Probability of Default Calibration for Low Default Portfolios: Revisiting the Bayesian Approach

by

Edward S. Venter

Thesis presented in partial fulfilment of the requirements for the degree of Master of Commerce in Financial Risk Management in the Faculty of Economics and Business Sciences at Stellenbosch University

Department of Statistics and Actuarial Sciences,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Supervisor: Prof. W.J. Conradie

March 2016

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature: ....

E.S. Venter

Date: ....March 2016.........

Abstract

Probability of Default Calibration for Low Default Portfolios: Revisiting the Bayesian Approach

E.S. Venter

Department of Statistics and Actuarial Sciences,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Thesis: MComm (Financial Risk Management)

December 2015

The Probability of Default is one of the fundamental parameters used in the quantification of credit risk. When estimating the Probability of Default for portfolios with a low default nature the Probability of Default will always be underestimated. Therefore, a need exists for calibrating the Probability of Default for Low Default Portfolios.

Various approaches have been considered in the literature review, with the main approaches being the Confidence Based Approach and Bayesian Approach. In this study the Bayesian Approach for calibrating the Probability of Default for portfolios of high grade credit is reconsidered. Two alternative prior distributions that can be used in the Bayesian Approach are proposed; these are an informative, Strict Pareto distribution and a non-informative Jeffreys prior. The performance of these proposals are then compared to existing calibration techniques by using real data.

Uittreksel

Die Waarskynlikheid van Wanbetaling vir Lae Wanbetaling Portefeuljes: 'n Heroorweging van die Bayesiaanse Benadering

E.S. Venter
Department of Statistics and Actuarial Sciences,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Tesis: MCom (Finansiële Risikobestuur)
Desember 2015

Die Waarskynlikheid van Wanbetaling is een van die fundamentele parameters in die beraming van kredietrisiko. Wanneer die Waarskynliheid van Wanbetaling beraam word vir 'n portefeulje met lae wanbetaling observasies in die historiese data, vind onderberaming altyd plaas. Dus bestaan daar 'n nood vir kalibrasie tegnieke vir die Waarskynlikhied van Wanbetling vir Lae Wanbetaling Portefeuljes.

'n Verskeidenheid van benaderings word in die literatuur voorgestel, waaronder die Vertroue Gebasseerde Benadering en die Bayesiaanse Benadering die bekendste is. In hierdie studie word die Bayesiaanse Benadering vir die kalibrasie van die Waarskynlikheid van Wanbetaling vir portefeuljes van hoë vlak krediet heroorweeg. Twee alternatiewe apriori verdelings word voorgestel om in die Bayesiaanse Benadering te gebruik. Hierdie apriori verdelings is die streng Pareto verdeling wat 'n inligting-gewende apriori verdeling is en die Jeffreys apriori verdeling wat 'n nie-inligting-gewende apriori verdeling is. Die prestasie van die tegnieke wat voortvloei uit die gebruik van die voorgenoemde twee apriori verdelings word dan vergelyk met bestaande kalibrasie tegnieke deur gebruik te maak van werklike data.

Acknowledgements

Firstly, I would like to acknowledge my supervisor, Professor W.J. Conradie, for his guidance throughout this research project and his inspiration througout the duration of my studies. Thank you for introducing me to the world of quantitative financial risk management. It was a pleasure working with and learning from you.

Also, Dr. JD van Heerden, thank you for the insight and knowledge you shared with me during my time at the University of Stellenbosch.

And, Dr. Tom Berning, thank you for giving me critical pointers as well as helping me with the MATLAB code used in the simulation study.

Furthermore, I would like to extend my sincerest gratitude to…

My guardian and friend, Estienne Pretorius, for his loving support and friendship.

My brother, Arnold Venter, for sticking with me through thick and thin.

My aunt and uncle, Charmagne and Christo Crous, for their support and for being my second family.

My grandmother, Ouma Nellie Venter, for her caring love, support and for always keeping me on my toes.

My girlfriend, Lara Kriege, for her love, support, motivation and friendship.

The gentleman of the quantitative risk management team of Deloitte LLP in London. Specifically Tim Thompson, Thomas Clifford and Krisztian Sebesteyn. Thank you for introducing me to the topic whilst on secondment in July 2013. The standard of work you guys produce is a true inspiration.

And finally, thanks to all my friends, specifically Wilhelm, Marthinus, Cornel and Ben, for their support and for all the good memories in Stellenbosch. You guys are awesome, cheers to all the memories made.

Hierdie werkstuk en graad word opgedra aan my ouers.

"*The power of equations lies in the philosophically difficult correspondence between mathematics, a collective creation of human minds, and an external physical reality. Equations model deep patterns in the outside world. By learning to value equations, and to read the stories they tell, we can uncover vital features of the world around us . . .*"

Ian Stewart, *Emeritus Professor of Mathematics at the University of Warwick, England.*

Contents

List of Figures

List of Tables

Nomenclature

Acronyms

AIVG	Basel Committee Accord Implementation Group's Validation Subgroup
AIRB	Advanced Internal Rating Based
BIPRU	Prudential Sourcebook for Banks, Building Societies and Investment Firms
BCBS	Basel Committee on Banking Supervision
EAD	Exposure at Default
FIRB	Foundation Internal Rating Based
GUI	Graphical User Interface
HPS	Highest Posterior Density
HSM	Half-Sample Mode
IRB	Internal Ratings Based
LDP	Low-Default Portfolio
LGD	Loss Given Default
PD	Probability of Default
RR	Recovery Rate
RWA	Risk Weighted Assets
S&P	Standard and Poor's

Mathematical Variables

κ	Maturity Adjustment
A	Company Asset Value
b	Default state
C	Credit rating
c	Number of Credit Rating Categories
D	Company Debt Market Value

DD	Distance to Default
d	Number of Defaults
σ	Volatility
E	Shareholder's Equity
M	Effective Maturity
EL	Expected Loss
F	Company Debt Face Value
K	Capital Requirement
L	Loss statistic
LR	Leverage Ratio
p	Probability of Default
Φ	Standard Normal Distribution Function
r_F	Risk-Free Rate
s	Credit portfolio size
τ	Intertemporal Correlation
ξ	Asset Correlation

Chapter 1

Introduction

> "*The United States can pay any debt it has because we can always print money to do that. So there is zero probability of default.*"
> **Alan Greenspan**, *Chairman of the Federal Reserve of the United States from 1987 to 2006.*

The Oxford English Dictionary defines a *default* as the failure to fulfil an obligation; in the financial sense this essentially is the failure to repay a loan. This leads to *credit risk* being defined as the risk of losses being incurred as a result of a borrower defaulting on his obligation to repay outstanding debt. One of the core inputs in managing credit risk is the Probability of Default (PD), i.e. the probability of a borrower failing to meet his/her financial obligation in repaying their debt. The PD is a measure of a borrowers credit quality and the accuracy of PD estimations has a direct relationship to credit risk model results.

A predominant challenge in determining the PD is the low number of defaults that is associated with good credit. This is especially the case for good credit rating grades as higher credit rating grades may experience years without any default observations. Even if a few defaults occur this would result in PD estimates being small and volatile over time. Such Low Default Portfolios (LDPs) account for a large share of total bank lending and these portfolios introduce numerous challenges for proper credit risk management.

Consider the following very simple example of a credit portfolio made up of 1 000 obligors that has experienced no defaults in the past year. If one attempts to follow an empirical calculation of the Probability of Default by dividing the total number of default observations by the total number of obligors in the portfolio, then the resulting PD is zero. This essentially means that there is complete certainty that none of the obligors in the portfolio will default on their debt. As the 2008 financial crisis highlighted when one of the largest financial institutions in the world, Lehman Brothers, went bankrupt; one can never be 100% certain that any individual or financial institution would be able to fulfil their debt obligations. Therefore, there is no such thing as a zero Probability of Default.

As stated by Tasche (2013), the Basel Committee on Banking Supervision (BCBS, 2005) took the challenges that low default portfolios introduce into account when paragraph 4 of the Basel II framework was published: "*In general, estimates of the Probability of Default, Loss Given Default, and Exposure at Default are likely to involve unpredictable errors. 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 errors. Where methods and data are less satisfactory and the likely range of errors is larger, the margin of conservatism must be larger*". It is, therefore, clear that in order to determine the PD of these Low Default Portfolios, calibration techniques are required.

One of the earliest approaches for incorporating a margin of conservatism in calibrating the PD estimates for low default portfolios is the Confidence Based Approach suggested by Pluto and Tasche (2011). This methodology, which is widely employed by banks, is based on using upper confidence bounds and a *most prudent estimation* approach. Some alternatives have been proposed in the form of rating systems or score functions for low default portfolios, see Erlenmaier (2011); Kennedy *et al.* (2012); Fernandes and Rocha (2011).

From all of these early propositions for calibrating the PD for LDPs, the Confidence Based Approach appears the favourite despite its criticism of generating overly conservative estimates. Tasche (2013) contributes this favouritism to the UK FSA's requirement, as stated in the Prudential sourcebook for Banks, Building Societies and Investment Firms (BIPRU, 2011), that: "*a firm must use a statistical technique to derive the distribution of defaults implied by the firm's experience, estimating PDs (the 'statistical PD') from the upper bound of a confidence interval set by the firm in order to produce conservative estimates of PDs...*".

Forrest (2005) and Benjamin *et al.* (2006) proposed certain adjustments to the Pluto and Tasche (2011) approach in order to address the inherent conservatism underlying the approach. These adjustments were also intended to facilitate the application of the approach. However, of all the alternatives and/or improvements proposed to the confidence based approach; the use of Bayesian Approaches seems the most promising as it eliminates the subjectivity present in selecting an appropriate upper confidence bound. Using Bayesian Approaches in estimating the PD for LDPs was first considered by Dwyer (2006), where the use of an Uniform prior distribution was proposed. These approaches was later explored in greater detail in Kiefer (2009), Kiefer (2010), and Kiefer (2011) where using prior distributions determined by expert judgement was considered.

The expert judgement considered by Kiefer is in the form of using a Beta distribution as a prior distribution; albeit this overcomes the subjectivity of the Confidence Based Approach, a new source of subjectivity is introduced in the form of the expert opinion. Tasche (2013) revisited the Bayesian approach, introducing a prior distribution in the form of a so called conservative prior that supposedly addresses the subjectivity issue. However, this approach receives some criticism in the context of its appropriateness in the LDP setting.

Clifford *et al.* (2013) considered an alternative method for incorporating expert judgement into the Bayesian setting by proposing a so called expert prior distribution. This approach once again received strong criticism for being overly subjective. Other authors that considered the application of the Bayesian approach are Chang and Yu (2014) and Kruger (2015). Clearly the Bayesian approach has received much attention and can be regarded as a valid method in addressing the problem of estimating the PD for high grade credit.

With this background in mind, the purpose of this study is to:

- Discuss in detail the theoretical background of PD estimation.
- Introduce and reconsider the problem of PD calibration for LDPs.
- Provide a detailed theoretical discussion of two of the main PD calibration methodologies, namely the Confidence Based Approach and the Bayesian Approach.

- Propose additional distributions as possible prior distributions for the Bayesian Approach.
- Evaluate and compare the performance of the PD calibration techniques considered in this study using real data.
- Illustrate the significance of the PD input in credit risk management.
- Indicate open questions for further research.

As an additional output of this study, the significance of using graphical user interfaces in the programming of complex financial models is illustrated.

In order to conduct a meaningful evaluation of PD calibration techniques for LDPs, it is essential that the importance of the PD input in credit risk management and its estimation in general is understood.

Chapter 2 is dedicated to this purpose. In this chapter the importance of credit risk management within financial risk management is highlighted in Section 2.1. The fundamental concepts used in this study, such as credit risk and the Probability of Default are defined in this section. The theoretical background underlying credit risk management is discussed in detail in Section 2.2. This section is divided into four important components which cover fundamental concepts such as Credit Ratings, the Expected Loss Function, Credit Risk Regulation and the Probability of Default. The aim of this chapter is to lay the foundation for Chapter 3 and 4, where PD calibration for LDP will be considered.

In **Chapter 3** the theory behind PD calibration for LDPs are discussed in detail. Starting off by defining what constitutes an LDP and considering background of these portfolios. Before discussing the calibration of the PD some of the industry concerns regarding LDPs are discussed. The main component of this chapter is discussing the theory behind PD calibration. The calibration techniques are divided into two main categories:

- The Confidence Based Approach, and
- The Bayesian Approach.

In the discussion on the Bayesian Approach the following distributions are considered as prior distributions; the Uniform distribution, the Beta distribution, the Conservative distribution and the Expert distribution. The aim of this chapter is to provide a theoretical background of PD calibration for LDPs.

Chapter 4 is a short but important chapter in this study. In the first part of this chapter critical comments are made on the prior distributions discussed in Chapter 3. In the second part it is illustrated that by considering a certain transformation of the PD parameter, an array of alternative distributions can be considered for PD calibration. The Strict Pareto distribution is then defined and discussed as an alternative to the prior distributions covered in Chapter 3.

In **Chapter 5**, all of the theory discussed in Chapters 2, 3 and 4 is applied. This chapter constitutes the empirical results and comparison of all of the PD calibration approaches discussed in Chapters 3 and 4. However, before the empirical results are discussed a comment is made on the Jeffreys prior as an alternative prior distribution in the Bayesian approach. In the first major section of Chapter 5, the simulation procedure that is applied in this study is discussed in detail. In the second section the PD calibration approaches are compared using real and fictitious portfolio data. The models are compared with regards to the generated empirical estimates, and the model sensitivities to asset correlation and intertemporal correlation inputs. The different prior distributions that can be used in the Bayesian approach are also compared with regards to their highest posterior density widths. In order to illustrate the practical significance of the PD estimate, risk weighted asset capital requirements are calculated using the PD estimates obtained in the empirical study. Finally, the models are discussed in comparison.

Chapter 6 is an additional chapter that lies outside of the direct focus of the study, but it should be stated that it contributes practical significance nonetheless. In chapter 6 the implementation of graphical user interfaces in the design of financial models is illustrated. It is shown that these models improve the accessibility of complex models and an argument of good practice is made for the design of graphical user interface models in quantitative finance.

In **Chapter 7**, the work carried out is summarised; the main contributions are highlighted and areas for further research is suggested.

Chapter 2

Literature Review: Credit Risk Management

"*All institutions, regardless of size, must resist the temptation to under-invest in the systems and controls they need to prevent greater risk and larger losses in the future.*"
Thomas Curry, *30th Comptroller of the Currency of the United States*

In this chapter the first component of the literature study is considered. The focus of this chapter is credit risk management. All the necessary theory required to better understand the problem of estimating the probability of default for low default portfolios is discussed in this chapter. The focus is to lay the theoretical foundation required to better understand the proposed PD calibration techniques discussed in Chapter 3.

In the first section risk management is discussed in general, the discussion attempts to paint credit risk management in the bigger picture of financial risk management. The second section is the focus of the chapter; in this section credit risk management techniques are discussed. The credit risk management section is divided into four important subsections. In the first subsection an overview of credit ratings is given, highlighting the scarcity in default observations for high-grade debt. In the following two subsections on expected loss and credit risk regulation the importance of estimating the probability of default is highlighted. In the final subsection the fundamental theory regarding estimation of the probability of default is discussed.

2.1 Financial Risk Management

The 1600's can be recognised as the early birth of modern risk management. This era marks the discovery of one of the most powerful risk management tools ever to be invented: the laws of probability. This powerful tool introduced a way of quantifying uncertainty (Bernstein, 1996).

Uncertainty is synonym to the word risk. The word risk originates from the Italian word *Riscare*, which means to dare. In this sense, risk is a choice rather than a fate (Bernstein, 1996). Extending this line of thought, the Concise Oxford English Dictionary defines the word risk as "hazard, a chance of bad consequences, loss or exposure to mischance" (McNeil *et al.*, 2015).

McNeil *et al.* (2015) gives the following definition for financial risk; "any event or action that may adversely affect an organization's ability to achieve it's objectives and execute it's strategies" or alternatively "the quantifiable likelihood of loss or returns that are lower than expected". It is further pointed out that in most cases people only associate downside with risk, forgetting that risk also includes the possibility of gain. In a financial context, risk has a strong relationship with return; therefore, companies seek risk.

This is especially the case when one considers for example a bank. Banks are considered a safeguard where individuals deposit their money. A part of the banks business model is to loan out a fraction of the holdings to other individuals or companies that require financing. Both the individuals trusting the bank with their money and the bank itself when approving a loan are taking on risk.

The individual with a deposit at the bank faces the risk of losing their deposit if the bank runs out of business. A common misconception is that banks will never go bankrupt. However, these unlikely events do occur. Take for example the 2008 financial crisis when the fourth largest bank in the United States, Lehman Brothers, filed for bankruptcy. Or as an example in the South African market; the recent bailout of African Bank Ltd.

The bank on the other hand faces the risk of never recouping the outstanding loan amounts. As it is part of its business a bank knows that it will never recover all the outstanding credit amounts. However it is fundamentally important for the bank to limit these losses. Therefore, it is important to understand and quantify the risks faced.

In order to better understand the nature of financial risks, it is critical to understand what the three main financial risk categories are. Firstly, consider the most well-known type of financial risk, called *market risk*.

Definition 2.1. ***Market Risk***: *Market risk is the risk that a change in the level of one or more market prices of commodities, equities, interest rates, credit instruments, foreign exchange, or other market factors will result in losses for a trading position or portfolio.*

The next important category, and also the focus of this research project, is called *credit risk*.

Definition 2.2. ***Credit Risk***: *Credit risk is defined as the risk of losses incurred as a result of a borrower defaulting on his obligation to repay outstanding debt.*

Examples of outstanding debt may be in the form of loans, bonds or other debt instruments (McNeil *et al.*, 2015). It is further noted in Meissner (2009) that there are essentially two parts of credit risk, namely *default risk* and *credit deterioration risk*.

Definition 2.3. ***Default Risk***: *Default risk is defined as the risk that a borrower does not repay part of, or his entire, financial obligation.*

In the event of a default the lender would only receive the amount recovered from the borrower, this is called the recovery rate.

Definition 2.4. ***Credit Deterioration Risk***: *Credit deterioration risk on the other hand is defined as the risk that the underlying borrowers credit quality decreases.*

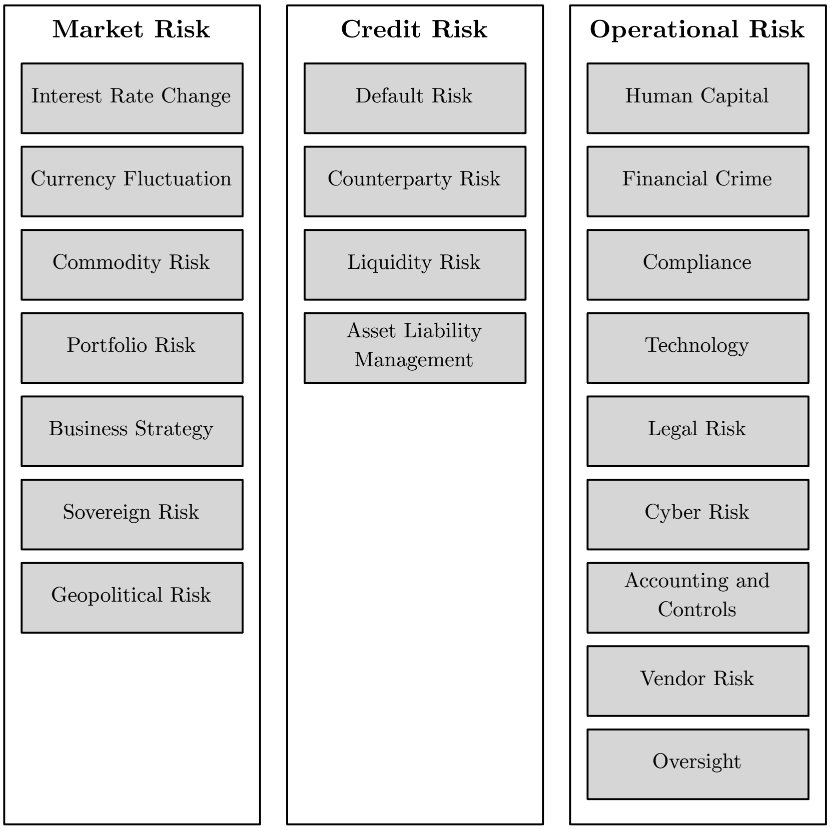


Figure 2.1: The Three Pillars of Financial Risk Management

Such a decrease would result in a decrease in the value of the assets of the lender. Subsequently this will result in financial loss. It is important to note that there is a relationship between default risk and credit deterioration risk since large credit deterioration is equivalent to default.

The third main financial risk category, which is somewhat disregarded at times, is *operational risk*.

Definition 2.5. ***Operational Risk***: *Operational risk is defined as the risk of financial loss resulting from either external events or failures in internal processes, people, and systems.*

The respective scopes of the three main financial risk categories are not always clearly defined. In Figure 2.1 an attempt is made to provide a break down of the three main risk categories by sorting the predominant financial risk factors into the three main categories. Note that, as described in McNeil *et al.* (2015) there exists certain notions of risk that are present in all three categories.

Aforementioned two main broad concepts are *liquidity risk* and *model risk*. Liquidity risk is a primary concern during periods of financial distress, it can be roughly translated into the risk derived from the shortage of marketability of an investment that subsequently cannot be bought or sold quickly enough to prevent or minimize an expected loss. Model risk is broadly defined as the risk of using an inappropriate model for measuring risk.

2.2 Credit Risk Management

Credit risk was defined in the previous section, it is intuitive that the simple banking example from Section 2.1 relates to this risk factor. In banking credit risk is a predominant issue as it is accounted for in nearly all parts of the banking book. The importance of this risk category however extends far beyond the banking sector. As an example consider the 2008 financial crisis, this is also known as the credit crisis. The main driver underlying the credit crisis was excessive risk taking on sub-prime loans and credit derivative instruments such as Collateralised Debt Obligations and Credit Default Swaps. For a comprehensive overview of the mechanics behind the crisis see Jarvis (2012).

A more recent example is the European Debt Crisis that has been confronting several Euro member countries since the end of 2009. This crisis followed the Credit Crisis of 2008 due to the excessive borrowing of Euro zone members. The European Debt crisis has the potential to bring global markets to a stand still due to the joint nature of the Euro economies. For an explanation of the mechanics behind this crisis see Jarvis (2015). These examples are extreme examples of what the mismanagement of debt can lead to.

If one considers an entire banking portfolio, it should be clear that losses due to the default of a borrower is part of the daily business of a bank. Hence, banks started to insure their debt. Insurance of this kind resulted in the birth of credit risk management (Bluhm *et al.*, 2010). Naturally, insurance is only one means for managing risk. Therefore, the credit risk exposures faced by an institution can also be managed by performing loss control, loss financing or internal risk reduction. Probably the most important financial risk management tool is having proper capital buffers in place to absorb losses. This is one of the essential drivers of Africa Bank's downfall; the bank had been providing loans that were not backed by assets.

Regardless of the risk management tool under consideration, all of the aforementioned risk management techniques require proper quantification of risk. One method used to capture the risk of credit exposures is credit ratings. This is discussed in the next section.

2.2.1 Credit Ratings

Credit ratings capture the creditworthiness of obligors or issuers of credit. Taking both quantitative and qualitative information into account, external rating bodies evaluate clients and assign a rating reflecting the grade of the client's credit. The issuer of the debt appoints a credit rating agency to assign the underlying debt instrument or the issuer itself a credit rating. Intuitively the aim of the borrower is to obtain the highest possible rating as the rating affects the interest rates that can be achieved by the company. Even though the rating body is appointed by the issuer, the agency needs to remain objective in it's opinion.

The three main external rating bodies are Moody's, Standard and Poor's (S&P) and Fitch. It is important to bear in mind that the ratings issued by these agencies is generally not an investment recommendation for a given security. Crouhy *et al.* (2001) gives the following extracts from *S&P's Corporate Rating Criteria* and *Moody's Credit Rating Research* that captures the central idea behind credit ratings. Firstly, consider the words of S&P:

> "*A credit rating is S&P's opinion of the general creditworthiness of an obligor, or the creditworthiness of an obligor with respect to a particular debt security or other financial obligation, based on relevant risk factors.*"

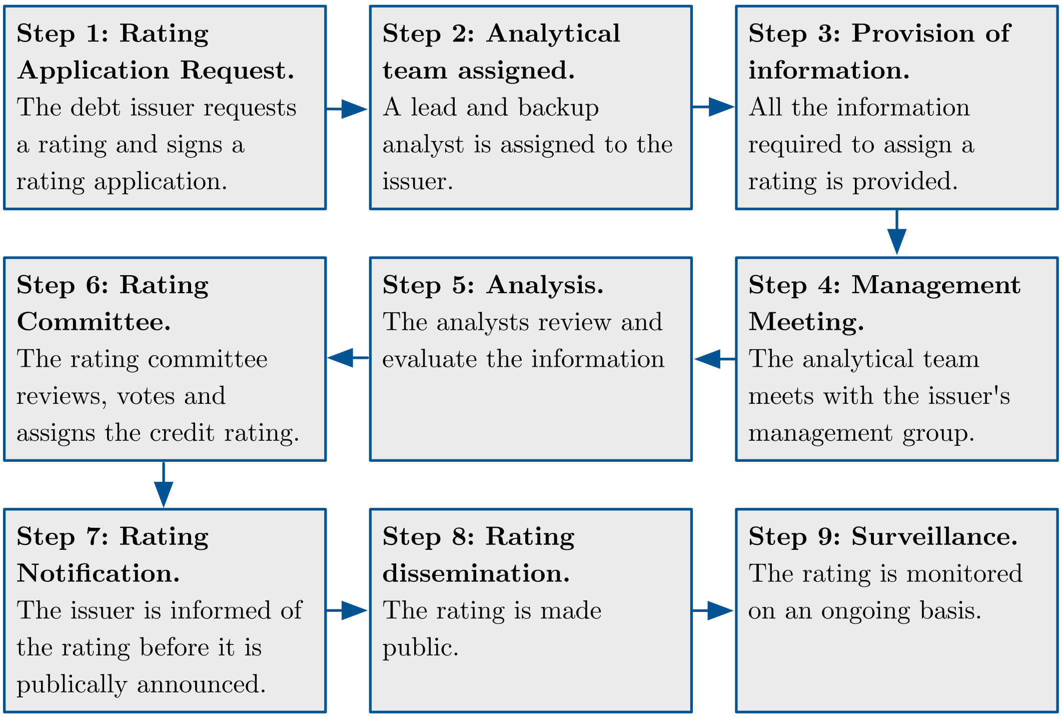


Figure 2.2: The Process of Obtaining a Credit Rating, adapted from *Moody's (2015)*

Moody's states the following regarding credit ratings:

> "*... an opinion on the future ability and legal obligation of an issuer to make timely payments of principal and interest on a specific fixed income security. Moody's ratings of industrial and financial companies have primarily reflected default probability, while expected severity of loss in the event of default has played an important secondary role. In the speculative-grade portion of the market, which has been developing into a distinct sector, Moody's ratings place more emphasis on expected loss than on relative default risk.*"

These external credit rating bodies are regarded as unbiased evaluators, and their ratings are widely accepted by market participants and regulatory agencies. Although, the 2008 financial crisis resulted in some questions regarding the trustworthiness and transparency of the ratings.

Nevertheless, credit ratings play an integral role in the management of credit risk. The external companies responsible for assigning a credit rating to a obligor have established an ordered scale of ratings in the form of a letter system describing the creditworthiness of rated companies. The three major rating agencies however implement a sightly different scale. The respective rating structures is compared in table 2.1, this is adapted from Hull (2012).

The process of mapping these rating grades to probability of default estimates is called calibration. This mapping essentially captures the probability of an obligor moving from a specific rating grade, e.g. an AA S&P rating, to a default within the risk horizon.

Table 2.1: Comparison of Rating Grades of Different Rating Agencies

Rating Description	Moody's	S&P	Fitch
Prime	Aaa	AAA	AAA
High grade	Aa1	AA+	AA+
	Aa2	AA	AA
	Aa3	AA-	AA-
Upper medium grade	A1	A+	A+
	A2	A	A
	A3	A-	A-
Lower medium grade	Baa1	BBB+	BBB+
	Baa2	BBB	BBB
	Baa3	BBB-	BBB-
Non-investment grade	Ba1	BB+	BB+
	Ba2	BB	BB
	Ba3	BB-	BB-
Highly speculative	B1	B+	B+
	B2	B	B
	B3	B-	B-
Substantial risks	Caa1	C	C
	Caa2	CCC	CCC
	Caa3	CCC-	CCC-
Extremely speculative	Ca	CC	CC
Default imminent	Ca	C	C
In default	C	RD	DDD
	/	SD	DD
	/	D	D

Table 2.2: Global Corporate Annual Default Rates as Percentages By Rating Category for S&P Ratings

Year	AAA	AA	A	BBB	BB	B	CCC/C
1981	0.00	0.00	0.00	0.00	0.00	2.27	0.00
1982	0.00	0.00	0.21	0.34	4.22	3.13	21.43
1983	0.00	0.00	0.00	0.32	1.16	4.58	6.67
1984	0.00	0.00	0.00	0.66	1.14	3.41	25.00
1985	0.00	0.00	0.00	0.00	1.48	6.47	15.38
1986	0.00	0.00	0.18	0.33	1.31	8.36	23.08
1987	0.00	0.00	0.00	0.00	0.38	3.08	12.28
1988	0.00	0.00	0.00	0.00	1.05	3.63	20.37
1989	0.00	0.00	0.18	0.60	0.72	3.38	33.33
1990	0.00	0.00	0.00	0.57	3.57	8.56	31.25
1991	0.00	0.00	0.00	0.55	1.69	13.84	33.87
1992	0.00	0.00	0.00	0.00	0.00	6.99	30.19
1993	0.00	0.00	0.00	0.00	0.70	2.62	13.33
1994	0.00	0.00	0.14	0.00	0.27	3.08	16.67
1995	0.00	0.00	0.00	0.17	0.99	4.58	28.00
1996	0.00	0.00	0.00	0.00	0.44	2.91	8.00
1997	0.00	0.00	0.00	0.25	0.19	3.49	12.00
1998	0.00	0.00	0.00	0.41	0.81	4.62	42.86
1999	0.00	0.17	0.18	0.20	0.95	7.28	33.33
2000	0.00	0.00	0.27	0.37	1.14	7.65	35.96
2001	0.00	0.00	0.27	0.33	2.93	11.45	45.45
2002	0.00	0.00	0.00	1.01	2.86	8.13	44.44
2003	0.00	0.00	0.00	0.23	0.58	4.02	32.73
2004	0.00	0.00	0.08	0.00	0.43	1.44	16.18
2005	0.00	0.00	0.00	0.07	0.31	1.72	9.09
2006	0.00	0.00	0.00	0.00	0.30	0.81	13.33
2007	0.00	0.00	0.00	0.00	0.20	0.25	15.09
2008	0.00	0.38	0.39	0.49	0.80	4.06	26.73
2009	0.00	0.00	0.22	0.54	0.73	10.80	48.94
2010	0.00	0.00	0.00	0.00	0.57	0.84	22.52
2011	0.00	0.00	0.00	0.07	0.00	1.63	16.06
2012	0.00	0.00	0.00	0.00	0.29	1.53	26.97
2013	0.00	0.00	0.00	0.00	0.09	1.60	23.42

The historic default frequencies for the years 1981 to 2013 is given in Table 2.2. The data given in this table is the global corporate annual default rates by rating category for S&P Ratings Services (Standard and Poor's, 2014). The descriptive statistics for the data in the aforementioned table is given in Table 2.3. From the descriptive statistics it is clear that for the highest rating grade, there is no default experience, however does this truly mean that the probability of default for AAA rated securities are zero? Also, note that the average default frequency for AA rated securities is 0.02%, however this increased to 0.38% during the financial distress of 2008. On the other hand, the historical default frequencies for the CCC rated securities remained close to the long-term average in 2008.

The results indicate the need for some calibration mechanism for the probability of default estimates of higher-grade or low-default securities. Later on in the literature study some of the available calibration techniques will be discussed in detail. However, before considering calibration of the probability of default it is necessary to understand the importance of this estimate in credit risk management. In the next section the expected loss function, the principle equation in credit risk management, is discussed.

Table 2.3: Descriptive Statistics On One-Year Global Default Rates as Percentages

(%)	AAA	AA	A	BBB	BB	B	CCC/C
Minimum	0.00	0.00	0.00	0.00	0.00	0.25	0.00
Maximum	0.00	0.38	0.39	1.01	4.22	13.84	48.94
Weighted long-term average	0.00	0.02	0.07	0.21	0.80	4.11	26.87
Average	0.00	0.02	0.06	0.23	0.98	4.61	23.76
Median	0.00	0.00	0.00	0.17	0.72	3.49	23.08
Standard deviation	0.00	0.07	0.11	0.26	1.03	3.34	12.10
2008 default rates	0.00	0.38	0.39	0.49	0.80	4.06	26.73

2.2.2 Expected Loss

Examples such as the one described in Section 2.2 illustrates the need for credit risk management. An important point already mentioned is that even good customers have the potential to default. This is also the case for the biggest companies or countries in the world. As an example Lehman Brother's has already been mentioned. Furthermore, note that since the late 90's there was a tremendous increase in Sovereign defaults with Greece defaulting twice in one year. Therefore banks not only face the need to manage the risk of the "bad" debt, but on all debt. As seen in the previous section, evaluation of the "good" debt poses problems. This is due to the lack of loss data (see the higher rating grades in Table 2.2). Before adressing this problem, it is necessary to understand the fundamentals of credit risk managemet.

As mentioned in Bluhm *et al.* (2010), banks are required to charge an appropriate risk premium for every loan issued. These premiums should then be pooled into an internal bank account, called the *expected loss reserve*. This reserve provides a capital buffer for the losses arising from defaults. The question now is, how does one determine an adequate capital buffer without historical loss data?

It is known that the risk premium on each individual loan should cover it's expected loss. The expected loss intuitively is the expected or mean value of the loss on the loan. As discussed in Van Gestel and Baesens (2009) the expected loss is dependent on three factors;

- the default risk of the borrower,

- the percentage of the loss in the event of a default and
- the exposure of the loan when a default is experienced.

Therefore the loss for a given time horizon can be expressed by the following stochastic variable:

$$Loss = EAD \times LGD \times L, \tag{2.2.1}$$

where the EAD is the Exposure at Default, LGD is the Loss Given Default and L is the default risk or default indicator. These elements are be formally defined below.

Definition 2.6. ***Exposure at Default (EAD)***: *This is defined as the exposure subject to be lost in the period under consideration given a default has occured.*

The EAD is regarded as a random or deterministic variable, where the random element is most important for credit cards and liquidity lines (Altman, 2006)(Van Gestel and Baesens, 2009).

Definition 2.7. ***Loss Given Default (LGD)***: *This is defined as the economic loss in the event of a default.*

When a default occurs a certain portion of the outstanding amount is recovered. This is called the recovery rate (RR) . To better understand the recovery rate, think of it in the context of bonds; the RR is defined as the bond's market value a few days after default as a percentage of its face value (Hull, 2012). Therefore the LGD is approximately equal to 1 minus the RR. Determining the LGD is however quite challenging as the recovery rates depend on many drivers such as the quality of collateral (securities, mortgages, guarantees, etc.), and on the seniority of the bank's claim on the borrower's assets (Dahlin and Storkitt, 2014).

Note that LGD is a random variable lying between 0% and 100%. Furthermore, due to the mathematical relationship between the LGD and the RR, the RR is also considered a random variable. Schuermann (2004) provides the following notes regarding the LGD:

- The RR as a percentage of the exposure is either high (approximately 70-80%) or low (approximately 20-30%). Therefore the loss distribution can be assumed to follow a bimodal distribution. This means that working with an average LGD can be deceiving.
- A determing factor of the recovery rate is the position of underlying claim in the capital structure. It follows that bonds usually have lower recovery rates than bank loans since bonds are lower down on the capital structure.
- During times of recession recovery rates are systematically lower, in other words losses are higher.
- The industry in which the obligor operates influences the recovery rate. Service sector companies typically have lower recovery rates than tangible asset-intensive industries.
- The size of the exposure appears to have no effect on the observed losses.

Finally, the default risk of an obligor is defined. The default event is defined as in paragraph 452 of Basel II, issued by the Basel Committee on Banking Supervision (BCBS, 2004), as follows.

Definition 2.8. ***The Basel II Definition of Default:*** *A default is considered to have occurred with regard to a particular 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 recourse by the bank to actions such as realising 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.*

In addition consider the words of S&P for a more general definition of a default, it is regarded as the failure to meet a principle or interest payment on the due date contained in the original terms of the debt issue. This results in the following intuitive definition of the probability of a default.

Definition 2.9. ***Probabilty of Default (PD)***: *The probability of default can broadly be defined as the probability of either or both of the events in the Basel II definition being realised.*

Taking above two definitions into consideration. Define a random variable L such that:

$$L = \begin{cases} 1 & \text{in the event of default} \\ 0 & \text{in the event of no default.} \end{cases} \tag{2.2.2}$$

The random variable L is a Bernoulli random variable with:

$$\begin{aligned} P(L=1) = P(\text{Default}) &= p \\ &= \text{Probability of Default}(PD) \end{aligned} \tag{2.2.3}$$

and:

$$P(L=0) = P(\text{No-Default}) = P(\text{Survival}) = 1-p. \tag{2.2.4}$$

The expectation and variance of L follows as:

- $\mathbb{E}(L) = 1 \cdot p + 0 \cdot (1-p) = p$,
- $\mathbb{E}(L^2) = 1 \cdot p = p$, and then:
- $Var(L) = p - (p)^2 = p(1-p)$.

Expression (2.2.1), is of fundamental importance for the management of credit risk. The parameters defined above is therefore regarded as the fundamental credit risk parameters. In estimating the expected loss, a holding period of one-year is typically taken. Assuming the three random variables EAD, LGD and L are independent random variables, then the expected value of the loss distribution is:

$$EL = \mathbb{E}(Loss) = \mathbb{E}(EAD) \times \mathbb{E}(LGD) \times \mathbb{E}(L) = \overline{EAD} \times \overline{LGD} \times p.$$

It follows that the expected loss is the expected exposure at default, times the expected loss in the event of default times the probability of default. In Van Gestel and Baesens (2009) it is noted that the expected loss is implemented in credit risk provisioning and/or calculation of the risk premium for a loan. Proportional to the exposure, the risk premium should in theory cover the LGD times the PD. This captures the desire to invest in loans with a low probability of default, low loss given default or both.

However, more importantly the risk parameters underlying the expected loss is of fundamental importance in calculating regulatory capital requirements. This is discussed in greater detail in the following section.

2.2.3 Credit Risk Regulation

The introduction of the Basel II framework in 2004 was marked by innovative proposals to improve the calculation of regulatory capital requirements. Note that the Basel Committee on Banking Supervision (BCBS) has since updated Basel II with Basel III released December 2010. These proposals however remain the same in the updated framework.

Pluto and Tasche (2010) highlights the importance of the new internal-ratings-based (IRB) approaches introduced in the Basel II framework. The framework allows banks to internally assess their credit risk exposure and dictate the amount of capital to be held against them. This is however subject to supervisory approval for which banks need to apply and fulfil a minimum set of requirements.

In the modern Basel III framework three approaches are available for calculating credit risk exposures, namely:

1. The Standardised approach,
2. The Foundation Internal Rating Based approach (FIRB), and
3. The Advanced Internal Rating Based approach (AIRB).

Before discussing the differences between these approaches, an overview is given of the general IRB approach. Under this approach banks are permitted to assess each risk exposure on a stand-alone basis. Risk estimates then serve as inputs for a supervisory credit risk model, which is defined by a set of risk weight functions. The risk weight functions determine the amount of capital that is regarded as sufficient to cover the credit risk of an exposure.

The main differences between the aforementioned approaches are in the prescription of the parameters. First of all distinguish between the IRB and standardised approaches. Until a bank has obtained supervisory approval for the entire banking book (or specific portfolios), the standardised approach needs to be applied. This is a simpler, less-risk sensitive approach for quantifying credit risk. In this approach minimum capital requirements are predominantly determined by the dependence on asset classes (sovereign, bank, corporate, or retail exposure) only. Where it is relevant the ratings provided by external rating agencies are utilized. The standardised approach therefore does not necessarily provide an accurate reflection of the risk of a specific credit portfolio (BCBS, 2011) (Pluto and Tasche, 2010).

The IRB approach is subdivided into FIRB and AIRB. Before distinguishing between the two subapproaches, consider the main characteristics underlying the IRB approach. Under the IRB approach banks are allowed to implement it's own estimates of the credit risk exposure. As mentioned in Section 2.2.2, the main risk parameters used in the quantification of credit risk is the PD, LGD and EAD.

The PD is typically determined by the banks internal historical default data. It is allowed to augment the historical data with external data. The PD risk parameter proposes several problems in the IRB approach since most financial institutions do not possess enough accessible data as in the Basel definition (Pluto and Tasche, 2010). At this stage it is important to highlight that there exists a need for the adjustment (or calibration) of this estimate where insufficient data exists. Calibration of the PD in such scenarios is the focus of this study.

The other fundametal parameters, the EAD and the LGD, is where the Basel framework differs from literature and practice. Industry credit rating systems regularly only take the PD into consideration. Whereas the Basel regulatory framework considers all three dimensions of credit risk. Hence, when a bank is authorised to apply the IRB approach, estimates for the PD, LGD and EAD needs to be provided for all credit risk exposures (Pluto and Tasche, 2010).

In the AIRB approach the bank is allowed to use it's own estimates of the LGD and the EAD in addition to the PD. Whilst in the FIRB approach only the PD requires estimation as supervisory estimates for the LGD and EAD may be implemented (Dahlin and Storkitt, 2014). Clearly the PD is especially important.

When dealing with corporate, sovereign, and bank exposures, the Basel committee proposed a formula for determing the risk weighted assets under Basel III. This is defined as follows.

Definition 2.10. ***The Basel Credit Risk Function:*** *The capital charge of an exposure is described by a closed form risk weight function. The risk weight function is derived from the capital requirement, defined as K per unit of currency of the exposure, where the general formulation of K is:*

$$K = LGD\left[\Phi\left(\frac{\Phi^{-1}(p)}{\sqrt{1-\xi}} + \Phi^{-1}(0.999)\sqrt{\frac{\xi}{1-\xi}}\right) - p\right]\frac{1+(M-2.5)\kappa}{1-1.5\kappa}. \tag{2.2.5}$$

The inputs p and LGD are measured in decimals, Φ is the standard normal cumulative distribution function, and M, is defined as the effective maturity. M is fixed at 1 year for retail exposures, and assumes values between 0 and 5 years for other exposures. Furthermore, κ is defined as the maturity adjustment given by:

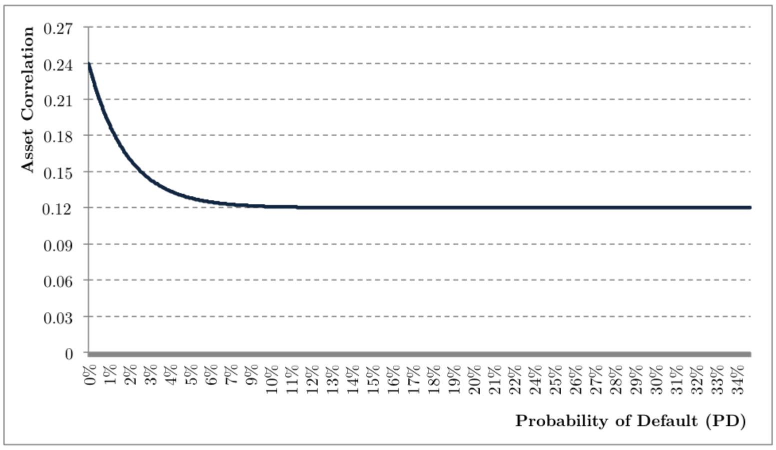
$$\kappa = (0.11852 - 0.05478 \times \ln(p))^2. \tag{2.2.6}$$

Finally, the risk weight functions for the different exposure classes mainly differ w.r.t. the asset correlation, ξ. The asset correlation can be interpreted as the default correlation between the obligors and is elaborated on in Section 2.2.4.2. For retail mortgage exposures and revolving retail credit exposures ξ is fixed at 15% and 4% respectively. Whereas for corporate, sovereign, and bank exposures ξ is dependent on the p estimate and is expressed as:

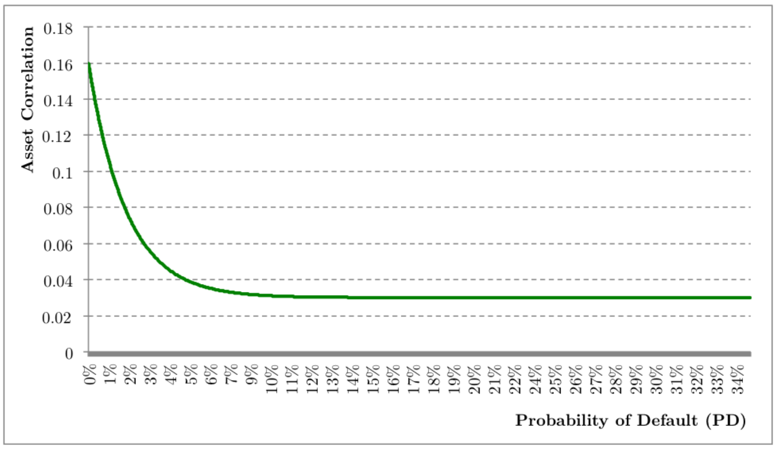
$$\xi = 0.12\left[\frac{1-\exp(-50p)}{1-\exp(-50)}\right] + 0.24\left[1 - \frac{1-\exp(-50p)}{1-\exp(-50)}\right] \tag{2.2.7}$$

and for other retail exposures as:

$$\xi = 0.03\left[\frac{1-\exp(-50p)}{1-\exp(-50)}\right] + 0.16\left[1 - \frac{1-\exp(-50p)}{1-\exp(-50)}\right]. \tag{2.2.8}$$



(a) Plot of the Sovereign, Corporate and Bank Exposure Asset Correlation Function (2.2.7)



(b) Plot of the Retail Exposure Asset Correlation Function (2.2.8)

Figure 2.3: The Basel Asset Correlation Functions

The capital charge on the risk, also known as the risk weighted assets (RWA), is given by:

$$RWA = K \times 12.5 \times EAD, \tag{2.2.9}$$

(BCBS, 2011).

For more information on how the Basel Credit Risk Function is derived, see Genest and Brie (2013). Note as stated in BCBS (2011), that the risk weighted assets follows from multiplying the minimum capital requirement (K) with the EAD and the reciprocal of the minimum capital ratio. The minimum capital ratio is 8%, the corresponding reciprocal is 12.5.

The two functions for the asset correlation is plotted in Figure 2.3. The function for corporate, sovereign, and bank exposures (2.2.7) is given in Figure 2.3a, note that this correlation is defined on $[0.12, 0.24]$. Whereas the function for retail exposures is defined on $[0.03, 0.16]$ as evident in Figure 2.3b. It is interesting to note from the plots that the Basel Asset Correlation functions exhibit a negative relationship between the PD and ξ. The functions suggest higher asset correlations for low PD levels and visa versa for high PD levels.

The Basel Credit Risk function illustrates the fundamental importance of the PD, LGD and EAD risk parameters in the context of regulation. Specifically when it comes to calculating

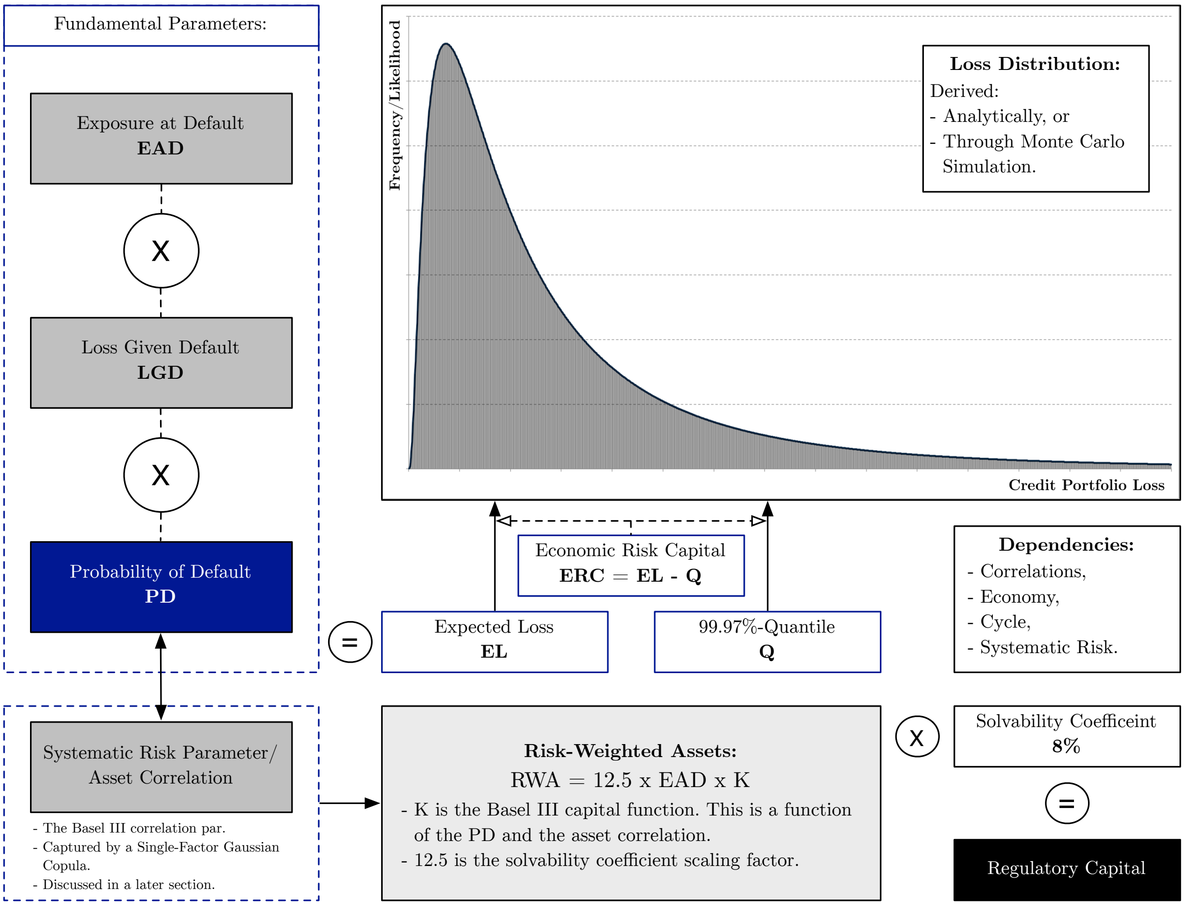


Figure 2.4: The Credit Risk Modelling Process

the minimum capital requirements for credit risk exposures under the IRB approach. Under the IRB approach the PD parameter requires estimation regardless of the sub-approach being implemented. It is clear that this is a fundamental input in the world of credit risk.

All the theory covered thus far regarding credit risk modelling can be integrated in the diagram adapted from Bluhm (2015), given in Figure 2.4.

2.2.4 Probability of Default

The probability of default (PD) was defined in Section 2.2.2, and as mentioned previously this is regarded as the principle estimate in credit risk management. In this section the mechanics of the PD estimate is discussed in greater detail.

