Estimation of Probability of Default in Low Default Portfolios

Nina Castor and Linnéa Gerhardsson

Department of Mathematical Statistics
Faculty of Engineering at Lund University

January 2017
Abstract

Estimation of probability of default (PD) is a fundamental part of credit risk modeling, and estimation of PD in low default portfolios is a common issue for banks and financial institutions. The Basel Committee on Banking Supervision requires banks and financial institutions to add an additional margin of conservatism to its PD estimates in the case of insufficient data, as in low default portfolios with few default observations. In addition, the Basel regulations also require banks to report PD estimates on grade level.

The purpose of this thesis is to study methods for estimation of probability of default in low default portfolios. In order to fulfill this purpose, two different models for estimation of probability of default in low default portfolios are considered. These are the Benjamin, Cathcart and Ryan (BCR) approach and a Bayesian approach. Because these models estimate PD on a portfolio level, different methods for allocation of portfolio PDs to rating grades are also considered. Lastly, methods to assign portfolio PDs to grade level for a portfolio consisting of several subportfolios are compared.

Keywords: Probability of default, PD, Low default portfolio, LDP, BCR, Bayesian, Vasicek, Monte Carlo, subportfolios, grade level estimates.
Acknowledgements

We would like to express our gratitude to our academic supervisor Magnus Wiktors-son for his help and valuable opinions. We would also like to thank Bujar Huskaj for the idea to this thesis and his guidance throughout the project.

Furthermore we would like to thank our families and friends for their support during our entire period of studies at LTH.

Lund, January 2017
Nina Castor
Linnéa Gerhardsson
## Abbreviations

A summary of all abbreviations used throughout the thesis.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-IRB</td>
<td>Advanced Internal Ratings Based</td>
</tr>
<tr>
<td>ARS</td>
<td>Adaptive Rejection Sampling</td>
</tr>
<tr>
<td>ARMS</td>
<td>Adaptive Rejection Metropolis Sampling</td>
</tr>
<tr>
<td>BCR</td>
<td>Benjamin, Cathcart &amp; Ryan</td>
</tr>
<tr>
<td>EAD</td>
<td>Exposure At Default</td>
</tr>
<tr>
<td>F-IRB</td>
<td>Foundation Internal Ratings Based</td>
</tr>
<tr>
<td>IRB</td>
<td>Internal Ratings Based</td>
</tr>
<tr>
<td>LDP</td>
<td>Low Default Portfolio</td>
</tr>
<tr>
<td>LGD</td>
<td>Loss Given Default</td>
</tr>
<tr>
<td>MC</td>
<td>Markov Chain</td>
</tr>
<tr>
<td>MCMC</td>
<td>Markov Chain Monte Carlo</td>
</tr>
<tr>
<td>PD</td>
<td>Probability of Default</td>
</tr>
<tr>
<td>PT</td>
<td>Pluto &amp; Tasche</td>
</tr>
</tbody>
</table>
# Contents

1 **Introduction** 1  
   1.1 Background 1  
   1.2 Problem formulation 2  
   1.3 Chapter outline 2  

2 **Credit Risk** 5  
   2.1 Credit Risk 5  
      2.1.1 Probability of Default (PD) 5  
      2.1.2 Loss Given Default (LGD) 6  
      2.1.3 Exposure At Default (EAD) 6  
      2.1.4 Expected Loss (EL) 6  
      2.1.5 Credit Risk Regulation 6  

3 **Theory and concepts** 9  
   3.1 Random number generation 9  
   3.2 The Vasicek model 10  
   3.3 Markov chains 13  
   3.4 Monte Carlo Simulation 13  
      3.4.1 Markov chain Monte Carlo methods 14  
      3.4.2 The Gibbs sampler 14  

4 **Confidence Based Approach** 17  
   4.1 Hypothesis testing 17  
   4.2 Independent case, no defaults 18  
   4.3 Independent case, few defaults 19  
   4.4 Dependent case 19  
   4.5 Multi-period case 20  
   4.6 Applications and Issues 21  

5 **The BCR Approach** 23  
   5.1 Theory 23  
      5.1.1 Independent case 23  
      5.1.2 Dependent case 24  
      5.1.3 Multi-period case 24  
   5.2 Estimation and simulation 25
Chapter 1

Introduction

1.1 Background

The probability of default (PD), the probability that a lender fails to meet his/her financial obligation, is a core input to credit risk modeling. Because of this, the accuracy of the PD estimate is directly linked to the quality of credit risk models.

One of the main obstacles connected to estimation of PD is when there is a low number of defaults, so-called low default portfolios (LDP). In these portfolios, the number of defaulted loans is much lower than the number of non-default loans. Examples of such portfolios are portfolios with an overall good quality of borrowers, like sovereigns or bank portfolios, and emerging market portfolios for up to medium size. In practice, a large part of a bank’s credit risk exposure is in LDPs, and therefore the problem of estimating PDs in LDPs is very important for a bank.

The problem with estimating PD in these low default portfolios arises from that the low number of defaults results in estimates that are small and very volatile over time. Therefore these estimates are unreliable in a statistical sense and may result in an inadequate assessment of the true risk.

The modeling of probability of defaults is addressed in the second of the Basel Accords, [1], which are recommendations on banking laws and regulations issued by the Basel Committee on Banking Supervision. They state that in general estimates of PDs involve some unpredictable errors and that banks should therefore add a margin of conservatism. They also indicate that if the data is less satisfactory, as in LDPs, a larger margin of conservatism should be added.

The topic of how to estimate PD in low default portfolios has been subject to much discussion, and one of the first attempts to address this issue in accordance with Basel’s recommendations was made by Pluto & Tasche [21]. This approach, referred to as the Confidence Based Approach, results in upper confidence bounds for the estimated PD by the most prudent estimation approach. Later, Benjamin, Cathcart and Ryan, [1], proposed revisions of this approach that handles the problem with inherent conservatism that was a main criticism of the original approach. This revised confidence based approach, referred to as the BCR method, is a quantitative method that produces upper confidence bounds for the estimates of PD on portfolio level,
Chapter 1. Introduction

and is today widely used in the industry. However, both methods have received some criticism, with one of the main points being the choice of input parameters, asset- and year-to-year-correlation which will be defined and explained later, as well as the choice of confidence level.

A method that avoids the problem of choosing confidence level is the Bayesian Approach first proposed by Dwyer [7] with a uniform prior distribution, and later revisited by Dirk Tasche, [23], with the suggestion of a conservative prior distribution. This approach has been discussed by numerous authors and can be considered as a valid method in addressing the above problem regarding LDPs. Further advantages with the Bayesian method discussed by Tasche is that the level of conservatism is reasonable and that the model is sensitive to correlation.

A problem with low default portfolios in general is how they are defined, since this is subjective and may vary from one bank to another. Another issue is how to spread portfolio wide estimates over grades and calculate grade level estimates for subportfolios in an aggregated portfolio.

1.2 Problem formulation

The aim with the thesis is to study methods for estimating PD in low default portfolios as well as to give a theoretical background on low default portfolios and a review of the current regulation on modeling of LDPs. The analysis of the models for estimation of PD in LDPs will focus on two problems that are of importance for banks in general.


- Compare methods for spreading portfolio-wide estimates of PD to grade-level estimates, and suggest a method for how to do this in an aggregated portfolio that consists of several subportfolios, which in turn consists of different grades.

1.3 Chapter outline

The following is the chapter outline for the thesis.

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

Chapter 3: In the third chapter, the theory used in later chapters is explained and defined. This includes theory of random number generation, the Vasicek model as well as on Markov chains and Markov chain Monte Carlo sampling.

Chapter 4: In Chapter 4, the Confidence Based approach proposed by Pluto & Tasche is presented in order to fully understand the later discussed BCR approach.

Chapter 5: This chapter describes the BCR approach for estimating PD. It has largely the same structure as the previous chapter. In addition, the estimation procedure is also described.
Chapter 6: The Bayesian approach is presented by first reviewing Bayesian statistics in a general setting. Then, the theory covering the Bayesian approach from previous articles is given. Lastly, the theory covering an alternative Bayesian approach with sampling of an additional factor is presented together with the estimation procedure.

Chapter 7: Four different methods to calculate the grade level PDs for a low default portfolio are presented. Further on three methods on how to calculate the grade level PDs for portfolios consisting of several subportfolios are suggested.

Chapter 8: The data used is presented, along with results for both the BCR and Bayesian alternative approach for different choices of model parameters. A comparison of the performance of the BCR and Bayesian model is performed. Furthermore, results for the grade level allocation methods are presented and analyzed.

Chapter 9: The results from the previous chapter are discussed as well as the methodologies and underlying assumptions in all the models in the thesis. Also, the model differences and their implications are deliberated, and suggestions for further research are given.

Chapter 10: A summary of the results in this thesis is given together with concluding remarks.
Chapter 2

Credit Risk

In this chapter, credit risk and the components to measure credit risk are defined. Further, the regulations for capital requirements due to credit risk are presented. These regulations are collected in a global regulation framework, the newest version being Basel III. The regulations consist of capital requirements for banks and financial institutions, and have grown increasingly strict in light of recent financial crises.

2.1 Credit Risk

To attract funds and to resell or invest is one of the core businesses of banking, but with investing comes risk. Risk is the probability that a negative event occurs, or in financial terms that there is a loss. For banks the risk is foremost connected to financial risk due to potential losses of financial products. Three main financial risks are represented in the Basel Capital Accord, which are recommendations on banking laws and regulations issued by the Basel Committee on Banking supervision. These are Credit risk, Market risk and Operational risk. Credit risk is the main focus of this thesis, and is defined as the risk of losses due to that a borrower defaults and does not meet its financial obligations on time. Credit risk is often represented by means of four factors: expected loss, default risk, loss risk and exposure risk. These are explained in the following sections.

2.1.1 Probability of Default (PD)

The default risk is defined as the probability that a default event occurs. This is called the probability of default. Since it is a probability it takes values in the interval \([0, 1]\) where the value zero represents the fact that there is no risk of a default and one represents that there will be a default. There are many definitions of a default event where the most common definition is a payment delay of at least three months. The following definition of a default event is made in Section 452 of [1].

**Definition 2.1.1. The Basel definition of default**

A default is considered to have occurred with regard to a specific obligor when either or both of the two following events have taken place

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

2.1.2 Loss Given Default (LGD)

The loss as a fraction of the exposure in case of default is determined by the loss risk. This is known as the loss given default (LGD) in Basel II. In other words when there is no loss the LGD is equal to 0 % and when the full exposure amount is lost the LGD is equal to 100 %. The LGD can also be larger than 100 % because of litigation costs and when there is almost zero recovery from the defaulted counterpart.

2.1.3 Exposure At Default (EAD)

The exposure at the time of default (EAD) may be known or not known beforehand since for some products like a bond or a straight loan this is a fixed amount but for some products like a credit card the amount varies. For a credit card the counterpart has a certain credit limit and can take credit up to this amount. This uncertainty of the exposure at a future default is the exposure risk [26].

2.1.4 Expected Loss (EL)

The Expected loss, the amount a bank can expect to loose due to credit events, is defined as

\[ EL = EAD \times LGD \times PD. \]

The expected loss connects the previous three measurements and is the basis for the capital requirement, the amount the bank is required to hold for this purpose.

2.1.5 Credit Risk Regulation

The current risk regulation framework, Basel III, was published by the Basel Committee of Banking Supervision in its original version in December 2010, and was later revised to its current version, published in June 2011. The document, as its predecessors Basel I and Basel II, consists of reforms to strengthen the global capital and liquidity rules in order to promote a more resilient banking sector. The objective of the reforms is to improve the banking sector’s ability to absorb shocks arising from financial and economic stress, and to thereby reduce the spillover of effects from the financial sector to the real economy [2].

As discussed in the previous section, the area of credit risk largely consists of models for estimation of the credit risk parameters PD, EAD and LGD. The Basel framework defines three possible ways to assess credit risk exposure: The Standardized Approach, the Foundation Internal Ratings Based Approach (F-IRB) and the Advanced Internal Ratings Based Approach (A-IRB), where F-IRB and A-IRB are sub-divisions of the Internal Ratings Based Approach (IRB) [12]. The Standardized approach is used by banks that are not allowed to use the IRB approaches, and is a simpler approach where minimum capital requirements are predominantly determined by dependence on asset classes (sovereign, bank, corporate or retail) [28]. The IRB approach allows banks to implement their own estimates of credit risk exposure. Under the F-IRB
2.1. Credit Risk

The methods for estimating PD in this thesis are those that are in accordance with the IRB approach, which is also the most widely used approach in the industry. Under the regulations of the IRB approach, the PD must be estimated regardless of which of the sub-divisions of the approach that is implemented. However, banks are allowed to implement their own rating systems for the estimation of PD as long as the system meets specified minimum requirements [9]. Important to note is that the IRB approach requires firms to make grade level estimates of PD.

In Basel’s document specifically handling the IRB approach, it is stated that each estimate of PD must represent a conservative view of a long-run average PD for the grade in question, and thus must be grounded on historical experience and empirical evidence. A conservative view can be described as a cautious or a non-optimistic view. Moreover, in Section 444 to 485 in Basel II, it is stated that PD estimates must be long-run averages of one-year default rates for each grade, and that internal estimates of PD must incorporate all relevant material and data as well as be grounded on historical and empirical evidence and not based purely on subjective or judgmental considerations. More specifically, in Section 461 to 463 of the same documents, it is stated that banks may use one or more of internal default experience, mapping to external data, and statistical default models when estimating the average PD for each rating grade. In the same sections it is also stated that regardless of what data sources a bank is using, that is internal, external, pooled data sources or a combination of some of them, for its calculation of the probability of default, the length of the underlying historical observation period used must be at least five years for at least one source. If the available observation period is over a longer period for any source, and this data are relevant and substantial, the longer period has to be used. Further it is declared that estimates of PDs in general are likely to involve unpredictable errors and in order to avoid over-optimism a bank must add to its estimates a margin of conservatism that is related to the likely range of error. Specifically for LDPs, where only limited data is available, it’s explained that the bank must add a greater margin of conservatism in its estimate of PD [1].

Lastly, in [8], it is stated that the PD of an exposure to a corporate or an institution shall be at least 0.03% and in [2] it is declared that the asset correlation between obligors should be in the interval between 12% and 24%.
Chapter 3

Theory and concepts

In order to understand the methods explained in later chapters, this chapter contains general theory and concepts. Among others, random number generation and the inversion sampling theorem is explained. Further, the Vasicek model for dependence between defaults used in all subsequent methods is introduced. Also, a definition of Markov chains and explanation of Markov chain Monte Carlo algorithms and the Gibbs sampler is presented. More comprehensive theory of Markov chains can be found in Appendix A.

3.1 Random number generation

Let $X$ be a random variable on $\mathbb{R}$, $X \subseteq \mathbb{R}$ with density function $f_X$ and distribution function $F_X$, where $F_X$ is invertible and the inverse distribution function is denoted $F_X^{-1}$.

The distribution function for $X$ conditional on that $X \in I$ where $I = [a, b]$ can be derived in the following manner, assuming that $P(X \in I) > 0$.

$$F_{X|X \in I}(x|X \in I) = P(X < x|X \in I) = \frac{P(X < x \cap X \in I)}{P(X \in I)} = \frac{P(a < X < x)}{P(a < X < b)} = \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)} \quad \text{for} \quad a < x < b.$$  

Similarly, when $x$ only has an upper bound $b$ the conditional distribution function is

$$F_{X|X \leq b} = P(X \leq x|X \leq b) = \frac{P(X \leq x \cap X \leq b)}{P(X \leq b)} = \frac{P(X \leq x)}{P(X \leq b)} = \frac{F_X(x)}{F_X(b)} \quad \text{for} \quad x \leq b.$$  

In contrast, when there only exists a lower bound $a$ for $x$

$$F_{X|X > a} = P(X \leq x|X > a) = \frac{P(X \leq x \cap X > a)}{P(X > a)} = \frac{P(X \leq x) - P(X \leq a)}{1 - P(X \leq a)} = \frac{F_X(x) - F_X(a)}{1 - F_X(a)} \quad \text{for} \quad x > a.$$
Chapter 3. Theory and concepts

Theorem 3.1.1. Inverse transform sampling
Let $X$ be a random variable with distribution function $F$, where $F$ is a strictly monotone and invertible function.

1. Draw $U \sim U(0,1)$
2. Set $X \leftarrow F^{-1}(U)$
Then $X$ has distribution $F$.

Proof.

$$P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x) \quad \text{since} \quad P(U \leq y) = y$$

Letting $X = F^{-1}(U)$ is obviously equivalent to $F(X) = U$, therefore by the inversion theorem it is possible to generate a random number from a sought distribution $F$ by letting $F(X) = U$ and solving for $X$. If $F(X)$ is the standard normal distribution function and $a < X < b$, the variable $X$ can therefore be sampled according to

$$\Phi(X) - \Phi(a) \over \Phi(b) - \Phi(a) = U \Leftrightarrow X = \Phi^{-1}(U \cdot (\Phi(b) - \Phi(a)) + \Phi(a)). \tag{3.1}$$

Likewise, when $x$ only has an upper bound, $x < b$, it can be sampled from

$$\Phi(X) \over \Phi(b) = U \Leftrightarrow X = \Phi^{-1}(U \cdot \Phi(b)) \quad \tag{3.2}$$

and when there only exists a lower bound as

$$\Phi(X) - \Phi(a) \over 1 - \Phi(a) = U \Leftrightarrow X = \Phi^{-1}(U \cdot (1 - \Phi(a)) + F(a)). \tag{3.3}$$

3.2 The Vasicek model

The Vasicek model, formulated by the mathematician Vasicek in 1997 [27], is a credit risk model with the same mechanism of default as in the Merton model, that a default occurs when the decrease in value is sufficiently large. The difference from the Merton model is that the change in value is driven by an obligor specific factor as well as a systematic factor, as opposed to the Merton model where the driving factor is only the obligor specific factor. For a description of the Merton model see [17].

Definition 3.2.1. The one period Vasicek Model

A loan defaults if the value of the borrower’s assets at the loan maturity time $T$ falls below the contractual value $B_i$ of its obligations payable. Let $A_i$ be the value of obligor $i$’s assets, described by the process

$$dA_i = \mu_i A_i dt + \sigma_i A_i dX_i$$

or equivalently

$$\ln A_i(T) = \ln A_i + \mu_i T - {1 \over 2} \sigma_i^2 T + \sigma_i \sqrt{T} X_i$$

where $\mu_i$ and $\sigma_i$ are the drift- and volatility rates in the underlying geometrical Brownian motion in the Black-Scholes model. The $X_i$s are jointly standard normally distributed with pairwise correlations $\rho$ and therefore can be represented as

$$X_i = \sqrt{\rho} Y + Z_i \sqrt{1 - \rho}$$

in which $Y$ and $Z_1, Z_2, \ldots, Z_n$ are mutually independent standard normal variables, which follows from the properties of the equally correlated normal distribution.
The change in asset value is driven by both a systematic and an obligor specific factor. The variable $Y$ describing the systematic factor can be interpreted as a portfolio common factor representing the state of the economy. The term $Z_i$ represents the company’s exposure to idiosyncratic risk, that is risk specific to the obligor in question.

According to the definition in the Merton model, a default occurs if $A_i(T) < B_i$, therefore the probability of default for obligor $i$, $p_i$ is given by

$$p_i = P[A_i(T) < B_i] = P[\ln A_i(T) < \ln B_i] =$$

$$P[\ln A_i + \mu_i T - \frac{1}{2} \sigma_i^2 T + \sigma_i \sqrt{T} X_i < \ln B_i] =$$

$$P[X_i < \frac{\ln B_i - \ln A_i - \mu_i T + \frac{1}{2} \sigma_i^2 T}{\sigma_i \sqrt{T}}] = P[X_i < c_i] = \Phi(c_i)$$

where $c_i = \frac{\ln B_i - \ln A_i - \mu_i T + \frac{1}{2} \sigma_i^2 T}{\sigma_i \sqrt{T}}$ and $\Phi(x)$ is the standard normal distribution function. The relation $p_i = \Phi[c_i]$ implies that $c_i = \Phi^{-1}[p_i]$.

