp.eurostat.ec.europa.eu/portal/page/portal/national\\_accounts/data/database](http://epp.eurostat.ec.europa.eu/portal/page/portal/national_accounts/data/database))<sup>1</sup>. It is displayed in Table C.1.

---

<sup>1</sup>Accessed on May 28th, 2013. Title: *Vierteljährliche Volkswirtschaftliche Gesamtrechnungen BIP und Hauptkomponenten - Jeweilige Preise (namq\_gdp\_c)*.

Table C.1: German GDP per capita in Euro. Observe that the values in February, May, August and and November were given whereas the GDP for the other months was calculated by linearization.

month	200803	200804	200805	200806	200807	200808	200809
GDP	7600	7600	7600	7567	7533	7500	7467
month	200810	200811	200812	200901	200902	200903	200904
GDP	7433	7400	7300	7200	7100	7133	7167
month	200905	200906	200907	200908	200909	200910	200911
GDP	7200	7233	7267	7300	7333	7367	7400
month	200912	201001	201002	201003	201004	201005	201006
GDP	7400	7400	7400	7467	7533	7600	7633
month	201007	201008	201009	201010	201011	201012	201101
GDP	7667	7700	7700	7700	7700	7767	7833
month	201102	201103	201104	201105	201106	201107	201108
GDP	7900	7900	7900	7900	7933	7967	8000
month	201109	201110	201111	201112	201201	201202	201203
GDP	7967	7933	7900	7933	7967	8000	8033
month	201204	201205	201206	201207	201208	201209	201210
GDP	8067	8100	8100	8100	8100	8100	8100
month	201211	201212	201301				
GDP	8100	8133	8167				

## Appendix D

# Simulation of Value-at-Risk in R

```
# Model: Calibration Curve
#####

PD=c(0.0041, 0.0069, 0.0115, 0.0191, 0.0317, 0.0526, 0.0872)
L=numeric(5000)
for(l in 1:length(L)){
  for(i in 1:n0){
    P=rbinom(1,1,PD[find_rating_class(Score[i])])
    L[l]=L[l]+EAD[i]*LGD[i]*P
  }
}
VaR01[v]=quantile(L,probs=c(0.99),na.rm=FALSE, names=TRUE)
VaR05[v]=quantile(L,probs=c(0.95),na.rm=FALSE, names=TRUE)
```

```
# Model: Normal Mixture
#####

mu=c(0,0, -2.37, -2.039, -1.53, -1.53, -1.53 )
sigma=-0.000997
L=numeric(5000)
PD_temp=numeric(7)
for(l in 1:length(L)){
  Z=rnorm(1)
  for(r in 3:7){
    PD_temp[r]=pnorm(mu[r]+sigma*Z)
  }
  PD_temp[1]=0.0003
  PD_temp[2]=0.0003
  for(i in 1:n0){
    P=rbinom(1,1,PD_temp[find_rating_class(Score[i])])
    L[l]=L[l]+EAD[i]*LGD[i]*P
  }
}
VaR01[v]=quantile(L,probs=c(0.99),na.rm=FALSE, names=TRUE)
VaR05[v]=quantile(L,probs=c(0.95),na.rm=FALSE, names=TRUE)
```

```
# Model: Beta Mixture
#####

a=c(0, 291, 291, 29.29, 3.02, 3.02, 54717 )
b=c(0, 56056, 56056, 1895.89, 81.41, 81.41, 1274054)
L=numeric(5000)
PD_temp=numeric(7)
for(l in 1:length(L)){
  for (r in 2:7){
    PD_temp[r]=rbeta(1,a[r],b[r])
  }
  PD_temp[1]=0.0003
  for (i in 1:n0){
    P=rbinom(1,1,PD_temp[find_rating_class(Score[i])])
    L[l]=L[l]+EAD[i]*LGD[i]*P
  }
}
VaR01=quantile(L, probs=c(0.99), na.rm=FALSE, names=TRUE)
VaR05=quantile(L, probs=c(0.95), na.rm=FALSE, names=TRUE)
```

# Bibliography

- [1] *Bank for International Settlements*, [www.bis.org](http://www.bis.org). Accessed June 13th, 2013.
- [2] Basel Committee on Banking Supervision. *International Convergence of Capital Measurement and Capital Standards*. July 1988.
- [3] Basel Committee on Banking Supervision. *International Convergence of Capital Measurement and Capital Standards: A Revised Framework - Comprehensive Version*. June 2006.
- [4] Braun, Allgeier, and Cremers. Ratingverfahren: Diskriminanzanalyse versus Logistische Regression. *Frankfurt School of Finance and Management, Working Paper No. 179*, December 2011.
- [5] Bundesanstalt für Finanzdienstleistungsaufsicht. *Mindestanforderungen an das Risikomanagement - MaRisk*. Rundschreiben 10/2012 (BA), December 14th, 2012.
- [6] Bundesanstalt für Finanzdienstleistungsaufsicht. *Liste der für die Bankaufsichtliche Risikogewichtung anerkannten Ratingagenturen samt Mapping*, [www.bafin.de/SharedDocs/Veroeffentlichungen/De/Auslegungsentscheidung/BA/ae\\_110509\\_ratingagenturen.html?nn=2696842](http://www.bafin.de/SharedDocs/Veroeffentlichungen/De/Auslegungsentscheidung/BA/ae_110509_ratingagenturen.html?nn=2696842), May 2011. Accessed April 8th, 2013.
- [7] Debeka Bausparkasse AG. *Ausfall*. November 2012.
- [8] Debeka Bausparkasse AG. *Ihr Bausparvertrag und was Sie darüber wissen sollten*. BS 9021, June 2010.
- [9] Debeka-Group. *Debeka eine Unternehmensgruppe mit Zukunft*. KV/LV/BS 37, June 2011.
- [10] Duffie and Singleton. *Credit Risk: Pricing, Measurement, and Management*. Princeton University Press, 2003.
- [11] Ernst&Young. *Fachkonzept PD-Schätzung nach Basel II/SolvV*, 2009.

- [12] Ernst&Young. *Fachkonzept LGD-Modell, Projekt Basel II, Version 1.6*, January 2007.
- [13] Ernst&Young. *Fachkonzept Validierung und Backtesting V2.1*, January 2012.
- [14] Ettmann, Wolff, and Wurm. *Kompaktwissen Bankbetriebslehre*. Bildungsverlag EINS, 2010.
- [15] Hartigan and Wong. A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 28, No. 1, pp. 100-108, 1979.
- [16] Hartmann-Wendels, Pfingsten, and Weber. *Bankbetriebslehre*. Springer Verlag, 2010.
- [17] Hult, Lindskog, Hammerlid, and Rehn. *Risk and Portfolio Analysis*. Springer, 2012.
- [18] McNeil, Frey, and Embrechts. *Quantitative Risk Management*. Princeton University Press, 2005.
- [19] *Solvabilitätsverordnung - SolvV*. [www.gesetze-im-internet.de](http://www.gesetze-im-internet.de), 2006. version of December 19th, 2012.
- [20] *Regulation governing the capital adequacy of institutions, groups of institutions and financial holding groups*. [http://www.bafin.de/SharedDocs/Aufsichtsrecht/EN/Verordnung/solvv\\_en\\_ba.html](http://www.bafin.de/SharedDocs/Aufsichtsrecht/EN/Verordnung/solvv_en_ba.html), 2006. Translation: Deutsche Bundesbank.