In order to build up some background knowledge regarding estimation of the probability of default, consider a scenario where there are s credit exposures in the bank's credit portfolio, each with a credit-rating $C_i, i = 1, \ldots, s$. There exists a probability of default corresponding to each rating, denoted by p_i. Let the range of the possible ratings be denoted by $\{0, 1, \ldots, b\}$, where 0 is the highest rating grade for any given rating agency in Table 2.1 and b denotes the default state (or bankruptcy). Hence

$$C_i \in \{0, 1, \ldots, b\} \text{ and } p_i = P(C_i = b).$$

Let $C \to C'$ represent a rating migration from C to C' over the risk horizon, which is taken as one-year.

Relate each credit-rating C_i to a random variable L_i, which is called the default risk or default indicator so that when $C_i = b$, $L_i = 1$ and when $C_i \neq b, L_i = 0$. This implies:

$$P(i\text{th credit exposure defaults}) = P(C_i = b) = P(L_i = 1) = p_i$$

and

$$P(i\text{th credit exposure does not default}) = P(C_i \neq b) = P(L_i = 0) = 1 - p_i$$

Given this formulation it is clear that L_i is a Bernoulli random variable and the Bernoulli distribution is an instinctive model to model the defaults (Bluhm *et al.*, 2010).

Hence, the observations of s credit exposures are written as $L = (L_1, \dots, L_s)$, with:

$$L_i \sim B(1; p_i), i = 1, \dots, s. \tag{2.2.10}$$

In the formulation above the default indicators and the corresponding p's are independent. In reality losses cannot be regarded as independent. As a simple example consider a recession economy, in such an economic climate an increase in unemployment is expected. As people lose their jobs it becomes increasingly difficult to keep up the mortgage repayments on their properties. Therefore people would start to default on their mortgage repayments. The financier of a pool of such mortgages is now at risk of experiencing more losses than it's capital buffers can withstand. Therefore this company is now also at risk of default. This so called credit contagion effect is elaborated on in Section 2.2.4.2.

It is clear that in order to find realistic approximations of the loss statistics, correlation between these loss statistics need to be taken into account. Capturing credit portfolio correlation is a predominant challenge in credit risk management. One of the earliest proposals of addressing the issue is through the standard binary mixture model from Joe (1997). This method will be discussed in greater detail below. However note that the modelling of correlated defaults is extended in a later section with Single-Factor Gaussian Copula models, which is the method that is mainly used in the industry.

Bernoulli mixture models in essence assumes a distribution for the underlying parameter, p. Therefore, the mixture model discussed in Bluhm *et al.* (2010) is in essence the application of Bayesian theory. Bayesian theory will be discussed in greater detail in Section 3.3.2.

Using the terminology as defined, the loss of a credit portfolio is given by the default indicators $\mathbf{L} = (L_1, \dots, L_s)$, where $L_i \sim B(1; p_i)$. However, in the Bayesian setting the probability of default is regarded as a random variable. Define $\mathbf{P} = (P_1, \dots, P_s)$, with joint probability density function $g_{P_1,\dots,P_s}(p_1, \dots, p_s)$ and $0 < p_i < 1, i = 1, \dots s$.

Assume that the variables $L_1, \dots, L_s$ are independent, conditional on the realisation of $\mathbf{p} = (p_1, \dots, p_s)$. This conditional independence can be written as:

$$L_i|_{P_i=p_i} \sim B(1; p_i), (L_i|_{\mathbf{P}=\mathbf{p}}) \text{ independent for } i = 1, \dots, s.$$

Hence, the conditional joint distribution of the L_i's, given the realisation $\mathbf{p} = (p_1, \dots, p_s)$ of probabilities of default, is given by:

$$P(L_1 = l_1, \dots, L_s = l|\mathbf{p}) = \prod_{i=1}^{s} p_i^{l_i}(1-p_i)^{1-l_i}. \tag{2.2.11}$$

So that the joint distribution of the L_i's and $\mathbf{p}$ is given by:

$$f_{L_1,\dots,L_s,\mathbf{p}}(l_1, \dots, l_s, p_1, \dots, p_s) = \prod_{i=1}^{s} p_i^{l_i}(1-p_i)^{1-l_i} g_{P_1,\dots,P_s}(p_1, \dots, p_s), \tag{2.2.12}$$

where $l_i = \{0, 1\}$. The unconditional distribution of the L_i's follows as:

$$f_{L_1,\dots,L_s}(l_1, \dots, l_s) = \int_0^1 \dots \int_0^1 \prod_{i=1}^{s} p_i^{l_i}(1-p_i)^{1-l_i} g_{P_1,\dots,P_s}(p_1, \dots, p_s)dp_1 \dots dp_s. \tag{2.2.13}$$

Let $g_{P_i}(p_i)$ be the marginal distribution of P_i, then the first and second moments of L_i follows from (2.2.13) through:

$$f_{L_i}(l_i) = \int_0^1 p_i^{l_i}(1-p_i)^{1-l_i} g_P(p_i)dp_i$$

as:

$$\begin{aligned}\mathbb{E}(L_i) &= 1 \times P(L_i = 1) \\ &= \int_0^1 p_i g_{P_i}(p_i)dp_i \\ &= \mathbb{E}(P_i),\end{aligned}$$

and:

$$\begin{aligned}Var(L_i) &= \mathbb{E}[L_i^2] - [\mathbb{E}[\mathbb{L}_\beth]]^2 \\ &= \mathbb{E}[P_i^2] - [\mathbb{E}[\mathbb{P}_\beth]]^2 \\ &= \mathbb{E}(P_i)(1 - \mathbb{E}(P_i)),\end{aligned}$$

for $i = 1, \dots, s$. The default indicators were assumed independent conditional on the realisation of $\mathbf{p} = (p_1, \dots, p_s)$. Intuitively, the covariance between the losses are dependent of the covariance between the p_i's. This follows through:

$$Cov(L_i, L_j) = \mathbb{E}[L_i.L_j] - \mathbb{E}[L_i]\mathbb{E}[L_j] = \mathbb{E}[P_i.P_j] - \mathbb{E}[P_i]\mathbb{E}[P_j] = Cov(P_i, P_j)$$

Therefore the default correlation in this Bayesian setting is:

$$Corr(L_i, L_j) = \frac{Cov(P_i, P_j)}{\sqrt{\mathbb{E}(P_i)(1 - \mathbb{E}(P_i))}\sqrt{\mathbb{E}(P_j)(1 - \mathbb{E}(P_j))}} \tag{2.2.14}$$

It is clear that the correlation structure of the default events can be fully captured by the covariance structure of the distribution on $\mathbf{P}$. Were the default indicators are only independent if and only if:

$$Cov(L_i, L_j) = Cov(P_i, P_j) = 0$$

When credit exposures are divided into different rating grades, then all the exposures obtaining a certain rating is assumed to be the same in terms of risk. Therefore it is assumed that all the obligor's that fall in a specific rating grade has the same PD. From expression (2.2.14), assumption of a homogeneous PD essentially means that the correlations between the obligors in that rating category is also consistent.

This can be generalised since credit portfolios can also be divided into the different types of exposures; examples include retail, corporate, banking etc. For a portfolio where the exposures in the portfolio are approximately the same in terms of risk the same equal PD assumption can be made.

Bluhm *et al.* (2010) mentions that these analytical approximations hold reasonably well for many retail as well as smaller banks portfolios. It is further stated in the literature that the assumption of a homogeneous default probability and a consistent correlation structure does not harm the outcome of the calculations with such a model.

The aforementioned equal probabilities assumption implies that:

$$P_1 = P_2 = \cdots = P_s = P,$$

where P is a random variable with probability density function $g_P(p)$ and $0 < p < 1$ so that the conditional probability distribution of L_i given $P = p$ is given by the Bernoulli distribution $B(1; p)$ i.e.

$$f_{L_i|p}(l_i|p) = p^{l_i}(1-p)^{1-l_i}, i = 1, \dots, s$$

so that:

$$f_{L_i}(l_i) = \int_0^1 p^{l_i}(1-p)^{1-l_i} g_P(p) dp. \tag{2.2.15}$$

Hence, the conditional probability distribution of the L_i's, given that they are independent conditional on p follows from (2.2.11):

$$\begin{aligned} f_{L_1,\dots,L_s}(l_1, \dots, l_s|P = p) &= \prod_{i=1}^{s} p^{l_i}(1-p)^{1-l_i} \\ &= p^{\sum_{i=1}^s l_i}(1-p)^{s-\sum_{i=1}^s l_i} \\ &= p^d(1-p)^{s-d} \end{aligned} \tag{2.2.16}$$

with $d = \sum_{i=1}^s l_i$ and $l_i = 0$ or 1.

The joint probability distribution of $L_1, \dots, L_s$ and P, follows from (2.2.12), as:

$$f_{L_1,\dots,L_s,P}(l_1, \dots, l_s, p) = p^d(1-p)^{s-d} g_P(p) \tag{2.2.17}$$

so that the unconditional distribution of the L_i's is given by:

$$P(L_1 = l_1, \dots, L_s = l_s) = f_{L_1,\dots,L_s}(l_1, \dots, l_s) = \int_0^1 p^d(1-p)^{s-d} g_P(p) dp. \tag{2.2.18}$$

Define $L = \sum_{i=1}^s l_i$, then the probability of exactly a defaults occurring is:

$$P(L = a) = \binom{s}{a} \int_0^1 p^a(1-p)^{s-a} g_P(p) dp. \tag{2.2.19}$$

From (2.2.15) it follows that:

$$P(L_i = 1) = \int_0^1 p g_P(p) dp = \mathbb{E}(P) = \bar{p}. \tag{2.2.20}$$

From the results above it follows that:

$$\mathbb{E}(L_i) = 1.P(L_i = 1) = \bar{p}, \tag{2.2.21}$$

$$\mathbb{E}(L_i^2) = 1.P(L_i = 1) = \bar{p},$$

$$Var(L_i) = \bar{p} - \bar{p}^2 = \bar{p}(1 - \bar{p}). \tag{2.2.22}$$

From (2.2.18) follows the unconditional joint distribution of L_i and L_j as:

$$f_{L_i,L_j}(l_i, l_j) = \int_0^1 p^d (1-p)^{s-d} g_P(p) dp, \tag{2.2.23}$$

so that:

$$\begin{aligned} \mathbb{E}(L_i.L_j) &= 1.P(L_i = 1, L_j = 1) \\ &= 1 \int_0^1 p^2 g_P(p) \\ &= \mathbb{E}(P^2), \end{aligned} \tag{2.2.24}$$

$$\begin{aligned} Cov(L_i, L_j) &= \mathbb{E}(L_i.L_j) - \mathbb{E}(L_i)\mathbb{E}(L_j) \\ &= \mathbb{E}(P^2) - \mathbb{E}(P)\mathbb{E}(P) \\ &= Var(P), \end{aligned} \tag{2.2.25}$$

$$\begin{aligned} \xi = Corr(L_i, L_j) &= \frac{Cov(L_i, L_j)}{\sqrt{Var(L_i)}\sqrt{Var(L_j)}} \\ &= \frac{Var(P)}{\bar{p}(1-\bar{p})}. \end{aligned} \tag{2.2.26}$$

This relation implies that the higher the volatility of the probability of default, the higher the correlation in the default indicators will be (Bluhm *et al.*, 2010). Furthermore, due to the fact that variances are always positive, it implies that the dependence between the L_i's are either positive or zero. The latter case would only be true in the event that the variance of P is zero, in turn implying that there is certainty regarding the PD (Bluhm *et al.*, 2010). Also note that the uniformity assumption is equivalent to assuming a uniform asset correlation.

For a given PD, p, expression (2.2.19) becomes:

$$P(L = d|p) = \binom{s}{d} p^d (1-p)^{s-d} \tag{2.2.27}$$

for a portfolio of s credit exposures, when d defaults have been observed. This can also be regarded as the likelihood function $L(p|d)$ (Kiefer, 2009). In order to illustrate the effect that small changes in the number of defaults would have on the likelihood, the expression above is

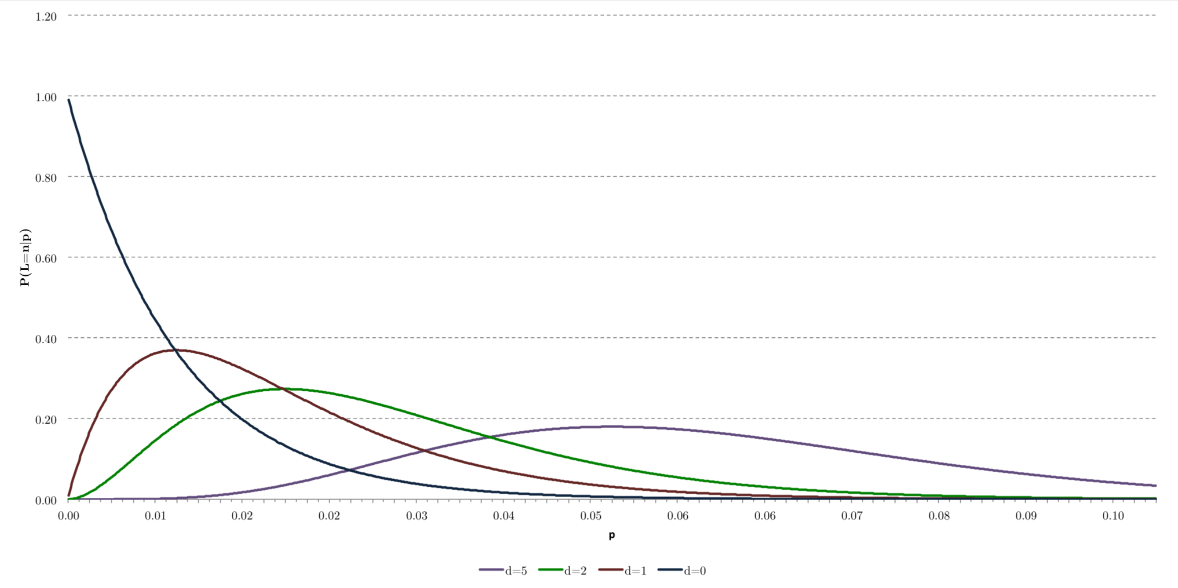


Figure 2.5: Likelihood plots, Expression (2.2.27), for a Portfolio of 100 Obligors and Varying Defaults.

plotted for different values of d. Figure 2.5, exhibits the likelihood functions for $s = 100$ and $d = 0, 1, 2, 5$.

The rest of this section gives an overview of some of two industry models that are used in the estimation of the probability of default. These models are not covered in full detail and only the fundamental models are covered. The first model under discussion is the original 1974 Merton Model.

2.2.4.1 The 1974 Merton Model

Robert Merton proposed a firm value model that can be used to estimate the value of the debt and the probability of default of a company in his paper, Merton (1974). This model allows estimating the probability of default through the company's equity price. In essence this model is a combination of the simple equation:

$$E = A - D,$$

and the Black-Scholes-Merton formula, where E denotes the shareholder's equity, A the value of the company's assets and D is the market value of the company debt (Meissner, 2009). The face value of the debt at time T including accrued interest is denoted by F.

Let r_F denote the risk-free rate then:

$$D_0 = F \exp(-r_F T).$$

As discussed in Crouhy *et al.* (2000), the debt obligation of the company is subject to credit risk. This is the risk that the value of the debt at time T will exceed the value of the firm's assets. In other words credit risk exists as long as $P(A_T < F) > 0$. Furthermore, credit risk is a function of the company's leverage ratio and volatility of the rate of return on the firms assets. The leverage ratio is given by:

$$LR = \frac{D_0}{A_0} = \frac{F\exp(-r_F T)}{A_0}, \tag{2.2.28}$$

In order to capture the credit risk of the loan, two assumptions are made. Firstly, assume that the loan is the only debt instrument of the company. Further assume that the only source of financing is equity (Crouhy *et al.*, 2000).

Hull (2012) describes the model by considering a company with a single zero-coupon bond outstanding, where the bond in question matures at time T. Define A_t as the value of the company's assets at time t, E_t as the value of the company's equity at time t and F is the value of the debt repayment due at time T. Furthermore, let σ_A and σ_E be the volatility of the assets and equity respectively, where the volatility of the assets is assumed constant.

Theoretically, when $A_T < F$ then the company defaults on it's debt at time T. In this event the value of the equity is zero. On the other hand when $A_T > F$, then the company pays back the outstanding bond at time T and $E_T = A_T - F$. The value of the equity at time T can therefore be written as:

$$E_T = \max(A_T - F, 0).$$

The aforementioned expression resembles the value of a call option on the value of the assets with a strike price equal to the value of the outstanding debt (Hull, 2012). Writing this in terms of the Black-Scholes-Merton framework, the current value of the equity follows as:

$$E_0 = A_0\Phi(d_1) - F\exp(-r_F T)\Phi(d_2), \tag{2.2.29}$$

where

$$d_1 = \frac{\ln\left[\frac{A_0}{F}\right] + (r_F + \frac{1}{2}\sigma_A^2)T}{\sigma_A\sqrt{T}},\ d_2 = d_1 - \sigma_A\sqrt{T}.$$

From(2.2.28), d_1 can also be written as:

$$\begin{aligned} d_1 &= \frac{\ln\left[\frac{A_0}{F}\right] + (r_F + \frac{1}{2}\sigma_A^2)T}{\sigma_A\sqrt{T}} \\ &= \frac{\ln\left[\frac{A_0}{F\exp(-r_F T)}\right] + (\frac{1}{2}\sigma_A^2)T}{\sigma_A\sqrt{T}} \\ &= \frac{\ln\left[\frac{A_0}{D}\right] + (\frac{1}{2}\sigma_A^2)T}{\sigma_A\sqrt{T}} \\ &= \frac{\ln\left[\frac{1}{LR}\right] + (\frac{1}{2}\sigma_A^2)T}{\sigma_A\sqrt{T}}. \end{aligned}$$

From the formulation above it is clear that the d_1 and d_2 underlying (2.2.29) is simply the company's leverage ratio adjusted for volatility (Dwyer, 2006). Furthermore, as discussed in Meissner (2009), equation (2.2.29) states that the value of a company's equity is related to the value of the company assets. In the event that the value of the assets increases, then there is unlimited upside potential on the value of the equity. However, if the value of the assets falls below the value of the debt, then the company will go bankrupt, in other words default.

It follows from Black-Scholes theory that:

$$P(A_T > F) = \Phi(d_2),$$

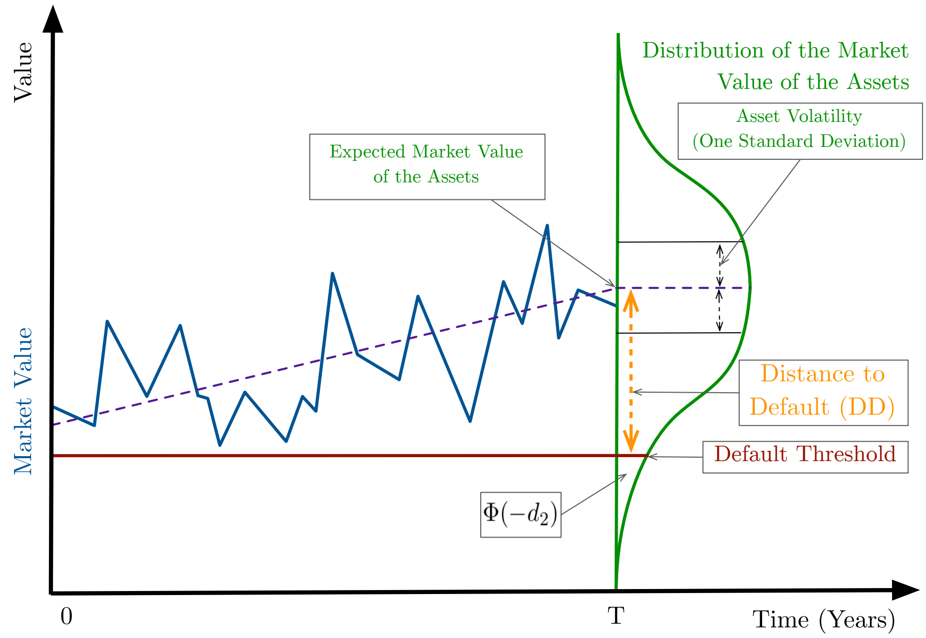


Figure 2.6: Illustration of Distance to Default

this is the risk-neutral probability of exercising the call option. Furthermore:

$$\begin{aligned} p &= P(F < A_T) \\ &= 1 - P(A_T > F) \\ &= 1 - \Phi(d_2) \\ &= \Phi(-d_2). \end{aligned}$$

Intuitively, this is the risk-neutral probability of not exercising the call option. The call option is not exercised when the value of the debt exceeds the value of the company assets. This is also the default state. Therefore the probability of default is given by $\Phi(-d_2)$.

The aforementioned theory can be summarised as follows, in the Merton model a default occurs when the value of the assets fall below some threshold, which is interpreted as the required debt payment. The difference (or distance) between the expected value of the asset and the default threshold is called the *distance to default*, defined as:

$$DD = d_2 = \frac{\ln\left[\frac{A_0}{F}\right] + (r_F - \frac{1}{2}\sigma_A^2)T}{\sigma_A\sqrt{T}} = -\Phi^{-1}(p). \tag{2.2.30}$$

The distance to default is graphically illustrated in Figure 2.6.

Thus far correlation between default events has only been addressed in its simplest from. Default correlations are elaborated on in the subsequent sections.

2.2.4.2 Default Correlation

The concept of default correlation was already introduced earlier in Section 2.2.4. A formal definition follows from Hull (2012).

Definition 2.11. ***Default Correlation***: *The tendency for two companies to default in about the same time.*

Default correlation exists for a number of reasons, the most obvious is that companies in the same industry may be affected by the same external factors and as a result may simultaneously experience financial difficulties. The same can be said for companies that do business in the same geographic location. Furthermore, default rates tend to be higher in times of economic distress (see Table 2.2 for the year 2008).

The *credit contagion effect* was already mentioned earlier. A formal definition follows as:

Definition 2.12. ***Credit Contagion***: *The effect of a default by one company or obligor constituting a default in another company or obligor.*

The credit contagion effect is another factor resulting in default correlation. This effect results in credit risk being impossible to be completely diversified away.

From (2.2.14) follows that the covariance structure of the probability of default distribution captures the correlation structure of the default indicators. Intuitively, default correlation plays an integral role in determining the distributions for the probability of default of the different exposures of a credit portfolio.

As mentioned in Hull (2012), default correlation can be addressed by utilising either reduced form models, or structural models. In this study a popular structural model will be considered in the form of the Gaussian Copula, this is also known as the Guassian One-Factor Model or the Vasicek Model as a reference to the Vasicek (2002) that introduced the model into the modelling default probabilities (Dwyer, 2006).

Structural models are based on a model similar to Merton's model, which was discussed in the previous subjection. In this framework a company defaults in the event where the value of the assets falls below a certain level. The default correlation between two companies X and Y is introduced by the assumption that the stochastic process followed by the value of the assets of company X is correlated with the stochastic process followed by the value of the assets of company Y.

2.2.4.3 The Single-Factor Gaussian Copula Model

A copula function is a function used to create a joint probability distribution for two or more marginal distributions. As mentioned by Romano (2002), a copula function in essence describes the dependence structure of a multivariate random variable. There are different types of copula functions. The most widely used copula is the Gaussian copula, which is also the copula function used in this study. The two other main copulas are the Student's t copula and the multivariate copula. The latter will be discussed in brief.

The formal definition of a copula is not considered as it lies beyond the scope of this study. A formal definition is however unnecessary for the application of copulas in the context in which it is applied below. For an explanation on how copula functions work, take note of the following discussion adapted from Hull (2006): Consider two correlated random variables Q_1 and Q_2. Regardless of the random variable it is possible to derive their respective unconditional distributions. However, without knowledge of the correlation structure the joint distribution

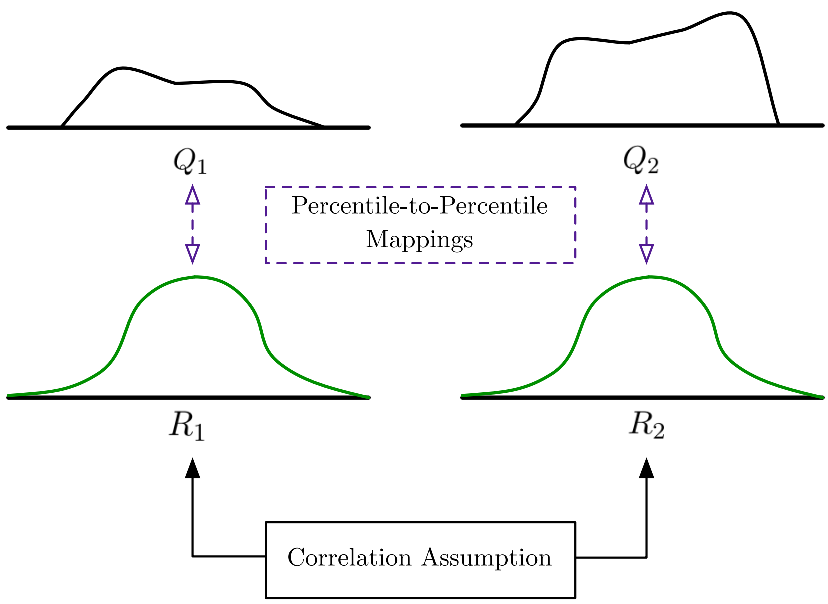


Figure 2.7: Defining a joint distribution with a Copula function

cannot be defined. Hence an assumption on the correlation structure is required. If these variables are normally distributed then it is intuitive to assume that the joint distribution is a bivariate normal distribution. Similar assumptions can be made for other known distributions, however there is no natural method for defining the correlation structure between two random variables. This is where copula functions are used.

In essence the Gaussian copula maps the random variables Q_1 and Q_2 into new random variables R_1 and R_2, where the new random variables follow a bivariate normal distribution. This mapping takes place on a percentile-to-percentile basis. As an example this means that the tenth-percentile point of the Q_1 distribution is mapped to the tenth-percentile point of the R_1 distribution, and so on for each percentile point of the underlying distribution. Since the newly mapped random variables R_1 and R_2 are standard normally distributed it can be assumed that the joint distribution is bivariate normal. This assumption implies a joint distribution between the original random variables Q_1 and Q_2. Hence the copula function indirectly implies a correlation structure between two random variables indirectly (Hull, 2006). The process of defining a joint distribution using a copula function is illustrated in Figure 2.7.

The random variables Q_1 and Q_2 can be mapped to any distributions from which a correlation structure easily follows, for example the Student's t distribution. Copulas can also be used to define the correlation structure between multiple variables.

The idea underlying the multivariate Gaussian copula is discussed in brief. Consider a set of random variables $Q_i, i = 1, \ldots, n$ and assume that the respective unconditional distributions are known. Apply the percentile-to-percentile mapping to each of these random variables yielding $R_i \sim \Phi(0, 1), i = 1, \ldots, n$. The assumption can now be made that the R_i's follows a multivariate normal distribution. In order to capture the correlation structure between the R_i's a factor model is typically used. Assuming that the R_i's are correlated through a single factor a Gaussian copula can be created by assuming:

$$R_i = Xa_i + Y_i\sqrt{1 - a_i^2},$$

where X and Y_i follows a standard normal distribution. The Y_i are independent and uncorrelated with X, this follows from a property of the equicorrelated normal distribution. The R_i's are now correlated through the single factor X (Hull, 2006). The a_i is a constant parameter defined on $[-1; 1]$ and this results in the correlation between Ri and R_j being defined by $a_i a_j$.

The theory discussed above will now be brought into the context of this study. Note that in the Gaussian Copula model it is assumed that all companies will eventually default. The model then attempts to quantify the correlation between the distributions of the probability of default between two companies.

Assuming there are s credit exposures in a portfolio each with a respective PD, $p_i, i = 1, \ldots, s$. Let p_i be the probability of default for company i and let p_j be the probability of default for company j. As previously mentioned, if the probability distributions of p_i and p_j were normal, then the joint distribution would be bivariate normal. This is where the Gaussian copula model is used. Transform p_i and p_j into new variables using the percentile-to-percentile mapping discussed above. This yields:

$$z_i = \Phi^{-1}(p_i) \text{ and } z_j = \Phi^{-1}(p_j).$$

By construction z_i and z_j are respective observations from a standardised normal random variable $Z_i, i = 1, \ldots, s, i \neq j$. The Gaussian copula assumption is that Z_i and Z_j are bivariate standard normal with a single correlation parameter, ξ_{ij}. This correlation parameter is called the copula correlation (Hull, 2012).

As discussed in Hull (2012), defining a different correlation for each pair of obligors in the portfolio can be avoided by using a single-factor model. In this case assume a common correlation, ξ, between all pairs of obligors. Therefore in the model above let $a_i = \sqrt{\xi}$. It follows that:

$$Z_i = X\sqrt{\xi} + Y_i\sqrt{1-\xi}, \tag{2.2.31}$$

where X and Y_i for $i = 1, \ldots, s$ are mutually independent standard normal random variables.

The variable X is the common factor affecting the PD of all the counterparties and $X\sqrt{\xi}$ is the specific counterparties exposure to this common factor. The $Y_i\sqrt{1-\xi}$ term is then the specific risk of the counterparty.

Under the Gaussian copula model default occurs when:

$$z_i < \Phi^{-1}(p_i).$$

To summarise, this implies that default occurs when:

$$\begin{aligned} &L_i = 1 \\ \Rightarrow &Z_i < \Phi^{-1}(p_i) \\ \Rightarrow &X\sqrt{\xi} + Y_i\sqrt{1-\xi} < \Phi^{-1}(p_i). \end{aligned}$$

Therefore, given the common factor X, the conditional probability of default on any single counterparty, i, with probability of default, p_i, follows through:

$$\begin{aligned} &P(X\sqrt{\xi} + Y_i\sqrt{1-\xi} < \Phi^{-1}(p_i)) \\ \Rightarrow &P\left(Y_i < \frac{\Phi^{-1}(p_i) - \sqrt{\xi}X}{\sqrt{1-\xi}}\right), \end{aligned}$$

as:

$$p_i(X) = P(L_i = 1|X) = \Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\xi}X}{\sqrt{1-\xi}}\right). \tag{2.2.32}$$

This result is the probability of default of an credit exposure in the portfolio underlying the common factor X. The unconditional probability of default then follows as the average of the conditional probabilities.

Upon introduction of expression (2.2.31), it is simple to see that default events are no longer independent for $\xi > 0$ by considering two creditors i and j. Creditor i defaults when $z_i < \Phi^{-1}(p_i)$, assuming that the probability of default is the same for i and j, then

$$P(\text{Creditors } i \text{ and } j\text{default}) = P(Z_i \cap Z_j) = \Phi_2(\Phi^{-1}(p), \Phi^{-1}(p); \xi) > p^2 = P(Z_i)P(Z_j)$$

where Φ_2 denotes a bivariate normal distribution with standardised marginals (Tasche, 2013).

2.3 Conclusion

This chapter constitutes the first part of the literature study. The aim of this chapter is to build an understanding of the importance of the probability of default as an input in credit risk management, as well as to lay the foundation for Chapter 3. In Section 2.1, financial risk management is discussed in general terms and some of the important definitions in financial risk management is defined. This is then extended in Section 2.2, which covers the important concepts underlying credit risk management. In Section 2.2.1, a brief overview of credit ratings is provided and the low default nature of high rated instruments is highlighted. The importance of the probability of default is then discussed in Section 2.2.2, where the three main parameters for credit risk modelling is covered. In Section 2.2.3, an overview of credit risk regulation is given, and the role of the probability of default in the regulatory capital function is illustrated. Finally, in Section 2.2.4, the foundation of probability of default estimation is discussed in detail.

The foundation discussed in Section 2.2.4, is elaborated on in Chapter 3 where probability of default calibration for low default portfolios is considered.

Chapter 3

Literature Review: Low Default Portfolio Calibration

"*If I owe you a pound, I have a problem; but if I owe you a million, the problem is yours.*"
John Maynard Keynes, *British Economist and father of Keynesian economics.*

Debt is one of the most important components of the modern financial world and as highlighted in Chapter 2, creditors need to have the necessary risk management instruments in place in order to quantify the risk of not recovering all of the credit exposures. As mentioned, the probability of default is one of the principal components in quantifying the risk of a credit exposure. However, once confronted with portfolios where limited historical loss data is available, then traditional techniques for estimating the probability of default are no longer sufficient. Therefore, some form of a calibration technique is required for these portfolios with low default characteristics.

The aim of this chapter is to address this problem by discussing some of the available calibration techniques. In Section 3.1, Low Default Portfolios are defined and discussed. The industry concerns regarding these portfolios are then mentioned in Section 3.2. The most important part of this chapter is, however, Section 3.3 where the PD calibration techniques for Low Default Portfolios are discussed in detail.

3.1 Low Default Portfolios

As mentioned in Section 2.2 the probability of default is a fundamental input in the quantification of a credit risk exposure. Intuitively, accurate PD estimates are of utmost importance in order to ensure high quality output from the credit risk models.

The PD is especially important for banks that wish to apply the IRB approach since these institutions then have to determine their own estimates of the probabilities of default for the obligors in their portfolios. A substantial part of the banking book however consists of high grade debt.

When confronted with portfolio's of high grade borrowers, certain obstacles arise. Portfolios mainly consisting of highly rated borrowers might go through many years without observing a default. Also, when a few defaults are observed this will result in a high degree of volatility in the observed default rate. Portfolios such as these are a key concern for regulators as the

PD might be underestimated due to data scarcity.

Such zero-default or low-default portfolios (LDP) are not uncommon in practice. Take for example sovereign or bank portfolios that exhibit an overall good quality of borrowers, or specialized lending that has high-volume-low-number characteristics (Pluto and Tasche, 2011).

Even though the term LDP is widely accepted in academic literature and in the industry, there exists no formal definition of what a low default portfolio is. In BCBS (2005) the Basel Committee Accord Implementation Group's Validation Subgroup (AIGV) addresses the validation of low default portfolios in the Basel II framework. In this newsletter the AIGV states that it does not believe that bank portfolios are low-default or high-default. The committee is convinced that there exists a link between these opposite ends of the spectrum implying that credit models should be able to quantify the risk of a credit portfolio regardless of the amount of observed defaults.

A portfolio is regarded closer to the LDP end of the spectrum when the bank's internal data systems have fewer loss events on records and therefore the lack of default data introduces obstacles in accurately quantifying the risk exposure. In the event were a low number of defaults are observed then the painless calculation based on historical losses observed in the portfolio would not be well grounded in estimating the PD, LGD or EAD (BCBS, 2005).

Even though there is no formal definition, a few informal definitions or interpretations of the term LDP exists. The following regulatory definition from the Prudential sourcebook for Banks, Building Societies and Investment Firms, also known as BIPRU (2011) is used in this research project;

Definition 3.1. ***Low Default Portfolio:*** *The number of defaults on a low default portfolio is so low that estimates of quantitative risk parameters based on historical default experience is unreliable or poor in some statistical sense.*

As stated in BIPRU (2011) paragraph 4.3.95, a LDP has the following characteristics:

- A firm's internal rating experience of exposures, of a type covered by a model or other rating system, is *20 defaults or fewer*, and
- In the firm's view, reliable estimates of PD cannot be derived from external sources of default data, including the use of market place related data, for all exposures covered in the rating system.

3.2 Industry concerns regarding LDPs

The BCBS (2005) states that there exists numerous industry concerns when it comes to LDP. Intuitively, the most significant concern is the lack of statistical data and therefore the difficulty in backtesting risk parameters would result in LDPs being excluded from IRB treatment under the Basel III framework.

This was however never explicitly stated by the Basel Committee and is a literal interpretation of industry participants. Industry participants are concerned that LDP doesn't meet the minimum requirements for IRB and would subsequently be obligated to use the simpler approaches for these portfolios. Since a very large portion of the banking book assets have

LDP characteristics, this is a noteworthy concern.

Challenges set aside, both the industry and the AIVG is of the opinion that LDPs should not be excluded from IRB treatment. This leaves room for the development of tools and techniques that can be implemented in assessing the risk of these portfolios in the absence of sufficient loss data.

As discussed in Kruger (2015), PD models are typically based on an historical data period of five years. In order to estimate and validate the PD models, the industry participant needs to have access to enough default data. It is discussed in Basel III that banks need to compare their PD estimates and realised default rates at single grade level. Furthermore, in order to address the problem for high grade exposures modelling assumptions need to be made. These assumptions can result in model risk and therefore a high level of conservatism needs to be applied.

This is where the concerns of the regulator and the industry practitioner differs. The regulator is concerned that credit risk will be underestimated and therefore the high levels of conservatism is subscribed. The bank on the other hand is concerned that the subscribed levels of conservatism will result in cynical PD estimates. The higher PD estimates will naturally impact their pricing and result in high regulatory capital reserves (Kruger, 2015).

3.3 PD Calibration Methods for LDP

In this section some of the PD calibration models proposed for the LDP problem will be discussed. In the literature a number of approaches has been proposed. However there is no consensus between academics or practitioners on which method is most appropriate.

Before discussing the proposed models, it is important to see why naïve estimators of the PD will fail.

Definition 3.2. ***The Naïve Estimator***: *In the naïve estimator the probability of default of rating grade i is given by:*

$$p_{i,t} = \frac{d_{i,t}}{s_{i,t-1}} \tag{3.3.1}$$

where $d_{i,t}$ is the number of defaults in rating grade i during the period t and $s_{i,t-1}$ is the number of obligors in the rating grade at the start of the period.

It should be intuitive to see that this estimator would fail as it would significantly underestimate the PD. Take for example a rating grade with 1000 obligors at the start of the period and 1 default during the period. Then the naïve estimator would derive at a PD estimate of 0.1%. Furthermore, what if there were no defaults in the historical data? Then the naïve estimate would be zero. However, it is impossible that this is the true PD estimates for the mentioned scenario.

Obtaining multi-period estimates from the one-period naive estimates follows through taking an average of the one-period default rates.

The following overview of some of the main proposed PD calibration methods is given:

- Pluto and Tasche (2011) proposed a confidence based approach. This approach is often used by banks to validate and model their LDP PD models and is discussed in further detail in Section 3.3.1.

- Forrest (2005) proposed a likelihood method that is in support of the aforementioned confidence based method. This will not be discussed in detail as this method will not be considered as it follows the same logic as the confidence based approach, the only difference is in the final step of the models application.

- Wilde and Jackson (2006) proposed an analytical estimator based on the CreditRisk+ model in the context of LDPs. This will not be discussed in detail as this method will not be considered as it is more of a calculation tool that eliminates the need of Monte Carlo Simulation rather than an outright credit risk model for LDPs.

- Bayesian approaches to LDPs have been proposed by numerous authors (see Dwyer (2006); Kiefer (2009); Orth (2011); Tasche (2013); Clifford *et al.* (2013); Chang and Yu (2014); Kruger (2015) to name a few) these approaches are discussed in detail in Section 3.3.2.

- Van Der Burgt (2008) proposed calibrating the PD for LDPs using a cumulative accuracy profile. This will not be discussed in detail as the underlying techniques used in this methodology is detached from the modelling techniques primarily considerd in this study.

Note that other methods has also been proposed, and that the aforementioned list only contains some of the main proposals.

Kruger (2015) mentions the following observations regarding the available methods in literature. First of all, some of the methods require default observations in the historical data and is subsequently not applicable to zero-default portfolios. Secondly, numerous methods are only single period. Due to the data scarcity, backtesting is nearly impossible. The last point is also highlighted in BCBS (2005). Finally, almost none of the proposed PD calibration models provide guidance on how to validate the PD model. Backtesting of PD models remains a significant obstacle.

It should be clear from Chapter 2 that a Bayesian approach should be the most appropriate method to address the problem. This is also evident in the literature. However, due to the Pluto and Tasche (2011) method being widely applied in the literature this method is also considered as a sort of “benchmark” methodology.

3.3.1 The Confidence Based Approach

The confidence based approach, introduced by Pluto and Tasche (2011) laid the foundation for calibrating the PD for low default portfolios. The underlying concept in their methodology is the *most prudent estimation principle*. In this section the theory underlying the confidence based approach will be covered in detail, where most of the underlying theory will be introduced using the similar examples as in the original Pluto and Tasche (2011) paper.

The confidence based approach is discussed as follows; first of all the most prudent estimation principle is introduced assuming no defaults and independence. This is secondly extended to the case of a few defaults, still assuming independence. Finally, default correlation is introduced before extending the method to a multi-period case.

No defaults, assuming independence. Consider a portfolio of credit exposures with sample size s. Let the portfolio's credit exposures be divided into three rating categories 1, 2, and 3, with respective sample sizes s_1, s_2, and s_3. Note that $s = s_1 + s_2 + s_3$. Intuitively, the grade with the highest credit rating is 1 and the lowest 3.

Assume that no defaults have been observed in any of the rating grades during the previous observation period. The respective rating grades corresponding PDs are p_1, p_2, and p_3. Therefore, rating grade 1 can be observed as a sub-portfolio of s_1 credit exposures and the overall PD for the rating grade is p_1.

Taking the decreasing credit-worthiness into account, then the following inequality makes intuitive sense:

$$p_1 \leq p_2 \leq p_3.$$

From the inequality, it is clear that the PD of rating 1 cannot be greater than that of rating 3. The *most prudent estimate* of Pluto and Tasche (2011) follows by setting $p_1 = p_3$, so that:

$$p_1 = p_2 = p_3. \tag{3.3.2}$$

Assuming the relation above holds, the following step is to determine a $(1-\alpha)100\%$ confidence region for p_1. From (3.3.2), it is clear that the three rating grades do not differ in their respective riskiness and hence a homogeneous sample of size $s_1 + s_2 + s_3$ is used. Under the assumption of unconditional independence between default events the probability of observing no defaults in the entire sample is $(1-p_1)^{s_1+s_2+s_3}$. This results in the following definition that can be used in order to determine the probability of default of the obligor's in rating grade 1.

Definition 3.3. ***A*** $(1-\alpha)\%$ ***Confidence Region for*** p_1: *Under the assumption of independence given no defaults have been observed. Pluto and Tasche (2011) defines the* $(1-\alpha)100\%$ *confidence region for* p_1 *as all the values of* p_1 *that satisfy:*

$$(1-p_1)^{s_1+s_2+s_3} \geq 1-\alpha. \tag{3.3.3}$$

Equation (3.3.3) implies that the region can also be written as:

$$p_1 \leq 1-(1-\alpha)^{1/(s_1+s_2+s_3)}, \tag{3.3.4}$$

from which a maximum value of the estimate $\hat{p}_1$ is obtained.

Similarly, the probability of default for rating grade 2 is obtained by setting $p_2 = p_3$. Rating grades 2 and 3 do not differ in their respective riskiness and a homogeneous sample of size $s_2 + s_3$ is considered. This results in a $(1-\alpha)100\%$ confidence region for p_2 as all the values of p_2 that satisfy:

$$(1-p_2)^{s_2+s_3} \geq 1-\alpha,$$

which implies that the confidence region for p_2 is written as:

$$p_2 \leq 1-(1-\alpha)^{1/(s_2+s_3)}. \tag{3.3.5}$$

Since there is no obvious upper bound for rating grade 3, only the observations in this rating grade is used when determining the confidence region. It follows using the same logic as before that the confidence region for p_3 is given by:

$$p_3 \leq 1-(1-\alpha)^{1/(s_3)}. \tag{3.3.6}$$

Consider the following fictitious example adapted from Pluto and Tasche (2011) as illustration.

Example 3.4. A bank has a portfolio of 1000 credit exposures in the form of regional mortgages. Of the 1000 mortgages, 250 obtained a 1 rating, 400 obtained a 2 rating and 350 obtained a 3 rating. Therefore:

$$s_1 = 250; s_2 = 400; s_3 = 350; S = 1000.$$

The results of estimating $\hat{p}_1, \hat{p}_2$ and $\hat{p}_3$ at different confidence regions using the sample sizes defined above is given in table 3.1.

Table 3.1: Upper Confidence Bound Estimates Example, No Defaults Observed.

α	50%	75%	90%	95%	99%
$\hat{p}_1$	0,07%	0,14%	0,23%	0,30%	0,46%
$\hat{p}_2$	0,09%	0,18%	0,31%	0,40%	0,61%
$\hat{p}_3$	0,20%	0,40%	0,66%	0,85%	1,31%

As evident in the results of the example the main driver of the p estimate apart from the confidence level is the sample size of the rating grade under consideration. Consider the p estimates at $\alpha = 99\%$ then;

$$\begin{aligned} \hat{p}_1 \leq 1 - (1 - 0,99)^{1/(250+400+350)} = 0,46\%; \\ \hat{p}_2 \leq 1 - (1 - 0,99)^{1/(400+350)} = 0,61\%; \\ \hat{p}_3 \leq 1 - (1 - 0,99)^{1/350} = 1,31\% \end{aligned}$$

It is clear that the smaller the sample size, the greater the upper confidence bound and therefore the greater the PD estimate. This effect is desirable due to the fact that the greater the credit-worthiness, the greater the number of creditors in the portfolio without any default. Also note that if the naive estimate, expression (3.3.1), of the PD was taken then all the PD estimates would have been 0.

Negligible defaults, assuming independence. The confidence based approach was introduced under the assumption that no defaults have been observed and that defaults are independent. Keeping the independence assumption. Let d_1, d_2 and d_3 denote the amount of defaults in rating grades 1, 2, and 3 respectively. Assume that:

$$d_1 = 0; d_2 = 2; d_3 = 3; d = d_1 + d_2 + d_3 = 5.$$

Implementing the *most prudent estimation* procedure in determining an estimate for p_1 allows consideration of the entire portfolio as a homogeneous sample of size $s_1 + s_2 + s_3$. Under the assumption that the defaults are independent, and that the number of defaults in the portfolio is binomially distributed. The probability of observing no more than five defaults is given by:

$$\sum_{k=0}^{5} \binom{s_1 + s_2 + s_3}{k} p_1^k (1 - p_1)^{(s_1+s_2+s_3)-k}. \tag{3.3.7}$$

Set the restriction that the probability in expression (3.3.7) must be larger than $1 - \alpha$. It follows that the $(1 - \alpha)100\%$ confidence region for p_1 is given as the set of all values for p_1 satisfying:

$$\sum_{k=0}^{5} \binom{s_1+s_2+s_3}{k} p_1^k (1-p_1)^{(s_1+s_2+s_3)-k} \geq 1-\alpha. \tag{3.3.8}$$

An appropriate Beta distribution can be used in order to analytically determine the tail of the binomial distribution, this follows from Lemma 3.5.

Lemma 3.5. As discussed in Wackerly *et al.* (2007), the cumulative distribution function for a beta random variable with parameters a and b is denoted by:

$$F(y) = \int_0^y \frac{t^{a-1}(1-t)^{b-1}}{B(a,b)} dt = I_y(a,b) \tag{3.3.9}$$

where

$$B(a,b) = \int_0^1 y^{a-1}(1-y)^{b-1} dy = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Equation (3.3.9) is commonly called the *incomplete beta function.* In the event where both the a and the b parameters are positive integers, $I_y(a,b)$ is analogous to the binomial probability function. It follows from integration by parts that, for $0 < y < 1$ and a, b both integers that:

$$F(y) = \int_0^y \frac{t^{a-1}(1-t)^{b-1}}{B(a,b)} dt = \sum_{i=a}^{a+b-1} \binom{a+b-1}{i} y^i (1-y)^{a+b-1-i}. \tag{3.3.10}$$

Note that the summation on the right-hand side resembles a cumulative binomial distribution with parameters $p = y$ and $n = a+b-1$ (Wackerly *et al.*, 2007)(Gupta and Nadarajah, 2004).

Thus from Lemma 3.5 using the beta distribution as an approximation of the tail of a binomial distribution, the p_1 in (3.3.8) can be determined analytically

Using the same logic as in the previous section. Implementing the *most prudent estimation* procedure in determining an estimate for p_2, the portfolio is considered as a homogeneous sample of size $s_2 + s_3$. Assume that the defaults are independent, and that the number of defaults in the portfolio is binomially distributed. The probability of observing no more than five defaults is given by:

$$\sum_{k=0}^{5} \binom{s_2+s_3}{k} p_2^k (1-p_2)^{(s_2+s_3)-k}. \tag{3.3.11}$$

If the restriction is set that the probability in expression (3.3.11) must be larger than $1-\alpha$; it follows that the $(1-\alpha)100\%$ confidence region for p_2 is given as the set of all values satisfying:

$$\sum_{k=0}^{5} \binom{s_2+s_3}{k} p_2^k (1-p_2)^{(s_2+s_3)-k} \geq 1-\alpha. \tag{3.3.12}$$

Finally, in determining the upper confidence bound for rating grade 3, this rating grade is once again considered as a stand alone portfolio with s_3 obligors. Under the same assumptions as before, the probability of observing no more than three defaults is given by:

$$\sum_{k=0}^{3} \binom{s_3}{k} p_3^k (1-p_3)^{s_3-k}. \tag{3.3.13}$$

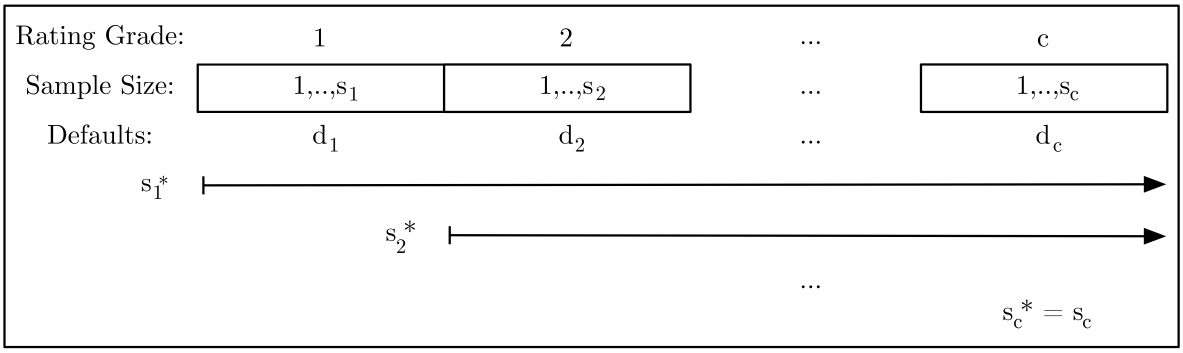


Figure 3.1: Determining the Sample Sizes for the Confidence Based Approach

Note that the defaults in the higher rating grade is not considered. If the restriction is set that the probability in expression (3.3.13) must be larger than $1-\alpha$; it follows that the $(1-\alpha)100\%$ confidence region for p_3, is as the set of all values satisfying:

$$\sum_{k=0}^{3} \binom{s_3}{k} p_3^k (1-p_3)^{s_3-k} \geq 1-\alpha. \tag{3.3.14}$$

Under the assumption that defaults are independent a general definition can be constructed. Assume that there are s obligor's in the portfolio and that these obligor's can be divided into c credit rating categories. Let s_i be the amount of obligor's falling in rating grade i for $i = 1, \ldots, c$. The probabilities of default for the respective rating grades is denoted by p_i and the amount of defaults in each rating grade is denoted by d_i for $i = 1, \ldots, c$.

Define the homogenoues portfolio size used for estimating p_1 as $s_1^* = \sum_{i=1}^{c} s_i$. Using the logic discussed earlier in the section, the homogenoues sample size used in the estimation of p_2 is $s_2^* = \sum_{i=2}^{c} s_i$. This is illustrated in Figure 3.1. In general, the sample size used for estimating the probability of default for rating grade j can be written as:

$$s_j^* = \sum_{i=j}^{c} s_i \tag{3.3.15}$$

where j denotes the position of the ordered rating grade. Note that 1 denotes the highest rating grade in the portfolio and c the lowest rating grade, hence $j = 1, \ldots, c$. Similarly the sum of the defaults taken into account when estimating the probability of default for rating grade j can be written as:

$$d_j^* = \sum_{i=j}^{c} d_i \tag{3.3.16}$$

where $j = 1, \ldots, c$ is as previously defined.