Applying $X_i = \sqrt{\rho} Y + Z_i \sqrt{1 - \rho}$ from the model and $c_i = \Phi^{-1}[p_i]$ the conditional probability of default in the Vasicek model given a realization of the systematic factor $Y$ is

$$\lambda(p_i, \rho, y) = P[X_i < c_i | Y = y] = P[\sqrt{\rho} y + Z_i \sqrt{1 - \rho} < \Phi^{-1}(p_i)] =$$

$$P[Z_i < \frac{\Phi^{-1}(p_i) - \sqrt{\rho} y}{\sqrt{1 - \rho}}] = \Phi[\frac{\Phi^{-1}(p_i) - \sqrt{\rho} y}{\sqrt{1 - \rho}}].$$

(3.4)

This is referred to as the Vasicek model or the Single Factor Gaussian Copula Model. In the multi-period case, simulation of the systematic and obligor specific factor is required for each time period, hence the change in asset value is dependent on the time $t$, $X_{i,t}$.

**Definition 3.2.2. The multi-period Vasicek model**

As above,

$$dA_{i,t} = \mu_i A_{i,t} dt + \sigma_i A_{i,t} dX_i$$

$$\ln A_i(T) = \ln A_i + \mu_i T - \frac{1}{2} \sigma_i^2 T + \sigma_i \sqrt{T} X_i$$

where

$$X_{i,t} = \sqrt{\rho} Y_t + Z_{i,t} \sqrt{1 - \rho}$$

in which the variables $Y_t$ and $Z_{i,t}$ are standard normally distributed and independent of each other. Furthermore, the joint distribution of the factors $Y_1, \cdots, Y_t$ is determined by the correlation matrix

$$\Sigma = \begin{pmatrix}
1 & r_{1,T} & \cdots & r_{1,T} \\
r_{2,1} & 1 & \cdots & r_{2,T} \\
\vdots & \vdots & \ddots & \vdots \\
r_{T,1} & r_{T,2} & \cdots & 1
\end{pmatrix}$$

where $r_{s,t} = \tau^{|s-t|}$ is the correlation between year $s$ and year $t$, $Y_t = \tau Y_{t-1} + \sqrt{1 - \tau^2} W_t$ and $\tau$ is the year-to-year, or inter-temporal, correlation.
As in the one period case, a default occurs for obligor \( i \) in year \( t \) if \( X_{i,t} \leq c_{i,t} = \Phi^{-1}[p_{i,t}] \), but in the multi-period also \( X_{i,1} > c_{i,1}, X_{i,2} > c_{i,2}, \ldots, X_{i,t-1} > c_{i,t-1} \) since the obligor can not have defaulted previously. The correlation structure is sometimes referred to as an AR(1)-correlation structure, as the dependence between the systematic factors in different years follow that of an AR(1)-process.

Since \( Y_t \) and \( Z_{i,t} \) are mutually independent standard normally distributed \( X_{i,t} \) is also normally distributed. The mean and variance of \( X_{i,t} \) are

\[
E[X_{i,t}] = E[\sqrt{p}Y_t + Z_{i,t}\sqrt{1-\rho}] = \sqrt{p}E[Y_t] + \sqrt{1-\rho}E[Z_{i,t}] = 0
\]

\[
V[X_{i,t}] = V[\sqrt{p}Y_t + Z_{i,t}\sqrt{1-\rho}] = (\sqrt{p})^2 V[Y_t] + (\sqrt{1-\rho})^2 V[Z_{i,t}] + 2\sqrt{p}\sqrt{1-\rho} Cov[Y_t, Z_{i,t}] = 1
\]

why \( X_{i,t} \) is clearly standard normally distributed. Therefore also \( P[X_{i,t} \leq c_{i,t}] = \Phi[c_{i,t}] \) and \( c_{i,t} = \Phi^{-1}[p_{i,t}] \). Because of the assumption that obligors in the same rating category have the same probability of default, \( p_{i,t} = p \) and \( c_{i,t} = c \) for all \( i \) and \( t \). Given this, the probability of default for obligor \( i \) given a realization of the systematic factors \( Y_1, \ldots, Y_T \), where \( Y = Y_1, \ldots, Y_T \) is

\[
\lambda(p, \rho, Y) = P[\text{Obligor } i \text{ defaults}|Y] = P\left[ \min_{t=1,\ldots,T} X_{i,t} \leq \Phi^{-1}(p)|Y \right] = 1 - P\left[ \min_{t=1,\ldots,T} X_{i,t} > \Phi^{-1}(p)|Y \right] = 1 - P[X_{i,1} > \Phi^{-1}(p), \ldots, X_{i,T} > \Phi^{-1}(p)|Y] = 1 - P[\sqrt{p}Y_1 + Z_{i,1}\sqrt{1-\rho} > \Phi^{-1}(p), \ldots, \sqrt{p}Y_T + Z_{i,T}\sqrt{1-\rho} > \Phi^{-1}(p)|Y] = 1 - P[Z_{i,1} > \frac{\Phi^{-1}(p) - \sqrt{p}Y_1}{\sqrt{1-\rho}}, \ldots, Z_{i,T} > \frac{\Phi^{-1}(p) - \sqrt{p}Y_T}{\sqrt{1-\rho}}]
\]

since \( Y \) is independent of the obligor idiosyncratic factors \( Z_{i,T} \). Because the idiosyncratic factors are independent of each other and standard normally distributed, the conditional probability becomes

\[
\lambda(p, \rho, Y) = 1 - \prod_{t=1}^{T} 1 - \Phi\left( \frac{\Phi^{-1}(p) - \sqrt{p}Y_t}{\sqrt{1-\rho}} \right). \quad (3.5)
\]

From the dependence structure of the systematic factors given by \( Y_t = \tau Y_{t-1} + \sqrt{1-\tau^2} W_t \) the conditional distribution of \( Y_t \) given the previous value \( Y_{t-1} \) and the year-to-year correlation \( \tau \) is normal distributed since it is a linear combination of two normally distributed variables. Consequently,

\[
f(y_t|y_{t-1}, \tau) \propto e^{-\frac{(y_t-y_{t-1})^2}{2(1-\tau^2)}} \quad (3.6)
\]

due to that

\[
E[y_t|y_{t-1}, \tau] = \tau y_{t-1} + \sqrt{1-\tau^2} \cdot E[W_t] = \tau y_{t-1}
\]

\[
V[y_t|y_{t-1}, \tau] = 0 + 1 - \tau^2 \cdot V[W_t] = 1 - \tau^2.
\]
3.3 Markov chains

**Definition 3.3.1. Markov chain**

A Markov chain on \( X \subset \mathbb{R}^d \) is a family of random variables \( \{X_k\}_{k \geq 0} \) taking values in \( X \) such that

\[
P(X_{k+1} \in A | X_0, \ldots, X_k) = P(X_{k+1} \in A | X_k).
\]

Subsequently, for a Markov chain the next state observed is only dependent on the present state. This is called the Markov property. See [20] and Appendix A for a more thorough theory of Markov processes.

For Markov chains, due to the Markov property,

\[
f_{X_1, \ldots, X_N}(x_1, \ldots, x_N) = f_{X_1}(x_1) \cdot f_{X_2|X_1}(x_2|x_1) \cdot f_{X_3|X_2}(x_3|x_2) \cdots f_{X_N|X_{N-1}}(x_N|x_{N-1}).
\]

Following from Bayes rule,

\[
f_{X_j|X_{-j}}(x_j|x_{-j}) = \frac{f_{X_1, \ldots, X_N}(x_1, \ldots, x_N)}{f_{x_{-j}}(x_{-j})} = \frac{f_{X_1}(x_1) \prod_{k=1}^N f_{X_k|X_{k-1}}(x_k|x_{k-1})}{\int f_{X_1}(x_1) \prod_{k=1}^N f_{X_k|X_{k-1}}(x_k|x_{k-1}) dx_j} \propto
\]

\[
\begin{cases} f_{X_1}(x_1)f_{X_2|X_1}(x_2|x_1) & \text{if } j = 1 \\ f_{X_{j+1}}(x_{j+1}|x_j)f_{X_j|X_{j-1}}(x_j|x_{j-1}) & \text{if } j \neq 1 \text{ and } j \neq N \\ f_{X_N|X_{N-1}}(x_N|x_{N-1}) & \text{if } j = N \end{cases}
\]

where \( x_{-j} = (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_N) \). This is due to that all factors except for those containing \( x_j \) will be present in both the nominator and denominator and can be moved outside the integral. Remaining in the nominator are the factors depending on \( x_j \), and in the denominator a factor which integrates to a constant, thereby the proportionality relationship is established.

### 3.4 Monte Carlo Simulation

The principal aim of Monte Carlo simulation, from Chapter 6.1 in [10], is to estimate some expectation

\[
\mu = \mathbb{E} [\phi(X)] = \int_A \phi(x) f(x) dx
\]

where

- \( X \) is a random variable taking values in \( A \subseteq \mathbb{R}^d \)
- \( f : A \rightarrow \mathbb{R}_+ \) is the probability density of \( X \), also called the target density
- \( \phi : A \rightarrow \mathbb{R} \) is a function such that the above expectation is finite.

Letting \( \mu_{MC} = \frac{1}{N} \sum_{i=0}^N \phi(X_i) \), when \( X_1, \ldots, X_N \) is a independent and identically distributed sample from \( f \), by the law of large numbers

\[
\mu_{MC} \rightarrow \mu \quad \text{as} \quad N \rightarrow \infty
\]
This means that for a large number of simulations \( N \), by the law of large numbers, the sought expected value or integral can be approximated by a mean of all the outcomes, a sample average. Thereby Monte Carlo methods enable a way of solving complex problems where an analytic solution is not available, which is useful in many different areas.

3.4.1 Markov chain Monte Carlo methods

As described in Chapter 7 of [10], the basic idea of Markov Chain Monte Carlo (MCMC) methods is that to be able to sample from a density \( f \), a Markov chain that has \( f \) as stationary distribution is constructed. The MCMC strategy is to construct an irreducible, aperiodic Markov chain for which the stationary distribution equals the target distribution \( f \). For sufficiently large \( t \), a realization \( X^{(t)} \) from this chain will have approximate marginal distribution \( f \). Markov chain Monte Carlo methods are the most common methods to sample from high dimension or complicated distributions \( f \), and a popular application of MCMC is for Bayesian inference where \( f \) is a Bayesian posterior distribution. A drawback of the method is however that samples will not be independent.

Similarly to the regular MC methods, the principle of the Markov chain Monte Carlo methods is that by the law of large numbers for Markov-chains, see Appendix A, it is possible to estimate some expectation

\[
\mu = \mathbb{E}[\phi(X)] = \int_A \phi(x) f(x) \, dx
\]

by simulation in \( N \) steps a Markov chain \( X_k \) with stationary distribution \( f \) and letting

\[
\mu_{MC} = \frac{1}{N} \sum_{k=1}^{N} \phi(X_k) \to \mu \quad \text{as} \quad N \to \infty.
\]

One type of algorithm that does this is the Gibbs sampler.

3.4.2 The Gibbs sampler

The Gibbs sampler is a MCMC algorithm specifically adapted for multi-dimensional target distributions. The Gibbs sampler constructs a Markov chain with stationary distribution equal to the target distribution \( f \) by sequentially sampling from univariate conditional distributions, which are often known, see Section 7.2 of [10]. Denote \( X = (X_1, X_2, \ldots, X_p) \) and \( X_{-i} = (X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_p) \), that is every component of the vector except for \( X_i \). Suppose that it is possible to sample from the univariate conditional density \( f(x_i|x_{-i}) \), then a general Gibbs sampler can be constructed with the following steps:

1. Set a starting value \( x^{(0)} \).
2. Generate, in turn
   \[
   x_1^{(t+1)} \sim f(x_1|x_2^{(t)}, \ldots, x_p^{(t)})
   \]
   \[
   x_2^{(t+1)} \sim f(x_2|x_1^{(t+1)}, x_3^{(t)}, \ldots, x_p^{(t)})
   \]
   \[
   \vdots
   \]
   \[
   x_p^{(t+1)} \sim f(x_p|x_1^{(t+1)}, \ldots, x_{p-1}^{(t+1)})
   \]
3. Increment \( t \) and repeat for a large number of simulations.

Thereby, the Gibbs sampler generates a Markov chain which has the sought distribution \( f \). Samples from the beginning of the chain may not give an accurate representation of the true distribution. For this reason, a burn-in period is usually set for the algorithm, and these samples are disregarded. The following algorithm describes the Gibbs sampler in pseudo-code.

**Algorithm 1** Gibbs sampler

**Require:** A sufficiently large number of simulations \( N \) and number of burn-in samples.

set \( \mathbf{x}^{(0)} = (X_1, \ldots, X_p) \)

for \( i = 1 \) to \( N + \) burn-in do

for \( j = 1 \) to \( p \) do

\( X_j^{(i)} \sim f(x_j|x_{-j}) \)

end for

end for
Chapter 4
Confidence Based Approach

In May 2005 Katja Pluto and Dirk Tasche proposed a method for calculating probability of default in the setting of none or very few observations of defaults, [21]. They suggested the most prudent estimation principle and used a one-sided upper confidence bound as an estimator of PD. This method forms the basis for the later presented BCR approach, and in this chapter the theoretical background and methodology of the Confidence Based Approach is presented.

4.1 Hypothesis testing

The theory in Pluto & Tasche [21] uses a hypothesis method described in another article by Tasche [23] that culminates in a one-sided upper confidence bound. The method begins by making an underlying assumption.

Assumption 4.1.1. At the beginning of the observation period (in practice often one year) there are $n > 0$ borrowers in the portfolio. Defaults of borrowers occur independently, and all have the same probability of default (PD) $0 < \lambda < 1$. At the end of the observation period $0 \leq d < n$ defaults are observed among the $n$ borrowers [23].

Henceforward the probability of default is labeled as $p$ and the number of defaults as $d$. The idea is to statistically test an estimate of the probability of default, $p_0$, with the null-hypothesis $H_0: p \geq p_0$ using historical data. The reason for choosing the null-hypothesis as $p \geq p_0$ and not $p \leq p_0$ is that if $p \geq p_0$ can be rejected the alternative hypothesis $H_1$ is true and an upper bound for the probability of default can be determined.

To test the hypothesis the p-value is calculated as $P[D \leq d]$ since defaults fewer than the observed number of defaults $d$ are the more extreme cases when $H_0$ is $p \geq p_0$. If $P[D \leq d] \leq \alpha$ for some predefined type I error of size $0 < \alpha < 1$ the outcome is an unlikely event under $H_0$ so the null-hypothesis should be rejected in favor of the alternative hypothesis, $H_1 : p < p_0$. The arbitrarily chosen upper PD bound, $p_0$, can be refined by instead finding the set of all values of $p_0$ so that $P[D \leq d] > \alpha$ has to be found because then $H_0$ is accepted for all the $p_0$ below $p_0^*$. 
Chapter 4. Confidence Based Approach

The first approach gives reasonable upper bounds when $H_0$ is rejected. The set when $H_0$ is rejected is an interval $[p_0^*, 1]$, and when $H_0$ is accepted $[0, p_0^*]$. The second interval is a one-sided confidence interval at confidence level $\gamma = 1 - \alpha$ for the probability of default $p$ and characterized by the upper confidence bound $p_0^*$. This can be shown by using the general duality theorem for statistical tests [5]. This statistical test is used in the Pluto & Tasche method where the upper confidence bound, $p_0^*$ is calculated.

4.2 Independent case, no defaults

In the elementary case independence between default events is assumed and there are no defaults in any of the three rating grades $A$, $B$ or $C$. The numbers $n_A$, $n_B$ and $n_C$ are the number of obligors in each grade. The following inequality holds

$$p_A \leq p_B \leq p_C$$

where $p_A$, $p_B$ and $p_C$ are the different probabilities of default for each rating grade. Since it is assumed that the PD increases with the grade level (4.1) is considered to be true. The most prudent estimate is considered where $p_A = p_C$ which, because of (4.1), leads to

$$p_A = p_B = p_C.$$ (4.2)

Since there are only two outcomes for the obligor, default or no default, the default event can be seen as a Bernoulli distributed variable with value one if the obligor defaults and value zero if it does not. Because of the assumption of independence between default events in the initial case, it follows that the they are binomially distributed with the probability density function

$$p_X(d) = P(X = d) = \binom{n}{d} p^d (1-p)^{n-d}$$

where $d$ is the number of defaults, $n$ is the number of obligors and $p$ the probability of default. Therefore the probability of observing no defaults among all the obligors is $(1-p_A)^{n_A+n_B+n_C}$. Using this fact and the hypothesis method a confidence region for $p_A$ at confidence level $\gamma$ can be obtained by solving

$$(1-\gamma) \leq (1-p_A)^{n_A+n_B+n_C}$$

with the result

$$p_A \leq 1 - (1-\gamma)^{1/(n_A+n_B+n_C)}.$$ 

Further on the probability of default for grade $B$ is investigated, where, as before, the most prudent estimate of $p_B$ is used and therefore $p_B = p_C$. But $p_A = p_B$ would violate the most prudent estimate of $p_B$ since $p_A$ is the lower bound of $p_B$. The sample of size $n_B + n_C$ is therefore the one considered and as for the case of grade level $A$ the confidence region of $p_B$ of confidence level $\gamma$ can be calculated from

$$(1-\gamma) \leq (1-p_B)^{n_B+n_C}.$$ (4.4)

The equation (4.4) indicates that the confidence region for $p_B$ is given by
\[ p_B \leq 1 - (1 - \gamma)^{1/(n_B + n_C)}. \]

For the last grade level \( C \) it is concluded that

\[ (1 - \gamma) \leq (1 - p_C)^{n_C} \quad (4.5) \]

and from (4.5) it follows that the confidence region for \( p_C \) and confidence level \( \gamma \) can be described by

\[ p_C \leq 1 - (1 - \gamma)^{1/n_C}. \]

### 4.3 Independent case, few defaults

The previous portfolio is again considered but in this case there are a few observed defaults. The number of defaults in each grade is \( d_A, d_B \) and \( d_C \). First looking at a most prudent confidence region for the PD \( p_A \) of A where (4.2) is assumed to be true. As before this means that the entire portfolio can be viewed as a homogenous sample of size \( n_A + n_B + n_C \). As explained in the previous section the number of defaults in the portfolio is binomially distributed and (4.3) leads to that the probability of not observing more than \( d_A + d_B + d_C \) defaults is given by the expression

\[
\sum_{i=0}^{d_A+d_B+d_C} \binom{n_A + n_B + n_C}{i} p_A^i (1-p_A)^{n_A+n_B+n_C-i}.
\]

The confidence region at level \( \gamma \) for \( p_A \) can then be obtained by calculating the set of all the values of \( p_A \) that satisfies the following inequality

\[ 1 - \gamma \leq \sum_{i=0}^{d_A+d_B+d_C} \binom{n_A + n_B + n_C}{i} p_A^i (1-p_A)^{n_A+n_B+n_C-i}. \]

This inequality can be solved numerically for \( p_A \). For \( p_B \) and \( p_C \), with the same argumentation as before

\[ 1 - \gamma \leq \sum_{i=0}^{d_B+d_C} \binom{n_B + n_C}{i} p_B^i (1-p_B)^{n_B+n_C-i} \]

and

\[ 1 - \gamma \leq \sum_{i=0}^{d_C} \binom{n_C}{i} p_C^i (1-p_C)^{n_C-i}. \]

### 4.4 Dependent case

Pluto & Tasche, [21] describe the dependence between the default events with the previously mentioned Vasicek model. By replacing the probability of default with the conditional probability from the Vasicek model in the case where there are no defaults, the probability of default is the solution to

\[
1 - \gamma \leq \int_{-\infty}^{\infty} \phi(y) \left( 1 - \Phi \left( \frac{\Phi^{-1}(p_A) - \sqrt{pB}}{\sqrt{1 - \rho}} \right) \right)^{n_A+n_B+n_C} \, dy. \quad (4.6)
\]
As previously the right hand side of (4.6) describes the one-period probability of not observing any default among \( n_A + n_B + n_C \) with \( p_A \) as probability of default. The right hand side is the expected value of

\[
\left( 1 - \Phi \left( \frac{-1(p_A) - \sqrt{\rho y}}{\sqrt{1-\rho}} \right) \right)^{n_A+n_B+n_C}
\]

with \( y \) as the stochastic variable. The formulas to calculate the confidence bounds of \( p_B \) and \( p_C \) are derived in the same way,

\[
1 - \gamma \leq \int_{-\infty}^{\infty} \phi(y) \left( 1 - \Phi \left( \frac{-1(p_B) - \sqrt{\rho y}}{\sqrt{1-\rho}} \right) \right)^{n_B+n_C} dy
\]

\[
1 - \gamma \leq \int_{-\infty}^{\infty} \phi(y) \left( 1 - \Phi \left( \frac{-1(p_C) - \sqrt{\rho y}}{\sqrt{1-\rho}} \right) \right)^{n_C} dy.
\]

The correlated model is then applied to the cases where there are few defaults. As before the number of defaults in each grade is \( d_A, d_B \) and \( d_C \). The confidence region at level \( \gamma \) for \( p_A \) is represented as the set of all values of \( p_A \) that satisfy the inequality

\[
1 - \gamma \leq \int_{-\infty}^{\infty} \phi(y) \sum_{i=0}^{d_A+d_B+d_C} \left( \binom{n_A + n_B + n_C}{i} \right) \left( \lambda(p_A, \rho, y) \right)^i \left( 1 - \lambda(p_A, \rho, y) \right)^{n_A+n_B+n_C-i} dy
\]

where

\[
\lambda(p, \rho, y) = \Phi \left( \frac{-1(p) - \sqrt{\rho y}}{\sqrt{1-\rho}} \right).
\]

For \( p_B \) and \( p_C \) the confidence regions are calculated by modifying (4.7) as in Section 2.2.

### 4.5 Multi-period case

In cases when data from several years is to be used to calculate the confidence region of the probability of default the previously described multi-period Vasicek model is applied. The obligors that were present in the beginning of the time period are considered and obligors entering the portfolio afterwards are neglected. The previously used methods are the same for the multi-period case, that is the most prudent estimation principle is again considered and \( p_A = p_B = p_C = p \) but the observation time is \( T > 1 \) and therefore the one-period Vasicek-model in (4.6) is replaced with (3.5) so that

\[
1 - \gamma \leq \mathbb{E} \left[ (1 - \pi(Y_1, \ldots, Y_T)^{n_A+n_B+n_C} \right]
\]

where

\[
\pi(Y_1, \ldots, Y_T) = 1 - \prod_{t=1}^{T} \left( 1 - \Phi \left( \frac{-1(p) - \sqrt{\rho Y_t}}{\sqrt{1-\rho}} \right) \right).
\]

The integral from the previous section is here replaced with the expected value with respect to \( Y \). This model is also used in cases where there are few defaults and (4.7) is replaced with
\[ 1 - \gamma \leq E \left[ \sum_{i=0}^{d_A + d_B + d_C} \left( n_A + n_B + n_C \right) \left( \pi(Y_1, \ldots, Y_T)^i (1 - \pi(Y_1, \ldots, Y_T)^{n_A + n_B + n_C - i} \right) \right]. \]

For \( p_B \) and \( p_C \) the confidence regions are calculated by modifying (4.8) as in Section 2.2.

4.6 Applications and Issues

One of the drawbacks that Pluto & Tasche, [21], mention in their article is that the most prudent estimation principle in the few defaults case leads to that the upper confidence bound PD estimates are higher than the average default rate of the overall portfolio. They propose to use a scaling factor \( K \) based on the probability of default for the portfolio so that the estimates move towards the central tendency. The scaling method used by Pluto & Tasche only works in cases where there is at least one default. They therefore suggest an alternative scaling method by using an upper confidence bound for the portfolio instead of the overall portfolio PD. This upper confidence bound can be calculated the same way as they calculate the bound for the highest grade level. The advantage of the latter method is that the degree of conservatism can be controlled by the appropriate choice of confidence level for the estimation of the upper confidence bound of the overall portfolio PD.

The authors suggest that the method might be of specific importance for portfolios where neither internal nor external default data meet the Basel requirements and that it could be used as an alternative for the Basel slotting approach. Further on they propose that the methodology might be used for all sorts of low default portfolios.

Pluto & Tasche mention further issues with their approach. One of the problems they write about is how to choose the appropriate confidence levels. It is described as an issue since there are many factors to include when choosing the confidence level and there are many arguments for both higher and lower values. They also discuss how to define a low default portfolio, for example at which number of defaults should their method be used. Lastly they discuss what happens if the amount of defaults in the highest rating grade is significantly higher than in the lower rating grades, for low default portfolios this could happen with only one or two additional defaults. This could lead to that the PD estimates no longer are monotone and they therefore question if this should be taken as an indication for the non-correctness of the original ranking.
Chapter 5

The BCR Approach

Benjamin, Cathcart and Ryan proposed adjustments to the Pluto & Tasche or so called Confidence Based Approach that is widely applied by banks. The BCR approach is a quantitative approach to produce conservative PD estimates. While Pluto & Tasche propose a methodology for generating grade level PDs by the most prudent estimation, the BCR approach concentrates on estimating a portfolio-wide PD that then apply to grade level estimates and result in capital requirements. The method, as the Confidence Based Approach, uses the Vasicek model to account for dependence between defaults, and dependence is introduced by a fixed correlation with a single latent variable. In the multi-period case the value of the latent variable is correlated in successive years, specified by a year-to-year correlation. As in Pluto & Tasche the inputs to the model are the asset correlation, the year-to-year correlation and the level of confidence. The final value of the portfolio PD estimate is obtained through Monte Carlo simulation.

5.1 Theory

5.1.1 Independent case

From the Confidence Based Approach proposed in Pluto & Tasche, for independent defaults, the probability of default $p_j$ in grade $j$ is given by the solution to

$$1 - \gamma \leq \sum_{k=0}^{d_j} \binom{s_j^*}{k} (p_j)^k (1 - p_j)^{s_j^* - k}$$

where $s_j^* = \sum_{i=j}^c s_i$ and $d_j^* = \sum_{i=j}^c d_i$, meaning that $s_j^*$ is the sum of the amount of obligors in grade $i$ and the grades with higher rating, and $d_j^*$ is the amount of defaults in this grade and the grades with higher rating.

The conservative portfolio estimate as in the BCR setting is therefore given by the solution to

$$1 - \gamma = \sum_{k=0}^{d} \binom{s}{k} p^k (1 - p)^{s - k}$$  \hspace{1cm} (5.1)

where $s$ is the amount of obligors and $d$ is the number of defaults in the portfolio.
5.1.2 Dependent case

Modeling dependence from the Vasicek model, it is assumed that there is a single normally distributed risk-factor \( Y \) to which all assets are correlated and that each asset has the correlation \( \sqrt{\rho} \) with the risk factor \( Y \). Then, for a given value of the common factor \( y = Y \) the conditional probability of default given a realization of the systematic factor \( Y \) is given by

\[
p = \Phi \left( \frac{\Phi^{-1}(p_i) - \sqrt{\rho}y}{\sqrt{1-\rho}} \right).
\]

Inserting this expression into (5.1) and integrating over all possible realizations of the systematic factor, the resulting expression is

\[
1 - \gamma = \int_{-\infty}^{\infty} \phi(y) \sum_{k=0}^{d} \left( \frac{s}{k} \right) \Phi \left( \frac{\Phi^{-1}(p_i) - \sqrt{\rho}y}{\sqrt{1-\rho}} \right)^k \left( 1 - \Phi \left( \frac{\Phi^{-1}(p_i) - \sqrt{\rho}y}{\sqrt{1-\rho}} \right) \right)^{s-k} dy.
\]

The right-hand side is recognized as an expected value, therefore the PD estimate is the solution to

\[
1 - \gamma = \mathbb{E} \left[ \sum_{k=0}^{d} \left( \frac{s}{k} \right) \Phi \left( \frac{\Phi^{-1}(p_i) - \sqrt{\rho}Y_t}{\sqrt{1-\rho}} \right)^k \left( 1 - \Phi \left( \frac{\Phi^{-1}(p_i) - \sqrt{\rho}Y_t}{\sqrt{1-\rho}} \right) \right)^{s-k} \right].
\]

Using Monte-Carlo simulation with \( N \) simulations, the expected value can be approximated according to

\[
1 - \gamma = \frac{1}{N} \sum_{i=1}^{N} \left[ \sum_{k=0}^{d} \left( \frac{s}{k} \right) \Phi \left( \frac{\Phi^{-1}(p_i) - \sqrt{\rho}Y_t}{\sqrt{1-\rho}} \right)^k \left( 1 - \Phi \left( \frac{\Phi^{-1}(p_i) - \sqrt{\rho}Y_t}{\sqrt{1-\rho}} \right) \right)^{s-k} \right] .
\]

To find the probability of default is equivalent to finding \( p \) such that the above is true. This can be done numerically through a binomial search method, described in Section 5.2.

5.1.3 Multi-period case

In the case of several years, as previously shown, the conditional probability of default given a realization of the systematic factor for \( t \) years as in the Vasicek model is

\[
\pi(Y_1, \ldots, Y_T) = 1 - \prod_{t=1}^{T} \left( 1 - \Phi \left( \frac{\Phi^{-1}(p_i) - \sqrt{\rho}Y_t}{\sqrt{1-\rho}} \right) \right).
\]

With the same reasoning as in the one-period case the probability of default \( p \) is given by the solution to

\[
1 - \gamma = \sum_{k=0}^{d} \left( \frac{s}{k} \right) \pi(Y_1, \ldots, Y_T)^k (1 - \pi(Y_1, \ldots, Y_T))^{s-k}.
\]

Once again integrating over all possible values of the systematic factor the probability of default can be obtained by the solution to

\[
1 - \gamma = \mathbb{E} \left[ \sum_{k=0}^{d} \left( \frac{s}{k} \right) \pi(Y_1, \ldots, Y_T)^k (1 - \pi(Y_1, \ldots, Y_T))^{s-k} \right] \Leftrightarrow
\]
1 − γ = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=0}^{d} \left( \begin{array}{c} s \\ k \end{array} \right) \pi(Y_i)^k(1 - \pi(Y_i))^{s-k} \tag{5.2}

where $Y_i$ is the $i$th simulation of the vector $Y = (Y_1, \ldots, Y_T)$. The solution to this problem can as in the one-period case be obtained by numerical methods.

### 5.2 Estimation and simulation

The estimation of the probability of default for the portfolio is computed in MATLAB

where $p$ is solved from (5.2) with numerical methods as mentioned before. This is executed in the following steps.

1. Draw $N$ samples from $N(\lambda, \Sigma)$ where $\lambda$ is a zero vector with the same length as the time period and $\Sigma$ is the correlation matrix mentioned in the Vasicek model.

2. Set $f(p) = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=0}^{d} \left( \begin{array}{c} s \\ k \end{array} \right) \pi(Y_i)^k(1 - \pi(Y_i))^{s-k} - (1 - \gamma)$.

3. Find $p$ such that $f(p)$ is close to zero using the following iteration.

   (a) Set the number of iterations to $n = \log_2((p_{\text{high}} - p_{\text{low}})/\delta)$ where $[p_{\text{low}}, p_{\text{high}}]$ is the interval $p$ is believed to be in and $\delta$ is the accepted error.

   (b) For $n$ number of iterations the midpoint, $p_{\text{mid}}$, of the interval is calculated. It is then checked if $f(p_{\text{mid}}) > 0$ or $< 0$. If it’s the first case the lower bound is set equal to the midpoint, in the second case the higher bound is set equal to the midpoint.

   (c) When the $n$ iterations are done the estimated probability of default is set to the final midpoint.

The following algorithm describes the calculations in pseudo-code.

**Algorithm 2** Multi-period BCR

**Require:** A sufficiently large number of simulations.

draw $N$ number of $Y_i \sim N(\lambda, \Sigma)$

set $f(p) = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=0}^{d} \left( \begin{array}{c} s \\ k \end{array} \right) \pi(Y_i)^k(1 - \pi(Y_i))^{s-k} - (1 - \gamma)$

for $j = 1$ to $n$
do

$p_{\text{mid}} = (p_{\text{high}} - p_{\text{low}})/2$

if $f(p_{\text{mid}}) < 0$
then
$p_{\text{high}} = p_{\text{mid}}$
else
$p_{\text{low}} = p_{\text{mid}}$
end if

end for
Chapter 6

Bayesian Approach

Modeling the probability of default for low default portfolios using the Bayesian approach was first considered by Dwyer in 2006 [7]. This method has gained more attention in later years with numerous articles written about the subject. One of the reasons it has received attention is that, unlike the confidence based approach, it avoids the issue of choosing confidence levels. This chapter will in the first part explain the use of Bayesian statistics in calculating PDs for LDPs, and in the second part describe two different methods for estimating PDs in the Bayesian setting, one of which is a new approach.

6.1 Bayesian statistics

In Bayesian statistics the unknown parameters are random variables of a known distribution, \( \pi_P(p) \), called the prior distribution. The prior distribution is a qualified guess on the likely value of the unknown parameters before the outcomes of the data \( L \) have been observed. In the setting of modeling PD in LDPs, the data \( L \) is a default indicator for the portfolio, and described if the obligor has defaulted or not, see Section 6.2. The probability density function of the data is decided given a value of the unknown parameter, \( f_{L|P}(l|p) \), this is called the likelihood function. The joint probability density function can be written as

\[
f_{L,P}(l,p) = f_{L|P}(l|p)\pi_P(p).
\]

(6.1)

From the definition of the marginal density function it is known that

\[
f_L(l) = \int f_{L,P}(l,p)dp = \int f_{L|P}(l|p)\pi_P(p)dp.
\]

(6.2)

The distribution of the probability given the data follows from Bayes’ rule.

**Definition 6.1.1. Bayes’ Rule**

*Let \( A \) and \( B \) be events. Then:*

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

From Bayes’ rule and (6.2) it is clear that

\[
f_{P|L}(p|l) = \frac{f_{L|P}(l|p)\pi_P(p)}{f_L(l)} = \frac{f_{L|P}(l|p)\pi_P(p)}{\int f_{L|P}(l|p)\pi_P(p)dp}.
\]

(6.3)
The result from (6.3) is called the posterior density function and can be interpreted as what is known about the unknown parameter when the data is given. In general, the aim is to find the posterior density of the parameter in order to obtain Bayesian estimates of the parameter.

A simpler version of the posterior can be written as

$$\text{Posterior Density } \propto \text{Likelihood } \times \text{Prior Density} \quad (6.4)$$

By this, to determine the posterior density up to a normalizing constant, the required components are the likelihood function and the prior distribution for the parameter. In general the posterior distributions are often too complex to work with analytically and therefore Monte Carlo methods are widely used.

### 6.2 Bayesian approach

To use Bayesian statistics in the setting of estimating a probability of default $p$, firstly the variable $L_i$ is defined as the default indicator for obligor $i$, $L = L_1, L_2, \ldots, L_n$. In the context of Bayesian statistics, $L$ is the data and represents what is known, that is if an obligor has defaulted or not. The variable $L_i$ is a Bernoulli distributed variable that can take the values 1 and 0,

$$L_i = \begin{cases} 
1 & \text{if obligor } i \text{ defaults.} \\
0 & \text{otherwise.} 
\end{cases} \quad (6.5)$$

Because $L_i$ is Bernoulli distributed, the density function for $L_i$, if the probability of default $p$ is known, is

$$f_{L_i|p}(l_i|p) = p_i^{l_i}(1 - p_i)^{1-l_i}. \quad (6.6)$$

Modeling dependence between defaults as in the multi-period Vasicek model,

$$X_{i,t} = \sqrt{\rho} Y_t + Z_{i,t}\sqrt{1-\rho}$$

and assuming that an obligor can not have defaulted previously, the probability of default at time $t$ for obligor $i$ is given by

$$p_{t,i} = P(L_{i,t} = 1) = P(X_{i,t} \leq c_{i,t}).$$

Also, assuming as in the BCR method that obligors in the same rating category have equal probability of default, the conditional probability of default given the asset correlation $\rho$ and a realization of the systematic factor $Y_t$ is given by

$$\lambda(p, \rho, Y_t) = P(X_{i,t} \leq c) = P(\sqrt{\rho} Y_t + Z_{i,t}\sqrt{1-\rho} \leq \Phi^{-1}(p)) =$$

$$P \left( Z_{i,t} \leq \frac{\Phi^{-1}(p) - \sqrt{\rho}}{\sqrt{1-\rho}} \right) = \Phi \left( \frac{\Phi^{-1}(p) - \sqrt{\rho}}{\sqrt{1-\rho}} \right). \quad (6.7)$$
6.2. Bayesian approach

6.2.1 Likelihood function

**Independent case**

The aim is to find the likelihood function, \( f_{L|P}(l|p) \), that is the density of the data when the probability of default is given. Assuming that the defaults between obligors are independent the distribution of the data given the probability of defaults \( p = p_1, p_2, \ldots, p_n \) is

\[
f_{L_1, L_2, \ldots, L_n | P}(l_1, l_2, \ldots, l_n | p) = \prod_{i=1}^{n} p_i^{l_i} (1 - p_i)^{1 - l_i}.
\]  

(6.8)

When the obligors fall into the same rating category it can be assumed that they have the same probability of default \( p = p_1 = p_2 = \ldots = p_n \), and from this it follows that

\[
f_{L_1, L_2, \ldots, L_n | P}(l_1, l_2, \ldots, l_n | p) = \prod_{i=1}^{n} p_i^{l_i} (1 - p_i)^{1 - l_i} = p^{\sum_{i=1}^{n} l_i} (1 - p)^{\sum_{i=1}^{n} 1 - l_i} = p^d (1 - p)^{n - d}
\]

(6.9)

where \( d = \sum_{i=1}^{n} l_i \) is the total number of defaults. Stating that \( L = \sum_{i=1}^{n} l_i \), the probability of exactly \( d \) defaults occurring when there are \( n \) obligors is

\[
P(L = d|P = p) = f_{L|P}(d|p) = \binom{n}{d} p^d (1 - p)^{n - d}
\]

(6.10)

and this is therefore the likelihood function in the independent case. From this it is concluded that the likelihood function can be written as

\[
f_{L|P}(d|p) = \binom{n}{d} p^d (1 - p)^{n - d}
\]

or

\[
f_{L|P}(l|p) = \prod_{i=1}^{n} p_i^{l_i} (1 - p_i)^{1 - l_i} = p^d (1 - p)^{n - d}.
\]

**Dependent case**

Incorporating dependence into the Bayesian setting, using the Vasicek model for dependence between defaults the conditional probability of default given a realization of the systematic factor \( Y \) is

\[
\lambda(p, \rho, Y) = \Phi \left( \frac{\Phi^{-1}(p) - \sqrt{\rho}Y}{\sqrt{1 - \rho}} \right).
\]

Following the same methodology as in the BCR approach, substituting the probability of default in the independent model with this expression and integrating over all possible values of the systematic factor, the resulting expression for the likelihood function in the dependent case is

\[
f_{L|P}(d|p) = \int_{-\infty}^{\infty} \binom{n}{d} \lambda(p, \rho, Y)^d (1 - \lambda(p, \rho, Y))^{n - d} \phi(y) dy.
\]
The likelihood function can, as in the independent case, also be written as

\[
f_{L|P}(L|p) = \prod_{i=1}^{n} \lambda(p, \rho, Y)^{l_i}(1 - \lambda(p, \rho, Y))^{1-l_i} = \\
\lambda(p, \rho, Y)^{d}(1 - \lambda(p, \rho, Y))^{n-d}.
\]

**Multi-period case**

Using the conditional probability of default in the time dependent setting from (6.2),

\[
\lambda_t(p, \rho, Y_t) = \Phi \left( \frac{\Phi^{-1}(p_t) - \sqrt{\rho}Y_t}{\sqrt{1-\rho}} \right).
\]

Furthermore, substituting into the likelihood function in the independent case and integrating over all possible realizations of the systematic factor \(Y_t\), one obtains the expression

\[
f_{L|P}(L_1 = d_1, \ldots, L_T = d_T|p) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \phi(y_1, \ldots, y_T) \cdot \\
\prod_{t=1}^{T} \binom{n_t}{d_t} \lambda_t(p, \rho, Y_t)^{d_t}(1 - \lambda_t(p, \rho, Y_t))^{n_t-d_t} dy_1 \cdots dy_T
\]

where \(d_t\) and \(n_t\) are the number of defaults and obligors in the time period \(t\). This is the likelihood function in the dependent multi-period setting. Similarly to the previous section, the likelihood function can also be written as

\[
f_{L|P}(L|p, \rho, Y) = \prod_{t=1}^{T} \prod_{i=1}^{n_t} \lambda_t(p, \rho, Y_i)^{l_i}(1 - \lambda_t(p, \rho, Y_i))^{1-l_i} = \\
\prod_{t=1}^{T} \lambda_t(p, \rho, Y_i)^{d_t}(1 - \lambda_t(p, \rho, Y_i))^{n_t-d_t}.
\]

(6.11)

**6.2.2 Prior distribution**

An important part of estimation of PDs in the Bayesian approach is the choice of prior distribution. The prior distribution reflects the belief of the distribution of the parameter \(p\) before the data \(L\) has been observed. To be able to derive posterior densities for the parameters, one must first choose a prior \(\pi_p(p)\).