Definition 3.6. ***General Upper Confidence Bound Assuming Independence:*** *Let α denote the confidence level and let s_j^* denote the portfolio size used in determining the probability of default for rating grade j. The number of defaults follows a binomial distribution so that the $(1-\alpha)100\%$ confidence region p_j follows from:*

$$1-\alpha \leq \sum_{k=0}^{d_j^*} \binom{s_j^*}{k} p_j^k (1-p_j)^{s_j^*-k}. \tag{3.3.17}$$

This definition is consistent with the theory covered in section 2.2.4.

To derive an analytical solution for p_j in equation (3.3.17) expression (3.3.10) is used i.e.:

$$F(y) = \int_0^y \frac{t^{a-1}(1-t)^{b-1}}{B(a,b)}dt = \sum_{i=a}^{a+b-1} \binom{a+b-1}{i} y^i(1-y)^{a+b-1-i}. \tag{3.3.18}$$

The last term is written as:

$$\sum_{i=a}^{a+b-1} \binom{a+b-1}{i} y^i(1-y)^{a+b-1-i} = 1 - \sum_{i=0}^{a-1} \binom{a+b-1}{i} y^i(1-y)^{a+b-1-i}.$$

Consider a random variable X that is binomially distributed with size parameter s and success probability p. It follows from Lemma 3.5 and expression (3.3.18) that for any interger $0 \leq d \leq s$ the following holds:

$$\begin{aligned} P[X \leq d] &= \sum_{k=0}^{d} \binom{s}{k} p^k(1-p)^{s-k} \\ &= 1 - \sum_{k=d+1}^{s} \binom{s}{k} p^k(1-p)^{s-k}. \end{aligned} \tag{3.3.19}$$

Let $a = d+1 \Rightarrow a-1 = d$ and $s = a+b-1 \Rightarrow b-1 = s-a = s_d - 1$, then:

$$\begin{aligned} 1 - \sum_{k=d+1}^{s} \binom{s}{k} p^k(1-p)^{s-k} &= 1 - \frac{\int_0^p t^d(1-t)^{s-d-1}dt}{\int_0^1 t^d(1-t)^{s-d-1}dt} \\ &= 1 - P[Y \leq p] \end{aligned} \tag{3.3.20}$$

where Y denotes a beta distributed random variable with parameters $a = d+1$ and $b = s-d$. The following proposition illustrates that a direct numerical solution can be found and also shows that a unique solution for (3.3.17) exists.

Proposition 3.7. Let X be a binomial random variable with parameters n and p. Let $0 \leq k < n$ be integers then:

$$\begin{aligned} P(X \leq k) &= \sum_{i=0}^{k} \binom{n}{i} p^i(1-p)^{n-i}, p \in (0,1). \\ &= f_{n,k}(p)(\text{say.}) \end{aligned} \tag{3.3.21}$$

$f_{n,k}(p)$ is a function such that $(0,1) \to \mathbb{R}$. Fix some $0 < v < 1$. Then the equation:

$$f_{n,k}(p) = v \tag{3.3.22}$$

has exactly one solution $0 < p = p(v) < 1$. Moreover, this solution $p(v)$ satisfies the inequalities:

$$1 - \sqrt[n]{v} \leq p(v) \leq \sqrt[n]{1-v}. \tag{3.3.23}$$

(Pluto and Tasche, 2011).

Proof. In Appendix A it is shown that:

$$\frac{df_{n,k}(p)}{dp} = -(n-k)\binom{n}{k}p^k(1-p)^{n-k-1} \tag{3.3.24}$$

From which it is clear that $f_{n,k}$ is a strictly decreasing function of p. This in turn implies a unique solution for (3.3.22). Furthermore, since $0 \leq k < n$, the inequalities:

$$f_{n,0}(p) \leq f_{n,k}(p) \leq f_{n,n-1}(p)$$

implies that a solution for (3.3.22) and that the inequalities (3.3.23) exists. □

Hence, under the assumption that defaults are independent equation (3.3.17) can be used to analytically determine a unique value for the PD that satisfies the confidence region.

No Defaults, Incorporating Dependence. As discussed in Chapter 2, the independence assumption is unrealistic as losses are not independent. Hence correlation needs to be introduced into the model. In Section 2.2.4 it was discussed that the assumption of a uniform PD between obligors essentially means that the correlations between the obligors in that rating category is also consistent (see equation (2.2.26)). It was also shown that a single factor Gaussian copula model can be used to capture the dependence structure.

Denote the asset correlation by ξ, and assume that the asset correlation is uniform between all borrowers. From equation (2.2.32), under the assumption of a single factor X and a uniform asset correlation the probability of default conditional on the realisation x of X can be written as;

$$p_i(X) = \Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\xi}X}{\sqrt{1-\xi}}\right) \text{ for } i = 1, \ldots, s \tag{3.3.25}$$

where $X \sim \Phi(0,1)$ represents the single factor introduced into the model. Given a realisation of X, expression (3.3.25) becomes;

$$p_i(X = x) = \Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\xi}x}{\sqrt{1-\xi}}\right) \text{ for } i = 1, \ldots, s$$

In the event that no defaults have been observed and assuming independence, a $(1-\alpha)100\%$ confidence region for the probability of default of rating grade 1 follows from the inequality defined by expression (3.3.3). This expression is:

$$(1-p_1)^{s_1+s_2+s_3} \geq 1-\alpha.$$

Incorporating dependence into this scenario through a single systematic factor X, a $(1-\alpha)100\%$ confidence region for the probability of rating grade 1 conditional on the realisation $X = x$ can be written as:

$$(1-p_1(X))^{s_1+s_2+s_3} = \left(1 - \Phi\left(\frac{\Phi^{-1}(p_1) - \sqrt{\xi}x}{\sqrt{1-\xi}}\right)\right)^{s_1+s_2+s_3} \geq 1-\alpha. \tag{3.3.26}$$

Assuming no defaults, and using a one factor model the $(1-\alpha)100\%$ confidence region for unconditional default probability is given by:

$$\int_{-\infty}^{\infty} \phi(x)\left(1 - \Phi\left(\frac{\Phi^{-1}(p_1) - \sqrt{\xi}x}{\sqrt{1-\xi}}\right)\right)^{s_1+s_2+s_3} dx \geq 1-\alpha \tag{3.3.27}$$

where ϕ represents the standard normal density function. The left-hand side can also be written as:

$$\mathbb{E}\left[\left(1-\Phi\left(\frac{\Phi^{-1}(p_1)-\sqrt{\xi}x}{\sqrt{1-\xi}}\right)\right)^{s_1+s_2+s_3}\right] \geq 1-\alpha.$$

Equation (3.3.27) can be solved numerically for a specified asset correlation. Recall that the bounds for the asset correlation as prescribed by Basel III is discussed in Section 2.2.3.

A similar result to Proposition 3.7 exists.

Proposition 3.8. For any probability $0 < p < 1$, any correlation $0 < \xi < 1$ and any real number x, define:

$$\lambda_\xi(p,x) = \Phi\left(\frac{\Phi^{-1}(p)+\sqrt{\xi}x}{\sqrt{1-\xi}}\right) \tag{3.3.28}$$

the same notation is used as in equation (3.3.27). Fix a value $0 < v < 1$ and a positive interger s. Then the equation

$$v = \int_{-\infty}^{\infty} \phi(x)(1-\lambda_\xi(p,x))^s dx \tag{3.3.29}$$

with ϕ denoting the standard normal density, has exactly one solution $0 < p = p(v) < 1$. This solution satisfies the inequality:

$$p(v) \geq 1-\sqrt[s]{v}. \tag{3.3.30}$$

(Pluto and Tasche, 2011).

(Note that there exists no obvious upper bound for the solution $p(v)$ of (3.3.29) as in (3.3.23).)

Proof. For fixed ξ and x, the function $F_\xi(p,x)$ is strictly increasing and continuous in p. Moreover,

$$\lim_{p\to 0} \lambda_\xi(p,x) = 0 \text{ and } \lim_{p\to 1} \lambda_\xi(p,x) = 1 \tag{3.3.31}$$

this implies the existence and the uniqueness of the solution of (3.3.29). Define a random variable Z:

$$U = \lambda_\xi(p,X) = \Phi\left(\frac{\Phi^{-1}(p)+\sqrt{\xi}X}{\sqrt{1-\xi}}\right), \tag{3.3.32}$$

where X is a random variable following a standard normal distribution. Then the random variable U is a Single Factor Gaussian Copula Model as discussed in Section 2.2.4.3, and:

$$\mathbb{E}[U] = p. \tag{3.3.33}$$

Using (3.3.32), equation (3.3.29) can be written as:

$$v = \mathbb{E}[(1-U)^s]. \tag{3.3.34}$$

The following result from probability theory is required. Jensen's inequality states that if Z is a random variable and g is a convex function, then:

$$g(\mathbb{E}(U)) \leq \mathbb{E}(g(U)).$$

Since $q(x) = (1-x)^s$ is a convex function for $0 < x < 1$ by (3.3.33) Jensen's inequality implies

$$v = \mathbb{E}[(1-U)^s] \geq (1-p)^s. \tag{3.3.35}$$

As the right hand side of (3.3.29) is decreasing in p,

$$p(v) \geq 1-\sqrt[s]{v}.$$

follows from (3.3.35), which proves the proposition. □

Note that by taking correlations into account the numerical complexity of the problem has increased dramatically (Pluto and Tasche, 2011). However, Proposition 3.8 shows that a unique solution does indeed exist. Similar equations to (3.3.27) can be derived for the lower rating grades using the same logic as at the start of this section.

Thus far it has been shown that an upper confidence bound for the PD estimate can be obtained for portfolio's with no historical default observations. This model can also be extended to incorporate correlations between defaults through a Single-Factor Gaussian copula. Hence it is shown that dependence can also be incorporated for portfolio's with few defaults.

Negligible Defaults, Incorporating Dependence. Consider the independent case where d defaults are present. The uniform probability of default for the assets falling in rating grade j, where the portfolio of s credit exposures has been divided into c rating grades can be determined by:

$$1 - \alpha \leq \sum_{k=0}^{d_j^*} \binom{s_j^*}{k} p_i^k (1 - p_i)^{s_j^* - k}. \tag{3.3.36}$$

where $s_j^* = \sum_{i=j}^{c} s_i$ and $d_j^* = \sum_{i=j}^{c} d_i$ for $j = 1, \ldots, c$.

Consider the same scenario as before where a portfolio of s credit exposures is considered. Assume that the dependence structure of the defaults can be captured by a single common systematic factor X; where X is a standard normal random variable. Assume that the asset correlation between any two pairs of counterparties is constant and let this be denoted by ξ. Furthermore, assume that the s obligor's in the portfolio can be divided into c credit rating categories.

As previously discussed, the probability of default conditional on the realisation of X is given by:

$$p_i(X) = \Phi\left(\frac{\Phi^{-1}(p) + \sqrt{\xi}X}{\sqrt{1-\xi}}\right) \text{ for } i = 1, \ldots, s. \tag{3.3.37}$$

Substituting (3.3.37) into (3.3.36) and integrating over all possible realisations of the systematic factor results in a general formula of the $100(1-\alpha)\%$ confidence region for p_j being given by:

$$\int_{-\infty}^{\infty} \phi(x) \sum_{k=0}^{d_j^*} \binom{s_j^*}{k} \Phi\left(\frac{\Phi^{-1}(p_j) - \sqrt{\xi}x}{\sqrt{1-\xi}}\right)^k \left(1 - \Phi\left(\frac{\Phi^{-1}(p_j) - \sqrt{\xi}x}{\sqrt{1-\xi}}\right)\right)^{s_j^* - k} dx \geq 1 - \alpha. \tag{3.3.38}$$

The probability of default can be estimated using Monte Carlo Simulation of the different states of the systematic factor. Furthermore (3.3.38) can be written as:

$$\mathbb{E} \sum_{k=0}^{d_j^*} \binom{s_j^*}{k} \Phi\left(\frac{\Phi^{-1}(p_j) - \sqrt{\xi}x}{\sqrt{1-\xi}}\right)^k \left(1 - \Phi\left(\frac{\Phi^{-1}(p_j) - \sqrt{\xi}x}{\sqrt{1-\xi}}\right)\right)^{s_j^* - k} \geq 1 - \alpha.$$

The process of Monte Carlo simulation entails generating a series of random outcomes representing the systematic factor, in other words generating values from a standard normal distribution. Integrating over the series of simulated occurrences provides an approximation of the systematic factor.

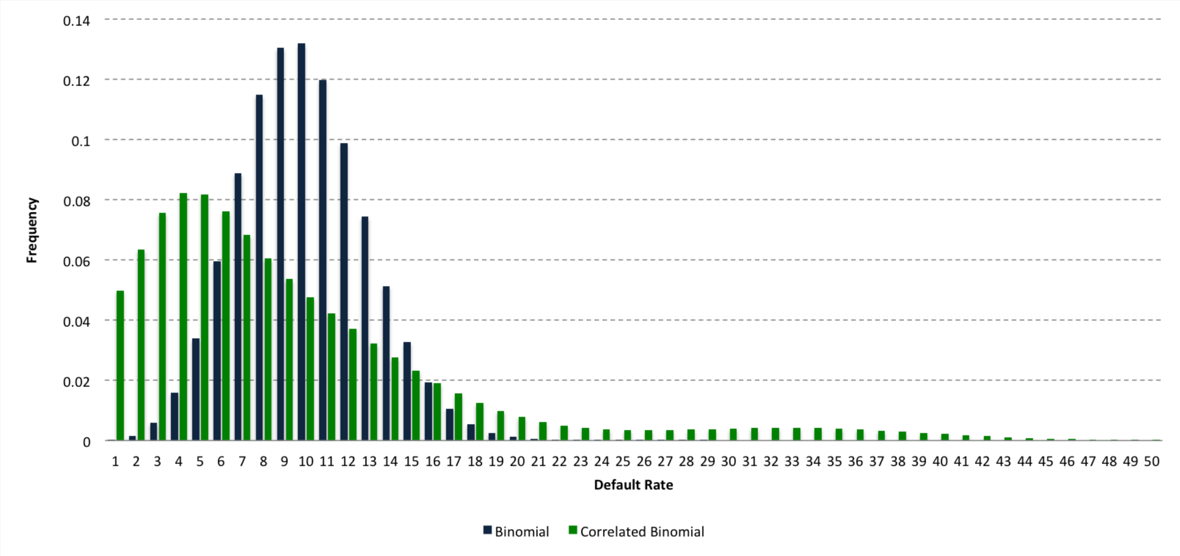


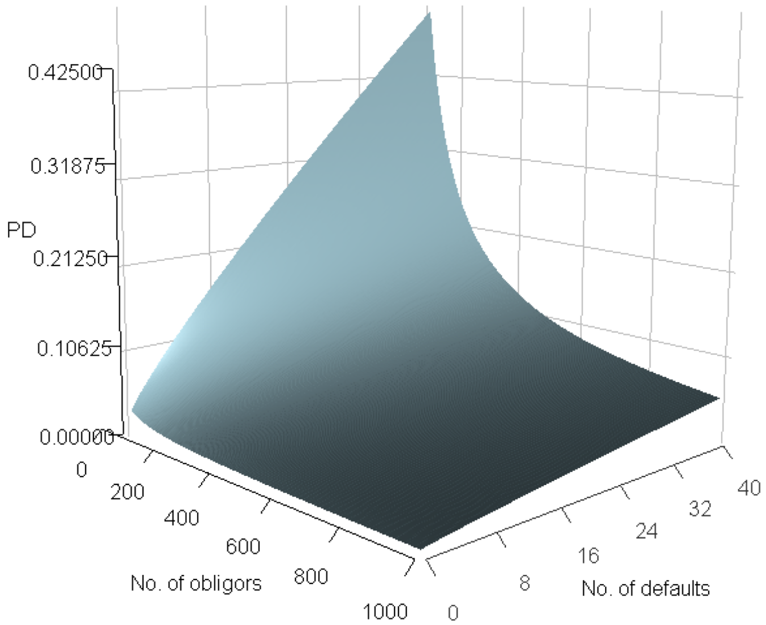
Figure 3.2: Effect of the Introduction of Correlation

The introduction of correlation is illustrated in Figure 3.2, this is a plot of the default rate against the associated frequency where the default rate is the amount of observed defaults. In the plot the probability of default, p, is taken as 10%, the correlation, ξ, is specified as 18%, a portfolio of $s = 100$ credit exposures is considered and 5 000 Monte Carlo simulations are used for the correlated binomial. It is clear from the plot that the introduction of correlation into the model increases the variance of the distribution. Also note that the central location of the distribution has shifted to the left.

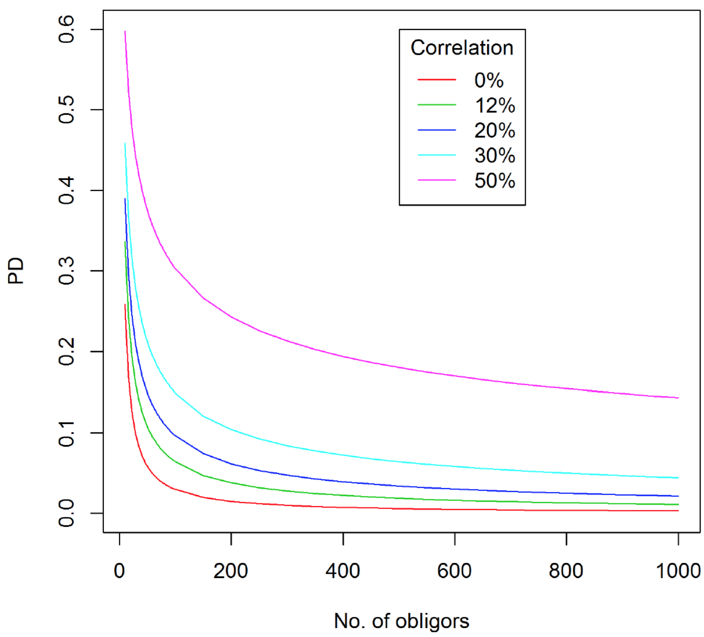
Consider Figure 3.3 from Frunza (2013), Figure 3.3a is a surface plot exhibiting different PD levels for different portfolio sizes and amount of portfolio defaults. Whereas Figure 3.3b shows the PD estimates for different correlation inputs and number of obligors in the portfolio in the event of no defaults.

Multi-period Extension. The theory covered thus far only discusses situations where estimation is carried out on a one-year data period. When a time series consisting of data from several years is available then the PDs for individual rating grades for these single year periods could be estimated. These estimations could then be used to calculate weighted averages of the PDs in order to make more efficient use of the data. Furthermore, as set out by paragraph 463 of Basel II (BCBS, 2005); banks that are applying the IRB approach need to use at least five years of historical default data for the estimation of the PD. In ideal circumstances, the time series would cover at least one full credit cycle (Tasche, 2013). Therefore, a multi-period approach is required. This would however result in the interpretation of the PD at some pre-defined upper confidence bound to be lost (Pluto and Tasche, 2011).

Discussion of the multi-period approach expands on the scenario used throughout this section. The only difference is the fact that the length of the observation period is $T > 1$. Only obligor's that were present at the start of the observation period are considered, subsequently it is important to take note that obligor's entering the portfolio afterwards are neglected for the purpose of the estimation exercise. However, this does not mean that the amount of obligor's are constant as a default would result in less obligor's in the pool.



(a) Surface Plot of the PD estimate for different porfolio sizes and number of defaults



(b) LDP Adjustment for a No-Default Portfolio for different correlation inputs

Figure 3.3: PD Calibration for LDP under Pluto and Tasche (2011).

In order to account for intertemporal correlations, the single-factor framework from Section 2.2.4.3 is extended into a dynamic factor framework. The dynamic factor model consolidates the idea that the asset value log-return of each credit exposure in the portfolio depends on a common dynamic latent factor.

Assume that one single rating category is being considered. All the creditors falling in the same rating category are assigned the same values of p, ξ and τ. Where τ is the intertemporal correlation, this is the correlation between time periods. If $Z_{t,i}$ denotes the asset value log-return of the ith obligor at time t, the dynamic factor model used by Pluto and Tasche (2011) can then be written as:

$$Z_{t,i} = X_t\sqrt{\xi} + Y_{t,i}\sqrt{1-\xi}, \tag{3.3.39}$$

$$X_t = \tau X_{t-1} + \sqrt{1-\tau^2}W_t \tag{3.3.40}$$

where $i = 1, \ldots, s_t$, $t = 1, \ldots, T$, $|\tau| < 1$, X_t is the latent factor common to all obligors, T is the number of years and s_t is the number of obligors in the portfolio at time t. Let:

$$\begin{aligned}
&Y_{t,i} \sim \Phi(0,1), i = 1, 2, \ldots, s_t,\\
&W_t \sim \Phi(0,1), t = 1, 2, \ldots, T, \text{ and}\\
&X_1 \sim \Phi(0,1).
\end{aligned}$$

Further assume that $Y_{i,t}$ is independent of w_t for all t and i. Also that X_1 is independent of W_t for all t. The parameter ξ denotes the common asset correlation, whereas the parameter τ denotes the autocorrelation of X_t. Finally, it follows that $Z_{i,t} \sim \Phi(0,1)$, for all t and i conditional on ξ and τ (Pluto and Tasche, 2011; Tasche, 2013; Chang and Yu, 2014; Kruger, 2015).

In the dynamic factor framework a single component X_t of the vector of systematic factors generate the intertemporal correlation of the default events at time t. The intertemporal correlation is affected by the dependence structure of the factors $X_1, \ldots, X_T$, written as a vector $\mathbf{X} = (X_1, \ldots, X_T)'$.

An assumption on the vector is that not only the individual components, but also the vector as a whole is normally distributed. The individual components are standardised and hence the joint distribution is completely determined by the correlation matrix:

$$Corr(\mathbf{X}\mathbf{X}') = \begin{pmatrix} 1 & \upsilon_{1,2} & \upsilon_{1,3} & \cdots & \upsilon_{1,n} \\ \upsilon_{2,1} & 1 & \upsilon_{2,3} & \cdots & \upsilon_{2,n} \\ \vdots & & \ddots & & \vdots \\ \upsilon_{T-1,1} & \cdots & \upsilon_{T-1,T-2} & 1 & \upsilon_{T-1,T} \\ \upsilon_{T,1} & \cdots & \upsilon_{T,T-2} & \upsilon_{T,T-1} & 1 \end{pmatrix}. \tag{3.3.41}$$

The cross sectional difference is however constant throughout the year for any pair of obligor's. The intertemporal dependence also weakens over time. The exponentially decreasing intertemporal dependence structure in the model is described by:

$$Corr(X_s, X_t) = \upsilon_{s,t} = \tau^{|s-t|}, s, t = 1, ..., T, s \neq t \tag{3.3.42}$$

for some appropriate $|\tau| < 1$, which is to be specified (Blochwitz *et al.*, 2004).

Recall from Section 2.2.4.3 that the ith obligor will default in the event that it's latent factor $Z_{t,i}$ falls below some threshold D_i. Therefore Obligor i defaults in yeat t if:

$$Z_{i,1} > D_{i,1}, Z_{i,2} > D_{i,2}, \ldots, Z_{i,t-1} > D_{i,t-1}, Z_{i,t} \leq D_{i,t}. \tag{3.3.43}$$

The probability of default is written as:

$$P(L_{t,i} = 1) = p_{t,i} = P(Z_{t,i} \leq D_{t,i})$$

This probability is however conditional on ξ and τ, therefore:

$$\begin{aligned} p_{t,i} &= P(L_{t,i} = 1|\xi, \tau) \\ &= P(Z_{t,i} \leq D_{t,i}|\xi, \tau) \\ &= \Phi(D_{t,i}) \end{aligned}$$

where Φ is the standard normal cumulative distribution function. As before:

$$D_{t,i} = \Phi^{-1}(p_{t,i}). \tag{3.3.44}$$

The uniformity assumption is once again made; it follows that when obligor's fall into the same rating category, then $p_{t,i} = p$ and $D_{t,i} = D$ for all t and i.

Consider the same scenario as before where a portfolio of s credit exposures is divided into c rating categories. Let $s_i, i = 1, \ldots, c$ be the sample size of rating grade i and let $d_i, i = 1, \ldots, c$ be the amount of defaults corresponding to that grade. Since the *most prudent estimation* approach is considered, construct the sample size and amount of defaults used for estimating p_j as before where $s_j^* = \sum_{i=j}^{c} s_i$ and $d_j^* = \sum_{i=j}^{c} d_i$ for $j = 1, \ldots, c$. It is assumed that the default of an obligor j in year $t = 1, \ldots, T$ is triggered if the change in value of their assets results in a value lower than some default threshold D.

Now assume that in a T year observation period, $d_i, i = 1, \ldots, c$, defaults were observed (note that these are the initial rating grades at the start of the observation period). Now for a specified $100(1-\alpha)\%$ confidence region, the maximum $\hat{p}$ of all the parameters p_j is to be determined such that;

$$P[\text{No more that } k \text{ defaults observed}] \geq 1 - \alpha \tag{3.3.45}$$

is satisfied. The left-hand side of this equation needs to be rewritten in order to derive a formulation accessible to numerical calculation. Firstly, an expression is developed for obligor j's conditional PD during the observation period given a realization of the systematic factors $X_1, \ldots, X_T$.

From (3.3.39), (3.3.43), (3.3.44), by using the independence of the $Z_{j,1}, \ldots, Z_{j,T}$, and given the systematic factors it follows that:

$$P[\text{Obligor } j \text{ defaults}|X_1, \ldots, X_T] = P[\min_{t=1,\ldots,T} Z_{j,t} \leq \Phi^{-1}(p)|X_1, \ldots, X_T]. \tag{3.3.46}$$

The expression above is the probability that the jth obligor defaults in the T year observation period, given the realisation of the systematic factors. This is equivalent to the probability that $Z_{j,t} \leq \Phi^{-1}(p)$ in the minimum amount of years given the systematic factors. This however requires further simplification, recall from (3.3.39) that $Z_{t,i} = X_t\sqrt{\xi}+Y_{t,i}\sqrt{1-\xi}$. Substituting this into the expression above results in:

$$\begin{aligned}
P[\text{Obligor } j \text{ defaults}|X_1,...,X_T] &= P[\min_{t=1,...,T} Z_{j,t} \leq \Phi^{-1}(p)|X_1,...,X_T] \\
&= P[\min_{t=1,...,T} Y_{j,t} \leq \lambda(p,\xi,X_t)|X_1,...,X_T] \\
&= 1 - P[Yj,1 > \lambda(p,\xi,X_1),...,Y_{j,T} > \lambda(p,\xi,X_T)|X_1,...,X_T] \\
&= 1 - \prod_{t=1}^{T}(1-\Phi(\lambda(p,\xi,X_t))). \qquad (3.3.47)
\end{aligned}$$

Where $\lambda(p,\xi,X_t) = \Phi\left(\frac{\Phi^{-1}(p)-\sqrt{\xi}X_t}{\sqrt{1-\xi}}\right)$. From construction of the model it follows that all the probabilities of default are equal for obligors falling in the same rating category. Therefore, for any obligor that falls in a specific rating category, define:

$$\pi(X_1,...,X_T) = P[\text{Obligor } j \text{ defaults}|X_1,...,X_T] = 1 - \prod_{t=1}^{T}(1-\Phi(\lambda(p,\xi,X_t))). \qquad (3.3.48)$$

Recall the independent case where d defaults are present. The uniform probability of default for the assets falling in rating grade j, where the portfolio of s credit exposures has been divided into c rating grades can be determined by:

$$1-\alpha \leq \sum_{k=0}^{d_j^*}\binom{s_j^*}{k}p_i^k(1-p_i)^{s_j^*-k}. \qquad (3.3.49)$$

where $s_j^* = \sum_{i=j}^{c} s_i$ and $d_j^* = \sum_{i=j}^{c} d_i$ for $j = 1,\ldots,c$.

The probability of default for obligor i in the T valuation given the realisation of a single systematic factor is given by (3.3.48). Substituting this into (3.3.36) and integrating over all possible realisations of the systematic factor results in an expression of the $100(1-\alpha)\%$ confidence region of the probability of default:

$$\int_{-\infty}^{\infty}\phi(x)\sum_{k=0}^{d_j^*}\binom{s_j^*}{k}\pi(X_1,...,X_T)^k\pi(X_1,...,X_T)^{s_j^*-k}dx \geq 1-\alpha. \qquad (3.3.50)$$

The left hand side of the expression above is the probability of observing no more that d_j^* defaults, which can also be written as:

$$\begin{aligned}
P[\text{No more that } d_j^* \text{ defaults observed}] &= \int_{-\infty}^{\infty}\phi(x)\sum_{k=0}^{d_j^*}\binom{s_j^*}{k}\pi(X_1,...,X_T)^k\pi(X_1,...,X_T)^{s_j^*-k}dx \\
&= \sum_{k=0}^{d_j^*}\mathbb{E}\big[P[\text{Exactly } k \text{ obligors default}|X_1,...,X_T]\big] \\
&= \sum_{k=0}^{d_j^*}\binom{s_j^*}{k}\mathbb{E}\big[\pi(X_1,...,X_T)^k(1-\pi(X_1,...,X_T))^{s_j^*-k}\big].
\end{aligned}$$

These probabilities can be estimated using Monte Carlo Simulation.

Final Comments. The *most prudent estimation* methodology developed by Pluto and Tasche (2011) can be used for a range of applications in the wider banking environment. The methodology is however constrained by the problem surrounding specification of an appropriate confidence level. In determining the confidence level, there is the risk of potential underestimation of the average PD.

The confidence based approach proposed by Pluto and Tasche is widely used in practice, Tasche (2013) makes the following comments on the method in question;

- Firstly, *most prudent estimation* methodology is criticised for delivering too conservative estimates. This view is supported by Chang and Yu (2014).
- The methodology is widely used in practice and this interest is expected to be encouraged by the FSA's requirement as stated in BIPRU (2011);

  > *"...a firm must use a statistical technique to derive the distribution of defaults implied by the firm's experience, estimating PDs from the upper bound of a confidence interval set by the firm in order to produce conservative estimates of PDs..."*

- A further point of criticism of the approach is the element of subjectivity introduced by the fact that three parameters need to be predefined in the multi-period approach.

The approach discussed in this section is supported by the likelihood method proposed by Forrest (2005). The likelihood approach resembles the confidence based approach in the fact that it also assumes a binomial model for modelling the observed default data. However in contrast to Pluto and Tasche (2011), Forrest (2005) proposes the use of likelihood and likelihood ratio to estimate the PD. The likelihood ratio method however does not avoid the problem of specifying a confidence level.

Researchers such as Forrest (2005) and Benjamin *et al.* (2006) suggested modifications of the confidence based approach discussed in Section 3.3.1. These modifications supposedly address the structural conservatism ingrained in this approach. As mentioned in Tasche (2013), many other researchers seek to find alternative methodologies to statistically based low default probability of default estimation. The need for specification of the confidence level can be avoided by considering the Bayesian methodology and hence this seems most promising possibility.

3.3.2 Bayesian Approach

Bayesian theory was already introduced into the framework of probability of default estimation in Section 2.2.4. In the stated section the Bayesian setting is however only mentioned in brief. Therefore, before proceeding in utilisation of Bayesian approaches to address the problem of estimating the probability of default for low default portfolios, a more detailed discussion of the Bayesian Approach is required.

In Bayesian statistics the true value of a parameter is regarded as a random variable. This random variable is assigned a probability distribution known as the prior distribution. As previously mentioned in Section 2.2.4, an obvious model to model the defaults, L, is the Bernoulli distribution, where:

$$L_i \sim B(1, p_i), i = 1, 2, \ldots, s$$

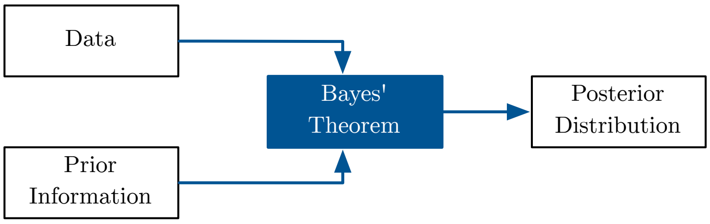


Figure 3.4: The Bayesian Method

and p_i is the probability of default for a portfolio of s credit exposures. Limited information is available on the defaults, however experts may have some knowledge of the probability of default, p_i.

As discussed in Rice (2007), in the Bayesian approach to parameter estimation the unknown parameter p is treated as a random variable, with a "prior" distribution, G, and a corresponding probability density function, $g_P(p)$. This distribution represents what is known regarding the parameter, before the data, L, has been observed. Then, for a given value, $P = p$, the probability density function of the data is denoted as, $f_{L|P}(l|p)$. The joint probability density of L and P follows as:

$$f_{L,P}(l,p) = f_{L|P}(l|p)g_P(p). \tag{3.3.51}$$

The marginal probability density function of the data, L, can then be obtained as:

$$\begin{aligned} f_L(l) &= \int f_{L,P}(l,p)dp \\ &= \int f_{L|P}(l|p)g_P(p)dp. \end{aligned} \tag{3.3.52}$$

The distribution of L given the data follows from Bayes' Rule, which is defined as:

Definition 3.9. Bayes' Rule: *Let A and $B_1, \ldots, B_n$ be events where the B_i are disjoint, with $\cup_{i=1}^n B_i = \Omega$ and $P(B_i) > 0$ for all i. Then:*

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^n P(A|B_i)P(B_i)} = \frac{P(A|B_j)P(B_j)}{P(A)},$$

Rice (2007).

Hence, from Bayes' rule follows:

$$\begin{aligned} f_{P|L}(p|l) &= \frac{f_{L,P}(l,p)}{f_L(l)} \\ &= \frac{f_{L|P}(l|p)g_P(p)}{\int f_{L|P}(l|p)g_P(p)dp.} \end{aligned}$$

The expression above is called the "posterior" density function, this denotes what is known about P after the data has been observed. A final note is that $f_{L|P}(l|p)$ is the well known likelihood function and this is a function of p, from which the result can be summarised as:

$$\begin{gathered} f_{P|L}(p|l) \propto f_{L|P}(l|p) \times g_P(p) \\ \text{Posterior Density} \propto \text{Likelihood} \times \text{Prior density.} \end{gathered} \tag{3.3.53}$$

Figure 3.4 attempts to provide a basic graphical illustration of how the Bayesian Approach combines the data with prior knowledge.

In order to extend the aforementioned theory into the context of probability of default estimation; assume that d defaults were observed in the portfolio of s credit exposures, where:

$$d = \sum_{i=1}^{s} l_i,$$

and:

$$L_i \sim B(1, p_i), i = 1, 2, \ldots, s.$$

The likelihood function, $f_{\mathbf{L}|P}(d, p)$, under the assumption of independent defaults, was already determined in Section 2.2.4. Recall from expression (2.2.27) that:

$$f_{\mathbf{L}|P}(d|p) = P(\mathbf{L} = d|P = p) = \binom{s}{d} p^d (1-p)^{s-d},$$

where $\mathbf{L} = (L_1, \ldots, L_s)$ is the complete default indicator of the portfolio, and recall that the probability of default for all obligors in the same rating grade is assumed equal. Therefore the joint probability density function follows as:

$$\begin{aligned} f_{\mathbf{L},P}(d, p) = P(\mathbf{L} = d, P \leq p) &= f_{\mathbf{L}|P}(d|p) g_P(p) \\ &= \int_0^p f_{\mathbf{L}|P}(d|\theta) g_P(\theta) d\theta \\ &= \int_0^p \binom{s}{d} \theta^d (1-\theta)^{s-d} g_P(\theta) d\theta. \end{aligned}$$

Since the likelihood function is known, the only outstanding component required to determine the Bayesian posterior density is the prior distribution and its prior density function. Before considering the prior distribution, the likelihood function will be extended to incorporate dependence between defaults, as well as multiple time periods.

As discussed in Section 2.2.4.3, dependence between default events can be incorporated through a Single-Factor Gaussian Copula. In the Single-Factor Gaussian Copula the dependence structure between default events can be captured through a single common systematic factor, $X \sim \Phi(0, 1)$. As before, assume that the asset correlation between any two pairs of obligors is constant and denoted by, ξ. The probability of default conditional on the realisation of the systematic factor was previously defined in Section 2.2.4.3 as:

$$\lambda_\xi(p, X) = \Phi\left(\frac{\Phi^{-1}(p) - \sqrt{\xi} X}{\sqrt{1-\xi}}\right).$$

Substituting this into the likelihood function for the independent case and integrating over all possible scenarios underlying the systematic factor, the likelihood function in the dependent setting follows as:

$$f_{\mathbf{L}|P}(d|p, \xi) = P(\mathbf{L} = d|P = p) = \int_{-\infty}^{\infty} \binom{s}{d} \lambda_\xi(p, x)^d (1 - \lambda_\xi(p, x))^{s-d} \phi(x) dx.$$

This results in the following joint density function:

$$\begin{aligned} f_{\mathbf{L},P}(d,p) = P(\mathbf{L}=d, P \le p) = f_{\mathbf{L}|P}(d|p)g_P(p) \\ &= \int_{-\infty}^{\infty}\int_0^p f_{\mathbf{L}|P}(d|\theta)g_P(\theta)dxd\theta \\ &= \int_{-\infty}^{\infty}\int_0^p \binom{s}{d}\lambda_\xi(\theta,x)^d(1-\lambda_\xi(\theta,x))^{s-d}g_P(\theta)\phi(x)dxd\theta. \end{aligned}$$

The theory can now be extended to a multi-period case using the same dynamic factor model discussed in the multi-period extension of Section 3.3.1.

As discussed earlier, the single component X_t of the vector of systematic factors generate the cross-sectional correlation of the default events at time t. The intertemporal correlation is affected by the dependence structure of the factors $X_1,\dots,X_T$, written as a vector $\mathbf{X} = (X_1,\dots,X_T)$. An assumption on the vector is that not only the individual components, but also the vector as a whole is multivariate normally distributed. The individual components are standardised and hence the joint distribution is completely determined by the correlation matrix:

$$\Sigma_\tau = \begin{pmatrix} 1 & \tau^{-1} & \tau^{-2} & \cdots & \tau^{1-T} \\ \tau^{1} & 1 & \tau^{-1} & \cdots & \tau^{2-T} \\ \vdots & & \ddots & & \vdots \\ \tau^{T-2} & \cdots & \tau^{1} & 1 & \tau^{-1} \\ \tau^{T-1} & \cdots & \tau^{-2} & \tau^{1} & 1 \end{pmatrix}. \tag{3.3.54}$$

It is known that the probability of the ith obligor defaulting in time t is defined as:

$$P(L_{t,i}=1) = p_{t,i} = P(Z_{t,i} \le D_i)$$

where D_i is the default threshold of obligor i. Since the probability is conditional on ξ and τ, write:

$$\begin{aligned} p_{t,i} &= P(L_{t,i}=1|\xi,\tau) \\ &= P(Z_{t,i} \le D_{t,i}|\xi,\tau) \\ &= \Phi(D_{t,i}) \end{aligned}$$

where Φ is the standard normal cumulative distribution function. As before:

$$D_{t,i} = \Phi^{-1}(p_{t,i}).$$

From the uniformity assumption it follows that when obligor's fall into the same rating category, then $p_{t,i} = p$ and $D_{t,i} = D$ for all t and i.

The complete default indicator of the portfolio in all time periods is defined as:

$$\mathbf{L} = (\mathbf{L}_1', \mathbf{L}_2', \dots, \mathbf{L}_T')',$$

where $\mathbf{L}_t = (L_{t,1},\dots,L_{t,s})'$. The unobserved vector of the realisations of the systematic factors is defined as $\mathbf{X} = (X_1, X_2, \dots, X_T)'$. It is known that the probability of default, conditional on p, ξ, τ and X_t can be written as:

$$\lambda_t(p,\xi,X_t) = \Phi\left(\frac{\Phi^{-1(p)} - \sqrt{\xi}X_t}{\sqrt{1-\xi}}\right).$$

It then follows that the joint conditional density function of $\mathbf{L}$ given p, ξ, τ and $\mathbf{X}$ can be written as:

$$
\begin{aligned}
f_{\mathbf{L}|P}(\mathbf{l}|p,\xi,\mathbf{X}) &= \prod_{t=1}^{T}\prod_{i=1}^{s_t}\lambda_t(p,\xi,X_t)^{l_{t,i}}(1-\lambda_t(p,\xi,X_t))^{1-l_{t,i}} && (3.3.55)\\
&= \prod_{t=1}^{T}\lambda_t(p,\xi,X_t)^{\sum_{i=1}^{s_t} l_{t,i}}(1-\lambda_t(p,\xi,X_t))^{\sum_{i=1}^{s_t}(1-l_{t,i})}\\
&= \prod_{t=1}^{T}\lambda_t(p,\xi,X_t)^{d_t}(1-\lambda_t(p,\xi,X_t))^{s_t-d_t} && (3.3.56)
\end{aligned}
$$

Furthermore, as highlighted in Kruger (2015), integrating over all possible outcomes of the systematic factor, the conditional probability of observing d_1 defaults in time 1, d_2 defaults in time 2, ... ,d_T defaults in time T, given ξ and τ can be written as:

$$
\begin{aligned}
P(L_1 = d_1,\dots,L_T = d_T|P=p) = \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\phi_\Sigma(x_1,\dots,x_T)\times\\
\prod_{t=1}^{T}\binom{s_t}{d_t}\lambda_t(p,\xi,x_t)^{d_t}(1-\lambda_t(p,\xi,x_t))^{s_t-d_t}dx_1\dots dx_T,
\end{aligned}
\qquad (3.3.57)
$$

where $\phi_\Sigma(x_1,\dots,x_T)$ is the joint probability density function of the multivariate normal distribution defined by the correlation matrix, Σ, given in (3.3.54). As mentioned, prior information on the probability of default can be incorporated by assuming a prior distribution on the parameter p with probability density function $g_P(p)$. The joint density function for the multi-period case assuming dependence between default events therefore follows as:

$$
\begin{aligned}
f_{\mathbf{L},P}(d,p) = P(\mathbf{L}=d, P\le p) &= f_{\mathbf{L}|P}(d|p)g_P(p)\\
&= \int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\int_0^p\phi_\Sigma(x_1,\dots,x_T)\times\\
&\prod_{t=1}^{T}\binom{s_t}{d_t}\lambda_t(\theta,\xi,x_t)^{d_t}(1-\lambda_t(\theta,\xi,x_t))^{s_t-d_t}g_P(\theta)d\theta dx_1\dots dx_T.
\end{aligned}
\qquad (3.3.58)
$$

For a specified prior distribution it is possible to derive a posterior distribution of the probability of default given the data. In the context of estimating the probability of default for low default portfolios, incorporating prior information as discussed in this section is regarded as an effective approach in overcoming the issue of data scarcity.

In the framework discussed thus far in this section, the prior distribution has only been defined in general terms. By specifying a prior distribution a belief regarding the PD is taken and the PD estimate is calibrated.

It is important to take note that the unconditional distribution of X_t is $\Phi(0,1)$, however from the dynamic factor model X_t is dependent on X_{t-1} and τ. So the conditional distribution of X_t given X_{t-1} and τ is:

$$X_t|X_{t-1} \sim \Phi(\tau x_{t-1}, 1-\tau^2).$$

From this the joint conditional density function of $\mathbf{X}$ given τ is:

$$
\begin{aligned}
f(\mathbf{X}|\tau) &= f(x_1)\times f(x_2|x_1,\tau)\times f(x_3|x_1,x_2,\tau)\times\dots\times f(x_T|x_1,x_2,\dots,x_{T-1},\tau)\\
&= f(x_1)\times f(x_2|x_1,\tau)\times f(x_3|x_2,\tau)\times\dots\times f(x_T|x_{T-1},\tau).
\end{aligned}
\qquad (3.3.59)
$$

3.3.2.1 The Choice of Prior Distributions

As pointed out in Berning (2010), two main categories of prior distributions exist. The fist category, and also the one that is predominantly considered in PD calibration for LDPs, is subjective prior distributions. For a subjective prior distribution, the manner in which the prior distribution is selected reflects the subjective opinion of the risk manager concerning the distribution of the defaults. Hence, it is said that subjective prior distributions are constructed with expert information, where the expert information is the risk managers experience.

It should be highlighted that subjectivity may vary from one expert to another, and therefore it is difficult to justify the choice of a subjective prior distribution. As such objective prior distributions have been constructed. An objective prior distribution assumes no subjective or expert information. These objective prior distributions are derived from the assumed probability density function of the data, $f_{\mathbf{L}|P}(d|p)$. There exists an array of different objective prior distributions, popular examples include Jeffrey's prior and the reference prior (Berning, 2010).

As mentioned, most of the prior distributions proposed thus far in the literature on PD calibration for LDPs falls into subjective prior distribution category. The reason for this being that risk managers have expert information and believe that this may assist in the calibration of the PD in the LDP setting. As further pointed out, different risk managers would have different views on the distribution of the PD for LDPs. This is why an array of prior specifications have been proposed in the literature. Some of the major proposals will be considered in the next subsection.

3.3.2.2 Prior Distributions

Naturally the posterior distributions discussed earlier are subject to specification of a prior distribution. The prior distribution allows specification of what the value of the PD may be, this belief is then updated by the observations. The posterior density functions corresponding to the prior distributions mentioned in this section will be derived in a later section.

i) The Uniform Prior Distribution: The first prior considered is the uniform prior proposed by Dwyer (2006), this specification treats the PD as being unknown and assumes an uninformed or objective prior distribution. Therefore, it is assumed that there is no knowledge of what the realised PD might be. Intuitively, the PD is assumed to be uniformly distributed between 0 and 1. Therefore:

$$P(P < p) = p \tag{3.3.60}$$

and

$$g_P(p) = 1. \tag{3.3.61}$$

When a uniform prior is specified, this reflects a position where there is no expectations about the distribution of the PD as all PDs are assumed equally possible.

Using the uniform distribution as a prior distribution would also allow the risk manager to take on a tailored approach, by specifying bounds for the PD (this introduces subjectivity). Say for example there exists a viewpoint that the PD will take on a value between p_l and p_u, i.e. $0 < p_l \leq p \leq p_u < 1$. In this case all PDs between these bounds are assumed equally likely and it follows that:

$$P(P < p) = \begin{cases} 0 & \text{for } p < p_l \\ \frac{p-p_l}{p_u-p_l} & \text{for } p_l \leq p < p_u \\ 1 & \text{for } p \geq p_u \end{cases} \tag{3.3.62}$$

and

$$g_P(p) = \begin{cases} \frac{1}{p_u-p_l} & \text{for } p_l \leq p < p_u \\ 0 & \text{for } p < p_l \text{ or } p \geq p_u. \end{cases} \tag{3.3.63}$$

In the LDP context specification of a lower bound, p_l, that is larger than 0 is highly unlikely. So the prior distribution of the PD can generally be regarded a uniform distribution on the interval $(0, p_u)$ where $0 < p_u \leq 1$. Tasche (2013) defines this as the $(0, p_u)$-constrained neutral Bayesian estimator of the PD, p. For $(p_u = 1)$ the unconstrained neutral Bayesian estimator is obtained this is an objective (on uninformative) prior distribution.

ii) The Beta Distribution: Returning to specification of the prior distribution as $g_P(p) = 1$ for $0 < p < 1$. As discussed in Kiefer (2009), this is an uninformative or unobjectionable prior distribution due to the fact that it assigns equal probability to equal length subsets of $[0, 1]$. The mean of the uniform distribution is 0.5, this is a highly unlikely prior expectation of the PD, especially for LDP.

As further discussed in Kiefer (2009), a generalisation of the uniform distribution that is commonly used for a parameter that is said to lie on $[0, 1]$ is the beta distribution. If it is assumed that $G \sim Beta(\alpha, \beta)$, the probability density function follows as:

$$g_P(p|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}. \tag{3.3.64}$$

The expected value of the Beta distribution is:

$$\mathbb{E}(p) = \frac{\alpha}{\alpha + \beta}.$$

The Beta distribution is an extremely flexible distribution and can take on various different shapes on the interval $[0, 1]$. The flexibility of the Beta-distribution is illustrated in Figure 3.5.

Kiefer (2009) discusses a four parameter extension of the distribution. This is essentially a modification of the distribution to have support on $[p_l, p_u]$ through a transformation. This four-parameter distribution allows flexibility within the range $[p_l, p_u]$, however in certain cases it may be regarded as too restrictive and hence this extension will not be considered.

A further generalisation mentioned in Kiefer (2009) is to consider a seven-parameter mixture distribution of two four-parameter distributions, however it is mentioned in the literature that it is unlikely to obtain sufficient expert information for such a representation. It is however noteworthy that the specification of enough beta-mixture terms would allow the approximation of an arbitrary continuous prior $g_P(p)$ for a Bernoulli parameter to be forged as arbitrarily accurate.

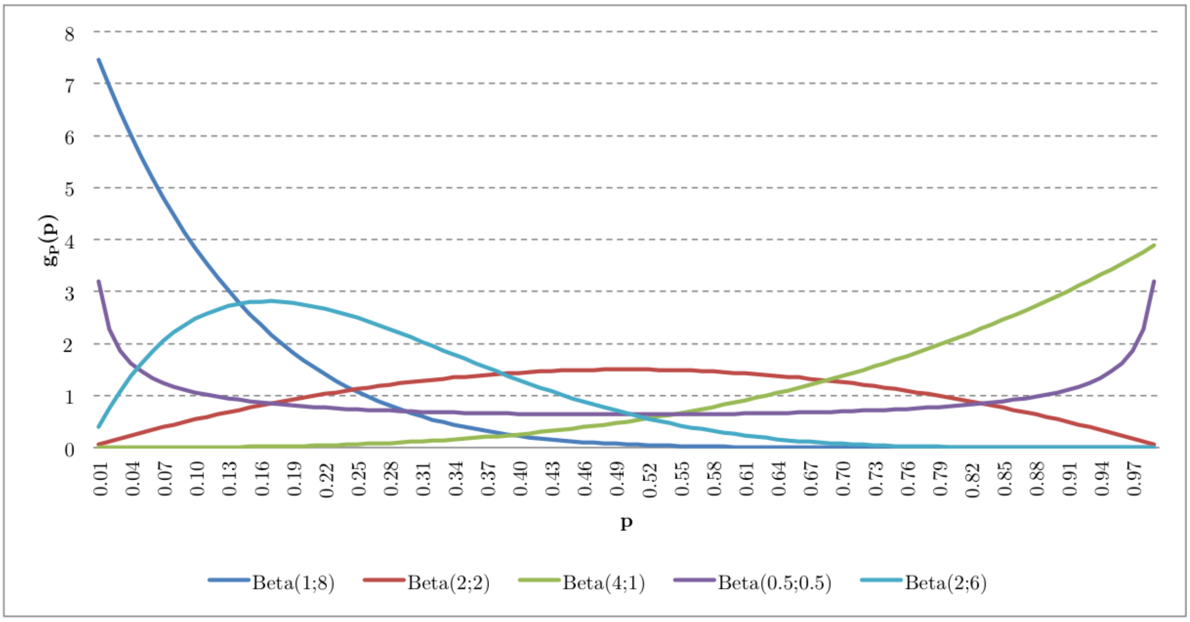


Figure 3.5: Plot of the Beta Prior for different parameters.

iii) The Conservative Prior: The conservative prior distribution was first proposed by Dwyer (2006), before being elaborated on in Tasche (2013). This prior distribution was proposed through a comment Dwyer (2006) made on the approach of Pluto and Tasche (2011), of a possible interpretation of the Pluto and Tasche (2011) independent one-period case in Bayesian terms.