**Uniform prior**

The uniform prior represents that nothing is known about the parameter, that is that \(p\) is uniformly distributed between 0 and 1 or between an upper and lower bound. This is therefore the most natural choice for a prior for \(p\), and is also referred to as the uninformative prior since it represents that nothing is known of the parameter beforehand. The uniform prior can be expressed as

\[
\pi_p(p) = \begin{cases} 
\frac{1}{p_u - p_l} & \text{for } p_l \leq p < p_u \\
0 & \text{elsewhere.}
\end{cases}
\]
6.2. Bayesian approach

6.2.3 Posterior density functions

As previously mentioned, the aim of the Bayesian approach for estimation of PDs is the derivation of the posterior density function for \( p \). From (6.3) and the dependence on \( Y \) in the expression for

\[
f_{L|p}(l|p, \rho, Y, \tau) \propto f_L(l) \prod_{t=1}^{T} \lambda_t(p, \rho, Y_t)^{d_t} (1 - \lambda_t(p, \rho, Y_t))^{n_t - d_t} \times \pi(p),
\]

where \( \pi(p, \rho, Y, \tau) \) is the joint prior density function. In practice, the parameters \( \rho \) and \( \tau \) are inputs to the model and hence specified, therefore the marginal posterior density function is

\[
f_{p|L}(p|l) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(p, \rho, Y, \tau|l) \, dy_1 \cdots dy_T.
\]

The analytic expression for the posterior density contains \( T \) integrals and is therefore very complex to evaluate. As such, Monte Carlo algorithms are a possible solution for this problem. Specifically, Markov chain Monte Carlo algorithms allow for the creation of a Markov chain that has the designated posterior density as stationary distribution, see Section 3.4.1. An algorithm that achieves this is the Gibbs sampler that is further explained in Section 3.4.2. To construct a Gibbs sampler, the marginal conditional posterior density function for both \( Y \) and \( p \) is required. These can be derived from the relation of the joint density,

\[
f(l, p, Y) = f(l|p, Y) \pi_p(p) f(Y|\tau)
\]

which follows from that given the realization of systematic risk factor \( \pi(p, Y) = \pi_p(p) f(Y|\tau) \) due to the conditional independence. In addition, it is known from (6.4) that the posterior density is proportional to the likelihood times the prior density.

Posterior density function for \( p \)

From the proportionality relationship given by (6.4), the conditional posterior density function for \( p \) can be derived as

\[
f(p|l, \rho, Y) \propto f(l|p, \rho, Y) \times \pi_p(p) = \prod_{t=1}^{T} \lambda_t(p, \rho, Y_t)^{d_t} (1 - \lambda_t(p, \rho, Y_t))^{n_t - d_t} \times \pi_p(p).
\]

Uniform prior

The conditional density for the uniform prior is given by

\[
f(p|l, \rho, Y) \propto \prod_{t=1}^{T} \lambda_t(p, \rho, Y_t)^{d_t} (1 - \lambda_t(p, \rho, Y_t))^{n_t - d_t} \times \frac{1}{p_u - p_l}.
\]

Posterior density function for the systematic factors \( Y \)

From the dependence structure of \( Y_t \) given by the multi-period Vasicek model, it is obvious that the sequence of the systematic factors in each time point is a Markov chain, since each \( Y \) is only dependent on the previous value. From Section 3.3 it is therefore known that the joint density function of the Ys has the structure of (3.7). Using this the conditional posterior density functions for the systematic factor can be...
attained from (6.12) removing everything that is not dependent on the \( y \) in question. This results in the following partial conditional densities.

\[
\begin{align*}
f(y_1|l, p, \rho, y_{-1}) &\propto f_L|P(l|p, \rho, Y) \cdot f_{Y_1}(y_1) \cdot f_{Y_2|Y_1}(y_2|y_1) \\
&\propto \prod_{t=1}^{T} \lambda_t(p, \rho, Y_t)^{d_t} (1 - \lambda_t(p, \rho, Y_t))^{n_t-d_t} \times e^{-\frac{y_t^2}{2}} \cdot e^{-\frac{(y_t-y_{t-1})^2}{2(1-\tau^2)}} \end{align*}
\]

\[
\begin{align*}
f(y_T|l, p, \rho, y_{-T}) &\propto f_L|P(l|p, \rho, Y) \cdot f_{Y_T|Y_{T-1}}(y_T|y_{T-1}) \\
&\propto \prod_{t=1}^{T} \lambda_t(p, \rho, Y_t)^{d_t} (1 - \lambda_t(p, \rho, Y_t))^{n_t-d_t} \times e^{-\frac{(y_T-y_T_{T-1})^2}{2(1-\tau^2)}} \end{align*}
\]

6.2.4 Estimation and simulation

Parameter estimate

In the Bayesian setting, there are at least three general ways to obtain a Bayesian estimator of a parameter based on three different loss functions. These are the square error loss, the absolute loss and the 0-1 loss. The loss functions correspond to the mean, the median and the mode of the marginal posterior distribution as the estimator of the sought parameter. The loss function considered in this thesis is the same as the one considered by Dwyer [7], Keifer [14], Tasche [23] and Chang & Yu [6], and is the square error loss, that is the mean as estimator. This is by far the most common estimator occurring in literature and therefore a natural choice.

Assuming a square error loss, the estimate of \( p \) is

\[
\hat{p} = \frac{1}{N} \sum_{i=1}^{N} p^{(i)}
\]

where \( p^{(i)} \) are values simulated from the marginal posterior distribution.

Estimation procedure

As previously mentioned, to obtain a sample from the posterior distribution for \( p \), a Gibbs sampler sampling from the conditional distribution of \( p \) and \( y \) can be constructed. The problem is thereby transformed to sampling from the conditional posterior densities for \( y \) and \( p \), given by (6.15) and (6.14) instead, something that unfortunately is more complicated than it first seems. Approaches mentioned by other authors to do this is adaptive rejection sampling (ARS) and adaptive rejection metropolis sampling (ARMS), [6], see [10] for a description of ARS and ARMS. However, implementation of this is deemed too complicated and time consuming to be an efficient method. This is due to that the first term of both conditional densities in the general case will be very small and also change with each simulation of the systematic factor, making it complicated to find a suitable rejection region. Hence, the algorithm to do this will be very complex and time consuming, and this method is for this reason ruled out as an efficient method. Instead, an alternative method to sample from the
conditional posterior densities in the Bayesian approach is considered, in which the idiosyncratic factor is also sampled.

### 6.2.5 Alternative approach

Arising from the issue of sampling from the complicated posterior densities in the previous section, an alternative approach is also considered, with the key feature that the idiosyncratic factor $Z_{i,t}$ is also sampled as well as that $c$, the normal inverse of $p$, is sampled instead of $p$. This type of approach, to the authors knowledge, has not been used for modeling of PD in LDPs before. As seen further on, because of the fact that in every time step all other components are known, this approach attends to the issue of sampling from the complicated posterior densities in the Bayesian setting by reducing the posterior densities to normal densities limited by some interval for the parameter. Sampling from a known density in a specified interval can be achieved, as seen in Section 3.1.

As in the general Bayesian approach, the probability of default is

$$
\lambda_t(p, \rho, Y_t) = P(L_{i,t} = 1) = P(X_{i,t} \leq c)
$$

and hence the probability of not defaulting is

$$
P(L_{i,t} = 0) = P(X_{i,t} > c).
$$

From the Vasicek model for dependence between defaults

$$
X_{i,t} = \sqrt{\rho}Y_t + Z_{i,t}\sqrt{1-\rho}
$$

where $Y_t$ and $Z_{i,t}$ are standard normally distributed variables independent of each other.

Given that both $Y_t$ and $Z_{i,t}$ in the Vasicek model are sampled and it is known that these follow a standard normal distribution, it is known that the conditional distribution of $Y_t$ and $Z_{i,t}$ given all other variables in the Vasicek model will be standard normal, limited to some interval derived below. The term $c = \Phi^{-1}(p)$ will because of the inversion sampling theorem in Section 3.1 also belong to the standard normal distribution when $p$ is uniform and therefore it is known that all three conditional distribution functions will be standard normally limited to some interval in the case of a uniform prior. For a variable which distribution function is known, in this case standard normal, and the interval to which it is limited is also known, it follows from Section 3.1 and the inverse sampling theorem that the variable can be sampled from (3.1). The boundaries for each variable can be derived from the fundamental relation that a default occurs if $P(X_{i,t} \leq c)$.

For the normal inverse of the probability of default, $c = \Phi^{-1}(p)$,

$$
\begin{align*}
X_{i,t} \leq c & \quad \text{if a default occurs} \\
X_{i,t} > c & \quad \text{if a default doesn’t occur}
\end{align*}
\quad \Leftrightarrow
\begin{align*}
\max_{L_{i,t} = 1} X_{i,t} & \leq c \\
\min_{L_{i,t} = 0} X_{i,t} & > c.
\end{align*}
$$
Because of the uniform prior on $p$, $p$ is limited to the interval $p_L < p < p_U$ and hence for $c$, $\Phi^{-1}(p_L) < c < \Phi^{-1}(p_U)$. Taking this into account, the boundaries for $c$ can be obtained as
\[
\begin{align*}
\max\left( \max_{l_i,t=1} X_{i,t}, \Phi^{-1}(p_L) \right) &\leq c \\
\min\left( \min_{L_{i,t}=0} X_{i,t}, \Phi^{-1}(p_U) \right) &> c \iff \left\{\begin{array}{ll}
\max(a_c, \Phi^{-1}(p_L)) &\leq c \\
\min(b_c, \Phi^{-1}(p_U)) &> c
\end{array}\right.
\end{align*}
\]
where $a_c = \max_{L_{i,t}=1} X_{i,t}$ and $b_c = \min_{L_{i,t}=0} X_{i,t}$.

Consequently, the posterior density function for $c$ is
\[
f(c|\rho, X, Y, L) = \frac{\Phi(c) - \Phi(a_c)}{\Phi(b_c) - \Phi(a_c)}
\]
and from (3.1), $c$ can be sampled from
\[
c = \Phi^{-1}(U \cdot (\Phi(\min(b_c, \Phi^{-1}(p_U))) - \Phi(\max(a_c, \Phi^{-1}(p_U)))) + \Phi(\max(a_c, \Phi^{-1}(p_L))))).
\]
(6.16)

For the idiosyncratic factor $Z_{i,t}$, the interval to which it is limited can be extracted in the same manner.
\[
\begin{align*}
\left\{ \begin{array}{ll}
X_{i,t} &\leq c &\text{if a default occurs} \\
X_{i,t} &> c &\text{if a default doesn’t occur}
\end{array}\right. \iff \\
\left\{ \begin{array}{ll}
\sqrt{\rho} Y_t + Z_{t,i} \sqrt{1 - \rho} &\leq c &\text{if } L_{i,t} = 1 \\
\sqrt{\rho} Y_t + Z_{t,i} \sqrt{1 - \rho} &> c &\text{if } L_{i,t} = 0
\end{array}\right. \iff \\
\left\{ \begin{array}{ll}
Z_{t,i} &\leq \frac{c - \sqrt{\rho} Y_t}{\sqrt{1 - \rho}} &\text{if } L_{i,t} = 1 \\
Z_{t,i} &> \frac{c - \sqrt{\rho} Y_t}{\sqrt{1 - \rho}} &\text{if } L_{i,t} = 0.
\end{array}\right.
\end{align*}
\]
Therefore, $Z_{i,t}$ has an upper bound $\eta = \frac{c - \sqrt{\rho} Y_t}{\sqrt{1 - \rho}}$ if a default occurs, and the same lower bound $\eta$ if there is no default. As a result, the conditional density for the systematic factor is
\[
f(z|c, \rho, Y, L) = \begin{cases} 
\Phi(z) &\text{if } L_{i,t} = 1 \\
\Phi(z) - \Phi(\eta) &\text{if } L_{i,t} = 0
\end{cases}
\]
and from the inversion sampling theorem and equation (3.2) in Section 3.1, the idiosyncratic factor can be sampled as
\[
\begin{align*}
\left\{ \begin{array}{ll}
Z_{t,i} = \Phi^{-1}(U \cdot \Phi(\eta)) &\text{if } L_{i,t} = 1 \\
Z_{t,i} = \Phi^{-1}(U \cdot (1 - \Phi(\eta)) + \Phi(\eta)) &\text{if } L_{i,t} = 0.
\end{array}\right. \quad (6.17)
\end{align*}
\]
Lastly, for the sampling of the systematic risk factors $Y_t$, one must also consider the dependence between the risk factors given by the multi-period Vasicek model. From Definition 3.2.2, the relation between the systematic factors is given as
\[
Y_t = \tau Y_{t-1} + \sqrt{1 - \tau^2} W_t.
\]
(6.18)
where
\[
f(y_t|y_{t-1}, \tau) \propto e^{-\frac{(y_t - \tau y_{t-1})^2}{2(1-\tau^2)}}.
\]
As in previous cases,
\[
\begin{align*}
\left\{ \begin{array}{ll}
X_{i,t} &\leq c &\text{if a default occurs} \\
X_{i,t} &> c &\text{if a default doesn’t occur}
\end{array}\right. \iff \\
\left\{ \begin{array}{ll}
\sqrt{\rho} Y_t + Z_{t,i} \sqrt{1 - \rho} &\leq c &\text{if } L_{i,t} = 1 \\
\sqrt{\rho} Y_t + Z_{t,i} \sqrt{1 - \rho} &> c &\text{if } L_{i,t} = 0
\end{array}\right. \iff \\
\left\{ \begin{array}{ll}
Z_{t,i} &\leq \frac{c - \sqrt{\rho} Y_t}{\sqrt{1 - \rho}} &\text{if } L_{i,t} = 1 \\
Z_{t,i} &> \frac{c - \sqrt{\rho} Y_t}{\sqrt{1 - \rho}} &\text{if } L_{i,t} = 0.
\end{array}\right.
\end{align*}
\]
conditional distributions are proportional to can be derived as
dependence from the multi-period Vasicek model above, distributions to which the proportionality relationships for a Markov chain given by (3.8) and by using the sequence of systematic factors is a Markov chain, from

\[
\begin{align*}
Y_t &\leq \min_{L+1=1} \frac{c - \sqrt{1 - \rho Z_{i,t}}}{\sqrt{\beta}} \\
Y_t &> \max_{L+1=0} \frac{c - \sqrt{1 - \rho Z_{i,t}}}{\sqrt{\beta}}.
\end{align*}
\]

Furthermore, using that the sequence of systematic factors is a Markov chain, from

\[
\begin{align*}
&f_{Y_1|Y_{-1}}(y_1|y_{-1}) \propto f_{Y_1}(y_1) \cdot f_{Y_1|Y_1}(y_2|y_1) = \\
&\exp \left( -\frac{x_1^2}{2} - \frac{\tau y_1}{2(1-\tau^2)} \right) \propto \exp \left( -\frac{(x_1 - \tau y_1)^2}{2(1-\tau^2)} \right),
\end{align*}
\]

\[
\begin{align*}
&f_{Y_t|Y_{t-1}}(y_t|y_{t-1}) \propto f_{Y_{t+1}|Y_{t}}(y_{t+1}|y_t) \cdot f_{Y_t|Y_{t-1}}(y_t|y_{t-1}) = \\
&\exp \left( -\frac{y_{t+1}^2 - \tau y_t}{2(1-\tau^2)} \right) \cdot \exp \left( -\frac{(y_{t+1} - \tau y_t)^2}{2(1-\tau^2)} \right) \propto \exp \left( -\frac{(y_{t+1} - \tau y_t)^2}{2(1-\tau^2)} \right),
\end{align*}
\]

\[
\begin{align*}
&f_{Y_T|Y_{T-1}}(y_T|y_{T-1}) \propto f_{Y_T|Y_{T-1}}(y_T|y_t) = \\
&\exp \left( -\frac{(y_T - \tau y_{T-1})^2}{2(1-\tau^2)} \right).
\end{align*}
\]

Due to this fact, it is known that for the distribution of the systematic factor the following holds

\[
\begin{align*}
Y_1|Y_{-1} &\sim \mathcal{N}(\tau y_2, 1 - \tau^2) \\
Y_t|Y_{t-1} &\sim \mathcal{N}(\tau(x_{t-1} + x_{t+1})_{t+1} - \tau^2 y_{t+1}, 1 - \tau^2) \\
Y_T|Y_{T-1} &\sim \mathcal{N}(\tau y_{T-1}, 1 - \tau^2).
\end{align*}
\]

Because the systematic factors are normally distributed with known expected value and standard deviation, in the general case the systematic factor \(Y_t\) can be expressed as \(Y_t = \mu_t + \sigma_t W_t\), where \(W_t\) is a standard normal variable. Therefore, from the previously obtained boundaries for \(Y_t\),
The following describes the Gibbs sampler.

\[
\begin{align*}
\mu_t + \sigma_t W_t &\leq \min_{L_{t,i}=1} \frac{c - \sqrt{1 - p} Z_{i,t}}{\sqrt{p}} \\
\mu_t + \sigma_t W_t &> \max_{L_{t,i}=0} \frac{c - \sqrt{1 - p} Z_{i,t}}{\sqrt{p}} \\
W_t &\leq \frac{\min_{L_{t,i}=1} \frac{c - \sqrt{1 - p} Z_{i,t}}{\sqrt{p}} - \mu_t}{\sigma_t} \\
W_t &> \frac{\max_{L_{t,i}=0} \frac{c - \sqrt{1 - p} Z_{i,t}}{\sqrt{p}} - \mu_t}{\sigma_t}.
\end{align*}
\]

As a result, \( W_t \) can be sampled as

\[
\Phi^{-1}(U \cdot (\Phi(b_y) - \Phi(a_y)) + \Phi(a_y)) \quad \text{where} \quad (6.22)
\]

\[
a_y = \begin{cases} 
\max_{L_{t,i}=0} \frac{c - \sqrt{1 - p} Z_{i,t}}{\sqrt{p}} - \tau y_2 & \text{if } t = 1 \\
\max_{L_{t,i}=0} \frac{c - \sqrt{1 - p} Z_{i,t}}{\sqrt{p}} - \frac{\tau(y_{t-1} + y_{t+1})}{1 + \tau^2} & \text{if } t \neq 1 \text{ and } t \neq T \\
\max_{L_{i,T}=0} \frac{c - \sqrt{1 - p} Z_{i,T}}{\sqrt{p}} - \tau y_{T-1} & \text{if } t = T
\end{cases}
\]

\[
b_y = \begin{cases} 
\min_{L_{t,i}=1} \frac{c - \sqrt{1 - p} Z_{i,t}}{\sqrt{p}} - \tau y_2 & \text{if } t = 1 \\
\min_{L_{t,i}=1} \frac{c - \sqrt{1 - p} Z_{i,t}}{\sqrt{p}} - \frac{\tau(y_{t-1} + y_{t+1})}{1 + \tau^2} & \text{if } t \neq 1 \text{ and } t \neq T \\
\min_{L_{i,T}=1} \frac{c - \sqrt{1 - p} Z_{i,T}}{\sqrt{p}} - \tau y_{T-1} & \text{if } t = T
\end{cases}
\]

and the conditional posterior density for \( W_t \) is

\[
f(w_t | c, \rho, \tau, Z, L) = \frac{\Phi(w_t) - \Phi(a_y)}{\Phi(b_y) - \Phi(a_y)}.
\]

From the sampled value of \( W_t \), it is then trivial to obtain a sample of \( Y_t \) as \( Y_t = \mu_t + \sigma_t W_t \), where

\[
\mu_t = \begin{cases} 
\tau y_2 & \text{if } t = 1 \\
\tau(y_{t-1} + y_{t+1}) & \text{if } t \neq 1 \text{ and } t \neq T \text{ and } \sigma_t = \begin{cases} 
1 - \tau^2 & \text{if } t = 1 \text{ or } t = T \\
\frac{1 - \tau^2}{1 + \tau^2} & \text{if } t \neq 1 \text{ and } t \neq T.
\end{cases}
\end{cases}
\]

(6.23)

**Estimation and simulation**

To estimate the probability of default, a sample of size \( N \) from its posterior density function is required. To obtain this sample, a Gibbs sampler can be constructed, where in each step the sampling is conducted in accordance with the above expressions. Each sample is obtained using the current sampled values of the other variables. The following describes the Gibbs sampler.

1. Set initial values; total number of samples, number of burn-in samples, the asset correlation \( \rho \), the year-to-year correlation \( \tau \) and the upper and lower bound of the uniform prior distribution for \( p, p_u \) and \( p_l \).

2. Set initial guesses of \( Y \) and \( c, Y^{(0)} \) and \( c^{(0)} \).
3. For each step in the Gibbs sampler
   (a) Sample the matrix $\mathbf{Z}^{(i)}$ containing all values of $Z_{i,t}^{(i)}$ from current sampled values of the systematic factor and $c$, $\mathbf{Y}^{(i-1)}$ and $c^{(i-1)}$ in (6.17).
   (b) Sample $\mathbf{Y}^{(i)}$ from $\mathbf{Z}^{(i)}$ and $c^{(i-1)}$ by first sampling a value of $W_1$, then $W_2, \ldots, W_{T-1}$ and lastly $W_T$ and putting $Y_t = \mu_t + \sigma_t W_t$ with $\mu_t$ and $\sigma_t$ as in (6.23).
   (c) Sample $c^{(i)}$ from (6.16) given $\mathbf{Y}^{(i)}$ and $\mathbf{Z}^{(i)}$.
   (d) Increment $i$.

4. Disregard the burn-in samples and the starting value of $c$.

5. Obtain a sample of $p$ from $p = \Phi(c)$.

6. Obtain an estimate of $p$ from its posterior distribution as $\tilde{p} = \sum_{i=1}^{N} p^{(i)}$. 
Chapter 7

Grade level allocation of PDs

As mentioned in Section [2.1.5] Basel regulations require firms to produce grade level estimates of PD. Both in the BCR and the Bayesian approach portfolio estimates of PD are obtained and therefore there is a need for a method to allocate the calculated portfolio PD estimate on to the different rating grades, resulting in PD estimates for each grade. This chapter presents several different methods to solve this problem, both for regular portfolios and portfolios consisting of several subportfolios.

7.1 Grade level estimates

7.1.1 Scaling

In BCR [4] a scaling method is proposed in order to obtain grade level estimates according to the calculated portfolio PD. To use this method a weighted portfolio PD needs to be calculated from some predefined grade level PDs. As mentioned in [4], these PDs may be derived subjectively or by using an analytical method. The weighted portfolio PD, $p_W$, is calculated by multiplying the amount of obligors in a grade with the PD for this grade and dividing the sum of these multiplications with the total amount of obligors over all the grades. The portfolio PD estimate $p_E$ from the BCR or Bayesian method is then used to calculate a scale-factor, $p_E/p_W$. If the weighted portfolio PD is less than the PD estimate the grade level PD estimates are adjusted upwards so that the weighted average PD is equal to the PD estimate. In the method proposed in [4], this is done by multiplying each grade level PD with the scale-factor. Thereby, the estimated portfolio PD is allocated on to grades, resulting in grade level PDs so that the weighted portfolio PD is equal to the estimated portfolio PD.

An alternative to using multiplication to scale up the grade level PDs so that an equality between the weighted portfolio PD and the estimated portfolio PD can be obtained is to add an additive term to each rating grade. As before this should only be done when the estimated PD is larger than the weighted portfolio PD. The idea is to calculate the difference between the two PDs, $p_E - p_W$, and add this result to each of the grade-level PDs so that the new weighted portfolio PD is the same as the estimated portfolio PD.

When deciding the scale-factor and the additive term the floor values presented in Section [2.1.5] must be accounted for, that is the grade level PD should be at least
0.03%. So the weighted portfolio PD with the accounted floor values are compared to the estimated portfolio PD since no scaling is necessary if this weighted portfolio PD is higher than the estimated portfolio PD.

### 7.1.2 Fitting

In the previous section it is mentioned that some predefined values of the grade level PDs are needed in order to scale PDs. These can be obtained from the available data, that is the amount of defaults and obligors in the different rating grades. Calculating the PDs for each rating grade from the data by dividing the number of defaults with the number of obligors in each rating grade results in an observed default rate. A curve is then constructed by fitting the observed default rate to the number of grades, where the PD increases with the rating grade, that is the best grade has the lowest number and the worst grade has the highest. The grade level PDs are then obtained by finding the point of the level on the fitted curve. These PDs are then adjusted by either of the above methods so that the weighted average PD is equal to the portfolio estimate of the PD.

### 7.1.3 Estimation and simulation

The estimations of the probabilities of default for the rating grades are computed in MATLAB using the above methods. This is executed in the following steps.

1. Estimate the portfolio PD, \( p_E \), using the BCR or Bayesian method presented in previous chapters, with input arguments of the asset correlation, \( \rho \), inter-temporal correlation, \( \tau \), and some confidence level, \( \gamma \).

2. Calculate the observed default rates by dividing the defaults per grade with the amount of obligors per grade over the whole time period.

3. Fit the observed default rates to a curve and find the values of the curve that matches the grade level. These are the fitted values of the PD, \( p_F \).

4. The observed default rates that are lower than 0.03% are replaced with this value and the weighted portfolio PD for the new PD values is calculated. If this is higher than the previously estimated portfolio PD this is the result, otherwise steps 5, 6 and 7 are executed on the observed default rates before the floor values are added.

5. When using a scale-factor the difference between the weighted portfolio PD and the estimated portfolio PD is

\[
f(y) = p_W - p_E = \sum_{i=1}^{g} \frac{n_i \cdot \max[p_{F_i}, y, 0.03%]}{n} - p_E
\]

and when using an additive term

\[
f(y) = p_W - p_E = \sum_{i=1}^{g} \frac{n_i \cdot \max[p_{F_i} + y, 0.03%]}{n} - p_E
\]

where \( y \) is the scaling factor respectively the additive term, \( n \) is the number of obligors, \( n_i \) is the number of obligors in the \( i \)th rating grade, \( p_{F_i} \) is the observed default rate for the \( i \)th obligor, \( p_E \) is the estimated portfolio PD and \( g \) is the number of grades.
6. Find \( y \) such that \( f(y) \) is close to zero using the following method.

   (a) Set the number of iterations to \( it = \log_2((y_{high} - y_{low})/\delta) \) where \([y_{low}, y_{high}]\) is the interval \( y \) is believed to be in and \( \delta \) is the accepted error.
   
   (b) For \( it \) number of iterations the midpoint, \( y_{mid} \), of the interval is calculated. It is then checked if \( f(y_{mid}) > 0 \) or \( < 0 \). If it is the first case the lower bound is set equal to the midpoint, in the second case the higher bound is set equal to the midpoint.
   
   (c) When \( it \) iterations are performed the estimated probability of default is set to the final midpoint.

7. The additive term or the scale-factor are either added or multiplied with each of the fitted grade level PDs and those under 0.03\% are replaced with the floor value.

The following algorithm describes the calculations in step 5 and 6 in pseudo-code.

**Algorithm 3** Grade level allocation

**Require:** An estimate of the portfolio PD.

\[
\text{set } f(y) = \sum_{i=1}^{g} \frac{n_i \cdot \max[p_{Fi} + y, 0.03\%]}{n} - p_E \\
\text{or set } f(y) = \sum_{i=1}^{g} \frac{n_i \cdot \max[p_{Fi} \cdot y, 0.03\%]}{n} - p_E
\]

for \( j = 1 \) to \( it \) do
    \( y_{mid} = (y_{high} - y_{low})/2 \)
    if \( f(y_{mid}) < 0 \) then
        \( y_{high} = y_{mid} \)
    else
        \( y_{low} = y_{mid} \)
    end if
end for

### 7.2 Grade level estimates over subportfolios

A possible scenario for a bank is to hold a portfolio that in turn consists of several subportfolios. Such a portfolio will in this thesis be called a main portfolio or an aggregated portfolio. An example of an aggregated portfolio is an international portfolio of exposures to obligors in several countries. The subportfolio’s grade level PDs can be calculated the same way as for the whole portfolio using the methods in explained in 7.2, that is using the fitting and scaling methods with an estimated portfolio PD for the subportfolio.

It is also possible to use the information in the aggregated portfolio to make an estimation of the subportfolios PD and their grade level PDs by assuming the same grade level PDs in the subportfolios as in the aggregated portfolio. That is simply setting the grade level subportfolio PDs equal to the grade level PD in the main portfolio. This can be motivated by that the grade level PDs are an estimation of the PDs for the obligors in each grade in the aggregated portfolio and the obligors in this portfolio are the same as in the subportfolios and they belong to the same grade in the subportfolio as in the aggregated portfolio.
A third method, that also uses the information from the aggregated portfolio, is to create scale-factors for each grade level in the subportfolios from information in the aggregated portfolio. These scale-factors are calculated from the observed default rate and the available grade level PDs in the aggregated portfolio, that is

\[ \text{scale-factor} = \frac{\text{grade level PD}}{\text{observed default rate}}. \]

This factor is calculated for each grade level and multiplied with the observed default rates in the subportfolios to calculate the final grade level PD. If the observed default rate is zero at some grade level the scale-factor is set to zero. After the calculations all grade level PDs that are below the floor value, that is 0.03%, are set to the floor value.
Chapter 8

Results

In this chapter, the results for the PD estimation and grade level allocation methods are presented. The results are simulated for three different low default portfolios as well as for different number of simulations and values of the correlations. A comparison of the results for the Bayesian and the BCR approach is also performed. Further on one of the low default portfolios consists of two subportfolios that are the underlying data for the results of grade level estimates over subportfolios.

8.1 Data selection

The results are simulated for three separate portfolios, one consisting of five year data and two of ten year data. All portfolios contain data of number of obligors and defaults for corporate portfolios. The data is presented on grade level for 7 different rating grades ranging from A to G, with A being the best grade and G the worst. Portfolio 1 is the same portfolio as in the original article by BCR, see [4], while Portfolio 2 and 3 consists of simulated data. Portfolio 3 is an aggregated portfolio and consists of two subportfolios, Subportfolio 3.1 and 3.2.

<table>
<thead>
<tr>
<th>Grade</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td>B</td>
<td>24</td>
<td>25</td>
<td>25</td>
<td>24</td>
<td>24</td>
<td>122</td>
</tr>
<tr>
<td>C</td>
<td>37</td>
<td>37</td>
<td>36</td>
<td>37</td>
<td>35</td>
<td>182</td>
</tr>
<tr>
<td>D</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>25</td>
<td>123</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>24</td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>99</td>
<td>103</td>
<td>101</td>
<td>99</td>
<td>98</td>
<td>500</td>
</tr>
</tbody>
</table>

Table 8.1: Number of obligors per rating grade at start of each year in Portfolio 1.
Table 8.2: Number of defaults per rating grade and year in Portfolio 1.

<table>
<thead>
<tr>
<th>Grade</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

The portfolio consists of 5-years of around 100 obligors, with a total of 4 defaults occurring. As is seen, defaults occur only in the three lowest rating grades, and with a maximum of one default per year. Since there are very few defaults, the portfolio is assumed to be LDP, as in the BCR article.

For Portfolio 2 and 3, the number of defaults and obligors per rating are presented in Tables 8.3 to 8.6.

Table 8.3: Number of obligors per rating grade and year in Portfolio 2.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Year 0</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
<th>Year 8</th>
<th>Year 9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>484</td>
<td>528</td>
<td>667</td>
<td>692</td>
<td>923</td>
<td>1133</td>
<td>1423</td>
<td>1435</td>
<td>1454</td>
<td>1570</td>
<td>10309</td>
</tr>
<tr>
<td>B</td>
<td>1448</td>
<td>1682</td>
<td>1817</td>
<td>2445</td>
<td>2500</td>
<td>2869</td>
<td>3686</td>
<td>4123</td>
<td>4296</td>
<td>5000</td>
<td>29866</td>
</tr>
<tr>
<td>C</td>
<td>3145</td>
<td>3155</td>
<td>3279</td>
<td>3296</td>
<td>3382</td>
<td>3392</td>
<td>3565</td>
<td>3584</td>
<td>3614</td>
<td>3750</td>
<td>34162</td>
</tr>
<tr>
<td>D</td>
<td>804</td>
<td>822</td>
<td>852</td>
<td>862</td>
<td>897</td>
<td>912</td>
<td>918</td>
<td>932</td>
<td>969</td>
<td>997</td>
<td>9065</td>
</tr>
<tr>
<td>E</td>
<td>827</td>
<td>940</td>
<td>1014</td>
<td>1178</td>
<td>1239</td>
<td>1291</td>
<td>1365</td>
<td>1421</td>
<td>1509</td>
<td>1925</td>
<td>12709</td>
</tr>
<tr>
<td>F</td>
<td>134</td>
<td>141</td>
<td>186</td>
<td>191</td>
<td>215</td>
<td>233</td>
<td>257</td>
<td>289</td>
<td>298</td>
<td>318</td>
<td>2262</td>
</tr>
<tr>
<td>G</td>
<td>49</td>
<td>59</td>
<td>66</td>
<td>117</td>
<td>141</td>
<td>153</td>
<td>172</td>
<td>206</td>
<td>223</td>
<td>227</td>
<td>1413</td>
</tr>
<tr>
<td>Total</td>
<td>6891</td>
<td>7327</td>
<td>7881</td>
<td>8781</td>
<td>9297</td>
<td>9983</td>
<td>11386</td>
<td>11990</td>
<td>12363</td>
<td>13887</td>
<td>99786</td>
</tr>
</tbody>
</table>

Table 8.4: Number of defaults per rating grade and year in Portfolio 2.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Year 0</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
<th>Year 8</th>
<th>Year 9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>19</td>
</tr>
</tbody>
</table>
## Table 8.5: Number of obligors per rating grade and year in Portfolio 3.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Year 0</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
<th>Year 8</th>
<th>Year 9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>484</td>
<td>528</td>
<td>667</td>
<td>692</td>
<td>923</td>
<td>1133</td>
<td>1423</td>
<td>1435</td>
<td>1454</td>
<td>1570</td>
<td>10309</td>
</tr>
<tr>
<td>B</td>
<td>1448</td>
<td>1682</td>
<td>1817</td>
<td>2445</td>
<td>2500</td>
<td>2869</td>
<td>3686</td>
<td>4123</td>
<td>4296</td>
<td>5000</td>
<td>29866</td>
</tr>
<tr>
<td>C</td>
<td>3145</td>
<td>3155</td>
<td>3279</td>
<td>3296</td>
<td>3382</td>
<td>3392</td>
<td>3565</td>
<td>3584</td>
<td>3754</td>
<td>35162</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>804</td>
<td>822</td>
<td>852</td>
<td>862</td>
<td>897</td>
<td>912</td>
<td>918</td>
<td>932</td>
<td>969</td>
<td>1097</td>
<td>9065</td>
</tr>
<tr>
<td>E</td>
<td>827</td>
<td>940</td>
<td>1014</td>
<td>1178</td>
<td>1239</td>
<td>1291</td>
<td>1365</td>
<td>1421</td>
<td>1509</td>
<td>1925</td>
<td>12709</td>
</tr>
<tr>
<td>F</td>
<td>134</td>
<td>141</td>
<td>186</td>
<td>215</td>
<td>233</td>
<td>257</td>
<td>289</td>
<td>298</td>
<td>318</td>
<td>2262</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>49</td>
<td>59</td>
<td>66</td>
<td>117</td>
<td>141</td>
<td>153</td>
<td>172</td>
<td>206</td>
<td>223</td>
<td>227</td>
<td>1413</td>
</tr>
<tr>
<td></td>
<td>6891</td>
<td>7327</td>
<td>7881</td>
<td>8781</td>
<td>9297</td>
<td>9983</td>
<td>11386</td>
<td>11990</td>
<td>12363</td>
<td>13887</td>
<td>99786</td>
</tr>
</tbody>
</table>

## Table 8.6: Number of defaults per rating grade and year in Portfolio 3.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Year 0</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
<th>Year 8</th>
<th>Year 9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>G</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>11</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>73</td>
</tr>
</tbody>
</table>

From the above, it can be concluded that for Portfolio 2 and 3, the number of obligors is significantly larger than for Portfolio 1, although the amount of defaults is still few in relation to the number of obligors. For Portfolio 2 the total amount of defaults from Table 8.4 is only 19 for a total of 99786 obligors. Portfolio 3 has a higher number of defaults, a total of 73 defaults for equally many obligors. It can also be noted that the distribution of the obligors across rating grades and years in the Portfolio 2 and 3 is identical. In this way, Portfolio 3 can be seen as a portfolio appropriate for robustness analysis of Portfolio 2, to see how results would change if there were more defaults. Furthermore, Portfolio 2 and 3 represent a more realistic real-world scenario for a bank or financial institution compared to Portfolio 1. For Portfolio 3 it can also be seen that the amount of defaults does not increase with the grade level. For example there are no defaults in grade level C while there are several in grade level A and B even though the amount of obligors in C is higher. This is of course strange since, as mentioned before, grade level A and B are supposed to be better than grade level C. All of the portfolios are assumed to be LDP.