This resulted in the specification of an unconditional prior cumulative distribution function on $0 < p < 1$ that is written as:

$$P(P \leq p) = G(P) = \int_0^p \frac{1}{1-\theta} d\theta = -\log(1-p), \tag{3.3.65}$$

so that the probability density function follows as:

$$g_P(p) = \begin{cases} \frac{1}{1-p} & \text{for } 0 < p < 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.3.66}$$

Tasche (2013) labelled the prior distribution a conservative prior distribution due to the fact that the distribution defined above appears biased towards higher values. This can be seen in the plot of the conservative prior density function, which is given in Figure 3.6.

iv) The Expert Distribution: Clifford *et al.* (2013) proposed a modification of the uniform prior distribution defined on $[p_l, p_u]$ in the form of a triangle distribution. This approach is defined as the expert distribution. In said approach expert judgement is used to specify a minimum PD, p_l, a maximum PD, p_u, and a most likely PD (or the mode PD), p_m, where $0 \leq p_p < p_m < p_l \leq 1$. The cumulative distribution function is given by:

$$P(P < p) = G(P) = \begin{cases} 0 & \text{for } 0 < p \leq p_l \\ \frac{(p-p_l)^2}{(p_u-p_l)(p_m-p_l)} & \text{for } p_l < p \leq p_m \\ 1 - \frac{(p_u-p)^2}{(p_u-p_l)(p_l-p_m)} & \text{for } p_m < p \leq p_u \\ 1 & \text{for } p_u < p \leq 1 \end{cases} \tag{3.3.67}$$

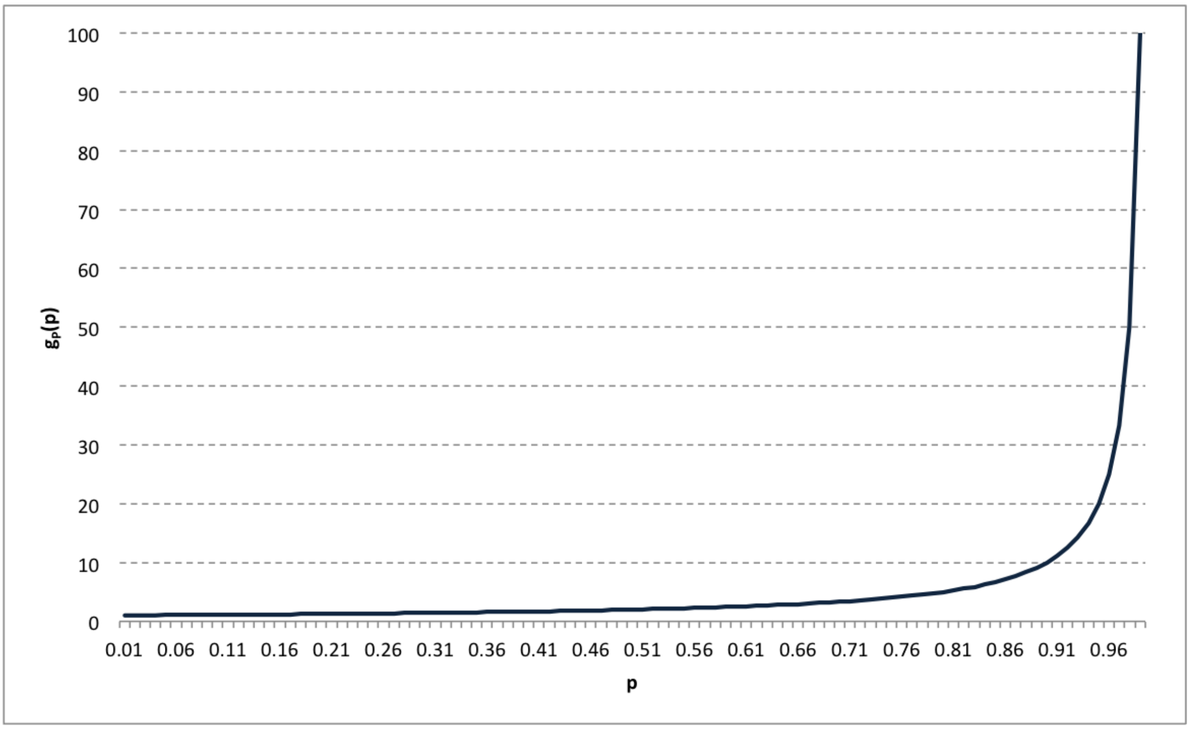


Figure 3.6: Plot of the Conservative Prior

with probability density function:

$$g_P(p) = \begin{cases} 0 & \text{for } 0 \leq p < p_l \\ \frac{2(p-p_l)}{(p_u-p_l)(p_m-p_l)} & \text{for } p_l \leq p < p_m \\ \frac{2}{(p_u-p_l)} & \text{for } p = p_m \\ \frac{2(p_u-p)}{(p_u-p_l)(p_l-p_m)} & \text{for } p_m < p \leq p_u \\ 0 & \text{for } p_u < p \leq 1 \end{cases} . \tag{3.3.68}$$

A plot of the prior distribution defined by the expert distribution is given in Figure 3.7, in the plot the parameters are specified as $p_l = 1\%, p_m = 4\%$ and $p_u = 6\%$. From the plot the flexibility of the distribution should be clear. Also, it is obvious that the expert distribution allows expert judgement to be incorporated in an intuitive manner. The expectation of the prior distribution follows as:

$$\mathbb{E}(p) = \frac{p_u + p_m + p_l}{3}. \tag{3.3.69}$$

As discussed by Clifford *et al.* (2013), the prior defined above can be used to incorporate management expectations through specification of the parameters using expert judgement. The prior can also be linked to industry benchmarks through these parameters. On the downside, the parameter estimates introduces a degree of subjectivity. Also, the prior may lead to less conservative results.

3.3.3 Estimation and Simulation

Numerous authors such as Kruger (2015), Chang and Yu (2014) and Tasche (2013) provide maximum likelihood expressions that can be used to obtain parameter estimates for the specified prior distribution. However, due to the numerical complexity involved in evaluating these

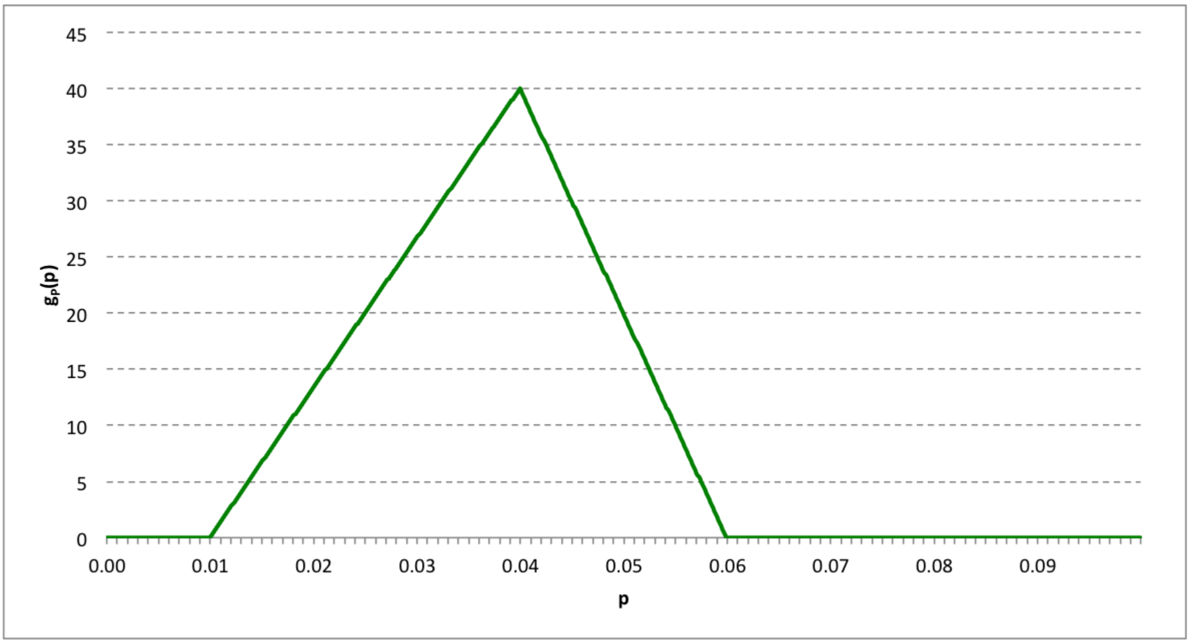


Figure 3.7: Plot of the Expert Distribution

expressions, it will not be considered. As a solution to this problem, Chang and Yu (2014) proposes use of the Generalised Method of Moments with Continuous Updating of Hansen (1982); Hansen *et al.* (1996). This will also not be considered and the initial parameters for the prior distributions will be specified by user as is common in practice.

At this stage it is important to take note that there are four different parameters in the model. These parameters are the probability of default, p, the asset correlation, ξ, the T- dimensional vector of systematic factors, $\mathbf{X}$, and the intertemporal correlation, τ. Basel provides guidelines for specifying the asset correlation (see Chapter 2), furthermore as discussed in Pluto and Tasche (2011) a reasonable specification of τ is 0.3 (this will be elaborated on in Chapter 5). Therefore, the parameters ξ and τ can be regarded as inputs.

Chang and Yu (2014) however proposes specifying prior distributions for the asset correlation and intertemporal correlation. Such an approach using multiple prior distributions over complicates the situation unnecessarily and hence this will not be considered.

Since two of the parameters are taken as inputs, there are $T+1$ parameters that needs estimation. As mentioned in Section 3.3.2, the realisation of the latent factor X_t depends on the specification of τ as well as the realisation in the previous time period, X_{t-1}. Furthermore, the estimate of p depends on the specification of ξ and τ as well as the realisation of the latent factor $\mathbf{X}$. Therefore, a joint posterior density of p, ξ, $\mathbf{X}$ and τ can be expressed as:

$$h(p, \xi, \mathbf{X}, \tau|\mathbf{l}) = \frac{f(\mathbf{l}|p, \xi, \mathbf{X}, \tau)g(p, \xi, \mathbf{X}, \tau)}{f(\mathbf{l})},$$

where $\mathbf{l} = (l_1, \ldots, l_t)'$. Dropping the parameters specified as inputs, this becomes:

$$h(p, \mathbf{X}|\mathbf{l}) = \frac{f(\mathbf{l}|p, \mathbf{X})g(p, \mathbf{X})}{f_{\mathbf{L}}(\mathbf{l})}, \tag{3.3.70}$$

where $g(p, \mathbf{X})$ is the joint prior density function of p and $\mathbf{X}$ and $f_{\mathbf{L}}(\mathbf{l})$ is the density function of the data $\mathbf{L}$. Based on expression (3.3.70), the marginal posterior density function of the probability of default, p, is given by:

$$h(p|\mathbf{l}) = \int_{-\infty}^{\infty} \dots \int_{\infty}^{\infty} h(p, \xi, \mathbf{X}, \tau|\mathbf{l}) dx_1 \dots dx_T. \tag{3.3.71}$$

This density contains T integrals, which makes it numerically difficult to evaluate. As such an Markov Chain Monte Carlo (MCMC) algorithm called the Gibbs Sampler is applied. The MCMC algorithm is used to estimate a posterior distribution from the parametric sampling distribution. The Gibbs Sampler then updates each of the two underlying variables by sampling from their conditional density functions, given the current sampled values of the other variable. This will be discussed in greater detail in Chapter 5. At this stage, it is however important to note that the partial conditional density function is an important component of the Gibbs sampler. The fundamental partial conditional density is the joint density function of the data and the unknown parameters, given by:

$$f(\mathbf{l}, p, \mathbf{X}) = f(\mathbf{L}|p, \mathbf{X}) g_P(p) f(\mathbf{X}|\tau). \tag{3.3.72}$$

The partial conditional function for the unknown parameters can now be derived. First, recall that:

$$f_{\mathbf{L}}(\mathbf{l}|p, \xi, \mathbf{X}) = \prod_{t=1}^{T} \prod_{i=1}^{s_t} \lambda_t(p, \xi, x_t)^{l_{t,i}} (1 - \lambda_t(p, \xi, x_t))^{1-l_{t,i}}.$$

Firstly, the marginal conditional posterior density for the latent variable $\mathbf{X}$ is required. The marginal conditional posterior density functions of $\mathbf{X}$ are considered separately for X_1 and $X_t, t = 2, \dots, T$. First of all for $X_1 \sim N(0, 1)$ where:

$$f_{X_1}(x_1) \propto \exp\left(-\frac{x_1^2}{2} \right)$$

the marginal conditional posterior density is given by:

$$\begin{aligned} h(x_1|\mathbf{L}, p, \xi, \mathbf{x}_{-1}) &\propto f_{\mathbf{L}}(\mathbf{l}|p, \xi, \mathbf{x}) \times f_{X_1}(x_1) \\ &\propto \left\{ \prod_{t=1}^{T} \prod_{i=1}^{s_t} \lambda_t(p, \xi, x_t)^{l_{t,i}} (1 - \lambda_t(p, \xi, x_t))^{1-l_{t,i}} \right\} \\ &\times \exp\left(-\frac{x_1^2}{2} \right). \end{aligned} \tag{3.3.73}$$

Then for $X_t = \tau X_{t-1} + \sqrt{1-\tau^2} w_t$, where $w_t \sim N(0, 1)$ and:

$$f_{X_t}(x_t|x_{t-1}, \tau) = \exp\left(-\frac{(x_t - \tau x_{t-1})^2}{2(1-\tau^2)} \right),$$

for $t = 2, \dots, T$ it follows that:

$$\begin{aligned} h(x_t|\mathbf{L}, p, \xi, \mathbf{x}_{-t},) &\propto f(\mathbf{L}|p, \xi, \mathbf{x}) \times f(x_t|\mathbf{x}_{-t}, \tau) \\ &\propto f(\mathbf{L}|p, \xi, \mathbf{x}) \times f(x_t|x_{t-1}, \tau) \\ &\propto \left\{ \prod_{t=1}^{T} \prod_{i=1}^{s_t} \lambda_t(p, \xi, x_t)^{l_{t,i}} (1 - \lambda_t(p, \xi, x_t))^{1-l_{t,i}} \right\} \\ &\times \exp\left(-\frac{(x_t - \tau x_{t-1})^2}{2(1-\tau^2)} \right), t = 2, \dots, T, \end{aligned} \tag{3.3.74}$$

where $\mathbf{x}_{-t} = (x_1, \dots, x_{t-1}, x_{t+1}, \dots, x_T)'$. The marginal conditional posterior density function of p depends on the choice of the prior distribution.

Marginal Conditional Posterior for the Uniform Prior Distribution: For:

$$P \sim Uniform(p_l, p_u)$$

it is known that:

$$g_P(p) = \begin{cases} \frac{1}{p_u - p_l} & \text{for } p_l \leq p < p_u \\ 0 & \text{for } p < p_l \text{ or } p \geq p_u. \end{cases}$$

Therefore the marginal conditional posterior density for the prior distribution in question follows as:

$$\begin{aligned} h(p|\mathbf{L}, \xi, \mathbf{x}) &\propto f(\mathbf{L}|p, \xi, \mathbf{x}) \times g_P(p) \\ &\propto \left\{ \prod_{t=1}^{T} \prod_{i=1}^{s_t} \lambda_t(p, \xi, x_t)^{l_{t,i}} (1 - \lambda_t(p, \xi, x_t))^{1-l_{t,i}} \right\} \\ &\times \frac{1}{p_u - p_l}, \end{aligned} \tag{3.3.75}$$

for $p_l \leq p < p_u$, where $h(p|\mathbf{L}, \xi, \mathbf{x}, \tau) = 0$ otherwise.

Marginal Conditional Posterior for the Beta Prior Distribution: The marginal conditional posterior density function of:

$$p \sim Beta(\alpha_p, \beta_p)$$

where:

$$g_P(p) = p^{\alpha_p - 1}(1 - p)^{\beta_p - 1},$$

is given by:

$$\begin{aligned} h(p|\mathbf{L}, \xi, \mathbf{x}) &\propto f(\mathbf{L}|p, \xi, \mathbf{x}) \times g_P(p) \\ &\propto \left\{ \prod_{t=1}^{T} \prod_{i=1}^{s_t} \lambda_t(p, \xi, x_t)^{l_{t,i}} (1 - \lambda_t(p, \xi, x_t))^{1-l_{t,i}} \right\} \\ &\times p^{\alpha_p - 1}(1 - p)^{\beta_p - 1}. \end{aligned} \tag{3.3.76}$$

Marginal Conditional Posterior for the Conservative Prior Distribution: The marginal conditional posterior density function of the conservative prior, where:

$$g_P(p) = \frac{1}{1 - p}.$$

is given by:

$$\begin{aligned} h(p|\mathbf{L}, \xi, \mathbf{x}) &\propto f(\mathbf{L}|p, \xi, \mathbf{x}) \times g_P(p) \\ &\propto \left\{ \prod_{t=1}^{T} \prod_{i=1}^{s_t} \lambda_t(p, \xi, x_t)^{l_{t,i}} (1 - \lambda_t(p, \xi, x_t))^{1-l_{t,i}} \right\} \\ &\times \frac{1}{1 - p}. \end{aligned} \tag{3.3.77}$$

Marginal Conditional Posterior for the Expert Prior Distribution: The marginal conditional posterior density function of the conservative prior, where:

$$g_P(p) = \begin{cases} 0 & \text{for } 0 \leq p < p_l \\ \frac{2(p-p_l)}{(p_u-p_l)(p_m-p_l)} & \text{for } p_l \leq p < p_m \\ \frac{2}{(p_u-p_l)} & \text{for } p = p_m \\ \frac{2(p_u-p)}{(p_u-p_l)(p_l-p_m)} & \text{for } p_m < p \leq p_u \\ 0 & \text{for } p_u < p \leq 1 \end{cases}.$$

is given by:

$$\begin{aligned} h(p|\mathbf{L},\xi,\mathbf{x}) &\propto f(\mathbf{L}|p,\xi,\mathbf{x}) \times g_P(p) \\ &\propto \left\{ \prod_{t=1}^{T} \prod_{i=1}^{s_t} \theta_t(p,\xi,x_t)^{l_{t,i}} (1-\theta_t(p,\xi,x_t))^{1-l_{t,i}} \right\} \\ &\times \frac{2(p-p_l)}{(p_u-p_l)(p_m-p_l)}, \end{aligned} \tag{3.3.78}$$

for $p_l \leq p \leq p_m$ and:

$$\begin{aligned} h(p|\mathbf{L},\xi,\mathbf{x}) &\propto f(\mathbf{L}|p,\xi,\mathbf{x}) \times g_P(p) \\ &\propto \left\{ \prod_{t=1}^{T} \prod_{i=1}^{s_t} \theta_t(p,\xi,x_t)^{l_{t,i}} (1-\theta_t(p,\xi,x_t))^{1-l_{t,i}} \right\} \\ &\times \frac{2(p_u-p)}{(p_u-p_l)(p_l-p_m)}, \end{aligned} \tag{3.3.79}$$

for $p_m < p \leq p_u$, where $h(p|\mathbf{L},\mathbf{x}) = 0$ otherwise.

Simulation of the derived marginal conditional posterior density functions will be discussed in Chapter 5. Simulation through MCMC methods yields the marginal posterior distribution, from which the parameter estimates can be obtained. Even though realisations of the latent factor $\mathbf{X}$ is simulated and therefore an estimate of this factor can be obtained, the only parameter of concern in this study is the probability of default estimate.

As is well known from literature, see for example Berning (2010), when a quadratic or a square loss is assumed, then the Bayesian estimate of p is the expected value (or mean) of its marginal posterior distribution, $h(p|\mathbf{L},\xi,\mathbf{X})$, this can be written as:

$$\hat{p} = \int p h(p|\mathbf{L},\xi,\mathbf{X}) dp.$$

When absolute error loss is assumed, then the Bayesian estimate of p is the median of the marginal posterior distribution. Finally, when $0-1$ loss is assumed, then the Bayesian estimate of p is the mode of the marginal posterior distribution. Further comments on the choice of the loss function will be made in Chapter 5.

3.4 Conclusion

This is a fundamental chapter of the study where the problem of calibrating the probability of default for low default portfolios is addressed. In Section 3.1, the concept of a Low Default

Portfolio is defined and the characteristics of these portfolios are discussed. The industry concerns regarding LDPs is then discussed in Section 3.2, before proceeding to the pivotal section of the literature study.

In Section 3.3, some of the main proposed PD calibration approaches are discussed in detail. The two main approaches that are considered in this study is the Confidence Based Approach as proposed by Pluto and Tasche (2011) (Section 3.3.1) and the Bayesian Approach (Section 3.3.2). The latter is reconsidered in Chapter 4, where the incorporation of an extremal distribution is proposed as an alternative to the prior distributions discussed in Section 3.3.2.2.

Chapter 4

Proposing the Pareto Distribution as a Prior Distribution

This chapter aims to provide deeper insight of the Strict Pareto distribution as an alternative prior distribution. In Section 4.1, the comments are made on the prior distributions discussed in Section 3.3.2.2. The Bayesian Approach is then reconsidered in Section 4.2; it is shown that the Pareto distribution can be considered as a possible prior distribution through applying a transformation on the PD.

4.1 Discussing the Prior Distributions from Section 3.3.2.2

Before proceeding in proposing a new prior distribution that may assist in capturing the low probability of default present in Low Default Portfolios, the distributions discussed in Section 3.3.2 is revisited.

Uniform Distribution: Firstly, consider the Uniform prior distribution. Specification of a Uniform prior distribution constrained on $(0, 1)$ makes sense from a statistical point of view in the event that no prior knowledge of the probability of default exists. However, it is known that the PD is very low and close to zero. This brings the use of a Uniform$(0, 1)$ prior distribution into question as the model places equal weight on higher realisations of the PD compared to lower realisations. Consequently, it would result in overly conservative estimates. Conservatism can be reduced by considering a Uniform distribution constrained on (p_l, p_u). This, however, this still places equal probability to all realisations specified within the bounds.

Beta Distribution: Secondly, the use of a Beta distribution as a prior distribution is considered. This is an extremely flexible distribution with support on $(0, 1)$. The distribution's flexibility is evident in Figure 3.5. As an example, consider specification of a $Beta(1, 1)$ distribution, which is equivalent to assuming a $Uniform(0, 1)$ distribution. Furthermore, note from Figure 3.5 that by for example specifying a $Beta(4, 1)$ distribution, a shape close to the distribution of the Conservative prior can be obtained. This flexibility is a great advantage of using a Beta prior distribution in a Bayesian analysis. However, the posterior distribution that follows the specification of a Beta prior is extremely dependent on the specification of the α and β parameters. Choosing initial parameter values for the distribution is challenging and introduces a great level of subjectivity into the model.

Conservative Distribution: Thirdly, the Conservative prior avoids the selection of initial parameters for the model. This distribution is, however, expected to produce overly conservative estimates. As illustrated in Figure 3.6, this prior distribution assumes that a 99% PD is the most likely outcome. Therefore, it is clear that this distribution is somewhat detached from reality as it is known that the probability of default is low.

Expert Distribution: Finally, the Expert distribution proposed by Clifford *et al.* (2013) is another distribution with great flexibility. This model allows incorporation of management expectations by specifying the lowest PD that is expected along with the most likely expected PD and then the highest expected PD. Naturally, this introduces high levels of subjectivity in the estimates. A great advantage of the Expert distribution, however, lies in its simplicity. It is simple to understand and, therefore, easy to explain to risk managers and obtain the required expert information. However, justification of the prior parameter selection can be challenging.
It is intuitive that the probability of default is a value between $(0, 1)$. This is the main reason why distributions with support on those bounds are considered. Due to the flexibility of the Beta distribution as a prior distribution, it is difficult to justify a proposal of any other distribution with support on said bounds aside from the Expert distribution. However, by making a simple transformation a door is opened to an array of new distributions, which can be considered as alternatives. This is elaborated on in the next section.

4.2 Extending the Bayesian Approach

Before proceeding, it is useful to reconsider some of the facts that are known on the probability of default of low default portfolios. As mentioned, the probability of default is a value between 0 and 1, this is intuitively obvious. The problem of estimating the probability of default for low default portfolios is that there exists a significant scarcity in default observations. This is evident in Table 2.2. In the attempt to calculate the PD from historic data in which there are no default observations, then an estimate of the PD equal to zero is expected. However, is it realistic that the probability of default is zero? This essentially means that it is certain that no defaults will occur, which is impossible. Therefore, calibration of the PD is required.

As discussed in Chapter 3, Bayesian approaches allows the incorporation of prior knowledge into the estimation process. Therefore, by specifying some form of a prior distribution the PD estimate can be calibrated. Prior knowledge would typically be in the form of expert knowledge, with which a risk professional can incorporate expectations of the PD. Essentially, three facts are known of the probability of default:

- The probability of default lies between 0 and 1.
- The probability of default for low default portfolios lies closer to 0.
- Even though historical data would suggest otherwise, the true probability of default for low default portfolio's cannot be equal to zero.

Due to the nature of the probability of default for low default portfolios being so small, there exists an element of extremity in the estimate. The question is, how does one capture this extreme nature?

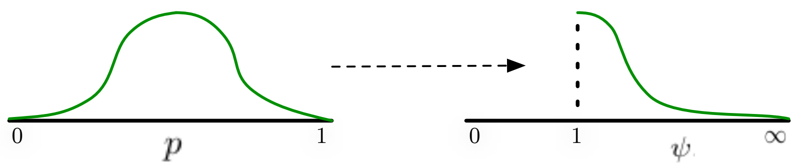


Figure 4.1: Illustration of the Transformation

Consider the following transformation:

$$p \in [0,1] \Rightarrow \psi = \frac{1}{p} \in [1, \infty).$$

It is intuitive that the larger the estimate of ψ, the smaller the probability of default p as shown through $\psi = \frac{1}{p} \Rightarrow p = \frac{1}{\psi}$, where $p = \frac{1}{\psi} \to 0$ as $\psi \to \infty$. In the context of low default portfolios ψ should have a heavy tail. The use of a heavy-tailed distribution as a prior distribution for ψ is therefore justified. The transformation is graphically illustrated in Figure 4.1.

The distribution that is considered as a prior distribution is the Strict Pareto Distribution, which is defined as follows:

Definition 4.1. ***The Strict Pareto Distribution***: *The strict Pareto distribution, Pareto*(γ)*, is a distribution with cumulative distribution function given by:*

$$G(\psi) = 1 - \psi^{-\frac{1}{\gamma}}, \text{ for } \psi \geq 1, \tag{4.2.1}$$

and the corresponding probability density function can be written as:

$$g_\Psi(\psi) = \frac{1}{\gamma}\psi^{-\frac{1}{\gamma}-1} \tag{4.2.2}$$

where $\gamma \in \mathbb{R}$.

A plot of the Pareto probability density function with $\gamma = 1$ is given in Figure 4.2. As illustrated this is a heavy tailed distribution and it is expected that the heavy right tail assists in capturing the low default probability as ψ tends to infinity. Note that as ψ only tends to infinity, p will never be equal to zero.

Since a transformation has been applied to p, the joint conditional density function of $\mathbf{L}$ given p, ξ, τ and $\mathbf{X}$ needs to be transformed to a joint conditional density function of $\mathbf{L}$ given ψ, ξ, τ and $\mathbf{X}$. Recall that the joint conditional density function of $\mathbf{L}$ given p, ξ, τ and $\mathbf{X}$ can be written as:

$$f_{\mathbf{L}}(\mathbf{l}|p, \xi, \mathbf{x}) = \prod_{t=1}^{T}\prod_{i=1}^{s_t} \lambda_t(p, \xi, x_t)^{l_{t,i}}(1 - \lambda_t(p, \xi, x_t))^{1-l_{t,i}}.$$

Let:

$$\psi = \frac{1}{p} \Rightarrow p = \frac{1}{\psi},$$

then:

$$f_{\mathbf{L}}(\mathbf{l}|\psi, \xi, \mathbf{X}) = f_{\mathbf{L}}(\mathbf{l}|p, \xi, \mathbf{X})\left|\frac{dp}{d\psi}\right|.$$

It follows that:

$$\left|\frac{dp}{d\psi}\right| = \frac{1}{\psi^2}.$$

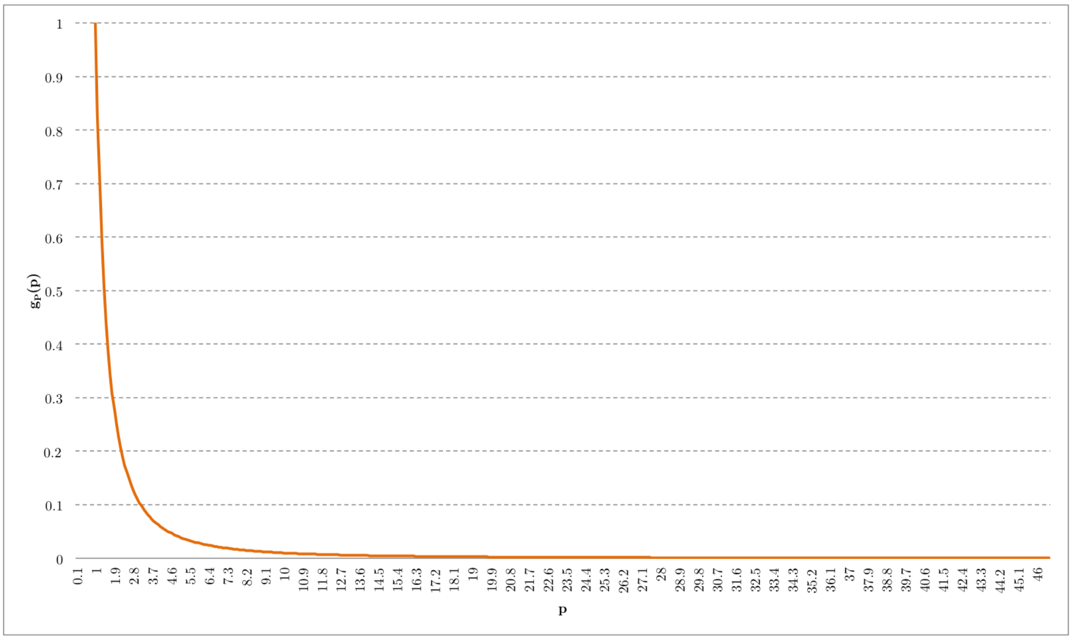


Figure 4.2: Plot of the Pareto(1) Distribution

So that:

$$\begin{aligned} f_{\mathbf{L}}(\mathbf{l}|\psi,\xi,\mathbf{X}) &= f_{\mathbf{L}}\Big(\mathbf{l}|\frac{1}{\psi},\xi,\mathbf{X}\Big)\Big(\frac{1}{\psi^2}\Big). \\ &= \prod_{t=1}^{T}\prod_{i=1}^{s_t}\lambda_t\Big(\frac{1}{\psi},\xi,x_t\Big)^{l_{t,i}}\Big(1-\lambda_t\Big(\frac{1}{\psi},\xi,x_t\Big)\Big)^{1-l_{t,i}}\Big(\frac{1}{\psi^2}\Big). \end{aligned} \tag{4.2.3}$$

The marginal conditional posterior density of $\mathbf{X}$, considered separately for X_1 and $X_t, t = 2,\ldots,T$, can now be rewritten. First of all for $X_1 \sim N(0,1)$ where:

$$f_{X_1}(x_1) \propto \exp\Big(-\frac{x_1^2}{2}\Big),$$

the conditional posterior is given by:

$$\begin{aligned} h(x_1|\mathbf{L},\frac{1}{\psi},\xi,\mathbf{x}_{-1}) &\propto f_{\mathbf{L}}(\mathbf{l}|\frac{1}{\psi},\xi,\mathbf{x}) \times f_{X_1}(x_1) \\ &\propto \Big\{\prod_{t=1}^{T}\prod_{i=1}^{s_t}\lambda_t\Big(\frac{1}{\psi},\xi,x_t\Big)^{l_{t,i}}\Big(1-\lambda_t\Big(\frac{1}{\psi},\xi,x_t\Big)\Big)^{1-l_{t,i}}\Big(\frac{1}{\psi^2}\Big)\Big\} \\ &\times \exp\Big(-\frac{x_1^2}{2}\Big). \end{aligned} \tag{4.2.4}$$

Then for $X_t = \tau X_{t-1} + \sqrt{1-\tau^2}\psi_t$, where $\psi_t \sim N(0,1)$ and:

$$f_{X_t}(x_t|x_{t-1},\tau) = \exp\Big(-\frac{(x_t-\tau x_{t-1})^2}{2(1-\tau^2)}\Big),$$

for $t = 2, \dots, T$ it follows that:

$$
\begin{aligned}
h(x_t|\mathbf{L}, \frac{1}{\psi}, \xi, \mathbf{x}_{-t}) &\propto f(\mathbf{L}|\frac{1}{\psi}, \xi, \mathbf{x}) \times f(x_t|\mathbf{x}_{-t}, \tau) \\
&\propto f(\mathbf{L}|\frac{1}{\psi}, \xi, \mathbf{x}) \times f(x_t|x_{t-1}, \tau) \\
&\propto \left\{ \prod_{t=1}^{T} \prod_{i=1}^{s_t} \lambda_t\left(\frac{1}{\psi}, \xi, x_t\right)^{l_{t,i}} \left(1 - \lambda_t\left(\frac{1}{\psi}, \xi, x_t\right)\right)^{1-l_{t,i}} \left(\frac{1}{\psi^2}\right)\right\}\right\} \\
&\times \exp\left(- \frac{(x_t - \tau x_{t-1})^2}{2(1-\tau^2)}\right), t = 2, \dots, T,
\end{aligned}
\tag{4.2.5}
$$

where $\mathbf{x}_{-t} = (x_1, \dots, x_{t-1}, x_{t+1}, \dots, x_T)'$.

The marginal posterior density of $\psi \sim Pa(\gamma)$ where:

$$g_\Psi(\psi) = \frac{1}{\gamma}\psi^{-\frac{1}{\gamma}-1},$$

is given by:

$$
\begin{aligned}
h(\frac{1}{\psi}|\mathbf{L}, \xi, \mathbf{x}) &\propto f(\mathbf{L}|\frac{1}{\psi}, \xi, \mathbf{x}) \times g_\Psi(\psi) \\
&\propto \left\{ \prod_{t=1}^{T} \prod_{i=1}^{s_t} \lambda_t\left(\frac{1}{\psi}, \xi, x_t\right)^{l_{t,i}} \left(1 - \lambda_t\left(\frac{1}{\psi}, \xi, x_t\right)\right)^{1-l_{t,i}} \left(\frac{1}{\psi^2}\right)\right\}\right\} \\
&\times \frac{1}{\gamma}\psi^{-\frac{1}{\gamma}-1}.
\end{aligned}
\tag{4.2.6}
$$

Simulation of the aforementioned marginal posterior density functions are discussed in Chapter 5. The performance of this prior distribution is compared to the prior distributions discussed in Chapter 3 through an empirical study in the same chapter.

4.3 Conclusion

In this short chapter the Bayesian Approach is reconsidered. In Section 4.1 some comments are made on the prior distributions discussed in Section 3.3.2.2. A different prior distribution is then proposed as an alternative in Section 4.2.

In the next chapter, Chapter 5, the different PD calibration techniques will be compared in an empirical study.

Chapter 5

Simulation Procedure and Empirical Study

In this chapter the empirical results and findings of this study are discussed. Before discussing the empirical results, the simulation procedure used is described in detail. This involves a detailed discussion on the Gibbs Sampling method and how this algorithm is applied in the study. Deliberation of the simulation procedure is also provided in the first part of this chapter.

In the second and most important section the empirical results of the study are discussed. In the empirical component the various different PD calibration models for Low Default Portfolios discussed in Chapters 3 and 4, are compared. The main models under consideration are:

- The Confidence Based Approach as proposed by Pluto and Tasche (2011) (Section 3.3.1).
- The Bayesian Approach using a:
  - Uniform prior distribution as proposed by Dwyer (2006) and elaborated on by Tasche (2013) (Section 3.3.2),
  - Beta prior distribution as proposed by Kiefer (2009) (Section 3.3.2),
  - Conservative prior distribution as proposed by Tasche (2013) (Section 3.3.2),
  - Expert prior distribution as proposed by Clifford *et al.* (2013), and
  - Pareto prior distribution as proposed in this study (Chapter 4).

Another prior distribution that is considered as a comparison to the Bayesian approaches is the Jeffreys prior. This is another example of an unobjectionable (or uninformative) prior distribution. The Jeffreys prior was proposed by Harold Jeffreys (1946).

Let p be the parameter of a distribution with P the corresponding random variable from a Bayes perspective, then the Jeffreys prior is defined as:

Definition 5.1. ***Jeffreys prior:*** *The idea underlying Jeffreys prior is to place a prior distribution on p, such that the probability density function can be defined as:*

$$g_P(p) \propto \sqrt{I(p)} \tag{5.0.1}$$

where $I(p)$ denotes Fischer's information criterion defined by:

$$I(p) = -\mathbb{E}_p\left[\frac{d^2 \log f(L|p)}{dp^2}\right] \tag{5.0.2}$$

and $f(L|p)$ is the likelihood.

As discussed in Chapter 2, the likelihood for the problem that is being addressed in this study is a binomial likelihood. Recall from expression (2.2.27) that:

$$f(L|p) = P(L = d|p) = \binom{s}{d} p^d (1-p)^{s-d}$$

where $d = \sum_{i=1}^{s} l_i$ and $L_i \sim B(1; p)$. Therefore it follows that:

$$\log f(L|p) = d \log p + (s-d)\log(1-p) + constant,$$

$$\frac{d}{dp} \log f(L|p) = \frac{d}{p} - \frac{s-d}{1-p},$$

so that:

$$\frac{d^2}{dp^2} \log f(L|p) = -\frac{d}{p^2} - \frac{s-d}{(1-p)^2}.$$

Since the likelihood can be regarded as a binomial random variable where $D \sim Bin(s; p)$ and $\mathbb{E}_P[D] = sp$, it follows that:

$$\begin{aligned} I(p) &= -\mathbb{E}_p\left[\frac{d^2 \log f(L|p)}{dp^2}\right] \\ &= \frac{sp}{p^2} + \frac{s-sp}{(1-p)^2} \\ &= \frac{s}{p} + \frac{s}{(1-p)} \\ &= \frac{s}{p(1-p)}. \end{aligned}$$

It follows from the above that the Jeffreys prior for the given scenario can be written as:

$$g_P(p) \propto \sqrt{I(p)} = p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}}. \tag{5.0.3}$$

The expression above is simply a $Beta(\frac{1}{2}, \frac{1}{2})$ distribution. Therefore, the same expressions from Section 3.3.3 for the Beta prior distribution can be applied.

The prior distribution discussed above, along with the other calibration models mentioned earlier, are compared using real and fictitious data. The results along with the findings will be discussed in the empirical results section.

5.1 Simulation Procedure

As mentioned in Chapter 3, the simulation technique used in this study to simulate the Bayesian posterior distributions is the MCMC method known as Gibbs Sampling. For a more detailed discussion on MCMC methods, refer to Appendix B. The details on Gibbs sampling discussed in this section include the estimation of the probability of default along with the simulation of the latent factor $\mathbf{X}$.

Consider the following Bayesian setting: Given the observations of the default events $L_1, L_2, ..., L_T$ over a T-year time period, the goal is to estimate the probability of default p through specification of a prior distribution with probability density function $g_P(p)$, where the dependence between the default events are captured through a latent factor $\mathbf{X}$.

A general overview of the Gibbs sampling procedure is provided below, this is followed by a more detailed discussion of each step. The discussion of the sampling procedure is adapted from Berning (2010):

1. Start with initial estimates of the parameter vector $p_{(0)}$ and the latent factor vector $\mathbf{X}_{(0)}$.
2. For a large number of repetitions $i = 1$ to M:
   a) Generate $p_{(i)}$, this is one simulated value of p from the conditional posterior of p given $X_{(i-1)}$. The simulated value is dependent on the specification of the prior distribution. The marginal conditional posterior densities of the different prior density functions under consideration are given in Section 3.3.3 and in Chapter 4.
   b) Generate $X_{1(i)}$, this is one simulated value of X_1 from the conditional posterior of X_1 given p.
   c) Generate $X_{j(i)}$ for $j = 2, ..., T$, this is one simulated value of X_j from the conditional posterior of X_j given p and X_{j-1}.
3. The resulting vectors of simulated values are:

$$\mathbf{p} = (p_{(1)}, p_{(2)}, \ldots, p_{(M)}),$$

$$\mathbf{X}_1 = (X_{1(1)}, X_{1(2)}, \ldots, X_{1(M)})$$

$$\mathbf{X}_j = (X_{j(1)}, X_{j(2)}, \ldots, X_{j(M)})$$

   for $j = 2, 3, \ldots, T$ where it then follows that $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_T)$. The simulated values can be regarded as draws from the marginal posterior distributions of p and $\mathbf{X}$. Estimating the parameters from the drawn values will be discussed in Section 5.1.1.

Before elaborating on the aforementioned steps, take note that some parts of the conditional posterior density functions are constant with respect to the parameter. These constant parts may be omitted for the purpose of the simulation. The steps of the Gibbs sampler are as follows:

- **Step 1:** Because the first draw is a value of p, $p_{(0)}$ does not need to be specified. The sampling method starts with drawing a value of p as it is unclear what the initial estimate should be. The initial values of the latent factor vector $\mathbf{X}_{(0)}$ are taken as zero, this is the midpoint on the interval $[-3; 3]$ (recall that the latent factors are simulated from a standard normal distribution).
- **Step 2:** In this step, the number of simulation repetitions, M, are chosen. This should ideally involve a large number of simulation repetitions such as $M =$100 000 or $M =$10 000. However, note that such large specifications results in a considerable amount of computational time. In this study the number of simulation repetitions is restricted to $M =$10 000, unless explicitly stated otherwise. Another important specification that is required when doing Gibbs sampling is the number of burn-in draws. As the arbitrary initial values of the parameters in the distribution may be inaccurate, the first draws from the marginal posterior distribution should be disregarded in order to compensate for the inaccuracy. Due to the element of randomness that is present in the simulation of the latent factor a high number of burn-in draws is selected. This ensures that the

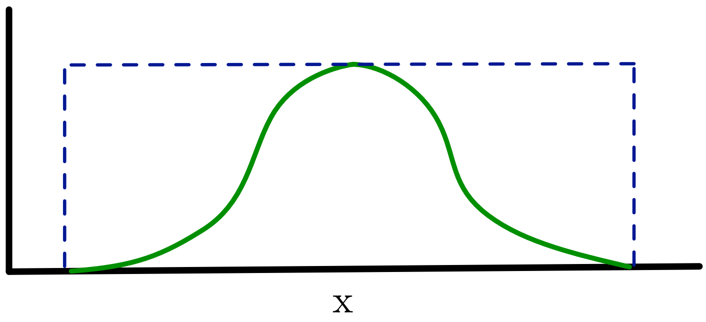


Figure 5.1: Illustration of the Rejection Step

first draw of p, which is dependent on the latent factor, is more accurate. The number of burn-in draws specified is 1000, which is discarded and then followed by 10 000 draws that are retained. It is expected that this results in a higher degree of accuracy in the estimates.

- **The Rejection Step:** In the sub-steps of the second step, a rejection method is required to simulate one value from the conditional posterior density under consideration. The rejection method that is considered follows from Rice (2007). For the sake of the discussion say that values need to be simulated from a given density function, $f(x)$, as graphically illustrated in the green plot in Figure 5.1.

  As discussed in Berning (2010), in order to apply the rejection method, the blue rectangle in Figure 5.1 needs to be constructed. It therefore follows that the maximum of the density function, along with the interval in which $f(x)$ is significant, needs to be known. The significant interval may be defined as all values for x for which $f(x) > 0.01 \max[f(x)]$, where 0.01 is the significant factor. This is also the factor used in the simulation algorithms.

  For every value that is to be simulated from the density $f(x)$, the following steps are performed:

  1. Simulate a value x^* from a $Uniform(x_{min}, x_{max})$ distribution.
  2. Independent of x^*, simulate y^* from a $Uniform(0, \max f(x))$ distribution.
  3. If $y^* \leq f(x^*)$, take x^* as the simulated value. Otherwise reject x^* and redo from the first step.

The steps listed above are elaborated on in the context of the conditional posterior distributions that are considered in this study. The rejection algorithm for the Pareto distribution as a prior distribution is considered below. The algorithms used for the other densities follow the same logic.

For the Pareto distribution as a prior distribution, the value of $\psi_{(k)}$, for $k = 1, \ldots, M$ is simulated from the conditional posterior:

$$
\begin{aligned}
h\left(\frac{1}{\psi_{(k)}}\middle|\mathbf{L}, \xi, \mathbf{X}_{(k)}\right) &\propto f\left(\mathbf{L}\middle|\frac{1}{\psi_{(k)}}, \xi, \mathbf{X}_{(k)}\right) \times g_{\Psi}(\psi) \\
&\propto \left\{\prod_{t=1}^{T}\prod_{i=1}^{s_t}\lambda_t\left(\frac{1}{\psi_{(k)}}, \xi, x_{t(k)}\right)^{l_{t,i}}\left(1-\lambda_t\left(\frac{1}{\psi_{(k)}}, \xi, x_{t(k)}\right)\right)^{1-l_{t,i}}\left(\frac{1}{\psi^2_{(k)}}\right)\right\} \\
&\times \frac{1}{\gamma}\psi_{(k)}{-}\frac{1}{\gamma}-1.
\end{aligned}
$$

For the sake of simplicity, let:

$$
f(\psi) = \left\{\prod_{t=1}^{T}\prod_{i=1}^{s_t}\lambda_t\left(\frac{1}{\psi}, \xi, x_t\right)^{l_{t,i}}\left(1-\lambda_t\left(\frac{1}{\psi}, \xi, x_t\right)\right)^{1-l_{t,i}}\left(\frac{1}{\psi^2}\right)\right\} \times \frac{1}{\gamma}\psi^{-\frac{1}{\gamma}-1}. \tag{5.1.1}
$$

As mentioned in Berning (2010), the normalizing constant is unknown and therefore it is advisable to first compute $\log(f(\psi))$. Then obtain another function (say $k(\psi)$) by scaling $\log(f(\psi))$ appropriately after which $\exp(k(\psi))$ is calculated. The log scaling is done for numerical considerations. From a mathematical point of view working with $f(\psi)$ and $\exp(k(\psi))$ produces the same result. However, from a numerical point of view $f(\psi)$ creates some problems in the simulation process.

Therefore, the following log-transformed function is used in the simulation procedure:

$$
\begin{aligned}
\log f(\psi) = &\sum_{t=1}^{T} s_t \log\left(\lambda_t\left(\exp\left(\frac{1}{\log\psi}\right), \xi, X_t\right)\right) \\
&+ \sum_{t=1}^{T}(n_t - s_t)\log\left(1-\lambda_t\left(\exp\left(\frac{1}{\log\psi}\right), \xi, X_t\right)\right) \\
&- \log\psi + \frac{1}{2(1+\gamma)}.
\end{aligned} \tag{5.1.2}
$$

Derivation of the expression above along with the derivations of the log scaled conditional posterior densities of the other considered prior distributions is provided in Appendix A. The simulation procedure follows as:

1. Choose a range of values of ψ, say $\psi = (1, 2, \ldots, 20)$.
2. Calculate $\log f(\psi)$.
3. Calculate $k(\psi) = \log f(\psi) - \max(\log f(\psi))$.
4. Calculate $\exp(k(\psi))$, which results in a vector of density values.
5. Retain the values of $\exp(k(\psi))$ that are greater than 0.01 and the corresponding values of ψ.
6. Use the result in the previous step and apply the rejection method to simulate a single observation from $h\left(\frac{1}{\psi_{(i)}}\middle|\mathbf{L}, \xi, \mathbf{X}_{(i)}\right)$.

The procedure discussed above is adopted from Berning (2010). All the simulations are done in MATLAB, where the code along with discussions are provided in Appendix C.

5.1.1 Inferencing with the Gibbs Sampler

At the end of Chapter 3 it is mentioned that there are three different loss functions that can generally be assumed in order to produce Bayesian estimates, namely:

- square error loss,
- absolute loss, or
- 0-1 loss.

The Bayesian estimators corresponding to the aforementioned loss functions are respectively the mean, the median and the mode of the simulated marginal posterior distribution. Without going into the detail of defining the loss function and its selection. It is simply noted that there are three general ways of obtaining a Bayesian estimator of a parameter. A decision on which one is best to use for the PD will be made by examining the simulation results in Section 5.2.

Note, however, that the estimator used by Dwyer (2006); Kiefer (2009); Tasche (2013); Clifford *et al.* (2013); Chang and Yu (2014) and Kruger (2015) is the mean of the marginal posterior distribution. In other words, the common loss function considered in the literature is the square error loss.

Calculating the mean of the marginal posterior distribution follows a simple calculation of taking the arithmetic mean of the simulated values. This results in the following Bayesian estimates of the PD:

$$\hat{p} = \frac{1}{M}\sum_{i=1}^{M} p_{(i)}, \tag{5.1.3}$$

where M is the number of simulations. The median follows effortlessly by simply calculating the median of the simulated values of the marginal posterior distribution. The mode, on the other hand, follows a more complicated calculation and is therefore discussed in the following subsection.

5.1.1.1 Half-Sample Mode

In this study, the method used to determine the mode of the distribution is the half sample mode (HSM) method. In Bickel and Frühwirth (2006) an array of estimators for the mode was evaluated in comparison and it was concluded that that the HSM shows strong performance under a variety of conditions.

As discussed in Berning (2010), this translates to the following: Find the 50% highest posterior density (HPD) region of the observations (this is the shortest interval containing at least 50% of the observations, as discussed in Section 5.1.1.2). Then, retain the observations in the 50% HPD and once again take the 50% HPD of the retained observations. This process is repeated until less than four observations are present. Now:

- If one observation remains, then this is the mode.
- If two observations remains, then the mode follows as the mean of the two remaining observations.

- If three observations remains, then the mode follows as the mean of the two observations that are closest together.

The HSM procedure discussed above is coded as a function in MATLAB that can be used to obtain the mode from the set of simulated observations. In order to obtain credible regions for the Bayesian estimates the Highest Posterior Density is used. This is discussed in detail in the next section.

5.1.1.2 Highest Posterior Density

The Bayesian intervals are estimated by selecting the bounds of the credible region to enclose $100(1-\alpha)\%$ of the MCMC random variables. Chang and Yu (2014) calculate the shortest possible interval enclosing $100(1-\alpha)\%$ of the MCMC random samples. This is the highest posterior density (HPD) credible region. The $\alpha\%$ highest posterior density is regarded as the shortest region in the parameter space that contains $\alpha\%$ of the posterior probability. The HPD is explained as follows; let $\{p_1, p_2, \ldots, p_M\}$ be the MCMC random samples of p, where M is the number of MCMC iterations. The $100(1-\alpha)\%$ HPD confidence interval of p is constructed as follows:

1. Sort $\{p_1, p_2, \ldots, p_M\}$ to obtain the ordered values:
$$p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(M)}.$$
2. Compute the $100(1-\alpha)\%$ credible region:
$$\big(p_{(j)}, p_{(j+[(1-\alpha)M])}\big), j = 1, 2, \ldots, M - [(1-\alpha)M].$$
3. The $100(1-\alpha)\%$ HPD credible region of p is the credible region with the smallest interval width among all credible regions in point 2.

Berning (2010) defines the HPD region as the shortest interval containing at least $100(1-\alpha)\%$ of the observations. It is important not to confuse the HPD, which is a form of a Bayesian credible region (or interval) with the confidence interval used by frequentists. In simple problems, the two may have an exact numerical correspondence. However, even though the solutions may be numerically identical, the interpretations are very different.