### 8.1.1 Subportfolios

Portfolio 3 consists of Subportfolio 3.1 and 3.2 presented below.
### Table 8.7: Number of obligors per rating grade and year in Subportfolio 3.1

<table>
<thead>
<tr>
<th>Grade</th>
<th>Year 0</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
<th>Year 8</th>
<th>Year 9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>265</td>
<td>267</td>
<td>348</td>
<td>353</td>
<td>466</td>
<td>605</td>
<td>723</td>
<td>725</td>
<td>729</td>
<td>753</td>
<td>5234</td>
</tr>
<tr>
<td>B</td>
<td>745</td>
<td>925</td>
<td>1047</td>
<td>1264</td>
<td>1267</td>
<td>1453</td>
<td>2143</td>
<td>2466</td>
<td>2547</td>
<td>2739</td>
<td>16596</td>
</tr>
<tr>
<td>C</td>
<td>1473</td>
<td>1745</td>
<td>1755</td>
<td>1764</td>
<td>1827</td>
<td>1832</td>
<td>1854</td>
<td>1863</td>
<td>1872</td>
<td>1883</td>
<td>18138</td>
</tr>
<tr>
<td>D</td>
<td>412</td>
<td>420</td>
<td>429</td>
<td>431</td>
<td>457</td>
<td>461</td>
<td>463</td>
<td>465</td>
<td>486</td>
<td>593</td>
<td>4617</td>
</tr>
<tr>
<td>E</td>
<td>476</td>
<td>529</td>
<td>554</td>
<td>664</td>
<td>672</td>
<td>744</td>
<td>763</td>
<td>849</td>
<td>978</td>
<td>6919</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>78</td>
<td>80</td>
<td>103</td>
<td>107</td>
<td>111</td>
<td>126</td>
<td>136</td>
<td>162</td>
<td>168</td>
<td>171</td>
<td>1242</td>
</tr>
<tr>
<td>G</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>61</td>
<td>73</td>
<td>86</td>
<td>114</td>
<td>128</td>
<td>131</td>
<td>766</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3744</td>
<td>3996</td>
<td>4271</td>
<td>4644</td>
<td>4873</td>
<td>5250</td>
<td>6149</td>
<td>6558</td>
<td>6779</td>
<td>7248</td>
<td>53512</td>
</tr>
</tbody>
</table>

### Table 8.8: Number of defaults per rating grade and year in Subportfolio 3.1

<table>
<thead>
<tr>
<th>Grade</th>
<th>Year 0</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
<th>Year 8</th>
<th>Year 9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>36</td>
<td></td>
</tr>
</tbody>
</table>

### Table 8.9: Number of obligors per rating grade and year in Subportfolio 3.2

<table>
<thead>
<tr>
<th>Grade</th>
<th>Year 0</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
<th>Year 8</th>
<th>Year 9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>219</td>
<td>261</td>
<td>319</td>
<td>339</td>
<td>457</td>
<td>528</td>
<td>700</td>
<td>710</td>
<td>725</td>
<td>817</td>
<td>5075</td>
</tr>
<tr>
<td>B</td>
<td>703</td>
<td>757</td>
<td>770</td>
<td>1181</td>
<td>1233</td>
<td>1416</td>
<td>1543</td>
<td>1657</td>
<td>1749</td>
<td>2261</td>
<td>13270</td>
</tr>
<tr>
<td>C</td>
<td>1402</td>
<td>1410</td>
<td>1524</td>
<td>1532</td>
<td>1555</td>
<td>1560</td>
<td>1711</td>
<td>1721</td>
<td>1742</td>
<td>1867</td>
<td>16024</td>
</tr>
<tr>
<td>D</td>
<td>392</td>
<td>402</td>
<td>423</td>
<td>431</td>
<td>440</td>
<td>451</td>
<td>455</td>
<td>467</td>
<td>483</td>
<td>504</td>
<td>4448</td>
</tr>
<tr>
<td>E</td>
<td>351</td>
<td>411</td>
<td>460</td>
<td>514</td>
<td>567</td>
<td>601</td>
<td>621</td>
<td>658</td>
<td>660</td>
<td>947</td>
<td>5790</td>
</tr>
<tr>
<td>F</td>
<td>56</td>
<td>61</td>
<td>83</td>
<td>84</td>
<td>104</td>
<td>107</td>
<td>121</td>
<td>127</td>
<td>130</td>
<td>147</td>
<td>1020</td>
</tr>
<tr>
<td>G</td>
<td>24</td>
<td>29</td>
<td>31</td>
<td>56</td>
<td>68</td>
<td>70</td>
<td>86</td>
<td>92</td>
<td>95</td>
<td>96</td>
<td>647</td>
</tr>
<tr>
<td></td>
<td>3147</td>
<td>3331</td>
<td>3610</td>
<td>4137</td>
<td>4424</td>
<td>4733</td>
<td>5237</td>
<td>5432</td>
<td>5584</td>
<td>6639</td>
<td>46274</td>
</tr>
</tbody>
</table>

### Table 8.10: Number of defaults per rating grade and year in Subportfolio 3.2

<table>
<thead>
<tr>
<th>Grade</th>
<th>Year 0</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
<th>Year 6</th>
<th>Year 7</th>
<th>Year 8</th>
<th>Year 9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>G</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>37</td>
</tr>
</tbody>
</table>

From the above it can be seen for the subportfolios, as in the aggregated portfolio, Portfolio 3, the amount of defaults does not increase with the grade level. It is also worth noticing that the total amount of defaults in both of the subportfolios are
almost the same even though the amount of obligors in Subportfolio 3.1 is significantly larger than the amount of obligors in Subportfolio 3.2.

8.2 Estimates of PD in the BCR approach

For the BCR approach, the results are first simulated for the data in Table 8.1 and Table 8.2, considering different year-to-year correlations \( \tau \) and confidence levels \( \gamma \) for a varying number of simulations \( N \). In the simulations, the average number of total obligors per year in Table 8.1 is considered, i.e. 100. The number of defaults used in the simulations is the total number of defaults across all years, given by the total in Table 8.2. The results for Portfolio 1 are presented in Table 8.11.

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( \gamma )</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau )</td>
<td>30%</td>
<td>80%</td>
<td>30%</td>
<td>80%</td>
<td>30%</td>
</tr>
<tr>
<td>( N = 10000 )</td>
<td>1.07%</td>
<td>1.24%</td>
<td>1.68%</td>
<td>2.17%</td>
<td>3.03%</td>
</tr>
<tr>
<td>( N = 15000 )</td>
<td>1.07%</td>
<td>1.23%</td>
<td>1.68%</td>
<td>2.17%</td>
<td>3.05%</td>
</tr>
<tr>
<td>( N = 20000 )</td>
<td>1.07%</td>
<td>1.23%</td>
<td>1.69%</td>
<td>2.18%</td>
<td>3.02%</td>
</tr>
</tbody>
</table>

Apparent from the results, the PD estimate is very similar when increasing the number of simulations \( N \). Therefore, in general \( N = 10000 \) is a sufficient number of simulations to achieve reasonable results. Also evident is that the PD estimate is higher for the higher correlation of \( \tau = 80\% \) compared to \( \tau = 30\% \). Due to that systematic factors are more correlated in the case of \( \tau = 80\% \) and therefore if a default occurs it is more probable that another default occurs, this is of course a very intuitive result.

Increasing the confidence level from 75\% to 90\% induces a rather large increase in the resulting estimate. As argued in BCR [4], a confidence level of higher than 75\% seems to produce overly conservative results and therefore most commonly the 75\% confidence level is used.

Similarly, results for Portfolio 2 and 3 are simulated for changing values of \( \gamma \) and \( \tau \). Here, since \( N = 10000 \) seems to give sufficient results, only simulations of \( N = 10000 \) is considered. The average number of total obligors for Portfolio 2 and 3 is 9979. The amount of defaults can be found in Tables 8.4 and 8.6.

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( \gamma )</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau )</td>
<td>30%</td>
<td>80%</td>
<td>30%</td>
<td>80%</td>
<td>30%</td>
</tr>
<tr>
<td>Portfolio 2</td>
<td>0.02%</td>
<td>0.03%</td>
<td>0.04%</td>
<td>0.06%</td>
<td>0.07%</td>
</tr>
<tr>
<td>Portfolio 3</td>
<td>0.09%</td>
<td>0.11%</td>
<td>0.13%</td>
<td>0.19%</td>
<td>0.21%</td>
</tr>
</tbody>
</table>
As in the results for Portfolio 1, the results for the higher of the year-to-year correlation is consecutively higher for all number of simulations and portfolios. This is an expected result. Also, it can be concluded that the resulting PD estimate for Portfolio 3 is the highest for all values of the year-to-year correlation and confidence level and the estimate for Portfolio 2 is quite below 3 for all the values. This is also expected since the observed default rate, $\frac{\text{total number of defaults}}{\text{total number of obligors}}$, is highest for Portfolio 3. The total observed default rate for Portfolio 3 is 0.073% compared to 0.0019% for Portfolio 2. Also, as for the results for Portfolio 1, the shift from 75% to 95% confidence level significantly increases the PD.

8.3 Estimates of PD in the Alternative approach

Estimates of PD in the Alternative approach are considered for different numbers of simulations and choices of year-to-year correlation $\tau$ for both data sets. The results are simulated for an asset correlation of $\rho = 12\%$ and year-to-year correlation of both $\tau = 30\%$ and $\tau = 80\%$. The estimated PD is obtained through the process described in Section 6.2.5 by implementing a Gibbs sampler and taking the PD as a mean of samples from the posterior distribution for $p$. The uniform prior in the interval [0,1] is considered, although implementation is adapted so that more restricted intervals could be considered.

<table>
<thead>
<tr>
<th>N</th>
<th>10000</th>
<th>15000</th>
<th>20000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau$</td>
<td>30%</td>
<td>80%</td>
<td>30%</td>
</tr>
<tr>
<td>PD</td>
<td>1.68%</td>
<td>2.33%</td>
<td>1.73%</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.000994</td>
<td>0.0018</td>
<td>0.000774</td>
</tr>
</tbody>
</table>

The PD estimates are in the same size order as those for a confidence level of 75% in the results for the BCR approach, see Table 8.11. The estimated values for $\tau = 80\%$ are obviously higher than those for $\tau = 30\%$, which is also consistent with the previously attained BCR results and to be expected. Also, the standard deviation of different runs is larger for $\tau = 80\%$, meaning that the results differ more and are not as consistent between different runs of the algorithm.
8.3. Estimates of PD in the Alternative approach

**Figure 8.1:** Plot of the Markov chain for $c$ in Portfolio 1 for $N = 10000$, burn in of 3000 and $\tau = 30\%$ to the left and $\tau = 80\%$ to the right.

**Figure 8.2:** Plot of the Markov chain for $c$ in Portfolio 1 for $N = 15000$, burn in of 3000 and $\tau = 30\%$ to the left and $\tau = 80\%$ to the right.
Chapter 8. Results

From the appearance of the Markov chains for \( c = \Phi^{-1}(p) \) it is evident that for all three scenarios and both year-to-year correlations the chain after a fairly small amount of simulations oscillates around a somewhat constant mean. There are no large jumps in the chain, and it can thereby be concluded that the algorithm finds the stationary value for the Markov chain. Since the Markov chains are stationary the resulting PD from the Gibbs sampler is a reliable estimate for the probability of default. Also, for all scenarios the method reaches the stationary state after only a couple of hundred samples, why a burn-in of 3000 is sufficient for all of the simulations.

Furthermore, as the results vary more for the higher year-to-year correlation and the Markov chains for \( \tau = 80\% \) also show evidence of this in that they contain more spikes and larger fluctuation in values, it can be concluded that the results are somewhat less stable for \( \tau = 80\% \). As the standard deviation decreases with a large number of simulations and the Markov chain for \( N = 20000 \) also show less spikes, this indicates that a larger number of simulations should be used for the case of \( \tau = 80\% \). However, the results are still deemed to be a good estimate of the PD, since the Markov chain obviously still is stationary.

Next, considering the Alternative approach for Portfolio 2 and 3 a number of \( N = 50000 \) simulations is used. This is due to that a higher number of simulations gives a somewhat more satisfactory appearance of the Markov chain.

The results are shown in the following table.

Table 8.14: PD estimates for Portfolio 2 and 3 with \( N = 50000 \), burn-in sample of 10000, \( \rho = 12\% \).

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( \tau )</th>
<th>30%</th>
<th>80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portfolio 2</td>
<td>0.05%</td>
<td>0.12%</td>
<td></td>
</tr>
<tr>
<td>Portfolio 3</td>
<td>0.18%</td>
<td>0.43%</td>
<td></td>
</tr>
</tbody>
</table>
8.3. Estimates of PD in the Alternative approach

From the results, it is obvious that for Portfolio 2 and 3 the mixing is worse in comparison to Portfolio 1. It is also clear that the results are less reliable than for Portfolio 1, as the Markov chain shows evidence of not being stationary, especially for Portfolio 3. For Portfolio 2, the results are a little more satisfying as the Markov chain, at least in the case of $\tau = 30\%$, looks to be varying around a fairly constant mean. Unfortunately, from the appearance of the Markov chain for $c$, for the other cases the results are concluded to be an unreliable estimate of the true PD since the Markov chain shows of large fluctuations and a non-stationary behavior. Results were also simulated for $N = 100000$ simulations but since the appearance of the Markov chains was not more satisfactory these results are not included.
8.4 BCR and Bayesian comparison

To form an opinion of the performance of both the BCR and Bayesian Alternative approach, the models are compared to each other. As an initial comparison, the results for Portfolio 1 for confidence level 75% are presented in the table below. The estimates in the Bayesian approach are the PD as a mean of the samples from the posterior distribution.

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( \gamma )</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau )</td>
<td>30%</td>
<td>80%</td>
<td>30%</td>
<td>80%</td>
<td>30%</td>
</tr>
<tr>
<td>( N = 10000 )</td>
<td>1.68%</td>
<td>2.17%</td>
<td>1.68%</td>
<td>2.33%</td>
<td></td>
</tr>
<tr>
<td>( N = 15000 )</td>
<td>1.68%</td>
<td>2.17%</td>
<td>1.73%</td>
<td>2.48%</td>
<td></td>
</tr>
<tr>
<td>( N = 20000 )</td>
<td>1.69%</td>
<td>2.18%</td>
<td>1.74%</td>
<td>2.48%</td>
<td></td>
</tr>
</tbody>
</table>

Undoubtedly, the estimates are of the same size order and only differ with a maximum of 0.87%, for \( N = 10000 \) and \( \tau = 80\% \), which demonstrates that both methods produce reasonable results. From the results, the Alternative approach is the more conservative of the two approaches and consequently produces somewhat higher estimates of PD. Comparing with the additional results for the other confidence levels from Table 8.11, the Bayesian results are more conservative than the BCR results for a 50% and 75% confidence level, but less conservative than BCR estimates obtained at the 95% and 99% level.

For the Alternative approach, it is also meaningful to present the estimate of PD as the quantile of the posterior distribution at a certain confidence level, in order to compare with the BCR approach. The reasoning behind this is further presented in Section 9.2. The quantile estimates at confidence level \( \gamma \) for the corresponding PD estimates as mean of the posterior distribution from the above table are presented in the following table.

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( \gamma )</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau )</td>
<td>30%</td>
<td>80%</td>
<td>30%</td>
<td>80%</td>
<td>30%</td>
</tr>
<tr>
<td>( N = 10000 )</td>
<td>1.39%</td>
<td>1.93%</td>
<td>2.21%</td>
<td>3.06%</td>
<td>4.04%</td>
</tr>
<tr>
<td>( N = 15000 )</td>
<td>1.42%</td>
<td>1.99%</td>
<td>2.19%</td>
<td>3.25%</td>
<td>3.99%</td>
</tr>
<tr>
<td>( N = 20000 )</td>
<td>1.49%</td>
<td>1.94%</td>
<td>2.26%</td>
<td>3.22%</td>
<td>3.86%</td>
</tr>
</tbody>
</table>

In conclusion, the estimates obtained as quantiles of the sample from the posterior distribution are more conservative than those obtained as a mean of the same posterior for every confidence level except 50%.
For Portfolio 2 and 3, the only reliable result for the Alternative approach as seen in Section 8.3, is for Portfolio 2 with a year-to-year correlation of $\tau = 30\%$. Thereby, the only meaningful comparison with the BCR approach is for this result.

Table 8.17: Comparison of PD estimates for the BCR and Alternative Approach for Portfolio 2 and $\tau = 30\%$.

<table>
<thead>
<tr>
<th>Method</th>
<th>BCR</th>
<th>Alternative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portfolio 2</td>
<td>0.03 %</td>
<td>0.05 %</td>
</tr>
</tbody>
</table>

The Bayesian Alternative approach, as for Portfolio 1, is the more conservative of the approaches and produces a slightly higher estimate. Similarly to Portfolio 1, the alternative approach estimate is less conservative than estimates at 95% and 99% level, but more conservative than the 50% and 75% estimates.

Table 8.18: Estimates of PD for Portfolio 2 as the quantile at confidence level $\gamma$.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>12%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau$</td>
<td>30%</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>50%</td>
</tr>
<tr>
<td>Portfolio 2</td>
<td>0.05 %</td>
</tr>
</tbody>
</table>

The quantile estimates for the alternative approach with $\tau = 30\%$ for Portfolio 2, as for Portfolio 1, are more conservative for every confidence level except for 50% where the estimate is the same as the mean.

### 8.5 Grade level allocation of PDs

The grade level estimates of the PD are calculated with the BCR results of the PD with $N = 10000$, $\tau = 30\%$, $\rho = 12\%$ and $\gamma = 75\%$ for all of the portfolios. As presented in the theory four different methods are used though it can be seen from the data given that the linear fit seems quite unlikely but it is still shown as an example.

Beginning with the results for Portfolio 1 where the portfolio PD from the BCR approach is calculated to 1.68% from Table 8.11. The observed default rates are calculated with the total amount of obligors and defaults from Table 8.1 and 8.2, that is $\frac{\text{total number of defaults}}{\text{total number of obligors}}$.

Table 8.19: Estimation of grade level PD in Portfolio 1 with linear fitting.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Observed default rate</th>
<th>Fitted PD</th>
<th>Added floor values</th>
<th>Scaling with addition</th>
<th>Scaling with multiplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.00%</td>
<td>-4.33%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>B</td>
<td>0.06%</td>
<td>-1.29%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>C</td>
<td>0.00%</td>
<td>1.75%</td>
<td>1.75%</td>
<td>1.75%</td>
<td>1.75%</td>
</tr>
<tr>
<td>D</td>
<td>0.00%</td>
<td>4.79%</td>
<td>4.79%</td>
<td>4.79%</td>
<td>4.79%</td>
</tr>
<tr>
<td>E</td>
<td>4.17%</td>
<td>7.83%</td>
<td>7.83%</td>
<td>7.83%</td>
<td>7.83%</td>
</tr>
<tr>
<td>F</td>
<td>7.14%</td>
<td>10.87%</td>
<td>10.87%</td>
<td>10.87%</td>
<td>10.87%</td>
</tr>
<tr>
<td>G</td>
<td>22.22%</td>
<td>13.91%</td>
<td>13.91%</td>
<td>13.91%</td>
<td>13.91%</td>
</tr>
<tr>
<td>Portfolio</td>
<td>0.80 %</td>
<td>2.21%</td>
<td>2.75%</td>
<td>2.75%</td>
<td>2.75%</td>
</tr>
</tbody>
</table>
Chapter 8. Results

For the linear fitting the weighted portfolio PD with the added floor values is higher than the calculated portfolio PD from the BCR method. This results in that there is no need for any up-scaling so the additive term is 0 and the scale-factor is 1. The only difference between the linear fitted PD is that a floor value is added in grade A and B according to the previously mentioned Basel regulations.

Table 8.20: Estimation of grade level PD in Portfolio 1 with exponential fitting.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Observed default rate</th>
<th>Fitted PD</th>
<th>Added floor Values</th>
<th>Scaling with addition</th>
<th>Scaling with multiplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.00%</td>
<td>0.04%</td>
<td>0.04%</td>
<td>0.59%</td>
<td>0.06%</td>
</tr>
<tr>
<td>B</td>
<td>0.00%</td>
<td>0.12%</td>
<td>0.12%</td>
<td>0.67%</td>
<td>0.17%</td>
</tr>
<tr>
<td>C</td>
<td>0.00%</td>
<td>0.33%</td>
<td>0.33%</td>
<td>0.88%</td>
<td>0.49%</td>
</tr>
<tr>
<td>D</td>
<td>0.00%</td>
<td>0.95%</td>
<td>0.95%</td>
<td>1.5%</td>
<td>1.41%</td>
</tr>
<tr>
<td>E</td>
<td>4.17%</td>
<td>2.71%</td>
<td>2.71%</td>
<td>3.26%</td>
<td>4.03%</td>
</tr>
<tr>
<td>F</td>
<td>7.14%</td>
<td>7.75%</td>
<td>7.75%</td>
<td>8.3%</td>
<td>11.52%</td>
</tr>
<tr>
<td>G</td>
<td>22.22%</td>
<td>22.14%</td>
<td>22.14%</td>
<td>22.60%</td>
<td>32.92%</td>
</tr>
<tr>
<td>Portfolio</td>
<td>0.80%</td>
<td>1.13%</td>
<td>1.13%</td>
<td>1.68%</td>
<td>1.68%</td>
</tr>
</tbody>
</table>

For the exponential fitting there is a slightly lower value for the weighted portfolio PD for the fitted values than the BCR portfolio PD. There is also no need to add any floor values since all fitted PDs are greater than the floor value of 0.03 %. Since the weighted portfolio PD for the fitted portfolio is lower than the portfolio PD calculated with the BCR method it needs to be scaled upwards. The additive term is 0.55% and the scale factor is 1.4867. After the scaling it is seen that the weighted portfolio PD for the fitted values are the same as the estimated BCR portfolio PD. When using the additive term PDs all the grade level PDs are equally increased but when multiplication is used the difference between the different grade level PDs increases.

Figure 8.6 represents how the grade level PD changes from the observed default rate with both linear and exponential fitting by plotting the values in Table 8.19 and 8.20 against the grade levels, here grade levels 1-7 represents grade levels A-G. For the linear fitting the grade level PDs are a bit more conservative than the exponential fitting since the portfolio PD is higher and it can clearly be seen that the only change after the linear fitting is that the floor values are added. It is also seen that the linear fitting is a bit above the observed default rate in a majority of the grade levels whilst the exponential fitting follows the points more accurately. It can also be seen from the plots that for the linear fitting in grade level G the observed default rate is above the estimated PD in grade level G. For the exponential fitting it can be seen that for grade level E the observed default rate is higher than the estimated grade level PD, though not the same difference as in grade level G for the linear fitting.
Moving on to Portfolio 2, which is the portfolio with fewer defaults, the portfolio has a calculated portfolio PD, from the BCR method, of 0.03% as seen in Table 8.12.

Table 8.21: Estimation of grade level PD in Portfolio 2 with linear fitting.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Observed default rate</th>
<th>Fitted PD</th>
<th>Added floor values</th>
<th>Scaling with addition</th>
<th>Scaling with multiplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.01%</td>
<td>-0.08%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>B</td>
<td>0.00%</td>
<td>-0.02%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>C</td>
<td>0.06%</td>
<td>0.04%</td>
<td>0.04%</td>
<td>0.04%</td>
<td>0.04%</td>
</tr>
<tr>
<td>D</td>
<td>0.02%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>0.1%</td>
<td>0.1%</td>
</tr>
<tr>
<td>E</td>
<td>0.02%</td>
<td>0.16%</td>
<td>0.16%</td>
<td>0.16%</td>
<td>0.16%</td>
</tr>
<tr>
<td>F</td>
<td>0.22%</td>
<td>0.22%</td>
<td>0.22%</td>
<td>0.22%</td>
<td>0.22%</td>
</tr>
<tr>
<td>G</td>
<td>0.42%</td>
<td>0.28%</td>
<td>0.28%</td>
<td>0.28%</td>
<td>0.28%</td>
</tr>
<tr>
<td>Portfolio</td>
<td>0.02%</td>
<td>0.04%</td>
<td>0.06%</td>
<td>0.06%</td>
<td>0.06%</td>
</tr>
</tbody>
</table>

For the linear fitting no scaling is necessary since the weighted portfolio PD for the fitted PDs with added floor values is higher than 0.03%. Therefore the resulted weighted portfolio PD is also higher than the previously estimated portfolio PD.

Table 8.22: Estimation of grade level PD in Portfolio 2 with exponential fitting.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Observed default rate</th>
<th>Fitted PD</th>
<th>Added floor values</th>
<th>Scaling with addition</th>
<th>Scaling with multiplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.01%</td>
<td>0.00%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>B</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>C</td>
<td>0.06%</td>
<td>0.01%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>D</td>
<td>0.02%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>E</td>
<td>0.02%</td>
<td>0.07%</td>
<td>0.07%</td>
<td>0.07%</td>
<td>0.07%</td>
</tr>
<tr>
<td>F</td>
<td>0.22%</td>
<td>0.17%</td>
<td>0.17%</td>
<td>0.17%</td>
<td>0.17%</td>
</tr>
<tr>
<td>G</td>
<td>0.42%</td>
<td>0.44%</td>
<td>0.44%</td>
<td>0.44%</td>
<td>0.44%</td>
</tr>
<tr>
<td>Portfolio</td>
<td>0.02%</td>
<td>0.03%</td>
<td>0.04%</td>
<td>0.04%</td>
<td>0.04%</td>
</tr>
</tbody>
</table>

The exponential fitting does not either require an upscaling but the weighted portfolio PD is still lower than the one of the linear fitting. In Figure 8.7 it can be seen in both plots that the only thing changing from the fitted plot is the added floor values. As for Portfolio 1 it can be seen from the graphs that the observed default rate in grade level G in the linear fitting and in grade level F for the exponential fitting is higher than the estimated grade level PD for the grade in question.
Lastly the results for Portfolio 3 are presented. The estimated portfolio PD with the BCR method is 0.13%, found in Table 8.12.

Table 8.23: Estimation of grade level PD in Portfolio 3 with linear fitting.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Observed default rate</th>
<th>Fitted PD</th>
<th>Added floor values</th>
<th>Scaling with addition</th>
<th>Scaling with multiplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.02%</td>
<td>-0.45%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>B</td>
<td>0.01%</td>
<td>-0.14%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>C</td>
<td>0.00%</td>
<td>0.17%</td>
<td>0.17%</td>
<td>0.17%</td>
<td>0.17%</td>
</tr>
<tr>
<td>D</td>
<td>0.07%</td>
<td>0.48%</td>
<td>0.48%</td>
<td>0.48%</td>
<td>0.48%</td>
</tr>
<tr>
<td>E</td>
<td>0.09%</td>
<td>0.79%</td>
<td>0.79%</td>
<td>0.79%</td>
<td>0.79%</td>
</tr>
<tr>
<td>F</td>
<td>0.84%</td>
<td>1.10%</td>
<td>1.10%</td>
<td>1.10%</td>
<td>1.10%</td>
</tr>
<tr>
<td>G</td>
<td>2.34%</td>
<td>1.41%</td>
<td>1.41%</td>
<td>1.41%</td>
<td>1.41%</td>
</tr>
<tr>
<td>Portfolio</td>
<td>0.07%</td>
<td>0.16%</td>
<td>0.26%</td>
<td>0.26%</td>
<td>0.26%</td>
</tr>
</tbody>
</table>

As for previous portfolios there is no required upscaling of the linear fitting.

Table 8.24: Estimation of grade level PD in Portfolio 3 with exponential fitting.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Observed default rate</th>
<th>Fitted PD</th>
<th>Added floor values</th>
<th>Scaling with addition</th>
<th>Scaling with multiplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.02%</td>
<td>0.00%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>B</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>C</td>
<td>0.00%</td>
<td>0.08%</td>
<td>0.08%</td>
<td>0.10%</td>
<td>0.09%</td>
</tr>
<tr>
<td>D</td>
<td>0.07%</td>
<td>0.24%</td>
<td>0.24%</td>
<td>0.27%</td>
<td>0.28%</td>
</tr>
<tr>
<td>E</td>
<td>0.09%</td>
<td>0.75%</td>
<td>0.75%</td>
<td>0.77%</td>
<td>0.88%</td>
</tr>
<tr>
<td>F</td>
<td>2.34%</td>
<td>2.35%</td>
<td>2.35%</td>
<td>2.38%</td>
<td>2.77%</td>
</tr>
<tr>
<td>Portfolio</td>
<td>0.07%</td>
<td>0.10%</td>
<td>0.11%</td>
<td>0.13%</td>
<td>0.13%</td>
</tr>
</tbody>
</table>

For the exponential fitting some upscaling is required but because the calculated weighted portfolio fitting for the fitted grade level PDs are near the estimated portfolio PD, using the BCR method, the values are quite small as the additive term is 0.03% and the scale-factor is 1.18.

The results are plotted below in Figure 8.8. As for the two previous portfolios the observed default rate in grade level G for the linear fitting is higher than the estimated grade level PD in the same grade, the same goes for grade F using the exponential fitting with the additive term.
In conclusion, for all three portfolios the exponential fitting method is proven to give more adequate results. The percental change in the portfolio PD due to upscaling is larger for Portfolio 2 than for Portfolio 3. This is consistent with expectations, since Portfolio 2 has fewer number of defaults and hence fewer observations and therefore should require a larger margin of conservatism and a larger upscaling.

### 8.5.1 Grade level estimates of subportfolios

There are three methods to calculate the grade level estimates of the subportfolios described in the theory section. For the subportfolios the portfolio PD is calculated with the BCR method with a confidence level at 75%, asset correlation of 12% and year-to-year correlation of 30%. This gives the result of a portfolio PD at 0.12% for Subportfolio 3.1 and 0.14% for Subportfolio 3.2.

The first method presented is when calculating the grade level estimates the same way as for the main portfolio which gives the results below for the two subportfolios.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Observed default rate</th>
<th>Fitted PD</th>
<th>Added floor values</th>
<th>Scaling with addition</th>
<th>Scaling with multiplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.04%</td>
<td>-0.38%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>B</td>
<td>0.00%</td>
<td>-0.11%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>C</td>
<td>0.00%</td>
<td>0.16%</td>
<td>0.16%</td>
<td>0.16%</td>
<td>0.16%</td>
</tr>
<tr>
<td>D</td>
<td>0.07%</td>
<td>0.42%</td>
<td>0.42%</td>
<td>0.42%</td>
<td>0.42%</td>
</tr>
<tr>
<td>E</td>
<td>0.09%</td>
<td>0.69%</td>
<td>0.69%</td>
<td>0.69%</td>
<td>0.69%</td>
</tr>
<tr>
<td>F</td>
<td>0.81%</td>
<td>0.95%</td>
<td>0.95%</td>
<td>0.95%</td>
<td>0.95%</td>
</tr>
<tr>
<td>G</td>
<td>1.96%</td>
<td>1.22%</td>
<td>1.22%</td>
<td>1.22%</td>
<td>1.22%</td>
</tr>
<tr>
<td>Portfolio</td>
<td>0.07%</td>
<td>0.15%</td>
<td>0.23%</td>
<td>0.23%</td>
<td>0.23%</td>
</tr>
</tbody>
</table>

As for previous portfolios there is no required upscaling of the linear fitting.
Chapter 8. Results

Table 8.26: Estimation of grade level PD in Subportfolio 3.1 with exponential fitting.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Observed default rate</th>
<th>Fitted PD</th>
<th>Added floor values</th>
<th>Scaling with addition</th>
<th>Scaling with multiplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.04%</td>
<td>0.00%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>B</td>
<td>0.00%</td>
<td>0.01%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>C</td>
<td>0.00%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.05%</td>
<td>0.03%</td>
</tr>
<tr>
<td>D</td>
<td>0.07%</td>
<td>0.08%</td>
<td>0.08%</td>
<td>0.11%</td>
<td>0.10%</td>
</tr>
<tr>
<td>E</td>
<td>0.09%</td>
<td>0.24%</td>
<td>0.24%</td>
<td>0.26%</td>
<td>0.28%</td>
</tr>
<tr>
<td>F</td>
<td>0.81%</td>
<td>1.69%</td>
<td>0.71%</td>
<td>0.80%</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>1.96%</td>
<td>1.98%</td>
<td></td>
<td>2.00%</td>
<td>2.30%</td>
</tr>
<tr>
<td>Portfolio</td>
<td>0.07%</td>
<td>0.09%</td>
<td>0.11%</td>
<td>0.12%</td>
<td>0.12%</td>
</tr>
</tbody>
</table>

For the exponential fitting some upscaling is necessary and the additive term is 0.02% and the scale-factor 1.13.

The results are plotted below in Figure 8.9. This subportfolio has the same pattern as the other portfolios where the observed default rate is higher than the estimated grade level PD in grade level G for the linear fitting and grade level F when using the exponential fitting with the additive term.

Further on the second subportfolio of Portfolio 3 is looked upon, that is Subportfolio 3.2. As previously mentioned the portfolio PD is 0.14% when \( \rho = 12\% \), \( \tau = 30\% \) and \( \gamma = 75\% \).

Table 8.27: Estimation of grade level PD in Subportfolio 3.2 with linear fitting.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Observed default rate</th>
<th>Fitted PD</th>
<th>Added floor values</th>
<th>Scaling with addition</th>
<th>Scaling with multiplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.00%</td>
<td>-0.54%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>B</td>
<td>0.02%</td>
<td>-0.17%</td>
<td>0.03%</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>C</td>
<td>0.00%</td>
<td>0.18%</td>
<td>0.18%</td>
<td>0.18%</td>
<td>0.18%</td>
</tr>
<tr>
<td>D</td>
<td>0.07%</td>
<td>0.55%</td>
<td>0.55%</td>
<td>0.55%</td>
<td>0.55%</td>
</tr>
<tr>
<td>E</td>
<td>0.09%</td>
<td>0.91%</td>
<td>0.91%</td>
<td>0.91%</td>
<td>0.91%</td>
</tr>
<tr>
<td>F</td>
<td>0.88%</td>
<td>1.27%</td>
<td>1.27%</td>
<td>1.27%</td>
<td>1.27%</td>
</tr>
<tr>
<td>G</td>
<td>2.78%</td>
<td>1.64%</td>
<td>1.64%</td>
<td>1.64%</td>
<td>1.64%</td>
</tr>
<tr>
<td>Portfolio</td>
<td>0.08%</td>
<td>0.17%</td>
<td>0.29%</td>
<td>0.29%</td>
<td>0.29%</td>
</tr>
</tbody>
</table>

As for previous portfolios there is no required upscaling of the linear fitting.
Table 8.28: Estimation of grade level PD in Subportfolio 3.2 with exponential fitting.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Observed default rate</th>
<th>Fitted PD</th>
<th>Added floor values</th>
<th>Scaling with addition</th>
<th>Scaling with multiplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.03%</td>
<td>0.04%</td>
<td>0.03%</td>
</tr>
<tr>
<td>B</td>
<td>0.02%</td>
<td>0.00%</td>
<td>0.03%</td>
<td>0.04%</td>
<td>0.03%</td>
</tr>
<tr>
<td>C</td>
<td>0.00%</td>
<td>0.02%</td>
<td>0.03%</td>
<td>0.06%</td>
<td>0.03%</td>
</tr>
<tr>
<td>D</td>
<td>0.07%</td>
<td>0.07%</td>
<td>0.07%</td>
<td>0.10%</td>
<td>0.08%</td>
</tr>
<tr>
<td>E</td>
<td>0.09%</td>
<td>0.23%</td>
<td>0.22%</td>
<td>0.27%</td>
<td>0.29%</td>
</tr>
<tr>
<td>F</td>
<td>0.88%</td>
<td>0.80%</td>
<td>0.80%</td>
<td>0.84%</td>
<td>1.01%</td>
</tr>
<tr>
<td>G</td>
<td>2.78%</td>
<td>2.79%</td>
<td>2.79%</td>
<td>2.83%</td>
<td>3.54%</td>
</tr>
<tr>
<td>Portfolio</td>
<td>0.08%</td>
<td>0.10%</td>
<td>0.11%</td>
<td>0.14%</td>
<td>0.14%</td>
</tr>
</tbody>
</table>

For the exponential fitting the fitted grade level PDs are too low compared to the calculated portfolio PD calculated with the BCR method when using scaling the additive term is 0.04% and the scale-factor is 1.27.

The results are plotted below in Figure 8.10. For the linear fitting the observed default rate is above the estimated grade level PD in grade level PD G as all of the other calculated grade level PDs for grade level G when using the linear fitting.

![Figure 8.10: Plot of grade level allocation of PDs for Subportfolio 3.2.](image)

To check if this method is valid the weighted average of the two subportfolios is calculated and compared to the portfolio PD of Portfolio 3. The portfolio PD from the exponential fitting is used since the linear fitting is not a good fit to the data. The weighted average portfolio PD from the subportfolios is 0.13% which is the same as the portfolio PD for Portfolio 3.

If it is assumed that the grade level PDs of Portfolio 3 are the same in the subportfolios the portfolio PD for the subportfolios can be calculated by taking the weighted portfolio PD. The results below uses the values from the exponential fitting from both scaling with addition and multiplication. In the calculations the total amount of obligors during the time period is used from Table 8.7 and Table 8.9.
Table 8.29: Estimation of portfolio PD in Subportfolio 3.1 and 3.2.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Scaling with addition</th>
<th>Scaling with multiplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>B</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>C</td>
<td>0.05%</td>
<td>0.03%</td>
</tr>
<tr>
<td>D</td>
<td>0.10%</td>
<td>0.09%</td>
</tr>
<tr>
<td>E</td>
<td>0.27%</td>
<td>0.28%</td>
</tr>
<tr>
<td>F</td>
<td>0.77%</td>
<td>0.88%</td>
</tr>
<tr>
<td>G</td>
<td>2.38%</td>
<td>2.77%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Subportfolio 3.1</th>
<th>Subportfolio 3.2</th>
<th>Subportfolio 3.1</th>
<th>Subportfolio 3.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted PD Subportfolio 3.1</td>
<td>0.13%</td>
<td>0.13%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted PD Subportfolio 3.2</td>
<td>0.12%</td>
<td>0.12%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted PD Portfolio 3</td>
<td>0.13%</td>
<td>0.13%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In this approach both subportfolios have the same grade level PDs but different weighted portfolio PD. It differs from the first method where Subportfolio 3.1 had a smaller portfolio PD than Subportfolio 3.2.

For the third method the scale-factors are presented below when using the final grade level PDs for both the additive and multiplicative method when using exponential fitting. As explained before the scale-factor is set to zero when the observed default rate is zero for a grade level. These scale-factors are not to be confused with the scale-factors used in previous methods.

Table 8.30: Estimations of scale-factors

<table>
<thead>
<tr>
<th>Grade</th>
<th>Additive method</th>
<th>Multiplicative method</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.55</td>
<td>1.55</td>
</tr>
<tr>
<td>B</td>
<td>5.22</td>
<td>4.48</td>
</tr>
<tr>
<td>C</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>D</td>
<td>1.56</td>
<td>1.35</td>
</tr>
<tr>
<td>E</td>
<td>3.06</td>
<td>3.25</td>
</tr>
<tr>
<td>F</td>
<td>0.92</td>
<td>1.05</td>
</tr>
<tr>
<td>G</td>
<td>1.02</td>
<td>1.19</td>
</tr>
</tbody>
</table>

In the table below the final values of the grade level PDs are presented.