In Bayesian theory probability distributions reflect a degree of belief and therefore the HPD can be interpreted as follows; given the observed data, there is an $(1-\alpha\%)$ chance that the true value of the parameter falls within the interval region. On the other hand, in frequentism the parameter is regarded as a fixed value and the data along with all quantities that are derived from the data is regarded as random variables. Therefore, when estimating the mean from the observed data, the confidence interval illustrates that there exists a $(1-\alpha)\%$ confidence that the true mean will fall into the confidence interval.

As such, in the current Bayesian setting the HPD is a statement of the probability of the parameter value given fixed bounds. In this study the HPD is used to compare the different prior distributions. From the definition of the HPD narrower HPD intervals can be used as a measure of accuracy of the estimate obtained from the sampled posterior distribution. The average confidence widths of the HPD credible regions are compared when the empirical results are discussed.

5.2 Empirical Study

In this section the models discussed in Chapters 3 and 4 are compared on an empirical basis. The models are tested on both real data and fictitious data. An important test with fictitious data is how the model results compare when moving away from the low default spectrum towards higher default levels. Furthermore, different sensitivity analyses are run on the models in order to compare the sensitivity to model inputs. Finally, the outcome of using the PD estimates obtained from the different models in calculating the risk weighted assets (RWA) under Basel III are also illustrated. Note, however, that the calculation of RWAs is not the focus of this study and that this discussion is presented on a very high level.

5.2.1 Initial model comparison

The first set of empirical results is an initial comparison of the models using a basic set of fictitious inputs. The default rate is gradually increased in order to see how the models perform when moving away from the zero default side of the spectrum.

In order to form an initial impression of how the models compare, a portfolio of 1 000 obligors is considered over a 1 year time period. The asset correlation is specified as, $\xi = 12\%$, and since only one year is considered there is no intertemporal correlation. The default rate (or naïve PD estimate) is then increased from 0% to 2%, i.e. defaults are increased from 0 to 20. This will give an initial impression on how the models calibrate the PD through the spectrum. A total of 10 000 simulation repetitions are used for all the models, where a burn-in of 1000 is used for the Bayesian approaches. Furthermore, note that for this initial consideration a square error loss function is assumed for the Bayesian approaches.

Before comparing the results of the different models an interesting feature of using the Pareto(γ) distribution as a prior distribution is highlighted. Figure 5.2, provides an indication of how the Pareto distribution calibrates the PD estimate at different specifications of the model parameter, γ, for the scenario discussed at the outset of this subsection. An interesting observation from Figure 5.2 is that the model is not sensitive to the model parameter. The parameter, γ, is tested from 0.001 through to 1 000 yielding extremely similar results. This is especially the case for very low default levels (between 0 and 5 defaults), where nearly identical estimates are obtained. Small differences are present as higher default levels. This insensitivity to the model parameter is a favourable attribute as it reduces subjectivity in selecting a value for γ. Note however that this does not mean that the Pareto prior is an uninformative or objective prior, because through selection of the Pareto distribution as a prior distribution it is already assumed that the PD is very low. Following the aforementioned results $\gamma = 1$ is chosen for all simulations with the Pareto distribution as a prior distribution.

Figure 5.3, shows the results of the PD estimates obtained for different calibration models given the scenario discussed at the outset of this section. As illustrated in the plot, the Confidence Based Method at 90% confidence intervals appears to be the most conservative calibration technique overall. The results, however, converges with some of the more conservative Bayesian approaches at higher default levels. The parameters for the Expert Distribution is chosen at 0, 3% and 10%, which is a very conservative stance, this is evident in Figure 5.3 (especially between 0 and 5 defaults). As the default rate is increased it appears as though the rate of calibration flattens out a bit. As expected the $Beta(1,1)$ distribution as a prior distribution has the same performance as the $Uniform(0,1)$ distribution. Also note that the

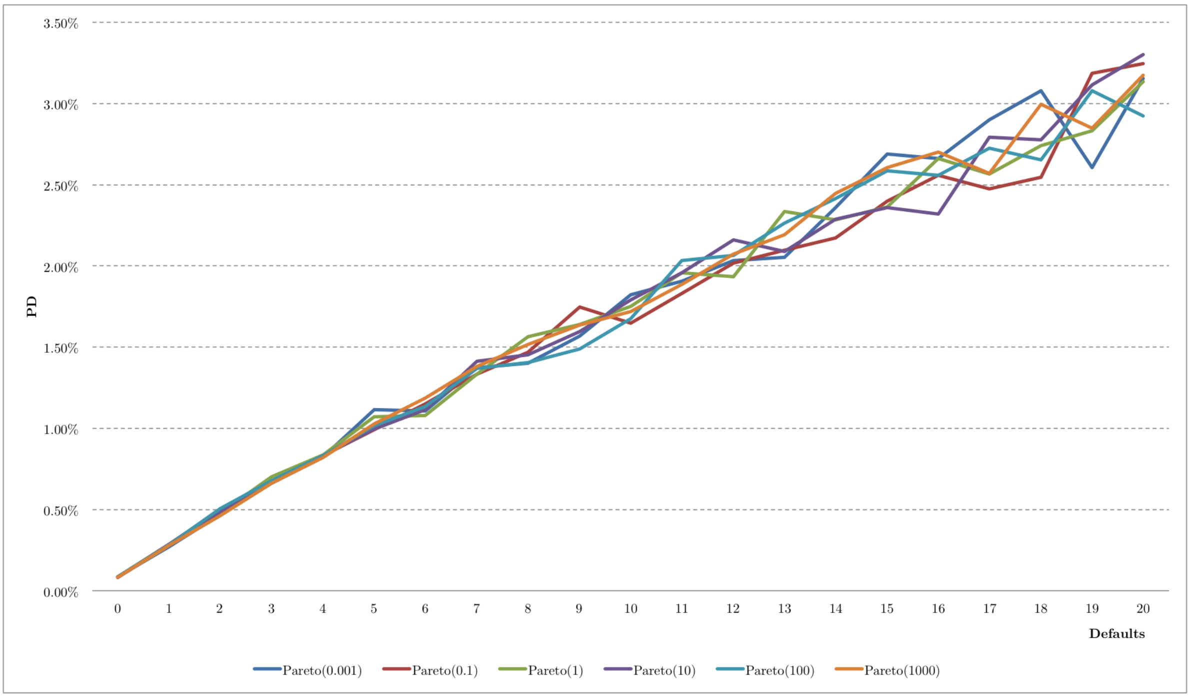


Figure 5.2: Comparison of PD Estimates at Different Parameter Inputs using the Pareto Distribution as a Prior Distribution.

Jeffreys prior ($Beta(0.5, 0.5)$) produces results comparable to the Confidence Based approach at the 75% confidence level. As evident, the $Beta(1, 400)$ as a prior distribution doesn't capture the PD and under calibrates the estimate as the default rate moves away from the low default side of the spectrum. This specification is used for some portfolios later in this Chapter and was included to indicate the effect of selecting the wrong model parameters for the scenario when using any informative prior distribution. Therefore, it is shown that a $Beta(1, 400)$ distribution is not the right model for the higher default rates. Finally, it appears as though the Pareto distribution is the least conservative approach out of all the models that performs a positive calibration, in other words a calibration upwards of the naïve approach. This may be due to the fact that the model accounts for the low default nature of the scenario with greater accuracy, which refers to less conservatism.

It should however be highlighted that more data needs to be utilised in the Bayesian approaches in order to gain full value from using these approaches.

5.2.2 Estimations Using Historical Data

Two different data sets are considered in this study. Each data set exhibits unique characteristics due to the type of credit instrument that is considered. However, a common feature of the data sets is that all of the portfolios can be considered as a low default portfolio. The two portfolios that are considered in this study are a corporate portfolio, and a retail portfolio.

Before continuing, the following should be noted regarding the selection of the parameters for the prior distribution. The parameters should be selected before considering the data. This is what constitutes prior knowledge. If the parameters are selected after considering the data, then the data influences the selection and the prior model can no longer be regarded as the

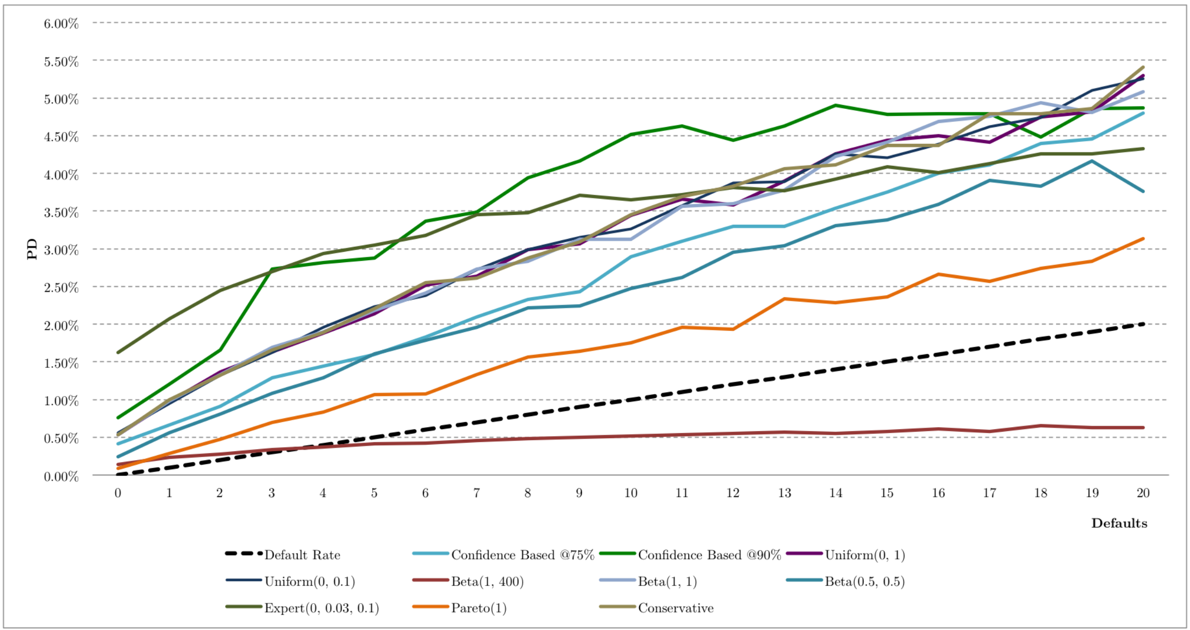


Figure 5.3: Comparison of PD Estimates for Different Calibration Models.

prior information. All the prior parameters used in this study are specified independent of the data, before the data set is considered in the analysis.

5.2.3 Corporate Portfolio

The Corporate Portfolio used in the study is the same portfolio of investment grade corporates as used in Kruger (2015). This data set is derived from Moody's (2014) default data. It is assumed that all the corporates fall in the same rating grade. The corporate portfolio is given in Table 5.1. Estimates are obtained using the full 10-year data period and then also for a shorter 5-year period.

Table 5.1: Corporate Portfolio

Year	Obligors	Defaults
2005	2710	2
2006	2738	0
2007	2742	0
2008	2709	14
2009	2600	11
2010	2481	2
2011	2522	1
2012	2498	1
2013	2560	1
2014	2643	2

5.2.3.1 Corporate Portfolio: 10-Year Estimates

The results for the entire 10-year data set is given in Table 5.2. All of the results in the table are produced using 5 000 simulations with a burn-in of 1 000 for the Bayesian approaches and 10 000 Monte Carlo Simulations for the Confidence Based approaches. The naïve PD estimate follows through applying expression (3.3.1) to each individual year in Table 5.1 and then taking the average of the one-year PD estimates. The naïve estimate for the portfolio under consideration is 0,128 9%. Naturally the naïve estimate does not account for intertemporal correlation or correlation between default events.

All of the calibration approaches considered account for asset correlation and intertemporal correlation. As discussed in Section 2.2.3, in the Basel II credit risk function, the asset correlation is defined on [0,12;0,24] for corporate portfolios. Therefore, in order to form a comparison of results at different correlation levels, simulation results are obtained at the 12%, 18% and 24% asset correlation levels. As discussed in Pluto and Tasche (2011), default events that are five-years apart can be regarded as nearly independent. From expression (3.3.41) the correlation between the systematic factors in years 1 and 5 is given by $Corr(X_1, X_5) = \tau^4$, setting $\tau = 30\%$ results in $Corr(X_1, X_5) = 0.81\%$. Therefore, the choice of $\tau = 30\%$ seems appropriate. For the sake of comparison, results are, however, also simulated for a higher intertemporal correlation of $\tau = 50\%$, for which it follows that $Corr(X_1, X_5) = 6.25\%$.

For the Confidence Based approaches results are obtained at the 75% and at the 90% confidence levels. From the results it is clear that increasing the asset correlation and/or increasing the intertemporal correlation increases the estimated PD. This is as expected. Also note that the PD estimates at the 75% confidence interval is lower than the estimates at the 90% confidence interval. This result is intuitive. Furthermore, when generally comparing the results of the PD estimates in the LDP framework, the results are referred to as "more conservative" when the model produces a higher PD estimate than the comparable models.

For the Bayesian approaches the HPD is calculated at a 90% confidence level and results are quoted for all three loss functions. Comparing the Bayesian estimators based on the different loss functions it is clear that the mean and the median as estimators are more or less in line, with the median being slightly smaller than the mean. The mode, on the other hand, generally produces noticeably lower estimates in most cases, apart from a few extreme estimates e.g. the mode PD estimate for the Extreme prior at $\xi = 24\%$ and $\tau = 30\%$. The fact that the following inequality:

$$\text{Mode PD Estimate} < \text{Median PD Estimate} < \text{Mean PD Estimate}$$

generally holds for results in question makes intuitive sense as it is expected that the posterior distribution would resemble the loss distribution in Figure 2.4, which is positive skew. In line with other authors on the subject, the mean is used as estimator for the PD i.e. a squared error loss is assumed. This is also the estimator that is discussed when comparing results. Although a strong case can be made for rather assuming an absolute loss than a squared error loss (i.e. using the median rather than the mean).

Firstly, an uninformative $Uniform(0, 1)$ prior distribution is considered. For this model at the lower asset correlation levels the PD estimates are lower than the corresponding estimates of the Confidence Based approach. At higher asset correlation levels results are more or less in line. At all correlation levels, the $Uniform(0, 1)$ results are more conservative than the

Confidence Based approach at the 75% confidence level, but less conservative than at the 90% confidence level. These observations are also in line with the early impression formed in Figure 5.3. For all the obtained PD estimates the estimate appears to fall within the 90% HPD confidence interval. There is also a clear widening of the HPD confidence intervals as the asset correlation is increased. An interesting observation however is the fact that the lower bounds stay more or less the same. This is expected due to the shape of the posterior distributions as illustrated later in this section.

Secondly, Jeffreys prior distribution is considered as an alternative uninformative prior distribution. The results for this model is in line with the results of the $Uniform(0, 1)$ distribution at lower asset correlations. However, it is noted that at higher levels of the asset correlation the results appear lower than the results of the $Uniform(0, 1)$ distribution.

Thirdly, a $Beta(1; 400)$ distribution is consider as a prior distribution. This is the same model that was fitted by Kruger (2015) for this data set. The mean of the aforementioned distribution is 0,249%, this provides some justification of the parameters. The results are in line with the previously discussed results for the Jeffreys prior, especially at lower asset correlation levels. At higher asset correlation levels it is, however, clear that the model produces lower estimates. Note that the Bayesian model with the chosen prior produces results that are less conservative than the Confidence Based approach and the Bayesian approach using a $Uniform(0, 1)$ prior. Furthermore, recall from Figure 5.3 that the $Beta(1; 400)$ model significantly underestimated the PD for that data as it was not a good fit for the higher default rates present in Figure 5.3. It appears as though the model is a good fit to the data in Table 5.1.

The fourth model considered is the Bayesian Approach with the Conservative prior distribution. It is expected that this model would produce more conservative results, however, the results prove the contrary. At lower asset correlation levels results are in line with the Bayesian approaches that were previously discussed. At higher asset correlation levels however the results are more conservative than the results for the Jeffreys prior and the $Beta(1; 400)$ distribution. The results are more in line with that for the $Uniform(0, 1)$ distribution.

The next model that is being compared is the Bayesian approach with the Expert distribution. In order to select the model parameters Table 2.3 is used. Considering the A-grade debt in this table it is clear that the minimum PD rate for Corporates over the 32-year period was 0%, the maximum was 0,39% and the average was 0,06%. Therefore let $p_l = 0, p_m = 0,0006$ and $p_u = 0,0039$. As evident in the results, this induces a more conservative calibration than the previous Bayesian approaches considered. The results are higher than the Confidence Based approach with a 75% confidence level and higher than the previously discussed Bayesian approaches at all asset correlation levels. However, estimates remain less conservative than the Confidence Based Approach at the 90% confidence level.

Finally, the model proposed in this study, the Bayesian Approach with the Pareto Distribution as a prior distribution is considered. The estimates obtained from the posterior distribution are estimates of $\hat{\psi}$ and the PD estimates then follow through the transformation $\hat{p} = 1/\hat{\psi}$. The estimates are more conservative than the Bayesian approaches that were previously discussed and the results are generally in line with the PD estimates obtained from the Confidence Based approach at the 90% confidence level.

Table 5.2: Corporate Portfolio Results: 10-year Data

Naïve PD Estimate:		0,1289%					
Asset Correlation:		12%	18%	24%	12%	18%	24%
Intertemporal Correlation:		30%	30%	30%	50%	50%	50%
Confidence Based Approach							
PD Estimate at 75%		0,21%	0,26%	0,29%	0,24%	0,29%	0,35%
PD Estimate at 90%		0,28%	0,41%	0,46%	0,35%	0,42%	0,54%
Bayesian: $Uniform(0,1)$ **Prior**							
Mean PD Estimate		0,15%	0,20%	0,38%	0,18%	0,29%	0,34%
Median PD Estimate:		0,14%	0,18%	0,33%	0,16%	0,25%	0,29%
Mode PD Estimate:		0,07%	0,11%	0,17%	0,10%	0,12%	0,23%
HPD:		(0,0594%; 0,26%)	(0,0518%; 0,37%)	(0,0584%; 0,7%)	(0,0532%; 0,3%)	(0,0491%; 0,52%)	(0,09%; 0,55%)
Bayesian: Jeffreys Prior							
Mean PD Estimate		0,17%	0,20%	0,27%	0,15%	0,23%	0,28%
Median PD Estimate:		0,15%	0,18%	0,24%	0,13%	0,20%	0,24%
Mode PD Estimate:		0,10%	0,15%	0,16%	0,10%	0,14%	0,11%
HPD:		(0,0541%; 0,28%)	(0,0549%; 0,35%)	(0,0723%; 0,5%)	(0,0474%; 0,26%)	(0,0436%; 0,41%)	(0,0457%; 0,5%)
Bayesian: $Beta(1,400)$ **Prior**							
Mean PD Estimate		0,15%	0,19%	0,25%	0,16%	0,21%	0,25%
Median PD Estimate:		0,14%	0,18%	0,23%	0,15%	0,18%	0,22%
Mode PD Estimate:		0,15%	0,10%	0,12%	0,10%	0,08%	0,14%
HPD:		(0,0524%; 0,25%)	(0,055%; 0,32%)	(0,0585%; 0,45%)	(0,0494%; 0,27%)	(0,0499%; 0,36%)	(0,0725%; 0,44%)
Bayesian: Conservative Prior							
Mean PD Estimate		0,17%	0,23%	0,33%	0,17%	0,25%	0,35%
Median PD Estimate:		0,15%	0,20%	0,30%	0,16%	0,22%	0,30%
Mode PD Estimate:		0,09%	0,23%	0,12%	0,11%	0,10%	0,07%
HPD:		(0,054%; 0,28%)	(0,0525%; 0,39%)	(0,0623%; 0,59%)	(0,0526%; 0,3%)	(0,06%; 0,41%)	(0,06%; 0,66%)
Bayesian: Expert Prior							
Mean PD Estimate		0,20%	0,31%	0,42%	0,23%	0,32%	0,49%
Median PD Estimate:		0,19%	0,21%	0,37%	0,21%	0,29%	0,42%
Mode PD Estimate:		0,17%	0,34%	1,09%	0,11%	0,10%	0,33%
HPD:		(0,0627%; 0,34%)	(0,0824%; 0,55%)	(0,01%; 0,073%)	(0,0806%; 0,39%)	(0,0832%; 0,55%)	(0,11%; 0,89%)
Bayesian: Pareto(1) Prior							
Mean Estimate:	ψ	*254,2611*	*225,693*	*204,4875*	*249,930*	*222,196*	*197,777*
	PD	0,39%	0,44%	0,489%	0,400%	0,450%	0,505%
Median Estimate:	ψ	*234,2563*	*208,419*	*188*	*232,180*	*205,13*	*182,489*
	PD	0,43%	0,48%	0,533%	0,431%	0,487%	0,548%
Mode Estimate:	ψ	*148,4132*	*105,572*	*148*	*148,413*	*148*	*148,413*
	PD	0,67%	0,95%	0,674%	0,674%	0,674%	0,673%
HPD:	ψ	*(401,4167; 126,507)*	*(351,04; 105,57)*	*(327,83; 90,7)*	*(392,48; 121,268)*	*(351,89; 106,951)*	*(310,11; 75,1)*
	PD	(0,249%; 0,791%)	(0,284%; 0,94%)	(0,305%; 1,103%)	(0,254%; 0,824%)	(0,284%; 0,935%)	(0,322%; 1,33%)

As an illustration of the posterior densities obtained for the different prior distributions, histograms of the posterior densities corresponding to the respective models discussed previously in this section are provided in Figure 5.4. The illustrated histograms are plotted for simulations run at a 18% asset correlation and a 30% intertemporal correlation. From the histogram plots it is clear why the lower bound of the HPD's for the different distributions are more or less in line and only the outer bounds shifts outwards. The histograms are also in line with the shape that is expected for the loss distribution for a portfolio of credits.

5.2.3.2 Corporate Portfolio: 5-Year Estimates

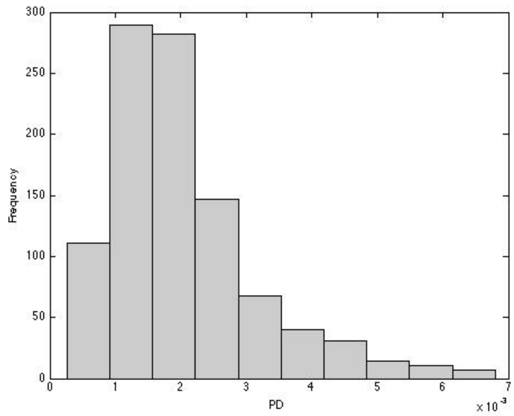
The 10-year historical data period discussed previously contains the 2008 credit crisis. Naturally this period resulted in numerous companies defaulting, as can be seen in the spike in defaults for 2008 and 2009 (Table 5.1). In order to form a better impression on how the models would perform when less defaults are present, a shorter data period is considered. Considering a 5-year historical data period, using the most recent data in Table 5.1, a naïve estimate of 0.0505% is obtained. The exclusion of the credit crisis clearly reduces the probability of default. The same inputs for the asset correlation and the default correlation are used as before.

Firstly, the confidence based PD estimates are considered on the 75% and the 90% confidence intervals. At this stage the same conclusions can be drawn as on the 10-year data. Increasing the asset correlation and intertemporal correlation increases the PD estimates. Also, it is clear that the confidence based approaches result in a significant upwards calibration of the PD estimate.

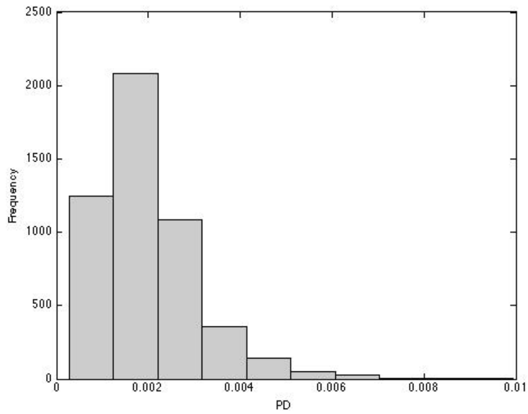
The Bayesian approaches are now considered. Following the previous discussion, the mean of the posterior distribution is used as the PD estimate. However, the results using the median and mode are also given. As before, the HPD is calculated at the 90% confidence interval. Using the uninformative $Uniform(0,1)$ distribution as a prior distribution results in PD estimates that are higher than the confidence based approach at the 75% confidence level, but lower than the PD estimates at the 90% confidence level. However, at the higher confidence levels results are closer to the confidence based results at the 90% confidence level. For the uniform distribution as a prior distribution the HPD widened comparing to the 10-year results. This suggests that due to the fact that less data is used, the prior distribution resulted in a more spread out posterior distribution.

Considering the results of Jeffreys prior, it appears as though the results are more in line with the confidence based approach at 75% confidence level for the lower intertemporal correlation input. Note however that the Jeffreys prior resulted in a higher PD estimate at the lower intertemporal correlation input and at higher asset correlation levels. Furthermore, at the higher intertemporal correlation input the Jeffreys prior is less conservative at all asset correlation levels. The HPD confidence bounds for the Jeffreys prior appears narrower than the corresponding HPD levels for the Uniform distribution.

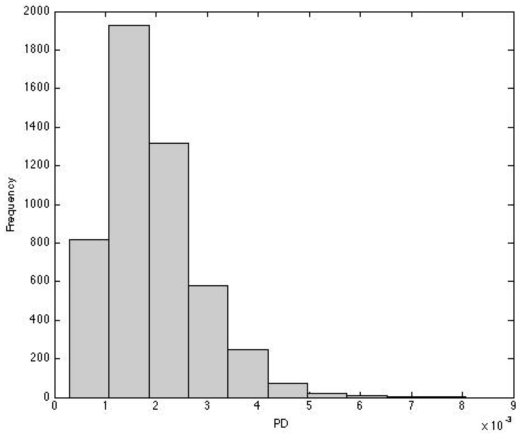
The $Beta(1,400)$ as a prior distribution is also in line with the results for the Jeffreys prior and the confidence based approach at the 75% confidence level. The $Beta(1,400)$ distribution as a prior distribution yields results in line with the $Beta(0.5,0.5)$ at the lower asset correlation inputs, but as the asset correlation increases to higher levels the $Beta(1,400)$ distribution yields less conservative results. The HPD confidence bounds for the $Beta(1,400)$ are also noticeably narrower than the bounds of the distributions that were previously considered.



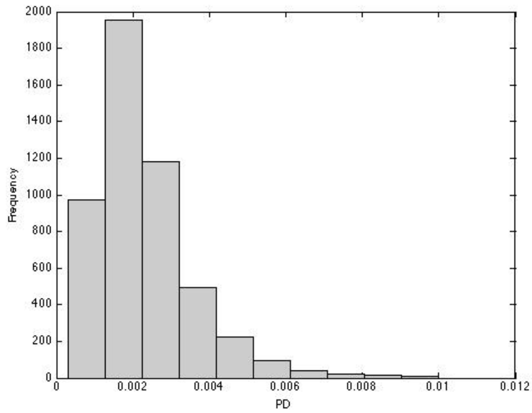
(a) The $Uniform(0, 1)$ as a prior distribution.



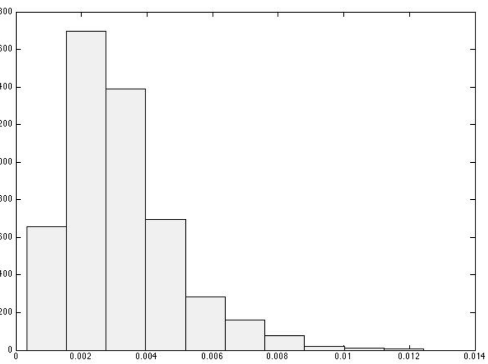
(b) The Jeffreys prior distribution.



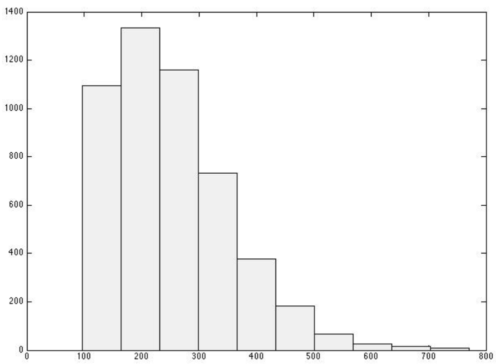
(c) The $Beta(1, 400)$ as a prior distribution.



(d) The Conservative prior distribution.



(e) The Expert prior distribution.



(f) The $Pareto(1)$ as a prior distribution.

Figure 5.4: Posterior Histograms of the PD for Different Prior Distributions.

The conservative prior distribution results in PD estimates in line with the Uniform distribution. The HPD confidence levels for these two priori are also in line with one another. It is interesting to observe how these two models have generated similar results at both the 10-year and 5-year data periods. In both cases the prior distribution resulted in an overly conservative PD estimate. This is, however, as expected.

The expert distribution used for this data set is the same as the one used for the previous data set. This is due to the fact that the prior knowledge is gained from the global corporate default rates. As previously mentioned, the most likely outcome assumed by this prior distribution is 0,06%, with the worst case PD assumed at 0,39%. It is expected that when data is taken into account that this approach would give a more realistic PD estimate since it accounts for the prior knowledge obtained from global historic corporate default trends. In the case for the 5-year data set, the expert distribution results in the least conservative estimates. This is in contrast to the 10-year data set which included more defaults for which the expert distribution as a prior distribution generated more conservative results than most of the other models. The HPD confidence levels for the expert distribution is the smallest of all the models.

Finally, the Pareto distribution is considered as a prior distribution. At the lower asset correlation inputs the results appear in line with the Confidence Based Approach at the 90% confidence level. Hence, generating results that are more conservative than the other Bayesian results. However, as the asset correlation input increases it appears as though the Confidence Based Approach at 90%, and the Bayesian Approaches with the $Uniform(0,1)$ prior and the conservative prior yields more conservative results than the pareto. The HPD confidence bounds of the Pareto distribution is the widest of all the considered distributions. This is, however, as expected due to the extreme nature of the distribution.

As an illustration of the iteration plots obtained for the different prior distributions, refer to Figure 5.5. The provided iteration plots are of the simulations run on the 5-year corporate portfolio data with an 18% asset correlation and a 30% intertemporal correlation. The simulations using the various prior distributions exhibit different degrees of variation . It is particularly noticeable that the simulation iterations using the Expert prior distribution shows high levels of variation between the iterations, where the unobjectionable priori and the $Beta(1,400)$ shows lower levels of variation between the iterations.

As mentioned in Section 5.1.1.2 the 90% highest posterior density can be regarded as the shortest interval in the parameter space that contains 90% of the of the posterior probability. The average credible width is defined as the width between the upper bound and the lower bound of the HPD. From the definition of the HPD, an estimate derived from a HPD with a smaller average credible width can be regarded as more reliable. The average credible widths of the 90% HPD confidence intervals quoted in Tables 5.2 and 5.3 are given in Tables 5.4 and 5.5.

Comparing the average confidence widths of the HPDs at 90% for the 10-year corporate portfolio there is no definitive model that has the narrowest HPD credible interval over all asset correlation and intertemporal correlation inputs. The models with the smallest interval over each combination is given in the penultimate row of Table 5.4 and the models with the widest interval are quoted in the final row of the table. When considering which model resulted in the widest HPD credible intervals of all the models, the $Uniform(0,1)$ and the Pareto models appear to produce the widest estimates as can be seen in Table 5.4. Generally the width of the credible intervals are however more or less in line for the 10-year data set.

Table 5.3: Corporate Portfolio Results: 5-year Data

Naïve PD Estimate:		0,0550%					
Asset Correlation:		12%	18%	24%	12%	18%	24%
Intertemporal Correlation:		30%	30%	30%	50%	50%	50%
Confidence Based Approach							
PD Estimate at 75%		0,13%	0,17%	0,23%	0,21%	0,29%	0,40%
PD Estimate at 90%		0,21%	0,28%	0,40%	0,24%	0,36%	0,51%
Bayesian: $Uniform(0,1)$ **Prior**							
Mean PD Estimate		0,17%	0,26%	0,41%	0,18%	0,27%	0,46%
Median PD Estimate:		0,14%	0,21%	0,33%	0,14%	0,21%	0,35%
Mode PD Estimate:		0,18%	0,08%	0,27%	0,12%	0,02%	0,21%
HPD at 90%:		(0,02%; 0,32%)	(0,02%; 0,51%)	(0,02%; 0,8%)	(0,01%; 0,35%)	(0,01%; 0,55%)	(0,02%; 0,93%)
Bayesian: Jeffreys Prior							
Mean PD Estimate		0,14%	0,20%	0,30%	0,15%	0,23%	0,36%
Median PD Estimate:		0,11%	0,26%	0,22%	0,08%	0,17%	0,26%
Mode PD Estimate:		0,08%	0,09%	0,38%	0,12%	0,04%	0,11%
HPD:		(0,02%; 0,27%)	(0,01%; 0,040%)	(0,02%; 0,62%)	(0,01%; 0,28%)	(0,01%; 0,48%)	(0,01%; 0,73%)
Bayesian: $Beta(1,400)$ **Prior**							
Mean PD Estimate		0,13%	0,18%	0,23%	0,14%	0,19%	0,23%
Median PD Estimate:		0,12%	0,15%	0,19%	0,12%	0,16%	0,19%
Mode PD Estimate:		0,05%	0,05%	0,32%	0,08%	0,06%	0,07%
HPD:		(0,02%; 0,25%)	(0,02%; 0,33%)	(0,03%; 0,45%)	(0,02%; 0,26%)	(0,02%; 0,37%)	(0,02%; 0,45%)
Bayesian: Conservative Prior							
Mean PD Estimate		0,17%	0,26%	0,43%	0,17%	0,29%	0,45%
Median PD Estimate:		0,14%	0,21%	0,33%	0,14%	0,23%	0,35%
Mode PD Estimate:		0,14%	0,11%	0,09%	0,21%	0,19%	0,04%
HPD:		(0,02%; 0,32%)	(0,01%; 0,51%)	(0,03%; 0,89%)	(0,02%; 0,33%)	(0,025%; 0,59%)	(0,022%; 0,92%)
Bayesian: Expert Prior							
Mean PD Estimate		0,13%	0,15%	0,18%	0,14%	0,16%	0,17%
Median PD Estimate:		0,12%	0,14%	0,17%	0,12%	0,15%	0,16%
Mode PD Estimate:		0,11%	0,09%	0,15%	0,11%	0,13%	0,12%
HPD:		(0,005%; 0,23%)	(0,006%; 0,27%)	(0,006%; 0,29%)	(0,006%; 0,25%)	(0,006%; 0,27%)	(0,006%; 0,29%)
Bayesian: Pareto(1) Prior							
Mean PD Estimate:	ψ	*440,021*	*358,080*	*279,29*	*438,510*	*334,94*	*254,480*
	PD	0,23%	0,28%	0,36%	0,23%	0,30%	0,39%
Median PD Estimate:	ψ	*360,306*	*282,280*	*218*	*344,320*	*262,95*	*194,230*
	PD	0,28%	0,35%	0,46%	0,29%	0,38%	0,51%
Mode PD Estimate:	ψ	*148,413*	*148,410*	*148*	*148,410*	*148*	*148,410*
	PD	0,67%	0,67%	0,67%	0,67%	0,68%	0,67%
HPD:	ψ	*(824,81; 113,62)*	*(695,06; 70,98)*	*(537,80; 58,71)*	*(844,39; 99,48)*	*(635,466; 56,68)*	*(496,53; 31,94)*
	PD	(0,12%; 0,88%)	(0,143%; 1,408%)	(0,185%; 1,72%)	(0,118%; 1,01%)	(0,157%; 1,76%)	(0,202%; 3,13%)

For the 5-year data the Expert distribution generally resulted in the narrowest average credible widths and the Pareto distribution clearly resulted in the widest intervals. Furthermore, from Table 5.5, apart from the results for the Pareto distribution, the results for the average confidence widths are generally closely matched.

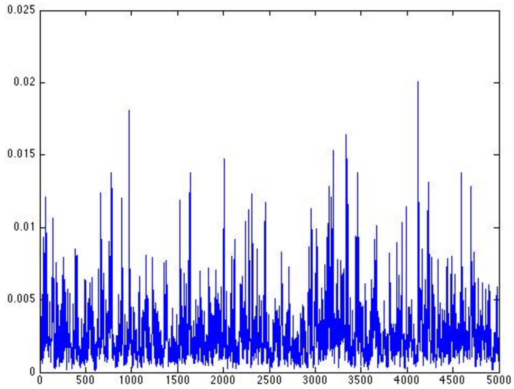
Table 5.4: Average HPD Credible Widths: 10-year Corporate Portfolio

Asset Correlation:	12%	18%	24%	12%	18%	24%
Intertemporal Correlation:	30%	30%	30%	50%	50%	50%
$Uniform(0,1)$	0,002006	0,003182	0,6416	0,2468	0,4709	0,46
Jeffreys Prior	0,002259	0,002951	0,004277	0,002126	0,003664	0,004543
$Beta(1,400)$	0,0025	0,00265	0,003915	0,002206	0,003101	0,003675
Conservative	0,0022600	0,0033750	0,0052770	0,0024740	0,0035	0,006
Expert	0,002773	0,004676	0,00063	0,003094	0,004668	0,0078
Pareto	0,00542	0,00656	0,00798	0,0057	0,00651	0,01008
Narrowest:	$Uniform(0,1)$	Jeffreys Prior	Expert	Jeffreys Prior	$Beta(1,400)$	$Beta(1,400)$
Widest	Pareto	Pareto	$Uniform(0,1)$	$Uniform(0,1)$	$Uniform(0,1)$	$Uniform(0,1)$

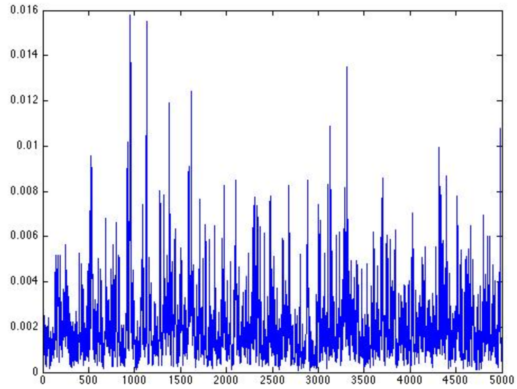
Table 5.5: Average HPD Credible Widths: 5-year Corporate Portfolio

Asset Correlation:	12%	18%	24%	12%	18%	24%
Intertemporal Correlation:	30%	30%	30%	50%	50%	50%
$Uniform(0,1)$	0,00012	0,0049	0,0078	0,0034	0,0054	0,0091
Jeffreys Prior	0,0025	0,0039	0,006	0,0027	0,0047	0,0072
$Beta(1,400)$	0,0023	0,0031	0,0042	0,0024	0,0035	0,0043
Conservative	0,003	0,005	0,0086	0,0031	0,00565	0,007
Expert	0,00225	0,00264	0,00284	0,00244	0,00264	0,00284
Pareto	0,0076	0,01265	0,01535	0,00892	0,01603	0,02928
Narrowest:	$Uniform(0,1)$	Expert	Expert	$Beta(1,400)$	Expert	Expert
Widest:	Pareto	Pareto	Pareto	Pareto	Pareto	Pareto

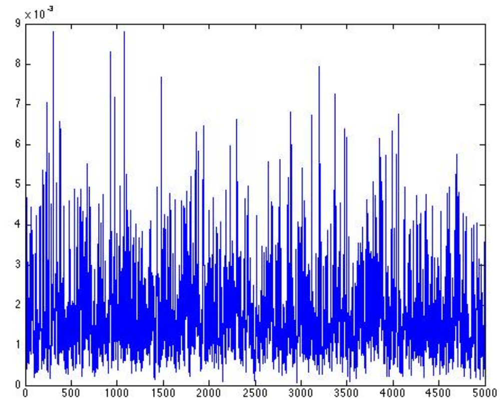
Overall, for the corporate portfolio considered in this study it is clear that all the models considered results in a positive calibration of the portfolio yielding more conservative results as the naïve estimate. This is a positive result, as it is what is expected from all the models and clearly all of the models succeed in their task. Furthermore, all the considered models react as expected when increasing the asset correlation and/or the intertemporal correlation. Generally the confidence based approaches yield the most conservative results. The Bayesian approaches on the other hand generally show mixed results across different correlation inputs. The uninformative prior introduced in this study, namely the Jeffreys prior, appears to be a good option for the data in question across both of the historical data horizons considered. Finally, extreme distribution introduced in this study, namely the Pareto distribution, shows to be a conservative approach for the portfolio in question.



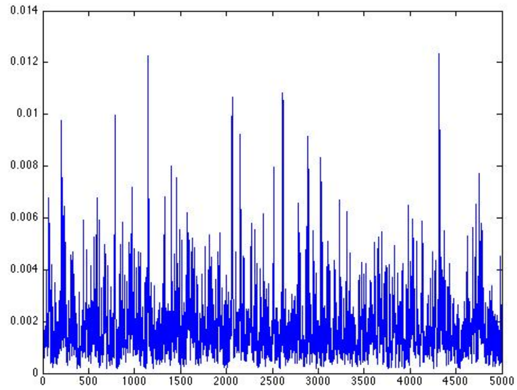
(a) The $Uniform(0, 1)$ as a prior distribution.



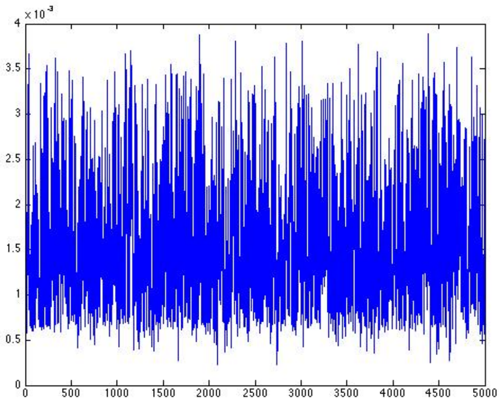
(b) The Jeffreys prior distribution.



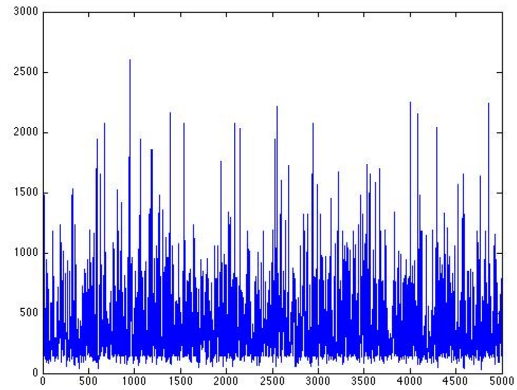
(c) The $Beta(1, 400)$ as a prior distribution.



(d) The Conservative prior distribution.



(e) The Expert prior distribution.



(f) $Pareto(1)$ as a prior distribution.

Figure 5.5: Iteration Plots for the Different Prior Distributions.

5.2.4 Retail Credit Portfolio

The retail credit portfolio is a real South African retail credit portfolio provided by a large South African bank. The data provider may not be disclosed due to confidentiality agreements. The portfolio consists of monthly data for the period 1 January 2011 to 31 July 2014. It is assumed that the portfolio comprises of prime credit. The Retail portfolio is given in Table 5.6.

Since the retail data, provided is monthly data higher levels of the intertemporal correlation need to be specified. Using the same arguments as before of the intertemporal correlation input for the corporate portfolio, the correlation between the systematic factors in months 1 and 5 is given by $Corr(X_1, X_5) = \tau^4$, setting $\tau = 80\%$ results in $Corr(X_1, X_5) = 40,96\%$. This would also result in a correlation between month 1 and 12 of 6,8%. This correlation specification would result in observations in years 1 and 5 to be nearly independent. For the sake of comparison, results are also simulated for a lower intertemporal correlation of 0.4. The lower intertemporal correlation input would translate to a 25,6% correlation between the systematic factors of months 1 and 5. The choice of the intertemporal correlation input would depend on the general economic climate, especially for monthly data.

It is important for the reader to note that corporate and retail portfolios exhibit very different characteristics. Corporate credit typically constitutes the financing of financial institutions. When dealing with corporate credit portfolios with low default characteristics the financial institutions in the portfolio would typically be institutions with high credit ratings. Retail credit portfolios, on the other hand, can be in the form of credit cards, retail mortgages, debit orders, etc. In practice the dynamics behind these different portfolios require different modelling approaches, especially when attempting to capture the systematic factor. This is, however, beyond the scope of this study and will not be discussed in greater detail. The retail credit portfolio considered in this study can, for example, be regarded as a growing regional mortgage portfolio as this is a retail credit portfolio that typically exhibits low default characteristics.

Recall from Section 2.2.3 that the asset correlation for retail credit portfolios are defined at $[3\%; 16\%]$. Therefore simulations are run with the asset correlation defined at the lower bound, at the midpoint (9.5%) and at the upper bound.

As evident in Table 5.6, the Retail portfolio in question is a zero-default portfolio, therefore the naïve estimate of the PD is 0%. As discussed in the literature study, the probability of default for any given portfolio can never be zero. This is one of the main reasons why calibration techniques such as the ones discussed in this study are required. The same number of simulation runs are performed as before. Results for the calibrations performed on Retail Portfolio A is given in Table 5.7.

All of the calibration techniques successfully calibrate the PD estimate to a positive value. The confidence based approaches once again results in a highly conservative calibration, especially at the 90% confidence interval, compared to the other models. The Bayesian Approach with the uninformative $Uniform(0, 1)$ distribution as a prior distribution generates results that are comparable to the confidence based approach at the 75% confidence level with the confidence based calibration being slightly more conservative.

The alternative uninformative prior distribution under consideration, the Jeffreys prior, does

Table 5.6: Retail Credit Portfolio

Month	Obligors	Defaults
Jan 2011	178	0
Feb 2011	109	0
Mar 2011	123	0
Apr 2011	179	0
May 2011	156	0
Jun 2011	155	0
Jul 2011	162	0
Aug 2011	174	0
Sep 2011	177	0
Oct 2011	147	0
Nov 2011	151	0
Dec 2011	185	0
Jan 2012	174	0
Feb 2012	177	0
Mar 2012	175	0
Apr 2012	181	0
May 2012	180	0
Jun 2012	199	0
Jul 2012	187	0
Aug 2012	188	0
Sep 2012	191	0
Oct 2012	189	0

Month	Obligors	Defaults
Nov 2012	194	0
Dec 2012	193	0
Jan 2013	190	0
Feb 2013	191	0
Mar 2013	188	0
Apr 2013	196	0
May 2013	198	0
Jun 2013	199	0
Jul 2013	189	0
Aug 2013	191	0
Sep 2013	186	0
Oct 2013	179	0
Nov 2013	178	0
Dec 2013	181	0
Jan-2014	181	0
Feb 2014	182	0
Mar 2014	183	0
Apr 2014	162	0
May 2014	194	0
Jun 2014	185	0
Jul 2014	194	0

not yield such a conservative calibration. The results generated by using the Jeffreys prior distribution are almost half the size of the PD estimates, which follow from the $Uniform(0, 1)$ distribution.

An industry expert in credit risk modelling was consulted in order to select the parameters for the $Beta$ and Expert prior distributions. Without any knowledge of the specific portfolio's under consideration the industry expert was asked to give his opinion on what the mean, mode, min and max PD estimates for a South African retail credit portfolio with low default characteristics can be. A mean of 2%, a mode of 1%, a min of 0%, and a max of 10% is therefore assigned. Based on this a $Beta(2, 100)$ prior distribution is used. The $Beta(2, 100)$ distribution has a mean of 1,96%. As evident in the results, this specification for the $Beta(2, 100)$ distribution is a conservative specification. The $Beta(2, 100)$ yields much more conservative results than the prior distributions discussed thus far for the current portfolio. Results are also more conservative than the confidence based approach at a 75% confidence level. The confidence based approach at the higher confidence level, however, still produces more conservative calibrated PD estimates.

Comparing the results of the Conservative distribution to those discussed thus far, it is clear that the model does not yield such a conservative calibration. The results generated with the conservative prior distribution are in line with the confidence based approach at the 75% confidence level as well as the results of the $Uniform(0, 1)$ distribution.

Table 5.7: Retail Portfolio Results

Naïve PD Estimate:		0%					
Asset Correlation:		3%	9,5%	16%	3%	9,5%	16%
Intertemporal Correlation:		40%	40%	40%	80%	80%	80%
Confidence Based							
PD Estimate @75%		0,21%	0,29%	0,37%	0,24%	0,35%	0,48%
PD Estimate @90%		0,38%	0,53%	0,71%	0,41%	0,66%	0,88%
Bayesian: $Uniform(0,1)$ **Prior**							
Mean PD Estimate		0,20%	0,24%	0,36%	0,18%	0,30%	0,40%
Median PD Estimate:		0,13%	0,15%	0,21%	0,12%	0,16%	0,23%
Mode PD Estimate:		0,00%	0,00%	0,39%	0,00%	0,39%	0,00%
HPD at 90%:		(0,04%; 0,47%)	(0,0454%; 0,6%)	(0,0454%; 0,87%)	(0,0454%; 0,42%)	(0,0454%; 0,7%)	(0,0454%; 0,36%)
Bayesian: Jeffreys Prior							
Mean PD Estimate		0,10%	0,13%	0,18%	0,10%	0,16%	0,25%
Median PD Estimate:		0,00%	0,06%	0,08%	0,05%	0,07%	0,09%
Mode PD Estimate:		0,005%	0,00%	0,00%	0,00%	0,00%	0,00%
HPD:		(0,0454%; 0,25%)	(0,0454%; 0,33%)	(0,0454%; 0,47%)	(0,0454%; 0,26%)	(0,0454%; 0,39%)	(0,0454%; 0,64%)
Bayesian: $Beta(2,100)$ **Prior**							
Mean PD Estimate		0,29%	0,41%	0,53%	0,30%	0,45%	0,59%
Median PD Estimate:		0,23%	0,32%	0,42%	0,34%	0,34%	0,44%
Mode PD Estimate:		0,38%	0,64%	1,05%	0,24%	0,64%	0,39%
HPD:		(0,0171%; 0,58%)	(0,0977%; 0,84%)	(0,0187%; 1,09%)	(0,0847%; 0,62%)	(0,0122%; 0,96%)	(0,0145%; 1,26%)
Bayesian: Conservative Prior							
Mean PD Estimate		0,17%	0,24%	0,37%	0,18%	0,30%	0,47%
Median PD Estimate:		0,11%	0,15%	0,21%	0,11%	0,16%	0,25%
Mode PD Estimate:		0,00%	0,00%	0,61%	0,00%	0,39%	0,39%
HPD:		(0,0454%; 0,40%)	(0,0454%; 0,58%)	(0,0454%; 0,90%)	(0,0454%; 0,41%)	(0,0454%; 0,69%)	(0,0454%; 1,11%)
Bayesian: Expert Prior							
Mean PD Estimate		0,40%	0,68%	0,97%	0,42%	0,75%	1,03%
Median PD Estimate:		0,30%	0,53%	1,01%	0,30%	0,62%	1,04%
Mode PD Estimate:		0,21%	1,05%	1,05%	1,05%	0,24%	1,05%
HPD:		(0,0128%; 0,9%)	(0,0265%; 1,35%)	(0,0197%; 1,74%)	(0,0178%; 1,03%)	(0,0188%; 1,45%)	(0,0402%; 1,87%)
Bayesian: Pareto(1) Prior							
Mean PD Estimate:	ψ	*2106,50*	*1713,10*	*1420,30*	*2128,60*	*1650,00*	*1331,90*
	PD	0,05%	0,06%	0,07%	0,05%	0,06%	0,08%
Median PD Estimate:	ψ	*889,22*	*669,20*	*477,02*	*891,94*	*605,84*	*408,05*
	PD	0,11%	0,15%	0,21%	0,11%	0,17%	0,25%
Mode PD Estimate:	ψ	*220,23*	*220,26*	*156,23*	*220,26*	*156,02*	*156,02*
	PD	0,45%	0,45%	0,64%	0,45%	0,64%	0,64%
HPD:	ψ	*(5131; 43,3277)*	*(3886,9; 32,93)*	*(3266,9; 17,49)*	*(5273,3; 46,7)*	*(3845,9; 19,39)*	*(3028; 16)*
	PD	(0,0194%; 2,307%)	(0,02572%; 3,036%)	(0,0306%; 5,717%)	(0,0189%; 2,142%)	(0,026%; 5,15%)	(0,033%; 6,075%)

For the Expert distribution the parameters are specified by expert opinion as previously mentioned. Therefore the parameters are $p_l = 0\%$, $p_m = 1\%$ and $p_u = 10\%$. Once again, a specification such as this is a conservative approach for a zero default portfolio. However, there was no knowledge that the portfolio under consideration would be a zero default portfolio and the parameters are specified for a general retail credit portfolio of good credit. As the results indicate, the specified Expert distribution yields the most conservative estimates of all the models considered on this portfolio.

Finally, the Pareto distribution, which is one of the models proposed in this study, yields the least conservative calibration compared to the other models. This result is in line with expectations since the model is expected to capture the low default nature with greater accuracy.

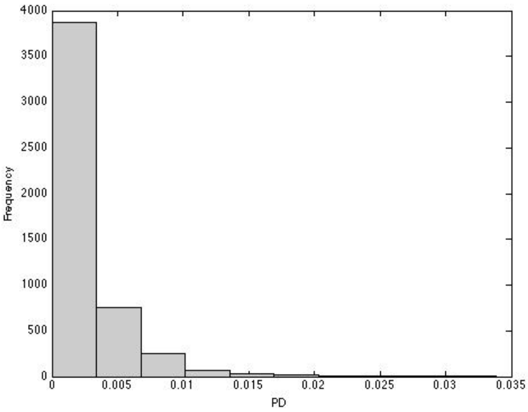
The histograms of the posterior distributions obtained through simulations at an asset correlation of 9.5% and an intertemporal correlation of 40% is given in Figure 5.6. The histograms for the zero default case exhibits a completely different shape than in the corporate scenario where defaults are present (Figure 5.4). This should, however, make intuitive sense as the probabilities are expected to be much lower in this scenario. Hence there is a much larger positive skewness. This is also clear when examining the mean, median and mode estimates of the PD in Table 5.7. Another interesting observation can be drawn form Figure 5.6e, where it is shown how the parameter estimates in question changed the shape of the distribution "pulling" it away from the lower end of the spectrum. This effect clearly results in the more conservative PD estimates. Furthermore, note in Figure 5.6f that there are quite a few extreme outliers in the right tail of the distribution.

The average credible widths of the 90% HPDs for the Retail results are given in Table 5.8. There are two clear trends in the results. First of all, considering the results for the smallest HPD widths, the uninformative priori yielded the best results; the Jeffreys prior specifically produced the narrowest HPD intervals across nearly all correlation combinations. On the other hand, the Pareto distribution clearly produced the widest HPD intervals across all correlation inputs. An interesting point to mention is that the average HPD credible widths for the retail portfolio are not so closely correlated as with the corporate portfolio.

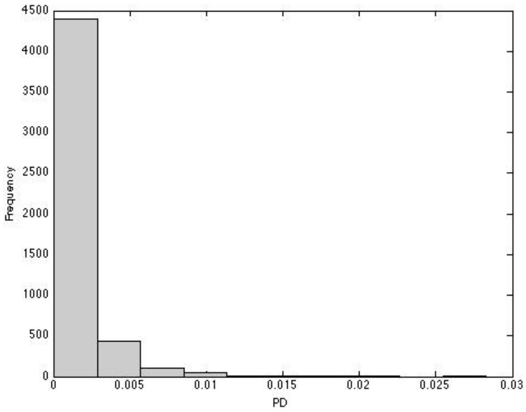
Table 5.8: Average HPD Credible Widths: Retail Portfolio

Asset Correlation:	12%	18%	24%	12%	18%	24%
Intertemporal Correlation:	30%	30%	30%	50%	50%	50%
$Uniform(0, 1)$	0,0043	0,005546	0,008246	0,003746	0,006546	0,003146
Jeffreys Prior	0,002046	0,002846	0,004246	0,002146	0,003446	0,005946
$Beta(2, 100)$	0,005629	0,007423	0,010716	0,005353	0,00838	0,012455
Conservative	0,003583	0,004577	0,00647	0,003207	0,004934	0,006509
Expert	0,008872	0,013235	0,017203	0,010122	0,014312	0,018298
Pareto	0,022876	0,030103	0,056864	0,021231	0,05124	0,06042
Narrowest:	Jeffreys	Jeffreys	Jeffreys	Jeffreys	Jeffreys	$Uniform(0, 1)$
Widest:	Pareto	Pareto	Pareto	Pareto	Pareto	Pareto

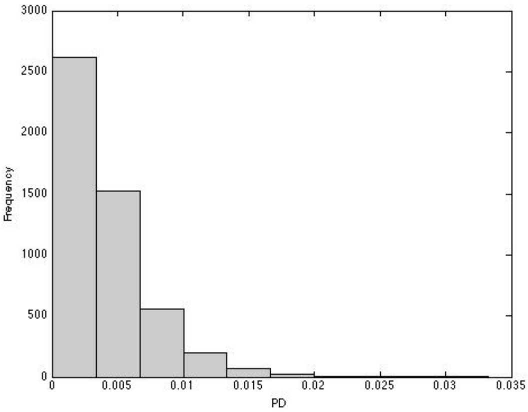
Overall all the models successfully calibrated the zero default portfolio and produced a strictly positive estimate. There are, however, various degrees of conservatism applied by the different models. The models that produce the estimates closest to the low default nature of the portfolio are the models proposed in this study. The Pareto distribution as a prior produced



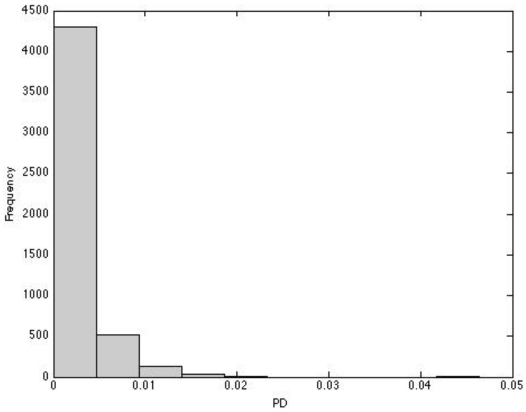
(a) The $Uniform(0, 1)$ as a prior distribution.



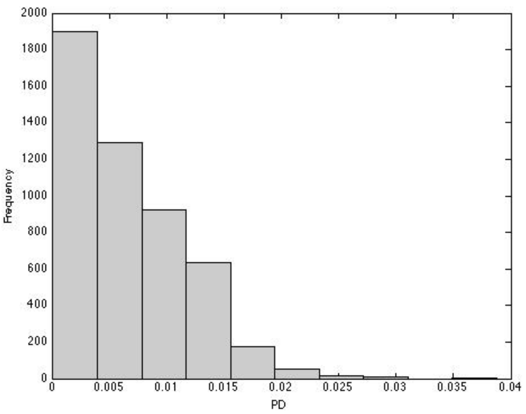
(b) The Jeffreys prior distribution.



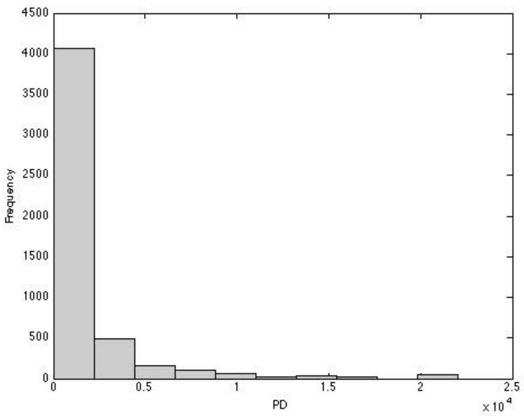
(c) The $Beta(2, 100)$ as a prior distribution.



(d) The Conservative prior distribution.



(e) The Expert prior distribution.



(f) The $Pareto(1)$ as a prior distribution.

Figure 5.6: Posterior Histograms of the PD for the Different Prior Distributions.

the smallest PD estimates, clearly capturing the low default nature of the portfolio. This result is in clear contrast to the corporate portfolio where defaults were present and the Pareto produces much more conservative calibrations. The Jeffreys prior appears to be a good fit to the data and produced the narrowest HPD credible intervals and results that are not overly conservative. The other models all generated much more conservative estimates, especially for a portfolio with a zero default nature.

Another important area of consideration when comparing models is the models' sensitivity to parameter inputs. In the next section the model sensitivities to these inputs is investigated in greater detail.

5.2.5 Model Sensitivities

In this section the respective models' sensitivities to the asset correlation and the intertemporal correlation are investigated. The asset correlation is varied across 10% and 50% and the intertemporal correlation is varied between 20% and 70%. All of the simulations are performed using the same amount of simulations as before and results are displayed in the form of a surface plot. The results for the corporate portfolio at the 10-year data horizon is given in Figure 5.7 and the results for the retail portfolio is given in Figure 5.8. The results in the plots are summarised in Tables 5.9 and 5.10. In the tabled summaries, the min refers to inputs $\xi = 10\%$ and $\tau = 20\%$, the base refers to inputs $\xi = 30\%$ and $\tau = 40\%$, and max refers to inputs $\xi = 50\%$ and $\tau = 70\%$. The models are then assigned a sensitivity ranking based on the size of the change from the min input to the max input.

For the corporate portfolio it is clear that the Confidence Based approach has the highest sensitivity to the correlation inputs. Another model that appears to exhibit a high level of sensitivity to the correlation parameters is the Bayesian model using the Expert distribution as a prior distribution. The Bayesian model with the $Beta(1, 400)$ prior specification, on the other hand, has the lowest sensitivity to the correlation inputs. The model using the $Pareto(1)$ distribution as a prior distribution also appears to have a low sensitivity. All other models are regarded as having a medium sensitivity to these parameters.

Table 5.9: Corporate Portfolio Sensitivity Summary

Model	Min	Base	Max	Sensitivity
Confidence Based at 90%	0,39%	0,68%	3,26%	High
$Uniform(0, 1)$	0,15%	0,47%	2,00%	Medium
Jeffreys Prior	0,14%	0,41%	1,55%	Medium
$Beta(1, 400)$	0,14%	0,30%	0,50%	Low
Conservative	0,16%	0,53%	1,83%	Medium
Expert	0,17%	0,67%	2,86%	High
Pareto	0,38%	0,59%	1,28%	Low

Similar results are obtained for the retail portfolio. The Confidence Based approach once again shows the highest sensitivity to the correlation inputs. Other models that show high levels of sensitivity are the Bayesian models with the $Uniform(0, 1)$ and the Conservative prior.

For the retail data the Pareto distribution shows the lowest sensitivity to the correlation inputs and the $Beta(2,100)$ distribution also shows low levels of sensitivity to these parameters. Other models show medium levels of sensitivity.

Table 5.10: Retail Portfolio Sensitivity Summary

Model	Min	Base	Max	Sensitivity
Confidence Based at 90%	0,37%	1,04%	3,58%	High
$Uniform(0,1)$	0,24%	0,77%	2,56%	High
Jeffreys Prior	0,13%	0,34%	1,18%	Medium
$Beta(2,100)$	0,40%	0,79%	1,25%	Low
Conservative	0,23%	0,78%	2,68%	High
Expert	0,84%	1,70%	2,46%	Medium
Pareto	0,06%	0,11%	0,29%	Low

Generally, when financial models are built, the correlation inputs always comes in the focus of attention. This is due to the fact that different modellers would always obtain different estimates of the correlation, whether it be an asset correlation or an intertemporal correlation. Therefore there is always an element of subjectivity when using a certain correlation input. Due to this fact it is undesirable to have a model that is overly sensitive to the correlation inputs. Nonetheless, the asset correlation and intertemporal correlations are both fundamental parameters and play a key role in estimating the PD. Therefore, it is also undesirable to work with a model that is insensitive to these inputs. Generally, a model that has moderate levels of sensitivity to parameter inputs is preferred. This is a topic that deserves a more detailed discussion, while it is, however, placed outside the scope of this study.

The next section provides a brief illustration of how the PD parameter estimates fall into calculating the RWA for credit portfolios under Basel. It should, however, be emphasised that this is not the focus of this study and, therefore, the following section should only be regarded as an illustration and not as guidance on RWA calculations under Basel III.

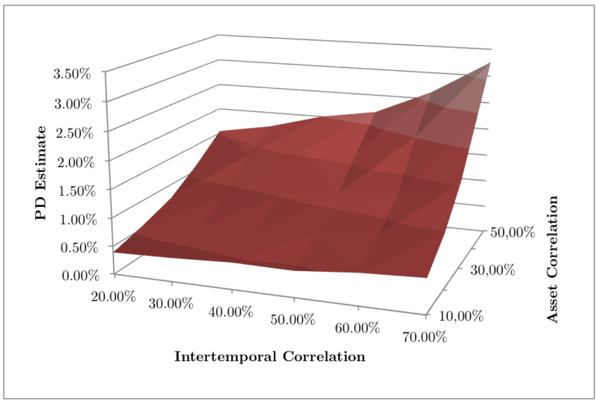
5.2.6 Comparing Risk Weighted Asset Estimates

In this section a practical implementation of the Basel Credit Risk function is illustrated using the different PD estimates obtained in the previous sections. Recall from Section 2.2.3 that the risk weight function is derived from the capital requirement, defined as K per unit of currency of the exposure, where the general formulation of K is:

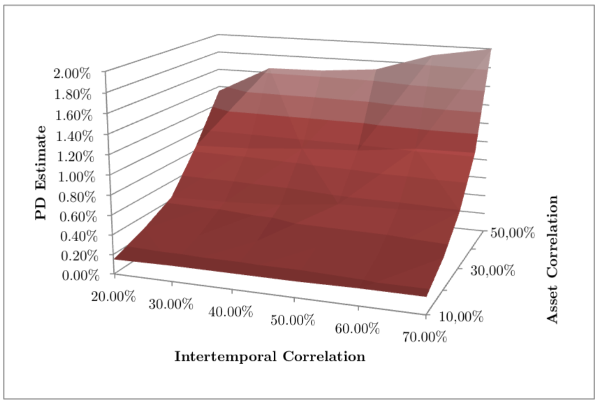
$$K = LGD\left[\Phi\left(\frac{\Phi^{-1}(p)}{\sqrt{1-\xi}} + \Phi^{-1}(0.999)\sqrt{\frac{\xi}{1-\xi}}\right) - p\right]\frac{1+(M-2.5)\kappa}{1-1.5\kappa}.$$

The inputs p and LGD are measured in decimals, Φ is the standard normal cumulative distribution function, and, M is defined as the effective maturity. The effective maturity is fixed at 1 year for retail exposures, and assumes values between 0 and 5 years for other exposures. Furthermore, κ is defined as the maturity adjustment given by:

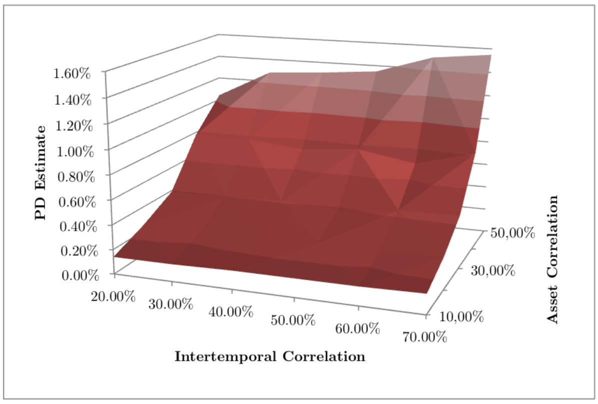
$$\kappa = (0.11852 - 0.05478 \times \ln(p))^2.$$



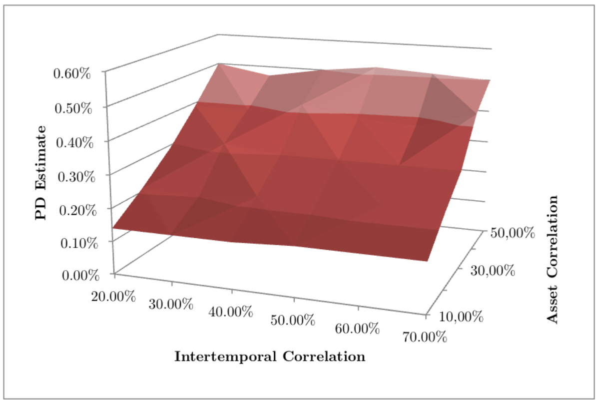
(a) The Confidence Based Approach at 90%.



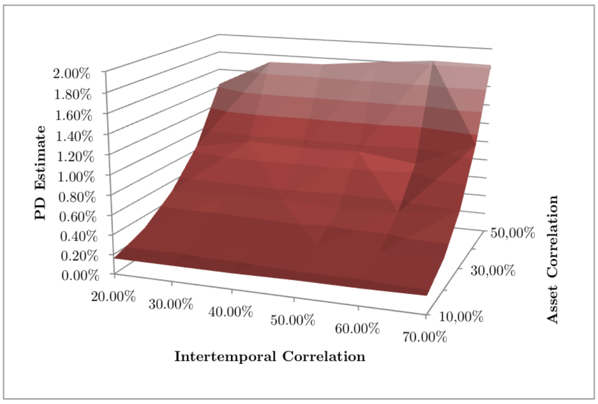
(b) The $Uniform(0, 1)$ as a prior distribution.



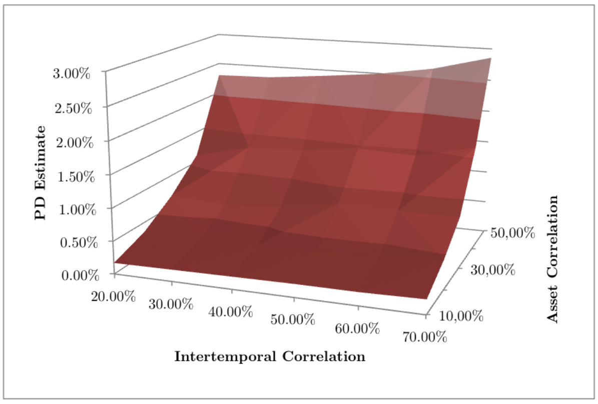
(c) The Jeffreys prior distribution.



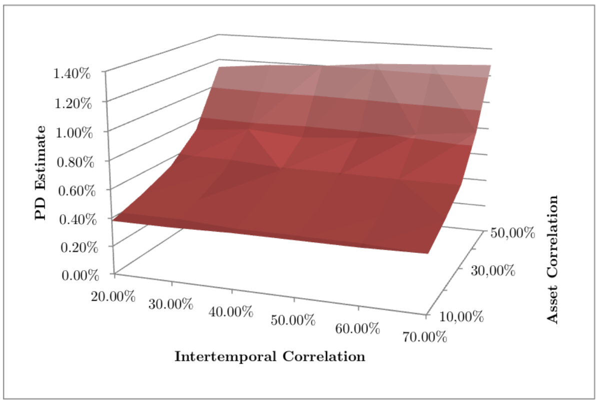
(d) The $Beta(1, 400)$ as a prior distribution.



(e) The Conservative prior distribution.

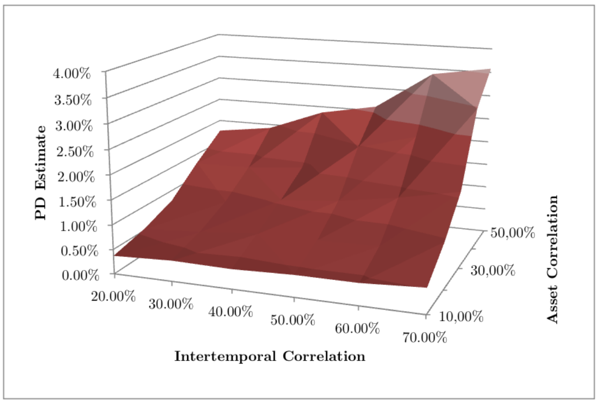


(f) The Expert prior distribution.

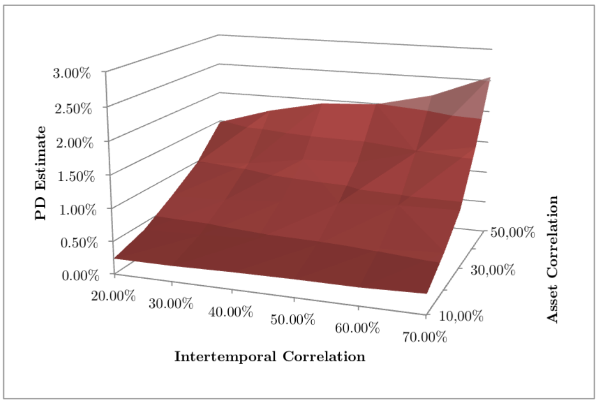


(g) The $Pareto(1)$ as a prior distribution.

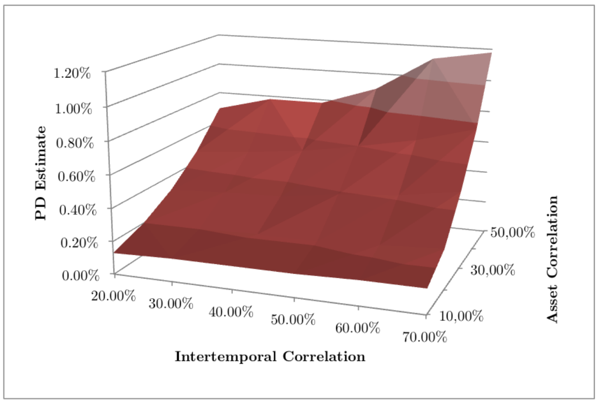
Figure 5.7: Sensitivity Analyses of the PD Estimate for the Corporate Data.



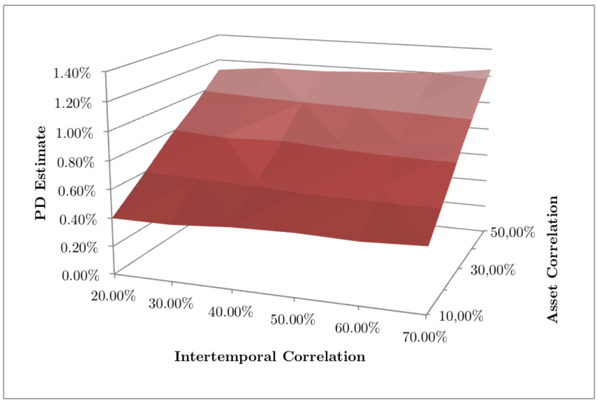
(a) The Confidence Based Approach at 90%.



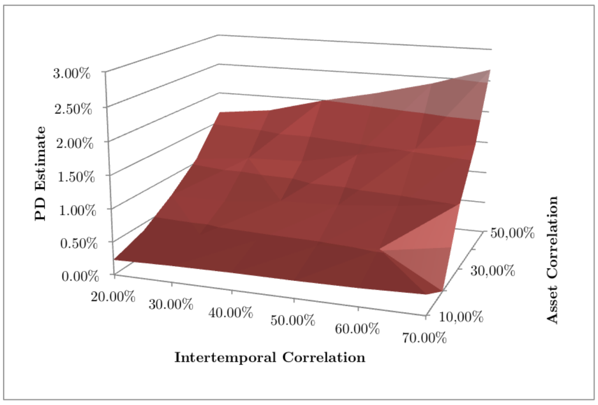
(b) The $Uniform(0, 1)$ as a prior distribution.



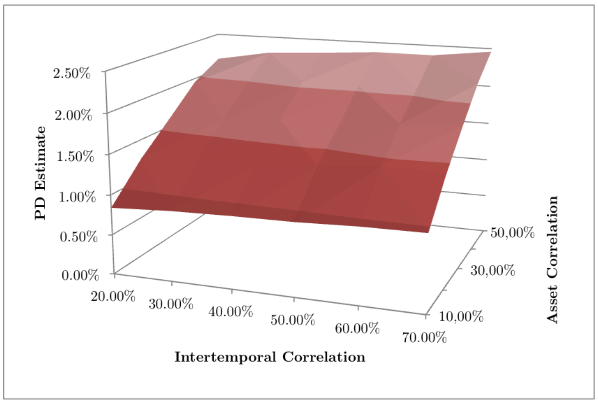
(c) The Jeffreys prior distribution.



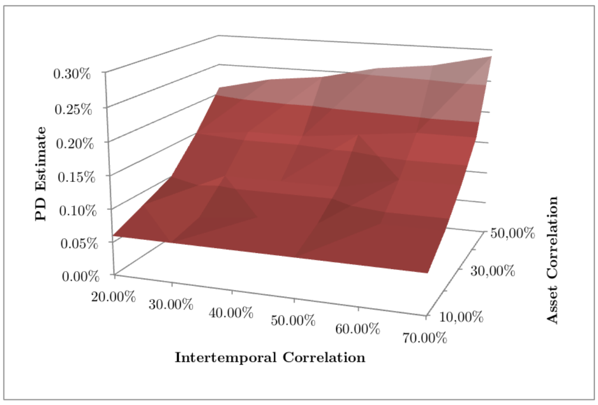
(d) The $Beta(1, 400)$ as a prior distribution.



(e) The Conservative prior distribution.



(f) The Expert prior distribution.



(g) The $Pareto(1)$ as a prior distribution.

Figure 5.8: Sensitivity Analyses of the PD Estimate for the Retail Data.

The capital charge on the risk, also known as the risk weighted assets (RWA), is given by:

$$RWA = K \times 12.5 \times EAD. \tag{5.2.1}$$

Firstly, the RWA for a fictitious portfolio of corporate credits held by a large South African Bank is calculated. Assume an EAD for the portfolio of R1bn. As mentioned, in Section 2.2.2, the LGD is bimodal, therefore, results are given for two respective LGD inputs. Initially, a smaller LGD of 45% is assumed, this is in line with Clifford *et al.* (2013). Then a larger LGD of 75% is considered. Define the effective maturity, M, as 5 years. Assume an asset correlation of 18% and an intertemporal correlation of 30%. The PD estimates from Table 5.2 corresponding to the assumed correlation inputs are used. The RWAs calculated using the inputs assumed above is given in Table 5.11.

Table 5.11: RWA Calculations for Corporate Portfolio

		LGD:45%			LGD:75%		
Model	PD Estimate (p)	κ	K	RWA	κ	K	RWA
Naïve PD Estimate	0,13%	0,2333	0,0314	R393 486,465	0,2333	0,0524	R655 810,80
Confidence Based at 75%	0,26%	0,1976	0,0459	R574 712,43	0,1976	0,0766	R957 853,94
Confidence Based at 90%	0,41%	0,1760	0,0585	R731 630,45	0,1760	0,0975	R1 219 384,11
Bayesian: $Uniform(0,1)$	0,20%	0,2106	0,0399	R499 196,77	0,2106	0,0665	R831 994,62
Bayesian: Jeffreys	0,20%	0,2106	0,03993	R499 196,77	0,2106	0,06655	R83 199,62
Bayesian: $Beta(1,400)$ Prior	0,19%	0,2132	0,03884	R48 557,79	0,2132	0,0647	R809 296,65
Bayesian: Conservative Prior	0,23%	0,2036	0,0430	R538 172,20	0,2036	0,0717	R896 953,61
Bayesian: Expert Prior	0,31%	0,1891	0,0504	R631 203,91	0,1891	0,0841	R1 052 006,53
Bayesian: $Pareto(1)$	0,44%	0,1725	0,0609	R761 967,98	0,1725	0,1015	R1 269 946,64

Finally, the RWA for a fictitious portfolio retail credits held by a large South African Bank is calculated. Assume an EAD for the portfolio of R10m. Once again two LGD levels are considered, initially an LGD of 15%, in line with Clifford *et al.* (2013), is assumed. Then an higher LGD of 55% is also considered. As per the Basel definition the effective maturity is set at 1 year. Assume an asset correlation of 9.5% and an intertemporal correlation of 80%. The PD estimates from Table 5.7 corresponding to the aforementioned correlation inputs will be used. The RWAs calculated using the inputs assumed above is given in Table 5.12.

Table 5.12: RWA Calculations for Retail Portfolio

		LGD:15%			LGD:55%		
Model	PD Estimate (p)	κ	K	RWA	κ	K	RWA
Naïve PD Estimate	0%	-	-	-	-	-	-
Confidence Based at 75%	0,35%	0,1834	0,0044	R559,75	0,0520	0,0164	R2 052,43
Confidence Based at 90%	0,66%	0,1548	0,0071	R895,56	0,0644	0,0262	R3 283,74
Bayesian: $Uniform(0,1)$	0,30%	0,1907	0,0039	R498,14	0,0491	0,0146	R1 826,53
Bayesian: Jeffreys	0,16%	0,2220	0,0024	R307,04	0,0381	0,0090	R1 125,80
Bayesian: $Beta(2,100)$ Prior	0,45%	0,1718	0,0054	R675,65	0,0568	0,0198	R2 477,39
Bayesian: Conservative Prior	0,30%	0,1907	0,0039	R498,14	0,0491	0,0146	R1 826,53
Bayesian: Expert Prior	0,75%	0,1494	0,0078	R982,47	0,0670	0,0288	R3 602,27
Bayesian: $Pareto(1)$	0,06%	0,2749	0,0011	R142,17	0,0234	0,0041	R521,29

The three principal credit risk inputs are the main drivers of the RWA calculation, this is also evident in the results. Furthermore, from these results it is clear how important the PD estimate is when calculating the RWA. Without going into detail, it can be seen in Table 5.11

that the PD estimate is a strong driver of the RWA calculation under Basel. It is clear that over-estimation of the PD estimate can result in large amounts of regulatory capital being tied up, and this can have a negative effect on business growth as the capital is taken out of the business cycle. Under-estimation on the other hand can result in inadequate regulatory capital being held for the risk exposures on the banking book. As a result, when losses are realised then the financial institution may not have enough capital buffers in place to absorb the losses. Note, however, that in practice the calculation of regulatory capital is much more complex than discussed above and that this is merely an example to illustrate the importance of the PD estimate in practical terms.

5.3 Comments on the PD Calibration Models

Section 5.2 compares all of the models discussed Chapters 3 and 4 on an empirical basis. In this section detailed and critical comments are made on the models reflecting on the results and the theory of the models.

From the results in the previous section it is clear that the Confidence Based Methodology from Section 3.3.1 successfully calibrates the PD estimate although the model can be criticised for being overly conservative. This can especially be noted for results at the 90% confidence level. The industry benchmark for calibrations using the Confidence Based Method is, however, to use a confidence level of 75% as highlighted by Clifford *et al.* (2013). The results on the 75% confidence level is generally comparable to some of the other more conservative approaches. The Confidence Based method is, however, highly sensitive to the correlation inputs, compared to the other models that were considered.

The Bayesian methodologies also successfully calibrate the PD estimate when the model parameters are correctly specified. Incorrect specification of the prior distribution parameters can result in distorted estimates. The need to specify prior parameters for the informative prior distributions that are used in the Bayesian approaches increases the flexibility of the models. It also increases the subjectivity of the models. This is one of the points of criticism from financial regulators on using Bayesian models with informative priori.

Firstly, consider the uninformative priori namely the Uniform and Jeffreys priori. Both the Bayesian models use these priori successfully calibrate the PD estimate. Also, when using a $Uniform(0,1)$ prior distribution, then the need for specification of a prior parameter is eliminated. The $Uniform(0,1)$ prior can however be criticised for being overly conservative. The results obtained from the $Uniform(0,1)$ prior distribution are generally comparable to the confidence based approaches at the 75% confidence level. Although the model is less sensitive to the correlation inputs than the confidence based model, it is still more sensitive to these inputs than the other models. As previously mentioned, it does not make intuitive sense to assume equal probabilities for all possible PD observations. Therefore, as an alternative unobjectionable prior distribution the Jeffreys prior is proposed in this study.

The Jeffreys prior for the current PD calibration scenario translates to a $Beta(0.5,0.5)$ prior distribution. The probability density function of this distribution is compared to that of the $Uniform(0,1)$ in Figure 5.9. From the figure this prior makes more intuitive sense as a prior distribution as it places more weight to the tails of the distribution. Considering the results of the Bayesian model using a Jeffreys prior, the model performs very well in calibrating the PD estimate. The calibration seems more balanced than that of the $Uniform(0,1)$ in the

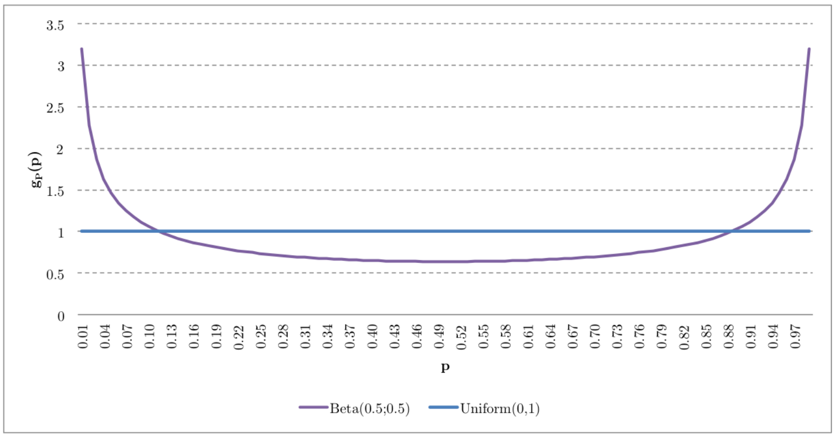


Figure 5.9: Comparison of the p.d.f.'s of the Jeffreys Prior and the Uniform Prior.

sense that it still applies a sound degree of conservatism when defaults are present in the data. However, much less conservatism is applied when there are no defaults in the data. This is a very favourable result. Furthermore, the model also does not appear overly sensitive to the correlation inputs which is favourable.

Moving to the informative priori, the $Beta(\alpha, \beta)$ is another possibility of using a Beta distribution as a prior distribution. As mentioned in Section 3.3.2, the Beta distribution as a prior distribution allows great flexibility. However, along with the greater flexibility, the risk of wrongful specification of the parameter is increased. As illustrated in Section 5.2.1, if the Beta distribution parameters are wrongly specified there is the risk that the model may under calibrate the PD. This is especially true when defaults are present in the data. However, the Beta parameter specifications that are used in the empirical study appears reasonable for the respective scenarios. In both the corporate and the retail cases the model performs successful calibrations of the PD estimate. The model, however, has low sensitivity towards the correlation inputs and, which can be regarded as an unfavourable characteristic.

The conservative prior introduced by Tasche (2013), did surprisingly not induce an overly conservative calibration for the 10-year corporate data set, when multiple defaults are present in the historical data period. For the 5-year corporate data and the retail data the model, however, resulted in much more conservative PD estimates. Therefore, the model disregarded the defaults present in the historical data period, which is an unfavourable observation. Another admonishing observation is that the model shows higher levels of sensitivity to the correlation inputs. The model has the advantage that no model parameters needs to specified and therefore this reduces subjectivity.

The expert prior introduced by Clifford *et al.* (2013) exhibits high levels of flexibility with the additional benefit of parameter inputs being simple to understand. The parameters, however, introduce high degrees of subjectivity in the model and as such a desired PD estimate can be tailored to a certain degree. The model, however, successfully calibrates the PD estimate with the degree of conservatism entirely depending on the specified parameters inputs. It appears as though the parameters specified in this study were overly conservative for the considered data sets as the results are in line (or exceeding) the confidence based approach at a 90% confidence level. This highlights the challenge of using this distribution and selecting the appropriate parameters. The model appears to show relatively high levels of sensitivity to the correlation inputs.

Finally, the extremal natured Pareto distribution that was introduced in this study shows very interesting results. As a first point of discussion, the model appears insensitive to the input parameter, γ. This is advantageous in that it reduces the risk of specifying the wrong parameter. However, this also means that there is a degree of inflexibility present in the model. Interesting observations can be made from the empirical results of the corporate and retail portfolio data. Recall that defaults are present in the corporate default data, it appears as though the model accounted for the defaults present in the data and as a result highly conservative PD estimates comparable to the 90% confidence based estimates are observed. The results for the retail data, however, show the contrary. For the zero default portfolio the Pareto model produces the least conservative estimates. Clearly, the data is a strong driver of the result produced by the model. This can be seen as favourable. With regards to the model sensitivity to the correlation inputs, the model appears very insensitive to these inputs. This can be regarded as an unfavourable attribute.

The Confidence Based approaches are widely used in industry. However as seen in the results this approach deserves the critique that it results in overly conservative PD estimates. The Bayesian approaches clearly serves as an alternative to the Confidence Based Approaches. As an uninformative prior, the Jeffreys prior showed extremely favourable results and can be recommended for the LDP setting. On the side of the informative priori, it should once again be highlighted that the parameter inputs of the prior distribution introduce some problems. The Pareto distribution introduced in this study accounts for the extremal nature of low default portfolios. However, the model is very conservative when defaults are present in the data and it may be criticised that it initiates too small of a calibration in the absence of defaults in the historical data.

5.4 Conclusion

This chapter constitutes a major part of this study. At the outset of the chapter an unobjectionable prior in the form of the Jeffreys prior is introduced. The simulation procedure used for producing the empirical results is then discussed in detail. Section 5.2 presents the comparison between all the models considered in the study and discusses the results in detail. As evident in the results, all of the PD calibration methods considered throughout this study successfully calibrate the PD estimate. There are, however, various degrees of conservatism applied by the respective models.

The models are compared with regards to their performance as the default rate is increased away from the low default spectrum. Then, using two data sets with different characteristics, the PD is estimated and results are compared for various correlation inputs. In this comparison the HPDs for the Bayesian models are considered. Another important element of comparison is the respective models sensitivities to the correlation inputs. As an illustration of the practical implementation of the PD estimates, risk weighted assets for fictitious portfolios are calculated and discussed. This chapter is then concluded with a comparison of all the considered models and corresponding results.

Chapter 6

An Interactive PD Calibration Model

In this chapter the MATLAB Graphical User Interface (GUI) that acts as a front end interface for the MATLAB code and the underlying model's are discussed. The aim of the MATLAB GUI is to increase the accessibility of the models and associated code. A front end graphical user interface would allow risk practitioners with limited MATLAB programming knowledge to utilise the PD calibration models that are discussed in this study.

The first part of this chapter discusses the advantages and disadvantages of using a GUI for a financial model. The second section is contributed to discussing the GUI that is designed for the models in this study. Note that this chapter does not discuss how a GUI is designed, but rather poses as a guide to using the GUI designed in this study and in the process attempt to highlight its benefits in the context of the study. The code for the GUI program will not be disclosed in the appendix along with the code for the models. The GUI code can, however, be requested from the author.

6.1 Advantages and Disadvantages

There are several advantages to building a GUI for a financial model that is designed in a programming language. The main advantage lies in the user-friendliness and the fact that it may speed up the user's work. Another strong advantage of well designed GUIs is that the model becomes more attractive to non technical people. In the modern era there is also a sense of professionalism and good practice that lies in designing a GUI for a financial model. This is especially the case in large financial institutions where models may be sold as business or where various departments or teams may use a specific model as a tool. Therefore, it is important to make coding tools accessible and a GUI generally improves accessibility when well designed.

The main disadvantage of designing a GUI is that when it is not properly built then it may be difficult to use and work with. It is expected that a GUI is designed such that its use is intuitive. Another disadvantage is the fact that a model running through a GUI generally requires more computing memory than non graphical models. It can also take up more hard drive space due to the visual nature.

6.2 The MATLAB Graphical User Interface

As previously mentioned, when building a financial model in a programming language it is generally good practice to build either front end code or a front end interface for the model in order to improve the usability of the model. All of the models discussed in this study has been coded in MATLAB and the GUI then groups the models into a single package. Upon running the PD calibration package designed in this study, the front end interface will open. This interface is displayed in Figure 6.1. After the interface window has been opened then the user needs to select which PD calibration methodology will be used.

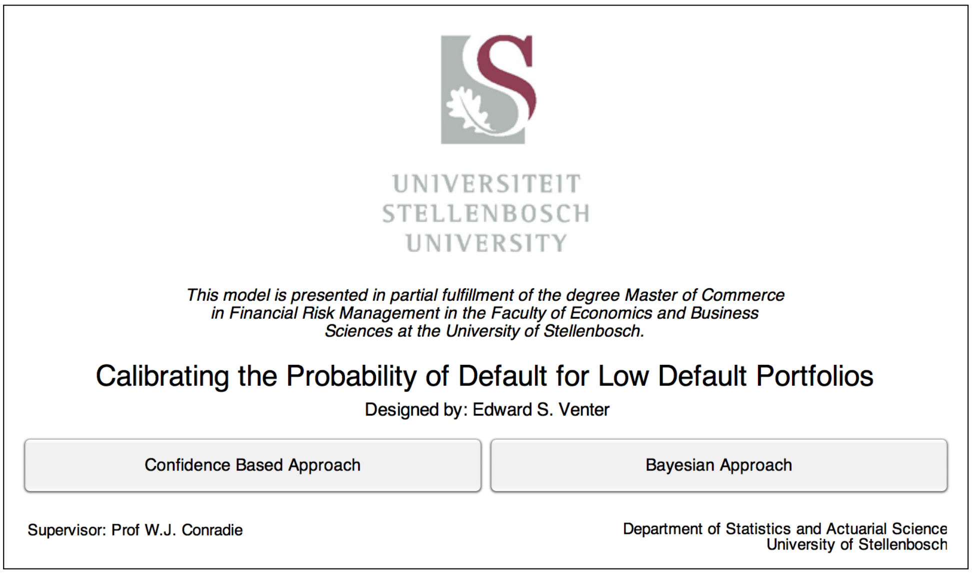


Figure 6.1: The Front End MATLAB GUI Interface

After selecting the PD calibration methodology that is to be applied, then a new window opens that represents the associated model. Firstly, the interface that governs the Confidence Based Approach will be discussed in Section 6.2.1. After which the interface that governs the Bayesian Approach will be discussed in Section 6.3.

6.2.1 Confidence Based Approach Interface

The interface window that opens when the Confidence Based Approach is selected, is displayed in Figure 6.2. As a first step, any user that is unfamiliar with the model and its inputs can click on the "Model Guide" button to open a user guide for the model. The model guide is displayed in Figure 6.3. The interface window allows the user to specify the model inputs by simply entering the inputs into the input boxes and then after the confidence interval underlying the simulation is selected, then the PD is calibrated by selecting the "Calculate" button.

Three additional buttons that generate the input from the original examples in Pluto and Tasche (2011) has been added in order to provide an illustration of how the model works. As an example of the input and output from the GUI consider Figure 6.4 which illustrates the functionality of the interface. An independent one-period case is generated by selecting the "Independent, One-Period" button. If the desired confidence level for the calibration is 90%, then this is selected from the drop down list and then the calculation is performed by executing

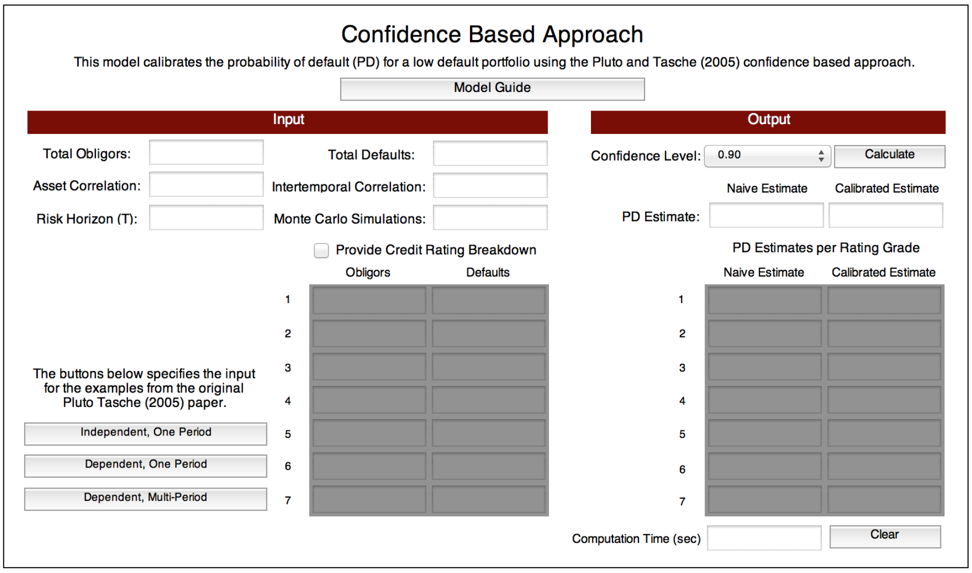


Figure 6.2: The Confidence Based Approach Interface

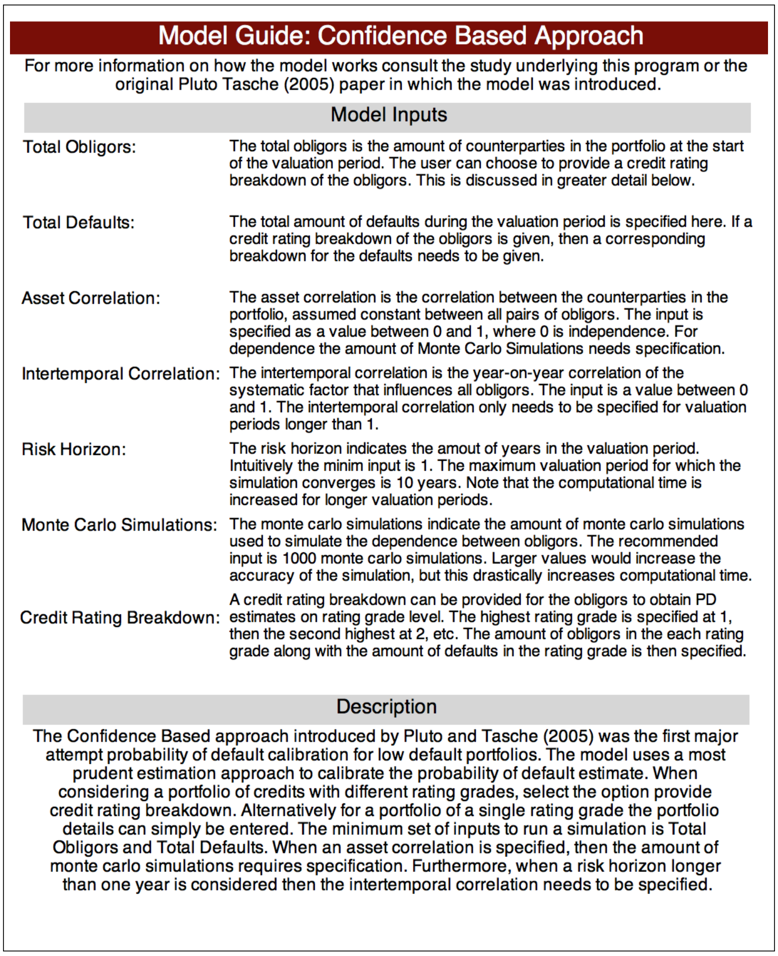
the "Calculate" button. An example of the output of this example is given in Figure 6.4a. Similarly, by selecting "Dependent, One-Period" the desired example is generated in the interface and upon executing the "Calculate" button the output given in Figure 6.4b is obtained. Finally, for an example of the dependent multi-period case the "Dependent, Multi-Period" button is selected and the desired input is generated. The output given in Figure 6.4c is then obtained by executing the calculation. A button has also been added that clears the interface window after a calculation has been performed.

As can be seen in the Computation Time output in Figure 6.4c, it can take some time to perform the calculation, therefore a time warning has been added to the interface. Upon executing the calculation the note window displayed in Figure 6.5a appears to inform the user that the simulation might be time consuming. Other warnings has also been added to the interface to improve user-ability. These include; a warning that not enough input has been specified (see Figure 6.5b), and warnings for when variable inputs are wrongly specified (e.g. the warning in Figure 6.5c for when an asset correlation outside the bounds of 0 and 1 has been specified).

6.3 Bayesian Approach Interface

When the Bayesian Approach is selected in the front end interface (Figure 6.1), then the Bayesian interface displayed in Figure 6.6 is opened. Once again the model has a model guide to assist any user that is unfamiliar with the model and its inputs. The model guide can be generated by executing the "Model Guide" button; the generated guide is displayed in Figure 6.7. The interface window has a similar functionality as the one for the Confidence based approach. The interface window allows the user to specify the model inputs by simply entering the inputs into the input boxes, the user then selects the prior distribution that will be considered and after entering the prior parameter inputs then the PD is calibrated by selecting the "Calculate" button.

A single example is included in the interface. By executing the "Example" button, the 5-year



Model Guide: Confidence Based Approach

For more information on how the model works consult the study underlying this program or the original Pluto Tasche (2005) paper in which the model was introduced.

Model Inputs

Total Obligors:	The total obligors is the amount of counterparties in the portfolio at the start of the valuation period. The user can choose to provide a credit rating breakdown of the obligors. This is discussed in greater detail below.
Total Defaults:	The total amount of defaults during the valuation period is specified here. If a credit rating breakdown of the obligors is given, then a corresponding breakdown for the defaults needs to be given.
Asset Correlation:	The asset correlation is the correlation between the counterparties in the portfolio, assumed constant between all pairs of obligors. The input is specified as a value between 0 and 1, where 0 is independence. For dependence the amount of Monte Carlo Simulations needs specification.
Intertemporal Correlation:	The intertemporal correlation is the year-on-year correlation of the systematic factor that influences all obligors. The input is a value between 0 and 1. The intertemporal correlation only needs to be specified for valuation periods longer than 1.
Risk Horizon:	The risk horizon indicates the amout of years in the valuation period. Intuitively the minim input is 1. The maximum valuation period for which the simulation converges is 10 years. Note that the computational time is increased for longer valuation periods.
Monte Carlo Simulations:	The monte carlo simulations indicate the amount of monte carlo simulations used to simulate the dependence between obligors. The recommended input is 1000 monte carlo simulations. Larger values would increase the accuracy of the simulation, but this drastically increases computational time.
Credit Rating Breakdown:	A credit rating breakdown can be provided for the obligors to obtain PD estimates on rating grade level. The highest rating grade is specified at 1, then the second highest at 2, etc. The amount of obligors in the each rating grade along with the amount of defaults in the rating grade is then specified.