Table 8.31: Estimations of grade level PDs and weighted portfolio PD for subportfolios

<table>
<thead>
<tr>
<th>Grade</th>
<th>Additive method</th>
<th>Multiplicative method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subportfolio 3.1</td>
<td>Subportfolio 3.2</td>
</tr>
<tr>
<td>A</td>
<td>0.06%</td>
<td>0.03%</td>
</tr>
<tr>
<td>B</td>
<td>0.03%</td>
<td>0.08%</td>
</tr>
<tr>
<td>C</td>
<td>0.03%</td>
<td>0.03%</td>
</tr>
<tr>
<td>D</td>
<td>0.10%</td>
<td>0.11%</td>
</tr>
<tr>
<td>E</td>
<td>0.27%</td>
<td>0.26%</td>
</tr>
<tr>
<td>F</td>
<td>0.74%</td>
<td>0.81%</td>
</tr>
<tr>
<td>G</td>
<td>1.99%</td>
<td>2.83%</td>
</tr>
</tbody>
</table>

|                  |                  |                      |                  |                  |
| Weighted Subportfolio PD | 0.99%            | 0.12%                | 0.10%            | 0.13%            |
| Weighted Portfolio PD     | 0.12%            | 0.12%                |                  |                  |
For this result the grade level PD does not increase with the grade level. But as in the first method the calculated portfolio PD for Subportfolio 3.1 is lower than the one for Subportfolio 3.2.
Chapter 9

Discussion

The results from Chapter 8 are discussed in this chapter, as well as the models and their implications. Assumptions, advantages and disadvantages of the models are reviewed. Further, suggestions for future research are made.

9.1 Choice of parameters

As a starting point, the choice of values for the two parameters that are inputs to both the BCR and the Alternative approach, the asset correlation $\rho$ and the year-to-year correlation $\tau$ is motivated. In the Basel II framework an asset correlation between 12% and 24% is assumed, and the lower limit is chosen in the calculations in this thesis as in several other articles, among them Pluto & Tasche [21]. For the year-to-year correlation the PD is calculated for two different values, 30% and 80%. In Pluto & Tasche [21] it is argued that the correlation between default events over 5 years is close to 0% and with a choice of $\tau = 30\%$ the correlation over 5 years is $0.3^4 = 0.81\%$, which therefore is a reasonable choice. The higher value of $\tau = 80\%$ is also used when calculating the PD estimates to see how much a higher $\tau$ effects the results. Important to note is that these values are purely examples and that in a real situation a bank or financial institution would need to justify their choice of parameters based on empirical evidence depending on the available data for the specific portfolio in question.

Moreover, a choice of the confidence level needs to be made. In simulating the results four different confidence levels are tested, 50%, 75%, 95% and 99%. From table 8.17 and 8.18 it can be seen that PD increases with increasing confidence level and the increment is larger the higher the confidence level, e.g. between 50% and 75% the increment is smaller than between 95% and 99% even though the difference in confidence level in the first case is 25% and in the latter only 5%. From table 8.17 it can also be seen that for higher confidence levels the value of the portfolio PD is not as stable as for the lower values which is a reason to choose a lower $\gamma$. In both Pluto & Tasche [21] and BCR [4] it is proposed to use a value below 95% and this is also a reason for why $\gamma$ is chosen as 75% in the grade level allocation methods. As for previous parameters this choice is to be done from some expert opinion of the specific portfolio in question.


9.2 Model comparison

9.2.1 Assumptions in the BCR approach

An open and known issue in the BCR approach is that one needs to assume a confidence level. This value greatly affects the outcome of the results as seen in Chapter 8. The choice of confidence level is therefore a large model sensitivity.

Other parameter and model assumptions include that PD behaves according to the Vasicek model, and that the probability of default is equal across all rating grades and does not change over time, see Section 9.2.3. Also, in the approach the confidence region for the PD estimate is calculated using data of past years. In this sense, the estimated PD value is not predictive. Further the mean number obligors across all years of data is used in the simulations. A drawback of the model could therefore be that the mean is not close enough to the true number of obligors if the number of obligors in the portfolios varies greatly.

The BCR method allows for use of data for any number of years, although the most common choice is a 5-year period in accordance with Basel’s regulations. The period of time can be extended to a longer time period, however the model does not take into account where in the economic cycle the data is from. This needs to be accounted for in calculating the portfolio PDs and is a possible improvement of the approach.

9.2.2 Assumptions in the Bayesian approach

In the Bayesian approaches, the prior distribution is introduced as a guess on the value of the estimated parameter before the data has been observed. The need to specify a prior distribution is a requirement of Bayesian statistics, therefore the main assumption made in the Bayesian approach is the choice of the prior distribution. How to choose the prior distribution is an evident element of uncertainty present in the Bayesian approaches compared to other methods. The uniform prior considered in this thesis is one of many prior distributions considered by other authors. Kiefer suggests using prior distributions determined by expert judgment [16], and Tasche suggests the conservative prior [23], among others. The uniform prior is unsubjective and represents that we know nothing of the likely value of the parameter before the data has been observed. It is possible that this is an unlikely assumption, and that the uniform prior in comparison with other prior distributions does not give the most reliable results. Therefore in order to correctly evaluate the full potential of the Bayesian approach other prior distributions should be considered. Unfortunately, due to lack of time this is not done in this thesis, however it is a suggestion for further development, see Section 9.4.

In the Bayesian approach, the resulting PD estimate is derived from the posterior distribution, that is the distribution of the parameter given the observed data, $f_{\theta|L}(\theta|l)$. The estimate of the PD is by this a PD derived from historical data. As in the BCR approach, it is not a prediction of the future, and in this sense it is only an estimate based on what has already happened, unless the future is an exact repetition of the past. This is a possible criticism of both approaches, since what should be desired to estimate is a forward-looking PD. However, predicting the future could be very complicated in this setting since this normally requires a large amount of data, and data is scarce in LDPs. Also in general in modeling of financial risk, the estimates
are based on historical data and not on predictions of the future.

Lastly, it is also assumed that $p_{i,t} = p$ for all $i$ and $t$, see Section 9.2.3.

### 9.2.3 The Vasicek model for dependence of defaults

As previously mentioned, the Vasicek model introduces dependence between defaults by a systematic factor representing the market risk, and an idiosyncratic representing the company’s specific risk, both to which all obligors are equally correlated. Both the correlation between the systematic factors in consecutive years and the asset correlation are fixed. A noticeable weakness of the model is thus that correlations can not vary neither between different obligors nor in different years. This is of course not persistent with the real world. For example, the asset value of two firms in the same industry should in reality be more highly correlated than that of two random firms, a feature which can not be achieved in this model.

Furthermore it is also assumed in the model that both the idiosyncratic and systematic factor are normally distributed and that they are independent of each other. This is a convenient assumption since the normal distribution is easy to sample from, however it might be a poor representation of a real world scenario.

In the derivation of the multi-period model, it is assumed that $p_{i,t} = p$ for all $i$ and $t$. Thus, it is assumed that obligors in the same rating category have the same probability of default, but also that the probability of default is constant over time. The first is consistent with the aim of estimating a portfolio-wide PD, however the second assumption is of course unrealistic since the probability of default in reality can change from year to year depending on many different factors, for example an economic downturn.

The Vasicek model is a simple model, and can be argued to not be very sophisticated. However, it allows for modeling of the key feature of dependence between defaults, although in a simple manner, and is widely used and accepted in the industry.

### 9.2.4 Model differences

The first and fundamental difference between the BCR and Bayesian approaches is of course that in a Bayesian context, the parameters are regarded as stochastic while in the BCR approach they are regarded as fixed. Thereby, in a Bayesian approach one can incorporate a belief of the parameter through a prior distribution, as a form of previous experience, which is not possible in the BCR approach. This enables more flexibility in the method since almost any distribution could in theory be chosen for the prior on $p$, depending on the belief of the parameter’s true value. This can be very useful when information that is relevant is available, for example if a bank has access to expert information.

Another model difference is that the Bayesian approach avoids the choice of confidence level when estimating the PD as the mean from the sampled values of the posterior distribution. Thus, the BCR approach requires a choice of an extra input parameter, the confidence level, in order to produce PD estimates. Although, as discussed next, in order to produce comparable results for the method, one should also specify a confidence level for the Bayesian approach, and in this context the number
of parameters that needs to be chosen beforehand is the same.

Furthermore, it is important to realize that the two methods resulting PD estimates are representation of different things. In the Bayesian setting, the probability of default is the mean of the sample of \( p \) from its posterior distribution \( f_{P|L}(p, l) \) simulated through a Gibbs sampler, and is thereby a point estimate of the sought probability. In the BCR approach based on the Confidence Based approach, however the PD estimate is the upper bound of a confidence level for the probability of default. In this sense, the two PD estimates are not equivalent. The Bayesian estimates does not intuitively correspond to a conservative estimate in the same way as in the BCR approach. Nonetheless, the results show that the Alternative approach in fact produces more conservative estimates than the BCR approach.

In light of this comparison, resulting PD estimates as quantiles of the posterior distribution are produced in addition to the estimates as the mean of the sample from the posterior in Section 8.3. The quantile estimates are the corresponding Bayesian estimates to the estimates as upper confidence bounds in the BCR approach. Applying this reasoning, the estimates of PD are comparable in the sense that they represent the same thing. In order to estimate the PD as the quantile of the posterior, a confidence level must be specified, and hence with this argumentation, the Bayesian approach depend upon both a choice of prior distribution and of the confidence level. This can be argued to introduce a too large element of uncertainty in the approach, and one might therefore prefer the BCR approach due to this fact.

### 9.3 Grade level allocation of PDs

#### 9.3.1 Grade level estimates

For the grade level estimates there are four different approaches presented. These are linear or exponential and addition or multiplication in four different combinations. Beginning with the linear fitting it can be seen that this method is quite conservative since in all four portfolios the weighted portfolio PD is higher than the estimated portfolio PD from the BCR approach. This is probably because the observed default rate does not follow a linear function very well, since for most of the portfolios there are barely no defaults in the higher grades. There is also a strong increase in defaults between grade level E and F, so the linear function lies above the observed default rate in many of the rating grades, that is all of the rating grades except for in the lowest grade level, G, in all of the portfolios. Therefore it can be concluded that a linear function is not a good fit. The exponential fit follows the curve of the observed default rates more accurately and is also less conservative than the linear approach according to the results, when comparing with the weighted portfolio PDs. For the exponential fit the increase in PD between the grades is larger the lower the grade, so there is a very small difference in the PDs between the best grades. Another remark for the exponential fit is that the observed default rate in some grades are higher than the estimated grade level PDs. This is also the case for grade level G in the linear fitting, as mentioned before. This is worth mentioning because in some way the methods purpose is to make the PDs more conservative. But since it is clearly seen from the plots that the linear fit does not follow the data as well as the exponential fit the latter is recommended between the two. Other fits could also be tested, as for example polynomials, a quadratic fit was tested though it provided negative values
for the lowest grade level PDs and was therefore ruled out.

Moving on to the differences in the two scaling methods, that is adding a term or multiplying with a factor. Since for the results in this thesis upscaling only is necessary when using the exponential fitting these are the cases that will be discussed. When using the addition, all of the grades are equally increased so it can be seen as quite conservative for the higher grades if these are close to zero. For example in portfolio 1 in grade level A the final PD, 0.59%, is more than 10 times the estimate for the fitted PD, that is 0.04%. If multiplication where to be used instead it is the lower grades that will increase the most, at least if the amounts that are added are looked upon. So for this situation it again depends on what seems reasonable for the portfolio that is discussed, but for the portfolios in this thesis it seems more accurate to choose the scaling with multiplication since when using addition the results are too conservative compared to the available data in the lower grades.

Issues that can be discussed is the fact that there is no data of defaults in the lower grades and therefore the estimates are built on very little or no information. The choice of fitting should depend on the underlying data. For example if the time period looked upon is during a crisis in the economic cycle and there are more defaults in the lower grades the linear fitting could be the better choice. That is if the PD increases more in the lower grades compared to how the exponential fitting would adjust to the data and the observed default rates seems to follow a linear function, though that scenario is quite unlikely. But as previously mentioned, for the data presented in this report the exponential fitting is a better choice.

9.3.2 Grade level estimates over subportfolios

For the first approach when calculating the subportfolios grade level PD with the same method as for the aggregated portfolio the conclusions are the same, with a recommendation of using the exponential fitting. In the results it can be seen that the weighted portfolio PD of the subportfolios is the same as the calculated PD of the aggregated portfolio with the BCR method which further gives some confirmation of the method. As mentioned before this method gives a higher portfolio PD for Subportfolio 3.2 than for Subportfolio 3.1.

Moving on to the second method, that is assuming the same grade level PDs in the subportfolios as in the aggregated portfolio. As seen in the results the weighted portfolio PD for Subportfolio 3.1 is higher than for Subportfolio 3.2 which differs from the first method. This is because this method does not take into account the amount of defaults in each subportfolio but only the amount of obligors in each grade. With the assumption that the grade level PDs are the same in the aggregated portfolio information can be lost if for example the obligors with the highest real PD value are moved to one subportfolio and the ones with lowest real PD value are moved to the other, then the whole grade level should be recalculated. So in conclusion this method does not use all of the information available.

For the last method both calculations for the aggregated portfolio and all the data available from the subportfolios are used. In the final results the grade level PDs do not increase with the grade level, that is they do not increase from the highest to the lowest grade, which contradicts the original credit rating. The weighted portfolio PD from the subportfolios is also lower than the calculated portfolio PD with the BCR
Chapter 9. Discussion

method for the aggregated portfolio which also implies that this method may not be valid.

From this argumentation and the results the first method is recommended even though it does not use any of the results from the aggregated portfolio in its calculations it can be seen that the weighted portfolio PD is the same as the calculated portfolio PD for the aggregated portfolio. Further criticism of the two latter methods are that they are not as conservative as the calculated PDs for the subportfolios using the BCR method, perhaps it would be reasonable to scale the grade level values as previously towards the calculated portfolio PD.

9.4 Suggestions for further research

In this thesis, two main models to estimate PD in low default portfolio were implemented and compared. The Bayesian model proposed had a new feature of an additional factor being sampled. In the Bayesian setting, only a uniform prior for the distribution of the parameter $p$ before the data is observed was considered. A natural extension of this work is therefore to consider several different priors and how these perform, both in comparison with each other and with the BCR approach. Suggestions for priors that could be considered are both objective and subjective priors, for example beta-distributions or extreme value distributions. This is expected to give a more complete picture of the performance of the Bayesian approaches in the context of estimating PD in low default portfolios, and could improve the results of the Alternative approach. However, it must be taken into consideration that for other priors than the uniform the result will not be a truncated normal distribution as in the case considered in this thesis. The sampling of $c$ will therefore be more troublesome, a proposition is therefore to consider only a limited interval and a rejection sampling method.

Further, a possible improvement of the Bayesian method proposed in this thesis is to implement a Metropolis-Hastings algorithm instead of the Gibbs sampler. Other possible extensions of this thesis is to incorporate dependence of other factors than the systematic and idiosyncratic factor in the model, and to more thoroughly study the choices for asset and year-to-year correlation as well as the definition of LDPs. Potential external factors that could improve the model are country and industry factors, as well as dynamic factors to represent for example business cycle. Further, a suggestion of improvement of the Vasicek model would be to allow varying correlations between assets and with the systematic factors to better represent the real world. Other models for dependence between defaults could also be considered.

Both the Bayesian and BCR approach estimate PD on a portfolio level which requires a grade level allocation method. A suggestion for further improvement is to modify the Alternative approach to instead estimate PD for each rating grade. Such a method would be more complex in calculations, but would not be overly complicated compared to the present method, and would eliminate the need to spread a portfolio estimate on to rating grades. Further, a method to calculate the PD on grade levels proposed by Nabil Iqbal & Syed Afraz Ali [13] is to use convolution to combine two probability distributions, the binomial distribution and the Poisson distribution, to produce a new distribution and where the model incorporates the relationship between the rating grades. This could be further investigated and compared to the
grade level allocation methods in this thesis.
Chapter 10

Summary and conclusion

The purpose of this thesis was to assess models for estimating probability of default in low default portfolios. To achieve this aim, models were compared on three different data sets of low default portfolios.

The models for PD estimation studied in this thesis were the BCR model along with Bayesian models, where sampling from posterior densities with rejection methods was found to be too complicated in order to produce an efficient method. Instead, a new Bayesian Alternative approach in which an additional component is sampled was implemented. This approach was successful in producing reliable estimates for Portfolio 1, but unfortunately did not produce as solid results for Portfolio 2 or Portfolio 3 in the case of a higher year-to-year correlation. A suggestion to improve the results through Metropolis Hastings sampling was made. In comparing the resulting PD estimates for Portfolio 1, the Alternative approach produced somewhat more conservative results.

Moreover, four different methods to allocate portfolio estimates of PD to grade levels were considered. The exponential fitting proved to give better results whereas the additive and multiplicative methods produced more similar results. For a portfolio consisting of subportfolios three suggestions on methods to allocate portfolio PDs to grades were made, where the first approach using the same method as the aggregated portfolio was shown to give the best results.

The conclusion of this thesis is that estimation of PD in low default portfolios is a complex problem with many issues. The Bayesian approach proposed in this thesis shows promise, but requires further improvement and extension to be fully evaluated. The BCR approach is an already widely used method which produces reliable results also for portfolios of many obligors, though both methods have significant drawbacks. Ultimately, further research on the subject is required.
Appendix A

More on Markov chains

The following theory of Markov chains follows the notation in [30]. For a more complete theory see Appendix A of [19] and Chapter 1 and 2 of [18].

Definition A.0.1. Transition density
The transition density \( q \) of \( X_k \) is the density of the distribution of \( X_{k+1} \) given \( X_k = x \),

\[
P(X_{k+1} \in A | X_k = x_k) = \int_A q(x_{k+1} | x_k) dx_{k+1}
\]

A distribution \( g(x) \) is said to be stationary if it satisfies local or global balance. The global balance condition is given by

\[
\int q(z|x)g(x)dx = g(z)
\]

The local balance condition is given by

\[
\lambda(x)q(z|x) = \lambda(z)q(x|z), \forall x, z \in X.
\]

If a Markov chain starts in the stationary distribution, the chain will always stay in the stationary distribution.

Definition A.0.2. Ergodicity
A Markov chain \( X_n \) with stationary distribution \( g \) is called ergodic if for all initial distributions \( \chi \),

\[
\sup_{A \subseteq X} |P(X_n \in A) - g(A)| \to 0 \text{ as } n \to \infty
\]

Theorem A.0.1. Geometric Ergodicity
Assume that there exists a density \( \mu(x) \) and a constant \( \epsilon \) such that for all \( x, z \in X \)

\[
q(z|x) \geq \epsilon \mu(z).
\]

Then the chain \( X_n \) is geometrically ergodic, that is there is \( \rho < 1 \) such that for all \( \chi \)

\[
\sup_{A \subseteq X} |P(X_n \in A) - g(A)| \leq \rho^n
\]
Appendix A. More on Markov chains

Geometric ergodicity means that the Markov chain forgets its initial distribution geometrically fast. For geometrically ergodic Markov chains, the states are only weakly dependent. For such Markov chains, there exists a Law of Large Numbers.

**Theorem A.0.2. Law of Large Numbers for Markov Chains**

Let $X_n$ be a geometrically ergodic Markov chain with stationary distribution $g$. Then for all $\epsilon > 0$,

$$P\left( \frac{1}{n} \sum_{k=1}^{n} \phi(X_k) - \int \phi(x) g(x) dx \geq \epsilon \right) \to 0 \text{ as } n \to \infty.$$

The Law of Large Numbers for Markov chains makes it possible to estimate expectation

$$\tau = E(\phi(X)) = \int_{X} \phi(x) f(x) dx$$

by simulating in $N$ steps a Markov chain $X_k$ with stationary distribution $f$ and letting

$$\tau_N = \frac{1}{N} \sum_{k=1}^{N} \phi(X_k) \to \tau \text{ as } N \to \infty.$$


[8] The European Unions Capital Requirements, Constitution 575/2013, Section 4, Article 160


[22] Tasche, D. (2009), *Estimating discriminatory power and PD curves when the number of defaults is small*, Lloyd’s Banking Group


[28] Venter, E. S. (2016), *Probability of Default Calibration for Low Default Portfolios: Revisiting the Bayesian Approach*, University of Stellenbosch