Description

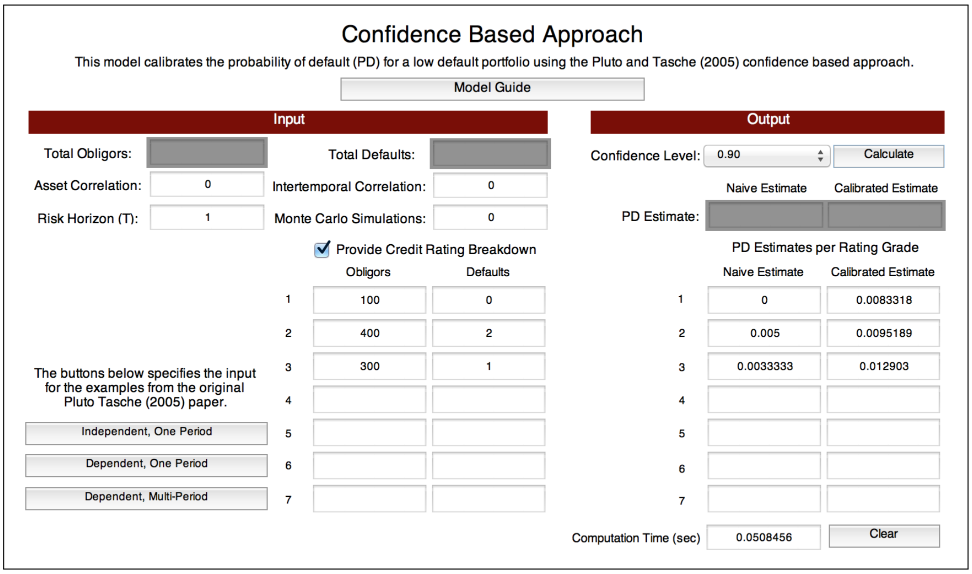
The Confidence Based approach introduced by Pluto and Tasche (2005) was the first major attempt probability of default calibration for low default portfolios. The model uses a most prudent estimation approach to calibrate the probability of default estimate. When considering a portfolio of credits with different rating grades, select the option provide credit rating breakdown. Alternatively for a portfolio of a single rating grade the portfolio details can simply be entered. The minimum set of inputs to run a simulation is Total Obligors and Total Defaults. When an asset correlation is specified, then the amount of monte carlo simulations requires specification. Furthermore, when a risk horizon longer than one year is considered then the intertemporal correlation needs to be specified.

Figure 6.3: The Confidence Based Approach Model Guide

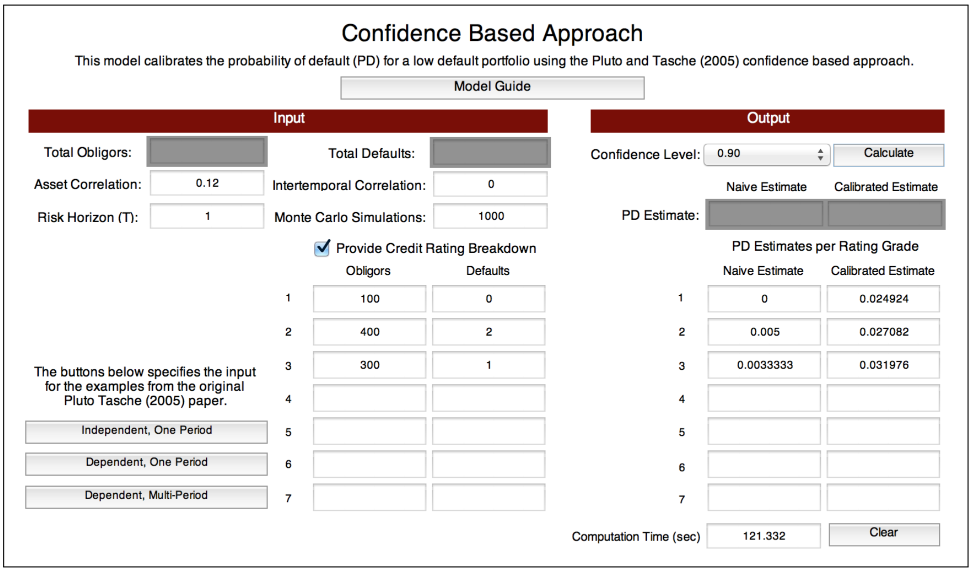
data from the corporate portfolio considered in this study is generated along with an asset correlation of 12%, an intertemporal correlation of 30%, a burn in of 100 and 1000 Monte Carlo Simulations. Lower simulation runs are considered in the example due to the simulation being extremely time consuming. After the example input has been generated the user needs to select the prior distribution from a drop down list, this is illustrated in Figure 6.8. An example of the output generated from selecting a $Uniform(0,1)$ prior, a Jeffreys prior and a $Pareto(1)$ prior is displayed in Figure 6.9. Note that both the iteration plot and histogram of the simulation is displayed in the interface. The interface has the same clearing function and warning system as the interface discussed previously.

6.4 Conclusion

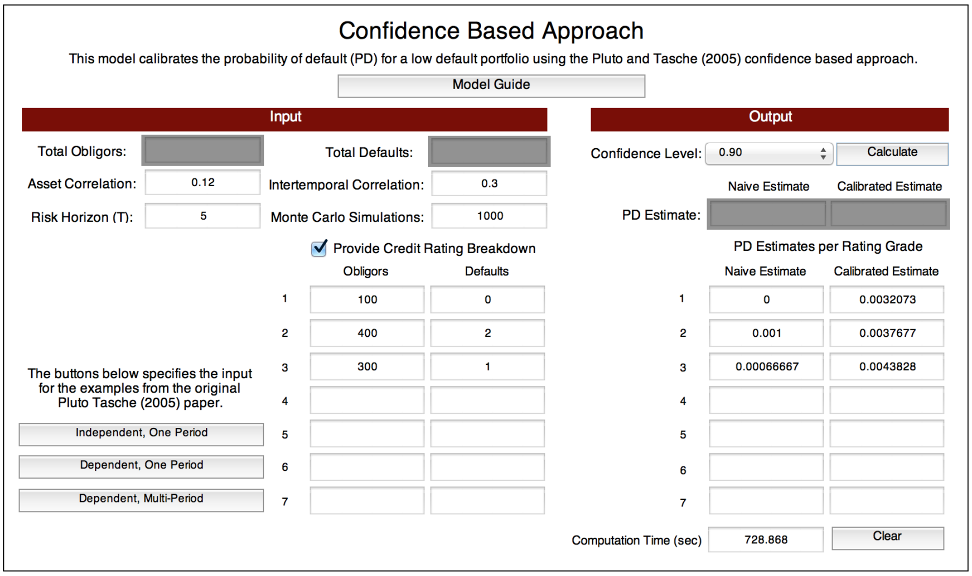
In this brief additional chapter the MATLAB GUI that was designed for the models discussed in Chapters 3 and 4 was presented. As illustrated in this chapter a well designed GUI can improve the usability of the models, especially for non technical risk practitioners.



(a) An example of the independent one-period case.

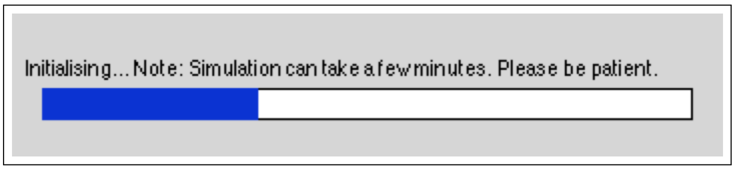


(b) An example of the dependent one-period case.

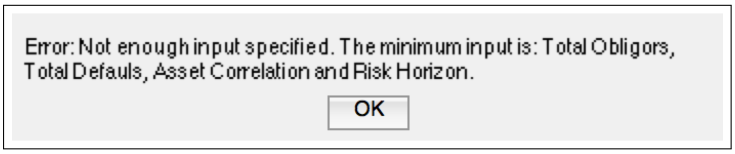


(c) An example of the dependent multi-period case.

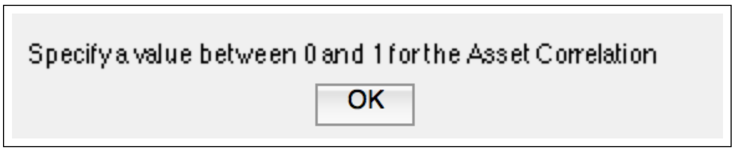
Figure 6.4: Example of the Confidence Based Interface Functionality.



(a) Simulation Time Warning.



(b) Warning that not enough input has been specified.



(c) Warning that the Asset Correlation lies outside the bounds of (0,1).

(d) Warning that the Intertemporal Correlation lies outside the bounds of (0,1).

Figure 6.5: Warnings in the Confidence Based Interface Interface.

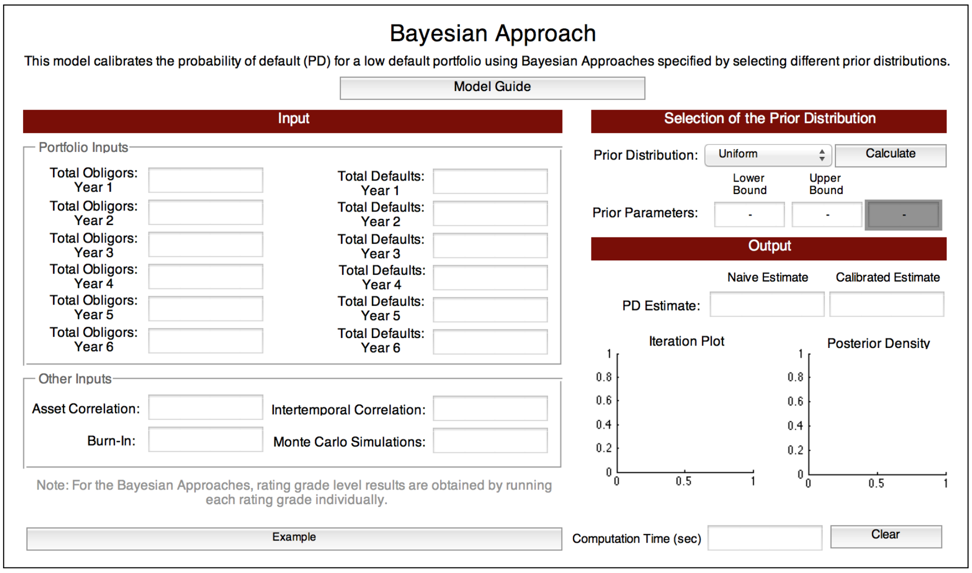


Figure 6.6: The Bayesian Approach Interface

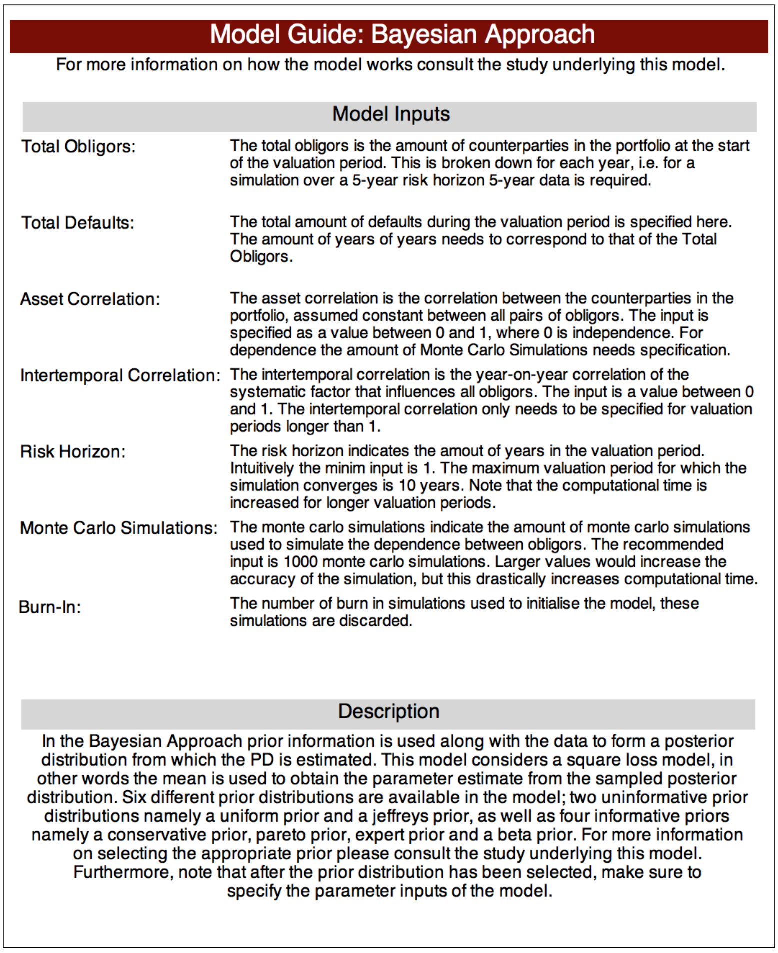


Figure 6.7: The Bayesian Approach Model Guide

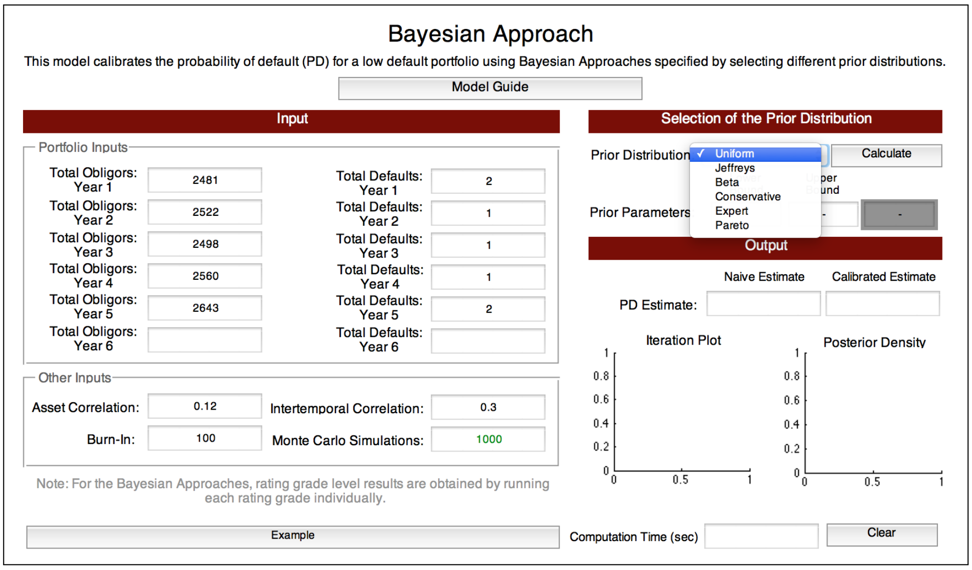
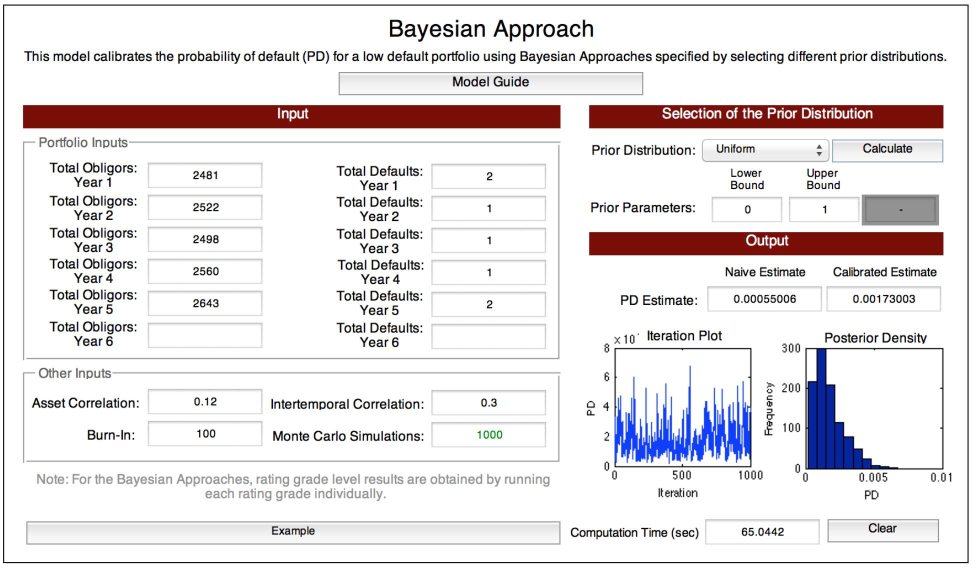
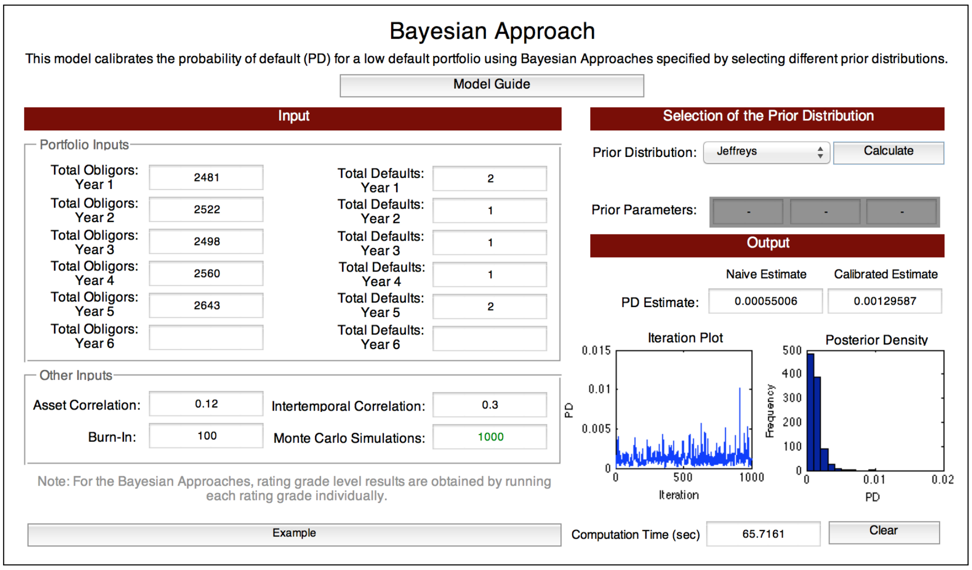


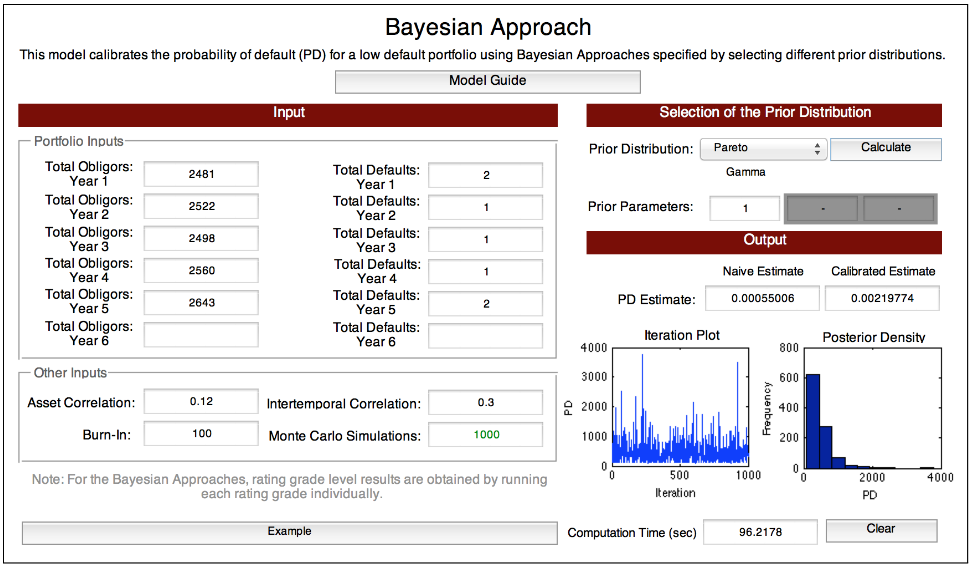
Figure 6.8: Illustration of Example Output and Prior Selection



(a) Example Output of a Simulation with the $Uniform(0, 1)$ Prior.



(b) Example Output of a Simulation with the Jeffreys Prior.



(c) Example Output of a Simulation with the $Pareto(1)$ Prior

Figure 6.9: Example of the Bayesian Interface Functionality.

Chapter 7

Conclusion

As stated in Chapter 1, one of the objectives in this study is to discuss in detail the theoretical background of PD estimation and highlight its importance in credit risk management. This is done in the first part of the literature study, in Chapter 2, where the focus of discussion is credit risk management. At the outset of Chapter 2 the importance of credit risk management within financial risk management is highlighted. The fundamental theory underlying credit risk management is then discussed in detail, the areas under consideration are; credit rating, the expected loss function, credit risk regulation, and the Probability of Default. The aim of this chapter is to highlight the significance of the PD estimate within credit risk management and discuss the fundamental principles underlying the estimation of this parameter. Elements such as the 1974 Merton Model, Default Correlation, and the Single-Factor Gaussian Copula are covered in detail as the understanding of these tools are required in order to better understand the PD calibration techniques considered in this study.

Other main objectives of this study are to introduce the problem of PD calibration for LDPs and to provide detailed discussions of the two main calibration approaches. Although the problem of LDPs is already touched upon in Chapter 2, it is only addressed in detail in Chapter 3, which marks the second component of the literature study. In Chapter 3 a formal definition of LDPs is provided and the industry concerns regarding these portfolios is discussed. The two main PD calibration approaches, namely the Confidence Based Approach and the Bayesian Approach, are discussed in detail, providing a detailed theoretical background of PD calibration for LDPs. In the discussion on the Bayesian Approach, all of the different prior distributions that have been proposed for the LDP problem are discussed. These are; the Uniform distribution, the Beta distribution, the Conservative distribution, and the Expert distribution.

One of the main objectives of this study is to propose additional distributions that can be considered as prior distributions for the Bayesian approach. It is observed in the theoretical discussion underlying Chapter 3 that all of the prior distributions that are considered for the PD are constrained on 0 and 1. This makes intuitive sense due to the definition of the PD, however, it is shown in Chapter 4 that through considering a small transformation of the PD parameter, distributions with support on 1 to infinity can now be considered. The distribution that is proposed as a possible prior distribution is the strict Pareto distribution. Due to the heavy-tailed nature of the distribution it is expected that by applying it to the LDP case the low default nature of the PD will be more accurately captured.

An additional well known uninformative prior distribution known as the Jeffreys prior is also discussed as an alternative option for use in the Bayesian Approach at the outset of Chapter 5. The main focus in Chapter 5 is, however, to address the evaluation and comparison of the PD calibration techniques discussed in this study. Before discussing the empirical results, the simulation and sampling algorithms used to produce the Bayesian posterior distributions are discussed. In this study a Gibbs sampling procedure with a simple rejection step is used and these are discussed in the first part of Chapter 5. In the second part the empirical results are discussed. Both real and fictitious data are considered for comparing the performance of the PD calibration models. The datasets considered are; a corporate portfolio derived from Moody's data, and a retail portfolio provided by a large South African bank. The models' PD estimates are compared at different correlation inputs, for the models falling under the Bayesian approach the highest posterior densities are considered, and then finally the respective models sensitivities to the correlation inputs are compared.

The significance of the PD estimate is also illustrated in Chapter 5 by providing a practical example of the parameter's use in the calculation of risk weighted asset capital requirements under Basel II. It is shown throughout Chapter 5 that all of the calibration techniques are successful in calibrating the PD estimate for LDP. There are, however, various degrees of conservatism applied by the different models. The levels of subjectivity present in some of the models is also evident in the empirical results. Furthermore, it is clear that the different models react with various degrees of sensitivity to the correlation parameter inputs. A detailed discussion of the respective PD calibration models discussed in this study is provided at the end of Chapter 5.

As an additional output of the study the user interface designed to produce the output in Chapter 5 is discussed in Chapter 6. This Chapter is not a direct benefit of the study, but rather aims to highlight the benefits of, and/or to make a case of good practice for the use of graphical user interfaces when designing complex financial models in any programming language. It is illustrated in the chapter that this makes the models more accessible to less technical individuals.

Finally, in this study the importance of the PD as a parameter in the modelling of credit risk is highlighted and the problem of estimating this parameter in the absence of sufficient default observations is addressed. It is shown that the problem introduced by LDPs in the estimation of the PD is real and that this can be overcome by using some form of a calibration technique. The PD calibration techniques that has been proposed over recent years are discussed in detail, these methodologies are grouped into two main approaches, namely the Confidence Based Approach and the Bayesian Approach. In this study the Bayesian Approach is reconsidered and two alternative prior distributions are successfully proposed as alternatives to the distributions already used in practice.

There are however areas of further research that are required in order to improve the quantification of the risk posed by LDPs. To this date, backtesting of the PD estimates of LDPs remain a significant issue and until this problem is solved the reliability of the parameters will always come into question. Furthermore, through using the transformation proposed in this study an array of new distributions can be considered as possible prior distributions in the Bayesian approach.

Appendices

Appendix A

Mathematical Derivations

This appendix contains all the mathematical derivations not given in the main body of the thesis.

A.1 Derivation of 3.3.24

In the proof of proposition 3.7 the following result was used;

$$\frac{df_{n,k}(p)}{dp} = -(n-k)\binom{n}{k}p^k(1-p)^{n-k-1}$$

where $f_{n,k}(p) = \sum_{i=0}^{k}\binom{n}{i}p^i(1-p)^{n-i}$, as defined in the proposition. The result follows by;

$$\begin{aligned}
\frac{df_{n,k}(p)}{dp} &= \frac{d}{dp}\sum_{i=0}^{k}\binom{n}{i}p^i(1-p)^{n-i} \\
&= \sum_{i=0}^{k}\binom{n}{i}(ip^{i-1}(1-p)^{n-i} + p^i(n-i)(1-p)^{n-i-1}(-1)) \\
&= \sum_{i=0}^{k}\binom{n}{i}(ip^{i-1}(1-p)^{n-i} - p^i(n-i)(1-p)^{n-i-1}) \\
&= \Bigg[\binom{n}{0}(-p^0(n)(1-p)^{n-1}) \\
&+ \binom{n}{1}(p^0(1-p)^{n-1} - p^1(n-1)(1-p)^{n-2}) \\
&+ \binom{n}{2}(2p^1(1-p)^{n-2} - p^2(n-2)(1-p)^{n-3}) \\
&+ \dots \\
&+ \binom{n}{k}(kp^{k-1}(1-p)^{n-k} - p^k(n-k)(1-p)^{n-k-1})\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&= \Big[(1)(-(n)(1-p)^{n-1}) \\
&+ (n)((1-p)^{n-1} - p^1(n-1)(1-p)^{n-2}) \\
&+ (\frac{1}{2}n(n-1))(2p^1(1-p)^{n-2} - p^2(n-2)(1-p)^{n-3}) \\
&+ \ldots \\
&+ \binom{n}{k}(kp^{k-1}(1-p)^{n-k} - p^k(n-k)(1-p)^{n-k-1})\Big] \\
&= \Big[((n)(1-p)^{n-1}) \\
&+ (n(1-p)^{n-1} - np^1(n-1)(1-p)^{n-2}) \\
&+ (n(n-1)p^1(1-p)^{n-2} - \frac{1}{2}n(n-1)p^2(n-2)(1-p)^{n-3}) \\
&+ \ldots \\
&+ \binom{n}{k}p^{k-1}(1-p)^{n-k} - \binom{n}{k}p^k(n-k)(1-p)^{n-k-1}\Big] \\
&= -(n-k)\binom{n}{k}p^k(1-p)^{n-k-1} *
\end{aligned}
$$

A.2 Derivations of the Log-transformed functions used in the Simulation Process

Log-Transform of the Marginal Conditional Posterior for the Uniform Prior Distribution:

For:

$$p \sim Uniform(p_l, p_u)$$

it is known that:

$$g_P(p) = \begin{cases} \frac{1}{p_u - p_l} & \text{for } p_l \leq p < p_u \\ 0 & \text{for } p < p_l \text{ or } p \geq p_u. \end{cases}$$

Therefore the marginal conditional posterior density for the prior distribution in question follows as:

$$
\begin{aligned}
h(p|\mathbf{L}, \xi, \mathbf{x}) &\propto f(\mathbf{L}|p, \xi, \mathbf{x}) \times g_P(p) \\
&\propto \left\{ \prod_{t=1}^{T}\prod_{i=1}^{s_t} \lambda_t(p, \xi, x_t)^{l_{t,i}}(1 - \lambda_t(p, \xi, x_t))^{1-l_{t,i}} \right\} \\
&\times \frac{1}{p_u - p_l},
\end{aligned}
$$

for $p_l \leq p < p_u$, where $h(p|\mathbf{L}, \xi, \mathbf{x}, \tau) = 0$ otherwise. Let:

$$v = \log p \Rightarrow p = e^v,$$

Then:

$$f(\mathbf{L}|v, \xi, x) = f(\mathbf{L}|e^v, \xi, x)|\frac{dp}{dv}|,$$

and:

$$|\frac{dp}{dv}| = e^v,$$

from which it follows that:

$$\begin{aligned} f(\mathbf{L}|v,\xi,x) &= f(\mathbf{L}|e^v,\xi,x)e^v \\ &= e^v \times \left\{ \prod_{t=1}^{T}\prod_{i=1}^{s_t} \lambda_t(e^v,\xi,x_t)^{l_{t,i}}(1-\lambda_t(e^v,\xi,x_t))^{1-l_{t,i}} \right\} \\ &\times \frac{1}{p_u - p_l}. \end{aligned}$$

Taking the log yields:

$$\begin{aligned} \log f(\mathbf{L}|v,\xi,x) = v &+ \left\{ \sum_{t=1}^{T}\sum_{i=1}^{s_t} l_{t,i}\log\lambda_t(e^v,\xi,x_t) + \sum_{t=1}^{T}\sum_{i=1}^{n_t-s_t}(1-l_{t,i})\log\lambda_t(e^v,\xi,x_t)) \right\} \\ &+ \log(1) - \log(p_u - p_l) \\ &= v + \left\{ \sum_{t=1}^{T} s_t l_t \log\lambda_t(e^v,\xi,x_t) + \sum_{t=1}^{T}(n_t - s_t)(1-l_t)\log\lambda_t(e^v,\xi,x_t)) \right\} \\ &- \log(p_u - p_l), \end{aligned}$$

for $p_l \leq p < p_u$, where $\log h(v) = 0$ otherwise. This is the function used in the simulation process.

Log-Transform of the Marginal Conditional Posterior for the Beta Prior Distribution:

The marginal conditional posterior density function of:

$$p \sim Beta(\alpha_p, \beta_p)$$

where:

$$g_P(p) = p^{\alpha_p - 1}(1-p)^{\beta_p - 1},$$

is given by:

$$\begin{aligned} h(p|\mathbf{L},\xi,\mathbf{x}) &\propto f(\mathbf{L}|p,\xi,\mathbf{x}) \times g_P(p) \\ &\propto \left\{ \prod_{t=1}^{T}\prod_{i=1}^{s_t} \lambda_t(p,\xi,x_t)^{l_{t,i}}(1-\lambda_t(p,\xi,x_t))^{1-l_{t,i}} \right\} \\ &\times p^{\alpha_p - 1}(1-p)^{\beta_p - 1}. \end{aligned}$$

Let:

$$v = \log p \Rightarrow p = e^v,$$

Then:

$$f(\mathbf{L}|v,\xi,x) = f(\mathbf{L}|e^v,\xi,x)|\frac{dp}{dv}|,$$

and:

$$|\frac{dp}{dv}| = e^v,$$

from which it follows that:

$$\begin{aligned} f(\mathbf{L}|v,\xi,x) &= f(\mathbf{L}|e^v,\xi,x)e^v \\ &= e^v \times \left\{ \prod_{t=1}^{T}\prod_{i=1}^{s_t} \lambda_t(e^v,\xi,x_t)^{l_{t,i}}(1-\lambda_t(e^v,\xi,x_t))^{1-l_{t,i}} \right\} \\ &\times e^{v(\alpha-1)}(1-e^v)^{\beta-1}. \end{aligned}$$

Taking the log yields:

$$\begin{aligned} \log f(\mathbf{L}|v,\xi,x) = v &+ \left\{ \sum_{t=1}^{T}\sum_{i=1}^{s_t} l_{t,i}\log\lambda_t(e^v,\xi,x_t) + \sum_{t=1}^{T}\sum_{i=1}^{n_t-s_t}(1-l_{t,i})\log\lambda_t(e^v,\xi,x_t)) \right\} \\ &+ v(\alpha-1) + (\beta-1)\log(1-e^v). \\ = v &+ \left\{ \sum_{t=1}^{T} s_t l_t \log\lambda_t(e^v,\xi,x_t) + \sum_{t=1}^{T}(n_t-s_t)(1-l_t)\log\lambda_t(e^v,\xi,x_t)) \right\} \\ &+ v(\alpha-1) + (\beta-1)\log(1-e^v). \end{aligned}$$

This is the function used in the simulation process. Note that the function above can also be used for the Jeffreys prior as this prior is simply a $Beta(0.5, 0.5)$ distribution for the current scenario.

Log-Transform of the Marginal Conditional Posterior for the Conservative Prior Distribution:

The marginal conditional posterior density function of the conservative prior, where:

$$g_P(p) = \frac{1}{1-p}.$$

is given by:

$$\begin{aligned} h(p|\mathbf{L},\xi,\mathbf{x}) &\propto f(\mathbf{L}|p,\xi,\mathbf{x}) \times g_P(p) \\ &\propto \left\{ \prod_{t=1}^{T}\prod_{i=1}^{s_t} \lambda_t(p,\xi,x_t)^{l_{t,i}}(1-\lambda_t(p,\xi,x_t))^{1-l_{t,i}} \right\} \\ &\times \frac{1}{1-p}. \end{aligned}$$

Let:

$$v = \log p \Rightarrow p = e^v,$$

Then:

$$f(\mathbf{L}|v,\xi,x) = f(\mathbf{L}|e^v,\xi,x)|\frac{dp}{dv}|,$$

and:

$$|\frac{dp}{dv}| = e^v,$$

from which it follows that:

$$\begin{aligned} f(\mathbf{L}|v,\xi,x) &= f(\mathbf{L}|e^v,\xi,x)e^v \\ &= e^v \times \left\{ \prod_{t=1}^{T}\prod_{i=1}^{s_t} \lambda_t(e^v,\xi,x_t)^{l_{t,i}}(1-\lambda_t(e^v,\xi,x_t))^{1-l_{t,i}} \right\} \\ &\times \frac{1}{1-e^v}.. \end{aligned}$$

Taking the log yields:

$$\begin{aligned}\log f(\mathbf{L}|v,\xi,x) &= v + \left\{\sum_{t=1}^{T}\sum_{i=1}^{s_t} l_{t,i}\log\lambda_t(e^v,\xi,x_t) + \sum_{t=1}^{T}\sum_{i=1}^{n_t-s_t}(1-l_{t,i})\log\lambda_t(e^v,\xi,x_t))\right\}\\ &+ \log\left(\frac{1}{1-e^v}\right)\\ &= v + \left\{\sum_{t=1}^{T} s_t l_t \log\lambda_t(e^v,\xi,x_t) + \sum_{t=1}^{T}(n_t-s_t)(1-l_t)\log\lambda_t(e^v,\xi,x_t))\right\}\\ &- \log(1-e^v).\end{aligned}$$

This is the function used in the simulation process.

Log-Transform of the Marginal Conditional Posterior for the Expert Prior Distribution:

The marginal conditional posterior density function of the conservative prior, where:

$$g_P(p) = \begin{cases} 0 & \text{for } 0 \le p < p_l \\ \frac{2(p-p_l)}{(p_u-p_l)(p_m-p_l)} & \text{for } p_l \le p < p_m \\ \frac{2}{(p_u-p_l)} & \text{for } p = p_m \\ \frac{2(p_u-p)}{(p_u-p_l)(p_l-p_m)} & \text{for } p_m < p \le p_u \\ 0 & \text{for } p_u < p \le 1 \end{cases}.$$

is given by:

$$\begin{aligned}h(p|\mathbf{L},\xi,\mathbf{x}) &\propto f(\mathbf{L}|p,\xi,\mathbf{x}) \times g_P(p)\\ &\propto \left\{\prod_{t=1}^{T}\prod_{i=1}^{s_t}\theta_t(p,\xi,x_t)^{l_{t,i}}(1-\theta_t(p,\xi,x_t))^{1-l_{t,i}}\right\}\\ &\times \frac{2(p-p_l)}{(p_u-p_l)(p_m-p_l)},\end{aligned}$$

for $p_l \le p \le p_m$ and:

$$\begin{aligned}h(p|\mathbf{L},\xi,\mathbf{x}) &\propto f(\mathbf{L}|p,\xi,\mathbf{x}) \times g_P(p)\\ &\propto \left\{\prod_{t=1}^{T}\prod_{i=1}^{s_t}\theta_t(p,\xi,x_t)^{l_{t,i}}(1-\theta_t(p,\xi,x_t))^{1-l_{t,i}}\right\}\\ &\times \frac{2(p_u-p)}{(p_u-p_l)(p_l-p_m)},\end{aligned}$$

for $p_m < p \le p_u$, where $h(p|\mathbf{L},\mathbf{x}) = 0$ otherwise.

Let:

$$v = \log p \Rightarrow p = e^v,$$

Then:

$$f(\mathbf{L}|v,\xi,x) = f(\mathbf{L}|e^v,\xi,x)|\frac{dp}{dv}|,$$

and:

$$|\frac{dp}{dv}| = e^v,$$

from which it follows for $\log p_l \leq v \leq \log p_m$ that:

$$\begin{aligned} f(\mathbf{L}|v,\xi,x) &= f(\mathbf{L}|e^v,\xi,x)e^v \\ &= e^v \times \left\{ \prod_{t=1}^{T} \prod_{i=1}^{s_t} \lambda_t(e^v,\xi,x_t)^{l_{t,i}} (1-\lambda_t(e^v,\xi,x_t))^{1-l_{t,i}} \right\} \\ &\times \frac{2(e^v - p_l)}{(\log p_u - p_l)(p_m - p_l)}. \end{aligned}$$

Taking the log yields:

$$\begin{aligned} \log f(\mathbf{L}|v,\xi,x) = v &+ \left\{ \sum_{t=1}^{T} \sum_{i=1}^{s_t} l_{t,i} \log \lambda_t(e^v,\xi,x_t) + \sum_{t=1}^{T} \sum_{i=1}^{n_t - s_t} (1-l_{t,i}) \log \lambda_t(e^v,\xi,x_t)) \right\} \\ &+ \log \left(\frac{2(e^v - p_l)}{(p_u - p_l)(p_m - p_l)} \right). \\ &= v + \left\{ \sum_{t=1}^{T} s_t l_t \log \lambda_t(e^v,\xi,x_t) + \sum_{t=1}^{T} (n_t - s_t)(1-l_t) \log \lambda_t(e^v,\xi,x_t)) \right\} \\ &+ \log(2(e^v - p_l)) - [\log(p_u - p_l) + \log(p_l - p_m)]. \end{aligned}$$

Also, for $\log p_m \leq v \leq \log p_u$ it follows that:

$$\begin{aligned} h(v) &= h(e^v)e^v \\ &= e^v \times \left\{ \prod_{t=1}^{T} \prod_{i=1}^{s_t} \lambda_t(e^v,\xi,x_t)^{l_{t,i}} (1-\lambda_t(e^v,\xi,x_t))^{1-l_{t,i}} \right\} \\ &\times \frac{2(p_u - e^v)}{(p_u - p_l)(p_l - p_m)} \end{aligned}$$

Taking the log yields:

$$\begin{aligned} \log h(v) = v &+ \left\{ \sum_{t=1}^{T} \sum_{i=1}^{s_t} l_{t,i} \log \lambda_t(e^v,\xi,x_t) + \sum_{t=1}^{T} \sum_{i=1}^{n_t - s_t} (1-l_{t,i}) \log \lambda_t(e^v,\xi,x_t)) \right\} \\ &+ \log \left(\frac{2(p_u - e^v)}{(p_u - p_l)(p_m - p_l)} \right). \\ &= v + \left\{ \sum_{t=1}^{T} s_t l_t \log \lambda_t(e^v,\xi,x_t) + \sum_{t=1}^{T} (n_t - s_t)(1-l_t) \log \lambda_t(e^v,\xi,x_t)) \right\} \\ &+ \log(2(p_u - e^v)) - [\log(p_u - p_l) + \log(p_l - p_m)]. \end{aligned}$$

The resulting functions above is used for simulation.

Appendix B

Additional Background Theory

This appendix contains all the definitions not given in the main body of the thesis.

B.1 Itô's Lemma

Definition B.1. Itô's Lemma From Hull (2012), suppose that the value of a variable y follows the Itô process:

$$dy = \alpha(y,t)dt + \beta(y,t)dz, \tag{B.1.1}$$

where α and β are functions of y and dz is a Wiener process. The variable y possesses a drift rate of α^2 and a variance rate of β^2. It then follows from Itô's Lemma that a function F of y and t follows the process:

$$dG = \left(\frac{\partial F}{\partial y}\alpha + \frac{\partial F}{\partial t} + \frac{\partial^2 F}{\partial y^2}\beta\right)dt + \frac{\partial F}{\partial y}\beta dz \tag{B.1.2}$$

where dz is the same Wiener process as in equation B.1.1. It follows that the function F follows an Itô process with a drif rate of:

$$\left(\frac{\partial F}{\partial y}\alpha + \frac{\partial F}{\partial t} + \frac{\partial^2 F}{\partial y^2}\beta\right)$$

and a variance rate of:

$$\left(\frac{\partial F}{\partial y}\right)^2 \beta^2.$$

B.2 Markov Chain Monte Carlo Method

As discussed in Rizzo (2007), many of the applications of MCMC methods are in problems arising from Bayesian inference (such is the case in Section 3.3.2). Huynh *et al.* (2011), defines a Markov Process as follows:

Definition B.2. Markov Process: Let $\{Z_i, i = 0, 1, 2, \dots\}$ be a given random process (this can be simulated). Suppose that the Z_i take on values in the state space, M. The process $\{Z_i, i = 0, 1, 2, \dots\}$ is then said to be a Markov Chain if the probability:

$$P(Z_{i+1} = z_{i+1} | Z_i = z_i, Z_{i-1} = z_{i-1}, \dots, Z_0 = z_0)$$

is equal to:

$$P(Z_{i+1} = z_{i+1}|Z_i = z_i).$$

The definition above is interpreted as follows, the probability that the process is in state z_{i+1} at $i+1$, given that it is in state z_i at i, is independent of everything that happened before i. The only points that are of importance are i and $i+1$.

It is also mentioned in Huynh *et al.* (2011) that a Markov Chain can be called homogeneous if:

$$P(Z_{i+1} = y|Z_i = x) = p_{x,y}.$$

This means that the probability of being in a state y given that the process was in a state x, is independent of i.

Generalise the Bayesian setting as follows, consider:

$$\mathbb{E}(h(\theta|z)) = \int h(\theta) f_{\theta|z}(\theta) d\theta = \frac{\int h(\theta) f_{z|\theta}(z) f_\theta(\theta) d\theta}{\int f_{z|\theta}(z) f_\theta(\theta) d\theta}$$

and write the above in more general terms;

$$\mathbb{E}(h(Z)) = \frac{\int h(t)\pi(t)dt}{\int \pi(t)dt}$$

where π is a density or likelihood.

Rizzo (2007) explains the MCMC method as follows; the Monte Carlo estimate of $\mathbb{E}(h(\theta)) = \int h(\theta) f_{\theta|z}(\theta) d\theta$ is the sample mean:

$$\bar{h} = \approx \frac{1}{M} \sum_{i=1}^{M} h(Z_i)$$

where the $z_1, \ldots, z_M$ is a sample from the distribution with density $f_{\theta|z}(\theta)$. If this is a random sample and the $z_1, \ldots, z_M$ are independent; the sample mean, $\bar{h}$, converges to $\mathbb{E}(h(\theta))$ by the law of large numbers for large M.

In the Bayesian framework it is difficult to obtain independent samples from a distribution with density $f_{\theta|z}(\theta)$. However Monte Carlo integration can be applied to a sample of dependent observations if their joint density roughly resembles the joint density of a random sample. As eloquently stated by Rizzo (2007), the MCMC method estimates the integral $\mathbb{E}(h(\theta)) = \int h(\theta) f_{\theta|z}(\theta) d\theta$ by Monte Carlo integration and then the Markov Chain provides the sampler that generates the random observations from the prior distribution.

It follows in general that, if $\{Z_i, i = 0, 1, 2, \ldots\}$ is the realization of a Markov Chain with a stationary distribution π. Then:

$$\overline{h(Z)}_M = \frac{1}{M} \sum_{t=0}^{M} h(Z_t) \tag{B.2.1}$$

converges to $\mathbb{E}(Z)$ as $M \to \infty$ with probability one, where Z has the stationary distribution π. Also, the expectation is taken with respect to π.

As discussed in Huynh *et al.* (2011), the most well known method for generating a Markov chain with π identical to $f_{\theta|z}(\theta)$ is the Metropolis-Hastings algorithm. The algorithm considered and applied in Chapter 5 is the Gibbs sampler, which is a special case of the Metropolis-Hastings algorithm.

From Rizzo (2007), the Gibbs sampler is often applied when the target distribution is a multivariate distribution. Assume that all the univariate conditional densities are fully specified, and that sampling from these distributions are relatively simple. The chain is then generated by sampling from the marginal distributions of the target distribution and every candidate point can therefore be accepted.

Let $Z = (Z_1, \ldots, Z_d)$ be a random vector in $\mathbb{R}^d$, and define the $d-1$ dimensional random vectors:

$$Z_{(-j)} = (Z_1, \ldots, Z_{j-1}, Z_{j+1}, \ldots, Z_d).$$

The univariate conditional density of Z_j given $Z_{(-j)}$ by $f(Z_j|Z_{(-j)})$. The Gibbs sampler then generates the chain by sampling from each of the d conditional densities $f(Z_j|Z_{(-j)})$.

It now follows from Rizzo (2007), that the algorithm for the Gibbs sampler can be defined as follows; let Z_t be denoted by $Z(t)$:

1. Initialise $Z(0)$ at time $t = 0$.
2. For each iteration, indexed $t = 1, 2, \ldots$ repeat:
   a) Set $z_1 = Z_1(t-1)$.
   b) For each coordinate $j = 1, 2, \ldots, d$:
      i. Generate $Z_j^*(t)$ from $f(Z_j|z_{(-j)})$.
      ii. Update $z_j = Z_j^*(t)$
   c) Set $Z(t) = (Z_1^*(t), Z_2^*(t), \ldots, Z_d^*(t))$.
   d) Increment t.

Appendix C

Program Code

This appendix lists all the program code used in the study. The programming language used is MATLAB.

The first block of code given is a simple user-defined function for the cumulative distribution function of the binomial distribution. This is one of the fundamental blocks of code in the thesis and this function is called in nearly all the applied methodologies.

Listing C.1: User-Defined Cumulative Binomial Function

```
function cumbin = cumbinomial(s,d,p)
% This function is an user-defined function for the cumulative
% distribution function of the binomial distribution function.
% The function corresponds to the right-hand-side of expression
% 2.4.16 in the literature study.
%
% Parameters:
%   s: amount of obligors in the specific rating grade.
%   d: amount of defaults in the specific rating grade.
%   p: probability of default.

warning('off', 'MATLAB:nchoosek:LargeCoefficient');

% Initialize output vector.
cumbin = 0;

for k = 0:d

    %RHS 2.4.1:
    cumbin =  cumbin + nchoosek(s,k)*(p^k)*((1-p).^(s-k));

end

end
```

In the following section of this appendix all of the code of the Confidence Based Approach is is presented.

C.1 Code for Confidence Based Approach

In this Section all the code used to apply the Pluto and Tasche (2011) Confidence Based PD calibration method for LDPs is listed. The first block of code is the simple case from Section 3.3.1, this is the case where negligible defaults have been observed under the assumption of independence.

Listing C.2: Confidence Based PD Calibration Assuming Independence

```
function PD = PlutoTascheIndependent(s,d,alpha)
%This function calculates the PD for a portfolio of obligors
%using Pluto and Tasche's Confidence Based Approach.
%   s       = Vector of Obligors by rating category, ordered
%             from highest to lowest rating category i.e. for
%             a credit portfolio where the obligors have been
%             placed into three rating grades 1, 2, and 3, the
%             input vector is [#1, #2, #3] where #1 is the amount
%             of obligors rated 1.
%   d       = Vector of Defaults by rating category corresponding
%             to the above.
%   alpha   = 100(1-alpha)% confidence region for the PD estimate.

%Initialize
NumberGrades = length(s);
PD = zeros(1, NumberGrades);

%"Reversing" the input vectors in order to apply the most prudent
%estimation principle.
Reverse_s = fliplr(s);
Reverse_d = fliplr(d);

%Determining s* and d* given by expressions 2.4.14 and 2.4.15.
%These are equivalent to cumulative sums of the vectors s and d.
s_star = fliplr(cumsum(Reverse_s));
d_star = fliplr(cumsum(Reverse_d));

for i = 1:NumberGrades

    b = @(x)cumbinomial(s_star(i),d_star(i),x)-(1-alpha);

    PD(i) = fzero(b, [0,1]);

end

end
```

Correlation is introduced into the model with the following block of code which incorporates the Sigle-Factor Gaussian copula. This piece of code is then called in the block thereafter where a Monte Carlo Simulation is run on the different states of the systematic factor variable.

Listing C.3: The Single-Factor Gaussian Copula

```
function cumbin = SimulationOnePeriod(PD, s, d, xi)
% This function is an user-defined function for the cumulative
%The Single-Factor Gaussian Copula used to introduce dependence.
%   PD   = probability of default.
%   s    = amount of obligors in the specific rating grade.
%   d    = amount of defaults in the specific rating grade.
```

```
%   xi   = Asset correlation between obligors, this is assumed
%          uniform between all pairs of obligors.

% Generate one sample of the systematic factor.
X = 0;
X = random('norm', 0,1);

% The probability of default as given by the Single-Factor
% Gaussian Copula.
PD_X = normcdf((norminv(PD) - X.*sqrt(xi))./sqrt(1-xi));

% The cumulative binomial distribution function given one
% simulation of X.
cumbin = cumbinomial(s,d,PD_X);

end
```

Listing C.4: Monte Carlo Simulation for the One-Period Case

```
function error = MonteCarloOnePeriod(PD, s, d, xi, conf, M)
% This function is an user-defined function for the cumulative
%Monte Carlo simulation of the systematic factor underlying the
%Gaussian Copula used to capture the dependence structure between
%obligors.
%   PD       = probability of default.
%   s        = Vector of Obligors by rating category, ordered from
%              highest to lowest rating category i.e. for a credit
%              portfolio where the obligors have been placed into
%              three rating grades 1, 2, and 3, the input vector is
%              [#1, #2, #3] where #1 is the amount of obligors rated 1.
%   d        = Vector of Defaults by rating category corresponding
%              to the above.
%   xi       = Asset correlation between obligors, this is assumed
%              uniform between all pairs of obligors.
%   conf     = 100(1-alpha)% confidence region for the PD estimate.
%   M        = Specifies the number of Monte Carlo simulations used
%              in simulating the systematic factor.

% Initialize Monte Carlo Simulation
simul(M,1) = 0;

parfor t = 1:M
    simul(t,1) = SimulationOnePeriod(PD, s, d, xi);
end

AverageSim = mean(simul);

error = abs(AverageSim-(1-conf));

end
```

The PD estimate is then obtained by using the two aforementioned functions in the following block of code.

Listing C.5: One-Period Confidence Based Approach Assuming Dependence

```
function PD = PlutoTascheDependentOnePeriod(s, d, xi, conf, M)
% This function is an user-defined function for the cumulative
```

```
%This function calculates the PD for a dependent portfolio of
%obligors using Pluto and Tasche's Confidence Based Approach
%assuming one period.
%   s       = Vector of Obligors by rating category, ordered from
%             highest to lowest rating category i.e. for a credit
%             portfolio where the obligors have been placed into
%             three rating grades 1, 2, and 3, the input vector
%             is [#1, #2, #3] where #1 is the amount of obligors
%             rated 1.
%   d       = Vector of Defaults by rating category corresponding
%             to the above.
%   xi      = Asset correlation between obligors, this is assumed
%             uniform between all pairs of obligors.
%   conf   = 100(1-alpha)% confidence region for the PD estimate.
%   M       = Specifies the number of Monte Carlo simulations used
%             in simulating the systematic factor.

%Initialize
NumberGrades = length(s);
PD = zeros(1, NumberGrades);

%"Reversing" the input vectors in order to apply the most prudent
%estimation principle.
Reverse_s = fliplr(s);
Reverse_d = fliplr(d);

%Determining s* and d* given by expressions 2.4.14 and 2.4.15.
% These are equivalent to cumulative sums of the vectors s and d.
s_star = fliplr(cumsum(Reverse_s));
d_star = fliplr(cumsum(Reverse_d));

h = waitbar(0,'Initialising... Note: Simulation can take a few minutes. ...
    Please be patient.');
hw=findobj(h,'Type','Patch');
set(hw,'EdgeColor',[0 0 0.8],'FaceColor',[0 0 0.8])

for i = 1:NumberGrades

    f = @(PD)MonteCarloOnePeriod(PD, s_star(i), d_star(i), xi, conf, M);

    PD(i) = fminbnd(f,-0.01,0.05,optimset('TolX',1e-6,'Display', 'off', ...
        'MaxIter', 100));

    waitbar(i/NumberGrades);

end

close(h);

end
```

Then finally extending this into a multi period case, the following three blocks of code provide Multi-Period extentions of the code above.

Listing C.6: Dynamic Factor Model

```
function cumbin = SimulationMultiPeriod(PD, s, d, xi, v, T)
% This function is an user-defined function for the cumulative
%The Single-Factor Gaussian Copula used to introduce dependence.
```

```
%   PD   = probability of default.
%   s    = amount of obligors in the specific rating grade.
%   d    = amount of defaults in the specific rating grade.
%   xi   = Asset correlation between obligors, this is assumed
%          uniform between all pairs of obligors.
%   v    = Intertemporal correlation.
%   T    = Amount of observation periods.

% Generate one sample of the systematic factor.

X(T,1) = 0;
X(1,1) = random('norm', 0,1);

if T>1
   for j=2:T
       X(j,1) = v*X(j-1)+sqrt(1-v^2)*random('norm', 0,1);
   end
end

%The probability of default for obligor i given the realisation
%of X1,...,XT systematic factors, as given in expression 2.4.46.
PD_X = 1-prod(1 - normcdf((norminv(PD)-X.*sqrt(xi))./sqrt(1-xi)));

cumbin = cumbinomial(s,d,PD_X);

end
```

Listing C.7: Monte Carlo Simulation for the Multi-Period Case

```
function error = MonteCarloMultiPeriod(PD, s, d, xi, v, T, conf, M)
% This function is an user-defined function for the cumulative
%Monte Carlo simulation of the systematic factor underlying the
%Gaussian Copula used to capture the dependence structure between
%obligors.
%   PD       = probability of default.
%   s        = Vector of Obligors by rating category, ordered from
%              highest to lowest rating category i.e. for a credit
%              portfolio where the obligors have been placed into
%              three rating grades 1, 2, and 3, the input vector is
%              [#1, #2, #3] where #1 is the amount of obligors rated 1.
%   d        = Vector of Defaults by rating category corresponding to
%              the above.
%   xi       = Asset correlation between obligors, this is assumed uniform
%              between all pairs of obligors.
%   conf     = 100(1-alpha)% confidence region for the PD estimate.
%   M        = Specifies the number of Monte Carlo simulations used in
%              simulating the systematic factor.

% Initialize Monte Carlo Simulation
simul(M,1) = 0;

for t = 1:M
    simul(t,1) = SimulationMultiPeriod(PD, s, d, xi, v, T);
end

AverageSim = 1 - mean(simul);

error = abs(AverageSim-(conf));
```

```
end
```

Listing C.8: Multi-Period Confidence Based Approach Assuming Dependence

```
function PD = PlutoTascheDependentMultiPeriod(s, d, xi, v, T, conf, M)
%This function is an user-defined function for the cumulative
%This function calculates the PD for a dependent portfolio of
%obligors using Pluto and Tasche's Confidence Based Approach
%assuming one period.
%   s       = Vector of Obligors by rating category, ordered
%             from highest to lowest rating category i.e. for
%             a credit portfolio where the obligors have been
%             placed into three rating grades 1, 2, and 3, the
%             input vector is [#1, #2, #3] where #1 is the amount
%             of obligors rated 1.
%   d       = Vector of Defaults by rating category corresponding
%             to the above.
%   xi      = Asset correlation between obligors, this is assumed
%             uniform between all pairs of obligors.
%   v       =
%   T       =
%   conf    = 100(1-alpha)% confidence region for the PD estimate.
%   M       = Specifies the number of Monte Carlo simulations used
%             in simulating the systematic factor.

%Initialize
NumberGrades = length(s);
PD = zeros(1, NumberGrades);

%"Reversing" the input vectors in order to apply the most prudent
%estimation principle.
Reverse_s = fliplr(s);
Reverse_d = fliplr(d);

%Determining s* and d* given by expressions 2.4.14 and 2.4.15. These
%are equivalent to cumulative sums of the vectors s and d.
s_star = fliplr(cumsum(Reverse_s));
d_star = fliplr(cumsum(Reverse_d));

h = waitbar(0,'Initialising... Note: Simulation can take a few minutes. ...
    Please be patient.');
hw=findobj(h,'Type','Patch');
set(hw,'EdgeColor',[0 0 0.8],'FaceColor',[0 0 0.8])

for i = 1:NumberGrades

    f = @(PD)MonteCarloMultiPeriod(PD, s_star(i), d_star(i), xi, v, T, ...
        conf, M);

    PD(i) = fminbnd(f,-0.01,0.05,optimset('TolX',1e-6,'Display', 'off', ...
        'MaxIter', 100));

    waitbar(i/NumberGrades);

end

close(h);

end
```

C.2 Code for the Bayesian Approach

The code for the Bayesian Approach was developed under consultation of Dr. Tom Berning. The code structure of the Bayesian Approach is more complicated than the Confidence Based Approach due to the level of sophistication that is present in the the simulation algorithms and marginal conditional posterior functions.

Before proceeding to the complicated set of MATLAB codes that perform the simulations for the Bayesian Posterior distributions, the code that calculated the Half-Sample Mode and the Highest Posterior Density from the sampled posterior distribution is displayed. The code for these two functions are given in the following two code blocks.

Listing C.9: Half-Sample Mode Function

```
% System file.

function Result = SysMode(Data);

% Similar to HPD.
% Data must be sorted column vector.

n = length(Data);
nd = ceil(0.5*n);
Pos = 1;
BestLen = inf;
for Count = 1:(n - nd + 1)
   ThisLen = (Data(Count + nd - 1) - Data(Count));
   if  ThisLen < BestLen
      BestLen = ThisLen;
      Pos = Count;
   end
end
Low = Data(Pos);
Up = Data(Pos + nd - 1);
Result = Data(find((Data >= Low)&(Data <= Up)));
```

Listing C.10: Highest Posterior Density Function

```
%HPD   Calculates highest posterior density region for Bayes analysis.
%   [LOW,UP] = HPD(DRAWS,LEVEL) returns in LOW and UP values such that
%   at least LEVEL*100% of the values in DRAWS falls between LOW and UP.
%   LEVEL: 0 < LEVEL < 1.

function [Low,Up] = HPD(Draws,Level);

if (length(Level) ~= 1)|(Level <= 0)|(Level >= 1)
   error('LEVEL must be a scalar between 0 and 1!')
end
if length(Draws) ~= prod(size(Draws))
   error('DRAWS must be a vector!')
end
n = length(Draws);
if n < 100
   error('DRAWS must contain at least 100 values!')
end
Data = sort(Draws);
```

```
nd = ceil(Level*n);
Pos = 1;
BestLen = inf;
for Count = 1:(n - nd + 1)
   ThisLen = (Data(Count + nd - 1) - Data(Count));
   if  ThisLen < BestLen
      BestLen = ThisLen;
      Pos = Count;
   end
end
Low = Data(Pos);
Up = Data(Pos + nd - 1);
```

The code for the Bayesian approach is presented in two sections, in the first section all the code for sampling from the Marginal Conditional Posterior Density functions of the Uniform, Beta, Conservative and Expert distributions are given. Then in the following section the code for the Pareto distribution is given.

C.3 Code for Sampling from the Marginal Conditional Posterior Density Functions of the Uniform, Beta, Conservative and Expert Distributions.

The following two blocks of code that is given is the code that calculates the Single-Factor Gaussian copula for the sampled PD and the sampled latent factors. These two blocks of code are used in the sampling of all of the marginal conditional posteriori density functions in this section.

Listing C.11: Single-Factor Gaussian Copula for the PD as a vector

```
% System file.

function Theta = SysPDefBayesa(p,xi,xt)

% Theta as function of row vector p

vec = Quantile('Normal',p);
vec2 = (vec - sqrt(xi)*xt)/sqrt(1 - xi);
Theta = Distr('Normal',vec2);
if prod(size(p)) ~= prod(size(Theta))
    error('Theta and p not same size!')
end
```

Listing C.12: Single-Factor Gaussian Copula for the latent variable as a vector

```
% System file.

function Theta = SysPDefBayesc(xt,p,xi)

% Theta as function of row vector xt

vec = (Quantile('Normal',p) - sqrt(xi)*xt)/sqrt(1 - xi);
Theta = Distr('Normal',vec);
```

The following block of code represents the marginal conditional posterior density function for the latent factor X. This is used in all of the sampling algorithms in this section.

Listing C.13: Marginal Conditional Posterior Density Function for the Latent Factor

```
% System file.

function h = SysPDefBayesf(xvec,Number,xi,p,xt,tau,nt,st)

% Posterior of the Number-th element of x, Number = 1, .., T, at range
%    specified by row vector xvec
% nt, st, xt are column vectors
% xt still of length T, but Number-th element value is irrelevant

T = length(xt);
nx = length(xvec);
xMat = xt*ones(1,nx);
xMat(Number,:) = xvec;
ThetaMat = zeros(T,nx);
for Count = 1:T
    % Theta = SysPDefBayesc(xt,p,xi)
    ThetaMat(Count,:) = SysPDefBayesc(xMat(Count,:),p,xi);
end
stMat = st*ones(1,nx);
ntMat = nt*ones(1,nx);
loglik = stMat.*log(ThetaMat) + (ntMat - stMat).*log(1 - ThetaMat);
if T > 1
    loglik = sum(loglik);   % Add columnwise
end
if Number == 1
   logprior = -xvec.^2/2;
else
   logprior = (xvec - tau*xt(Number - 1)).^2;
   logprior = -0.5*logprior/(1 - tau^2);
end
logh = loglik + logprior;
logh = logh - max(logh);   % Rescales so that max(h) = 1
h = exp(logh);
if max(h) == 0
    error('Posterior density zero everywhere!')
end
```

The following block of code represents the rejection step applied in all of the sampling algorithms.

Listing C.14: The Rejection Step

```
%System file.

function Result = SysPDefBayesh(x,f)

% Same as Reject.m, but without some error testing in order to save time:
%REJECT   Simulates a single value from density, using the Rejection method.
%   ONEDRAW = REJECT(X,F) returns a single value simulated from vector of ...
    density values F, calculated
%      at points in the vector X.
%   X: Vector of x-values (area of support).
%   F: Density values corresponding to X. F may be a marginal posterior, ...
    that is not
```

```
%      normalized to integrate to 1.

Randomise;
n = length(x);
maxf = max(f);
if isinf(maxf)
   error('Density may not have inf as maximum!')
end
if maxf == 0
   error('Density zero everywhere!')
end
Result = [];
while isempty(Result)
    uni = ceil(rand*n);
    uni2 = rand*maxf;
  if uni2 < f(uni)
    Result = x(uni);
  end
end
```

C.3.1 Sampling from the Conditional Posterior Density Function of the Uniform Prior Distribution

The following block of code represents the marginal conditional posterior density function of the Uniform prior distribution.

Listing C.15: Marginal Conditional Posterior Density Function for the Uniform Prior Distribution

```
% System file.

function h = SysPDefBayesUniform(v,xi,xt,nt,st,alphaP,betaP)

% Posterior of V = log(P) at range specified by row vector v
% nt, st, xt are column vectors

if sum(v > 0) > 0
  error('v cannot be positive!')
end
T = length(xt);
nv = length(v);
ThetaMat = zeros(T,nv);
for Count = 1:T
    % Theta = SysPDefBayesa(p,xi,xt)
    ThetaMat(Count,:) = SysPDefBayesa(exp(v),xi,xt(Count));
end
stMat = st*ones(1,nv);
ntMat = nt*ones(1,nv);
loglik = stMat.*log(ThetaMat) + (ntMat - stMat).*log(1 - ThetaMat);
if T > 1
    loglik = sum(loglik);   % Add columnwise
end
logprior = v + log(1/(betaP-alphaP));
logh = loglik + logprior;
logh = logh - max(logh);   % Rescales so that max(h) = 1
h = exp(logh);
if max(h) == 0
    error('Posterior density zero everywhere!')
```

```
end
```

The sampling algorithm used in this study is then applied using a Uniform prior distribution in the following block of code.

Listing C.16: Applying the Sampling Method to the Uniform Prior Distribution

```
function Draws = BayesUniform(nt,st,NReps,NBurn, xi, tau, lowP, highP)
%The BayesUniform function calculates draws of a marginal posterior
%distribution for the probability of default using a Uniform prior and the
%Gibbs sampler.
%   nt      = Obligors at the start of each year considerd in the risk
%             horizon, this is a vector of length T, where T is the number
%             of years in the historical data period.
%   st      = Number of defaults in the portfolio in each year.
%   NReps   = The number of MCMC draws that are retained.
%   NBurn   = The number of Burn-in draws that are unretained.
%   lowP    = The lower bound of the uniform prior distribution, for
%             standard uniform distribution let lowP = 0.
%   highP   = The upper bound of the uniform prior distribution, for the
%             standard uniform distribution let highP = 1.
%
%In order to obtain the Bayesian estimate of the PD from the draws of the
%posterior distribution issue one of the following commands in the viewer
%window:
%   mean(DRAWS) assuming squared error loss (standard),
%   median(DRAWS) assuming absolute loss, and
%   mode(DRAWS) assuming 0-1 loss.

T = length(nt);

%Prelimanary tests and warnings.
if (length(NReps) ~= 1)|(length(NBurn) ~= 1)
   error('NREPS and NBURN should be scalars!')
end

if (NReps < 100)|(NBurn < 0)
   error('Minimum value for NREPS is 100 and NBURN 0!')
end

if (fix(NReps) ~= NReps)|(fix(NBurn) ~= NBurn)
   error('NREPS and NBURN should be integers!')
end

if NReps <= NBurn
   error('NREPS must be greater than NBURN!')
end

if prod(size(nt)) ~= length(nt)
    error('NT should be a vector!')
end

if sum(nt < 10) > 0
    error('Minimum value in NT is 10!')
end

if sum(fix(nt) == nt) < T
    error('NT should contain only integers!')
end
```

```
nt = nt(:);

if prod(size(st)) ~= length(st)
    error('ST should be a vector!')
end

if sum(st < 0) > 0
    error('Minimum value in ST is 0!')
end

if sum(fix(st) == st) < T
    error('ST should contain only integers!')
end

st = st(:);

if length(st) ~= T
    error('NT and ST should be same length!')
end

if sum(nt <= st) > 0
    error('Values in NT should all be larger than those in ST!')
end

% Initialize
xpars = zeros(T,1);

% Initialize draw vectors
P = zeros(NReps,1);
XPARS = zeros(NReps,T);

% Initialize limits
vLower = -10;
vUpper = -0.1;
xAllLower = -3*ones(T,1);
xAllUpper = 3*ones(T,1);

t0 = clock;

for Count = (-NBurn + 1):NReps
   % Draw a value of v (or p)
   Inc = (vUpper - vLower)/20;
   vVec = vLower:Inc:vUpper;
   f = SysPDefBayesUniform(vVec,xi,xpars,nt,st,lowP,highP);
   DoAgain = 0;
   if (f(1) > 0.01)&(vLower ~= -10)
       DoAgain = 1;
       vLower = -10;
   end
   if (f(end) > 0.01)&(vUpper ~= -0.1)
       DoAgain = 1;
       vUpper = -0.1;
   end
   if DoAgain
     Inc = (vUpper - vLower)/20;
     vVec = vLower:Inc:vUpper;
     f = SysPDefBayesUniform(vVec,xi,xpars,nt,st,lowP,highP);
   end
   PosVec = find(f >= 0.01);
```

```
    if PosVec(1) ~= 1
       PosVec = [(PosVec(1) - 1) PosVec];
    end
    if PosVec(end) ~= length(f)
       PosVec = [PosVec (PosVec(end) + 1)];
    end
    vVec = vVec(PosVec);
    f = f(PosVec);
    if length(PosVec) < 10
       Inc = (vVec(end) - vVec(1))/20;
       vVec = vVec(1):Inc:vVec(end);
       f = SysPDefBayesUniform(vVec,xi,xpars,nt,st,lowP,highP);
    end
    v = SysPDefBayesh(vVec,f);
    p = exp(v);
    vLower = vVec(1) - 0.2*(vVec(end) - vVec(1));
    if vLower < -10
        vLower = -10;
    end
    vUpper = vVec(end) + 0.2*(vVec(end) - vVec(1));
    if vUpper > -0.1
        vUpper = -0.1;
    end
    if Count > 0
       P(Count) = p;
    end

    % Draw a value of xt
    for CountT = 1:T
      xLower = xAllLower(CountT);
      xUpper = xAllUpper(CountT);
      Inc = (xUpper - xLower)/20;
      xVec = xLower:Inc:xUpper;
      f = SysPDefBayesf(xVec,CountT,xi,p,xpars,tau,nt,st);
      Changed = 0;
      Lastf = f(end);
      PosMax = find(f == max(f));
      PosMax = PosMax(1);
      xMax = xVec(PosMax);
      while f(1) > 0.01
          Changed = 1;
          xLower = xLower - 3;
          f = SysPDefBayesf([xLower xMax],CountT,xi,p,xpars,tau,nt,st);
      end
      while Lastf > 0.01
          Changed = 1;
          xUpper = xUpper + 3;
          f = SysPDefBayesf([xMax xUpper],CountT,xi,p,xpars,tau,nt,st);
          Lastf = f(end);
      end
      if Changed
        Inc = (xUpper - xLower)/20;
        xVec = xLower:Inc:xUpper;
        f = SysPDefBayesf(xVec,CountT,xi,p,xpars,tau,nt,st);
      end
      PosVec = find(f >= 0.01);
      if PosVec(1) ~= 1
        PosVec = [(PosVec(1) - 1) PosVec];
      end
      if PosVec(end) ~= length(f)
```

```
      PosVec = [PosVec (PosVec(end) + 1)];
    end
    xVec = xVec(PosVec);
    f = f(PosVec);
    if length(PosVec) < 10
      Inc = (xVec(end) - xVec(1))/20;
      xVec = xVec(1):Inc:xVec(end);
      f = SysPDefBayesf(xVec,CountT,xi,p,xpars,tau,nt,st);
    end
    x = SysPDefBayesh(xVec,f);
    xpars(CountT) = x;
    xLower = xVec(1) - 0.2*(xVec(end) - xVec(1));
    xUpper = xVec(end) + 0.2*(xVec(end) - xVec(1));
    xAllLower(CountT) = xLower;
    xAllUpper(CountT) = xUpper;
    if Count > 0
      XPARS(Count,CountT) = x;
    end
  end

  % Pause and display
   if etime(clock,t0) > 30
     t0 = clock;
     disp(strcat(num2str(round(1000*(Count + NBurn)/(NReps + ...
         NBurn))/10),'%'))
       pause(1)
   end

end

%Plot a histogram of the marginal posterior density of the draws of the
%Probability of Default.
%hist(Draws(:,1));
Draws = [P XPARS];
```

C.3.2 Sampling from the Conditional Posterior Density Function of the Beta Prior Distribution

The following block of code represents the marginal conditional posterior density function of the Beta prior distribution.

Listing C.17: Marginal Conditional Posterior Density Function for the Beta Prior Distribution

```
% System file.

function h = SysPDefBayesd(v,xi,xt,nt,st,alphaP,betaP)

% Posterior of V = log(P) at range specified by row vector v
% nt, st, xt are column vectors

if sum(v > 0) > 0
  error('v cannot be positive!')
end
T = length(xt);
nv = length(v);
ThetaMat = zeros(T,nv);
for Count = 1:T
    % Theta = SysPDefBayesa(p,xi,xt)
```

```
    ThetaMat(Count,:) = SysPDefBayesa(exp(v),xi,xt(Count));
end
stMat = st*ones(1,nv);
ntMat = nt*ones(1,nv);
loglik = stMat.*log(ThetaMat) + (ntMat - stMat).*log(1 - ThetaMat);
if T > 1
    loglik = sum(loglik);   % Add columnwise
end
logprior = alphaP*v + (betaP - 1)*log(1 - exp(v));
logh = loglik + logprior;
logh = logh - max(logh);   % Rescales so that max(h) = 1
h = exp(logh);
if max(h) == 0
    error('Posterior density zero everywhere!')
end
```

The sampling algorithm used in this study is then applied using a Beta prior distribution in the following block of code.

Listing C.18: Applying the Sampling Method to the Beta Prior Distribution

```
function Draws = BayesBeta(nt,st,NReps,NBurn, xi, tau, alphaP, betaP)
%The BayesUniform function calculates draws of a marginal posterior
%distribution for the probability of default using a Uniform prior and the
%Gibbs sampler.
%   nt      = Obligors at the start of each year considerd in the risk
%             horizon, this is a vector of length T, where T is the number
%             of years in the historical data period.
%   st      = Number of defaults in the portfolio in each year.
%   NReps   = The number of MCMC draws that are retained.
%   NBurn   = The number of Burn-in draws that are unretained.
%   lowP    = The lower bound of the uniform prior distribution, for
%             standard uniform distribution let lowP = 0.
%   highP   = The upper bound of the uniform prior distribution, for the
%             standard uniform distribution let highP = 1.
%
%In order to obtain the Bayesian estimate of the PD from the draws of the
%posterior distribution issue one of the following commands in the viewer
%window:
%   mean(DRAWS) assuming squared error loss (standard),
%   median(DRAWS) assuming absolute loss, and
%   mode(DRAWS) assuming 0-1 loss.

T = length(nt);

%Prelimanary tests and warnings.
if (length(NReps) ~= 1)|(length(NBurn) ~= 1)
   error('NREPS and NBURN should be scalars!')
end

if (NReps < 100)|(NBurn < 0)
   error('Minimum value for NREPS is 100 and NBURN 0!')
end

if (fix(NReps) ~= NReps)|(fix(NBurn) ~= NBurn)
   error('NREPS and NBURN should be integers!')
end

if NReps <= NBurn
```

```
   error('NREPS must be greater than NBURN!')
end

if prod(size(nt)) ~= length(nt)
    error('NT should be a vector!')
end

if sum(nt < 10) > 0
    error('Minimum value in NT is 10!')
end

if sum(fix(nt) == nt) < T
    error('NT should contain only integers!')
end

nt = nt(:);

if prod(size(st)) ~= length(st)
    error('ST should be a vector!')
end

if sum(st < 0) > 0
    error('Minimum value in ST is 0!')
end

if sum(fix(st) == st) < T
    error('ST should contain only integers!')
end

st = st(:);

if length(st) ~= T
    error('NT and ST should be same length!')
end

if sum(nt <= st) > 0
    error('Values in NT should all be larger than those in ST!')
end

% Initialize
xpars = zeros(T,1);

% Initialize draw vectors
P = zeros(NReps,1);
XPARS = zeros(NReps,T);

% Initialize limits
vLower = -10;
vUpper = -0.1;
xAllLower = -3*ones(T,1);
xAllUpper = 3*ones(T,1);

t0 = clock;

for Count = (-NBurn + 1):NReps
   % Draw a value of v (or p)
   Inc = (vUpper - vLower)/20;
   vVec = vLower:Inc:vUpper;
   f = SysPDefBayesd(vVec,xi,xpars,nt,st,alphaP,betaP);
   DoAgain = 0;
```

```
if (f(1) > 0.01)&(vLower ~= -10)
     DoAgain = 1;
     vLower = -10;
end
if (f(end) > 0.01)&(vUpper ~= -0.1)
     DoAgain = 1;
     vUpper = -0.1;
end
if DoAgain
  Inc = (vUpper - vLower)/20;
  vVec = vLower:Inc:vUpper;
  f = SysPDefBayesd(vVec,xi,xpars,nt,st,alphaP,betaP);
end
PosVec = find(f >= 0.01);
if PosVec(1) ~= 1
   PosVec = [(PosVec(1) - 1) PosVec];
end
if PosVec(end) ~= length(f)
   PosVec = [PosVec (PosVec(end) + 1)];
end
vVec = vVec(PosVec);
f = f(PosVec);
if length(PosVec) < 10
   Inc = (vVec(end) - vVec(1))/20;
   vVec = vVec(1):Inc:vVec(end);
   f = SysPDefBayesd(vVec,xi,xpars,nt,st,alphaP,betaP);
end
v = SysPDefBayesh(vVec,f);
p = exp(v);
vLower = vVec(1) - 0.2*(vVec(end) - vVec(1));
if vLower < -10
     vLower = -10;
end
vUpper = vVec(end) + 0.2*(vVec(end) - vVec(1));
if vUpper > -0.1
     vUpper = -0.1;
end
if Count > 0
   P(Count) = p;
end

% Draw a value of xt
for CountT = 1:T
  xLower = xAllLower(CountT);
  xUpper = xAllUpper(CountT);
  Inc = (xUpper - xLower)/20;
  xVec = xLower:Inc:xUpper;
  f = SysPDefBayesf(xVec,CountT,xi,p,xpars,tau,nt,st);
  Changed = 0;
  Lastf = f(end);
  PosMax = find(f == max(f));
  PosMax = PosMax(1);
  xMax = xVec(PosMax);
  while f(1) > 0.01
      Changed = 1;
      xLower = xLower - 3;
      f = SysPDefBayesf([xLower xMax],CountT,xi,p,xpars,tau,nt,st);
  end
  while Lastf > 0.01
      Changed = 1;
```

```
          xUpper = xUpper + 3;
          f = SysPDefBayesf([xMax xUpper],CountT,xi,p,xpars,tau,nt,st);
          Lastf = f(end);
      end
      if Changed
        Inc = (xUpper - xLower)/20;
        xVec = xLower:Inc:xUpper;
        f = SysPDefBayesf(xVec,CountT,xi,p,xpars,tau,nt,st);
      end
      PosVec = find(f >= 0.01);
      if PosVec(1) ~= 1
        PosVec = [(PosVec(1) - 1) PosVec];
      end
      if PosVec(end) ~= length(f)
        PosVec = [PosVec (PosVec(end) + 1)];
      end
      xVec = xVec(PosVec);
      f = f(PosVec);
      if length(PosVec) < 10
        Inc = (xVec(end) - xVec(1))/20;
        xVec = xVec(1):Inc:xVec(end);
        f = SysPDefBayesf(xVec,CountT,xi,p,xpars,tau,nt,st);
      end
      x = SysPDefBayesh(xVec,f);
      xpars(CountT) = x;
      xLower = xVec(1) - 0.2*(xVec(end) - xVec(1));
      xUpper = xVec(end) + 0.2*(xVec(end) - xVec(1));
      xAllLower(CountT) = xLower;
      xAllUpper(CountT) = xUpper;
      if Count > 0
        XPARS(Count,CountT) = x;
      end
    end

    % Pause and display
     if etime(clock,t0) > 30
       t0 = clock;
       disp(strcat(num2str(round(1000*(Count + NBurn)/(NReps + ...
           NBurn))/10),'%'))
         pause(1)
     end

end

%Plot a histogram of the marginal posterior density of the draws of the
%Probability of Default.
%hist(Draws(:,1));
Draws = [P XPARS];
```

C.3.3 Sampling from the Conditional Posterior Density Function of the Conservative Prior Distribution

The following block of code represents the marginal conditional posterior density function of the Conservative prior distribution.

Listing C.19: Marginal Conditional Posterior Density Function for the Conservative Prior Distribution

```
% System file.

function h = SysPDefBayesCon(v,xi,xt,nt,st)

% Posterior of V = log(P) at range specified by row vector v
% nt, st, xt are column vectors

if sum(v > 0) > 0
  error('v cannot be positive!')
end
T = length(xt);
nv = length(v);
ThetaMat = zeros(T,nv);
for Count = 1:T
    % Theta = SysPDefBayesa(p,xi,xt)
    ThetaMat(Count,:) = SysPDefBayesa(exp(v),xi,xt(Count));
end
stMat = st*ones(1,nv);
ntMat = nt*ones(1,nv);
loglik = stMat.*log(ThetaMat) + (ntMat - stMat).*log(1 - ThetaMat);
if T > 1
    loglik = sum(loglik);   % Add columnwise
end
logprior = v + log(1)-log(1 - exp(v));
logh = loglik + logprior;
logh = logh - max(logh);   % Rescales so that max(h) = 1
h = exp(logh);
if max(h) == 0
    error('Posterior density zero everywhere!')
end
```

The sampling algorithm used in this study is then applied using a Beta prior distribution in the following block of code.

Listing C.20: Applying the Sampling Method to the Conservative Prior Distribution

```
function Draws = BayesConservative(nt,st,NReps,NBurn, xi, tau)
%The BayesUniform function calculates draws of a marginal posterior
%distribution for the probability of default using a Uniform prior and the
%Gibbs sampler.
%   nt      = Obligors at the start of each year considerd in the risk
%             horizon, this is a vector of length T, where T is the number
%             of years in the historical data period.
%   st      = Number of defaults in the portfolio in each year.
%   NReps   = The number of MCMC draws that are retained.
%   NBurn   = The number of Burn-in draws that are unretained.
%   lowP    = The lower bound of the uniform prior distribution, for
%             standard uniform distribution let lowP = 0.
%   highP   = The upper bound of the uniform prior distribution, for the
%             standard uniform distribution let highP = 1.
%
%In order to obtain the Bayesian estimate of the PD from the draws of the
%posterior distribution issue one of the following commands in the viewer
%window:
%   mean(DRAWS) assuming squared error loss (standard),
%   median(DRAWS) assuming absolute loss, and
%   mode(DRAWS) assuming 0-1 loss.

T = length(nt);
```

```
%Prelimanary tests and warnings.
if (length(NReps) ~= 1)|(length(NBurn) ~= 1)
   error('NREPS and NBURN should be scalars!')
end

if (NReps < 100)|(NBurn < 0)
   error('Minimum value for NREPS is 100 and NBURN 0!')
end

if (fix(NReps) ~= NReps)|(fix(NBurn) ~= NBurn)
   error('NREPS and NBURN should be integers!')
end

if NReps <= NBurn
   error('NREPS must be greater than NBURN!')
end

if prod(size(nt)) ~= length(nt)
    error('NT should be a vector!')
end

if sum(nt < 10) > 0
    error('Minimum value in NT is 10!')
end

if sum(fix(nt) == nt) < T
    error('NT should contain only integers!')
end

nt = nt(:);

if prod(size(st)) ~= length(st)
    error('ST should be a vector!')
end

if sum(st < 0) > 0
    error('Minimum value in ST is 0!')
end

if sum(fix(st) == st) < T
    error('ST should contain only integers!')
end

st = st(:);

if length(st) ~= T
    error('NT and ST should be same length!')
end

if sum(nt <= st) > 0
    error('Values in NT should all be larger than those in ST!')
end

% Initialize
xpars = zeros(T,1);

% Initialize draw vectors
P = zeros(NReps,1);
XPARS = zeros(NReps,T);
```

```
% Initialize limits
vLower = -10;
vUpper = -0.1;
xAllLower = -3*ones(T,1);
xAllUpper = 3*ones(T,1);

t0 = clock;

for Count = (-NBurn + 1):NReps
   % Draw a value of v (or p)
   Inc = (vUpper - vLower)/20;
   vVec = vLower:Inc:vUpper;
   f = SysPDefBayesCon(vVec,xi,xpars,nt,st);
   DoAgain = 0;
   if (f(1) > 0.01)&(vLower ~= -10)
       DoAgain = 1;
       vLower = -10;
   end
   if (f(end) > 0.01)&(vUpper ~= -0.1)
       DoAgain = 1;
       vUpper = -0.1;
   end
   if DoAgain
     Inc = (vUpper - vLower)/20;
     vVec = vLower:Inc:vUpper;
     f = SysPDefBayesCon(vVec,xi,xpars,nt,st);
   end
   PosVec = find(f >= 0.01);
   if PosVec(1) ~= 1
      PosVec = [(PosVec(1) - 1) PosVec];
   end
   if PosVec(end) ~= length(f)
      PosVec = [PosVec (PosVec(end) + 1)];
   end
   vVec = vVec(PosVec);
   f = f(PosVec);
   if length(PosVec) < 10
      Inc = (vVec(end) - vVec(1))/20;
      vVec = vVec(1):Inc:vVec(end);
      f = SysPDefBayesCon(vVec,xi,xpars,nt,st);
   end
   v = SysPDefBayesh(vVec,f);
   p = exp(v);
   vLower = vVec(1) - 0.2*(vVec(end) - vVec(1));
   if vLower < -10
       vLower = -10;
   end
   vUpper = vVec(end) + 0.2*(vVec(end) - vVec(1));
   if vUpper > -0.1
       vUpper = -0.1;
   end
   if Count > 0
      P(Count) = p;
   end

   % Draw a value of xt
   for CountT = 1:T
     xLower = xAllLower(CountT);
     xUpper = xAllUpper(CountT);
```

```
      Inc = (xUpper - xLower)/20;
      xVec = xLower:Inc:xUpper;
      f = SysPDefBayesf(xVec,CountT,xi,p,xpars,tau,nt,st);
      Changed = 0;
      Lastf = f(end);
      PosMax = find(f == max(f));
      PosMax = PosMax(1);
      xMax = xVec(PosMax);
      while f(1) > 0.01
          Changed = 1;
          xLower = xLower - 3;
          f = SysPDefBayesf([xLower xMax],CountT,xi,p,xpars,tau,nt,st);
      end
      while Lastf > 0.01
          Changed = 1;
          xUpper = xUpper + 3;
          f = SysPDefBayesf([xMax xUpper],CountT,xi,p,xpars,tau,nt,st);
          Lastf = f(end);
      end
      if Changed
        Inc = (xUpper - xLower)/20;
        xVec = xLower:Inc:xUpper;
        f = SysPDefBayesf(xVec,CountT,xi,p,xpars,tau,nt,st);
      end
      PosVec = find(f >= 0.01);
      if PosVec(1) ~= 1
        PosVec = [(PosVec(1) - 1) PosVec];
      end
      if PosVec(end) ~= length(f)
        PosVec = [PosVec (PosVec(end) + 1)];
      end
      xVec = xVec(PosVec);
      f = f(PosVec);
      if length(PosVec) < 10
        Inc = (xVec(end) - xVec(1))/20;
        xVec = xVec(1):Inc:xVec(end);
        f = SysPDefBayesf(xVec,CountT,xi,p,xpars,tau,nt,st);
      end
      x = SysPDefBayesh(xVec,f);
      xpars(CountT) = x;
      xLower = xVec(1) - 0.2*(xVec(end) - xVec(1));
      xUpper = xVec(end) + 0.2*(xVec(end) - xVec(1));
      xAllLower(CountT) = xLower;
      xAllUpper(CountT) = xUpper;
      if Count > 0
        XPARS(Count,CountT) = x;
      end
    end

    % Pause and display
     if etime(clock,t0) > 30
       t0 = clock;
       disp(strcat(num2str(round(1000*(Count + NBurn)/(NReps + ...
           NBurn))/10),'%'))
         pause(1)
     end

end

%Plot a histogram of the marginal posterior density of the draws of the
```

```
%Probability of Default.
%hist(Draws(:,1));
Draws = [P XPARS];
```

C.3.4 Sampling from the Conditional Posterior Density Function of the Expert Prior Distribution

The following block of code represents the marginal conditional posterior density function of the Expert prior distribution.

Listing C.21: Marginal Conditional Posterior Density Function for the Expert Prior Distribution

```
% System file.

function h = SysPDefBayesExpert(v,xi,xt,nt,st, low, mid, up)

% Posterior of V = log(P) at range specified by row vector v
% nt, st, xt are column vectors

if sum(v > 0) > 0
  error('v cannot be positive!')
end
T = length(xt);
nv = length(v);
ThetaMat = zeros(T,nv);
for Count = 1:T
    % Theta = SysPDefBayesa(p,xi,xt)
    ThetaMat(Count,:) = SysPDefBayesa(exp(v),xi,xt(Count));
end
ThetaMat(find(ThetaMat==0))=0.000001;
stMat = st*ones(1,nv);
ntMat = nt*ones(1,nv);

loglik = stMat.*log(ThetaMat) + (ntMat - stMat).*log(1 - ThetaMat);
if T > 1
    loglik = sum(loglik);   % Add columnwise
end

logh = zeros(1,nv);
logup = log(up);
logmid = log(mid);
loglow = log(low);

for i = 1:nv

    if v(i) > logup

        logh(i) = -inf;

    elseif v(i) > logmid

        logprior = v(i) + log((exp(logup)-exp(v(i))));

        logh(i) = loglik(i) + logprior;

    elseif v(i) > loglow
```

```
        logprior = v(i) + log((exp(v(i))-exp(loglow)));

        logh(i) = loglik(i) + logprior;

    else

        logh(i) = -inf;

    end

end

logh = logh - max(logh);   % Rescales so that max(h) = 1

h = exp(logh);

if max(h) == 0
    error('Posterior density zero everywhere!')
end
```

The sampling algorithm used in this study is then applied using a Beta prior distribution in the following block of code.

Listing C.22: Applying the Sampling Method to the Expert Prior Distribution

```
function Draws = BayesExpert(nt,st,NReps,NBurn, xi, tau, low, mid, up)
%The BayesUniform function calculates draws of a marginal posterior
%distribution for the probability of default using a Uniform prior and the
%Gibbs sampler.
%   nt      = Obligors at the start of each year considerd in the risk
%             horizon, this is a vector of length T, where T is the number
%             of years in the historical data period.
%   st      = Number of defaults in the portfolio in each year.
%   NReps   = The number of MCMC draws that are retained.
%   NBurn   = The number of Burn-in draws that are unretained.
%   lowP    = The lower bound of the uniform prior distribution, for
%             standard uniform distribution let lowP = 0.
%   highP   = The upper bound of the uniform prior distribution, for the
%             standard uniform distribution let highP = 1.
%
%In order to obtain the Bayesian estimate of the PD from the draws of the
%posterior distribution issue one of the following commands in the viewer
%window:
%   mean(DRAWS) assuming squared error loss (standard),
%   median(DRAWS) assuming absolute loss, and
%   mode(DRAWS) assuming 0-1 loss.

T = length(nt);

%Prelimanary tests and warnings.
if (length(NReps) ~= 1)|(length(NBurn) ~= 1)
   error('NREPS and NBURN should be scalars!')
end

if (NReps < 100)|(NBurn < 0)
   error('Minimum value for NREPS is 100 and NBURN 0!')
end

if (fix(NReps) ~= NReps)|(fix(NBurn) ~= NBurn)
```

```
   error('NREPS and NBURN should be integers!')
end

if NReps <= NBurn
   error('NREPS must be greater than NBURN!')
end

if prod(size(nt)) ~= length(nt)
    error('NT should be a vector!')
end

if sum(nt < 10) > 0
    error('Minimum value in NT is 10!')
end

if sum(fix(nt) == nt) < T
    error('NT should contain only integers!')
end

nt = nt(:);

if prod(size(st)) ~= length(st)
    error('ST should be a vector!')
end

if sum(st < 0) > 0
    error('Minimum value in ST is 0!')
end

if sum(fix(st) == st) < T
    error('ST should contain only integers!')
end

st = st(:);

if length(st) ~= T
    error('NT and ST should be same length!')
end

if sum(nt <= st) > 0
    error('Values in NT should all be larger than those in ST!')
end

% Initialize
xpars = zeros(T,1);

% Initialize draw vectors
P = zeros(NReps,1);
XPARS = zeros(NReps,T);

% Initialize limits
vLower = -10;
vUpper = -0.1;
xAllLower = -3*ones(T,1);
xAllUpper = 3*ones(T,1);

t0 = clock;

for Count = (-NBurn + 1):NReps
   % Draw a value of v (or p)
```

```
Inc = (vUpper - vLower)/20;
vVec = vLower:Inc:vUpper;
f = SysPDefBayesExpert(vVec,xi,xpars,nt,st,low, mid, up);
DoAgain = 0;
if (f(1) > 0.01)&(vLower ~= -10)
    DoAgain = 1;
    vLower = -10;
end
if (f(end) > 0.01)&(vUpper ~= -0.1)
    DoAgain = 1;
    vUpper = -0.1;
end
if DoAgain
  Inc = (vUpper - vLower)/20;
  vVec = vLower:Inc:vUpper;
  f = SysPDefBayesExpert(vVec,xi,xpars,nt,st,low, mid, up);
end
PosVec = find(f >= 0.01);
if PosVec(1) ~= 1
   PosVec = [(PosVec(1) - 1) PosVec];
end
if PosVec(end) ~= length(f)
   PosVec = [PosVec (PosVec(end) + 1)];
end
vVec = vVec(PosVec);
f = f(PosVec);
if length(PosVec) < 10
   Inc = (vVec(end) - vVec(1))/20;
   vVec = vVec(1):Inc:vVec(end);
   f = SysPDefBayesExpert(vVec,xi,xpars,nt,st,low, mid, up);
end
v = SysPDefBayesh(vVec,f);
p = exp(v);
vLower = vVec(1) - 0.2*(vVec(end) - vVec(1));
if vLower < -10
    vLower = -10;
end
vUpper = vVec(end) + 0.2*(vVec(end) - vVec(1));
if vUpper > -0.1
    vUpper = -0.1;
end
if Count > 0
   P(Count) = p;
end

% Draw a value of xt
for CountT = 1:T
  xLower = xAllLower(CountT);
  xUpper = xAllUpper(CountT);
  Inc = (xUpper - xLower)/20;
  xVec = xLower:Inc:xUpper;
  f = SysPDefBayesf(xVec,CountT,xi,p,xpars,tau,nt,st);
  Changed = 0;
  Lastf = f(end);
  PosMax = find(f == max(f));
  PosMax = PosMax(1);
  xMax = xVec(PosMax);
  while f(1) > 0.01
      Changed = 1;
      xLower = xLower - 3;
```

```
        f = SysPDefBayesf([xLower xMax],CountT,xi,p,xpars,tau,nt,st);
      end
      while Lastf > 0.01
        Changed = 1;
        xUpper = xUpper + 3;
        f = SysPDefBayesf([xMax xUpper],CountT,xi,p,xpars,tau,nt,st);
        Lastf = f(end);
      end
      if Changed
        Inc = (xUpper - xLower)/20;
        xVec = xLower:Inc:xUpper;
        f = SysPDefBayesf(xVec,CountT,xi,p,xpars,tau,nt,st);
      end
      PosVec = find(f >= 0.01);
      if PosVec(1) ~= 1
        PosVec = [(PosVec(1) - 1) PosVec];
      end
      if PosVec(end) ~= length(f)
        PosVec = [PosVec (PosVec(end) + 1)];
      end
      xVec = xVec(PosVec);
      f = f(PosVec);
      if length(PosVec) < 10
        Inc = (xVec(end) - xVec(1))/20;
        xVec = xVec(1):Inc:xVec(end);
        f = SysPDefBayesf(xVec,CountT,xi,p,xpars,tau,nt,st);
      end
      x = SysPDefBayesh(xVec,f);
      xpars(CountT) = x;
      xLower = xVec(1) - 0.2*(xVec(end) - xVec(1));
      xUpper = xVec(end) + 0.2*(xVec(end) - xVec(1));
      xAllLower(CountT) = xLower;
      xAllUpper(CountT) = xUpper;
      if Count > 0
        XPARS(Count,CountT) = x;
      end
    end

    % Pause and display
     if etime(clock,t0) > 30
       t0 = clock;
       disp(strcat(num2str(round(1000*(Count + NBurn)/(NReps + ...
           NBurn))/10),'%'))
         pause(1)
     end

end

%Plot a histogram of the marginal posterior density of the draws of the
%Probability of Default.
%hist(Draws(:,1));
Draws = [P XPARS];
```

C.4 Sampling from the Conditional Posterior Density Function of the Pareto Prior Distribution

Recall from Chapter 4 that the transformation $\psi = 1/p$ allows the use of the Pareto distribution as a prior distribution. Due to this transformation the marginal conditional posterior of the latent factor takes a new shape. The code for the Single Factor Gaussian copula will also differ from the code mentioned before.

The first two blocks of code is the Single-Factor Gaussian copula for ψ and the latent factor variable following the transformation.

Listing C.23: Single-Factor Gaussian Copula for the ψ as a vector

```
% System file.

function Theta = SysPDefBayesPareto(w,xi,xt)

% Theta as function of row vector p

vec = Quantile('Normal',1./w);
vec2 = (vec - sqrt(xi)*xt)/sqrt(1 - xi);
Theta = Distr('Normal',vec2);
if prod(size(w)) ~= prod(size(Theta))
    error('Theta and p not same size!')
end
```

Listing C.24: Single-Factor Gaussian Copula for the latent variable as a vector

```
% System file.

function Theta = SysPDefBayesPareto3(xt,w,xi)

% Theta as function of row vector xt

vec = (Quantile('Normal',1./w) - sqrt(xi)*xt)/sqrt(1 - xi);
Theta = Distr('Normal',vec);
```

The following block of code represents the marginal conditional posterior density function of the Pareto prior distribution.

Listing C.25: Marginal Conditional Posterior Density Function for the Pareto Prior Distribution

```
% System file.

function h = SysPDefBayesPareto2(v,xi,xt,nt,st,gammaP)

% Posterior of V = log(1/P) at range specified by row vector v
% nt, st, xt are column vectors

%if sum(v > 0) > 0
%  error('v cannot be positive!')
%end
T = length(xt);
nv = length(v);
```

```
ThetaMat = zeros(T,nv);
for Count = 1:T
    % Theta = SysPDefBayesa(p,xi,xt)
    %w=1./exp(v);
    ThetaMat(Count,:) = SysPDefBayesPareto(exp(v),xi,xt(Count));
end
stMat = st*ones(1,nv);
ntMat = nt*ones(1,nv);
loglik = stMat.*log(ThetaMat) + (ntMat - stMat).*log(1 - ThetaMat);
if T > 1
    loglik = sum(loglik);   % Add columnwise
end
logprior = -v + 1/(2*(1+gammaP)); %alpha = gamma (eintlik alpha=1/gamma) ...
   logprior = -v + 1/(2*(1+alphaP));
logh = loglik + logprior;
logh = logh - max(logh);   % Rescales so that max(h) = 1
h = exp(logh);
if max(h) == 0
    error('Posterior density zero everywhere!')
end
```

Due to the transformation the marginal conditional posterior of the latent factor also takes on a new shape, this is given by the following block of code.

Listing C.26: Marginal Conditional Posterior Density Function for the Latent Factor in the Pareto Setting

```
% System file.

function h = SysPDefBayesPareto4(xvec,Number,xi,p,xt,tau,nt,st)

% Posterior of the Number-th element of x, Number = 1, .., T, at range
%    specified by row vector xvec
% nt, st, xt are column vectors
% xt still of length T, but Number-th element value is irrelevant

T = length(xt);
nx = length(xvec);
xMat = xt*ones(1,nx);
xMat(Number,:) = xvec;
ThetaMat = zeros(T,nx);
for Count = 1:T
    % Theta = SysPDefBayesc(xt,p,xi)
    ThetaMat(Count,:) = SysPDefBayesPareto3(xMat(Count,:),p,xi);
end
stMat = st*ones(1,nx);
ntMat = nt*ones(1,nx);
loglik = stMat.*log(ThetaMat) + (ntMat - stMat).*log(1 - ThetaMat);
if T > 1
    loglik = sum(loglik);   % Add columnwise
end
if Number == 1
   logprior = -xvec.^2/2;
else
   logprior = (xvec - tau*xt(Number - 1)).^2;
   logprior = -0.5*logprior/(1 - tau^2);
end
logh = loglik + logprior;
logh = logh - max(logh);   % Rescales so that max(h) = 1
```

```
h = exp(logh);
if max(h) == 0
    error('Posterior density zero everywhere!')
end
```

The sampling algorithm used in this study is then applied using a Pareto prior distribution in the following block of code. The code for the rejection step is the same as before.

Listing C.27: Applying the Sampling Method to the Pareto Prior Distribution

```
function Draws = BayesPareto(nt,st,NReps,NBurn, xi, tau, gammaP)
%The BayesUniform function calculates draws of a marginal posterior
%distribution for the probability of default using a Uniform prior and the
%Gibbs sampler.
%   nt      = Obligors at the start of each year considerd in the risk
%             horizon, this is a vector of length T, where T is the number
%             of years in the historical data period.
%   st      = Number of defaults in the portfolio in each year.
%   NReps   = The number of MCMC draws that are retained.
%   NBurn   = The number of Burn-in draws that are unretained.
%   lowP    = The lower bound of the uniform prior distribution, for
%             standard uniform distribution let lowP = 0.
%   highP   = The upper bound of the uniform prior distribution, for the
%             standard uniform distribution let highP = 1.
%
%In order to obtain the Bayesian estimate of the PD from the draws of the
%posterior distribution issue one of the following commands in the viewer
%window:
%   mean(DRAWS) assuming squared error loss (standard),
%   median(DRAWS) assuming absolute loss, and
%   mode(DRAWS) assuming 0-1 loss.

T = length(nt);

%Prelimanary tests and warnings.
if (length(NReps) ~= 1)|(length(NBurn) ~= 1)
   error('NREPS and NBURN should be scalars!')
end

if (NReps < 100)|(NBurn < 0)
   error('Minimum value for NREPS is 100 and NBURN 0!')
end

if (fix(NReps) ~= NReps)|(fix(NBurn) ~= NBurn)
   error('NREPS and NBURN should be integers!')
end

if NReps <= NBurn
   error('NREPS must be greater than NBURN!')
end

if prod(size(nt)) ~= length(nt)
    error('NT should be a vector!')
end

if sum(nt < 10) > 0
    error('Minimum value in NT is 10!')
end
```

```
if sum(fix(nt) == nt) < T
    error('NT should contain only integers!')
end

nt = nt(:);

if prod(size(st)) ~= length(st)
    error('ST should be a vector!')
end

if sum(st < 0) > 0
    error('Minimum value in ST is 0!')
end

if sum(fix(st) == st) < T
    error('ST should contain only integers!')
end

st = st(:);

if length(st) ~= T
    error('NT and ST should be same length!')
end

if sum(nt <= st) > 0
    error('Values in NT should all be larger than those in ST!')
end

% Initialize
xpars = zeros(T,1);

% Initialize draw vectors
P = zeros(NReps,1);
XPARS = zeros(NReps,T);

% Initialize limits
wLower = 1;%-10;
wUpper = 20;%5; %-0.1;
xAllLower = -3*ones(T,1);
xAllUpper = 3*ones(T,1);

t0 = clock;

for Count = (-NBurn + 1):NReps
   % Draw a value of w (or 1/p)
   Inc = (wUpper - wLower)/20;
   wVec = wLower:Inc:wUpper;
   f = SysPDefBayesPareto2(wVec,xi,xpars,nt,st,gammaP);
   %h = SysPDefBayesPareto2(v,xi,xt,nt,st,gammaP)
   DoAgain = 0;
   if (f(1) > 0.01)&(wLower ~= 1)
       DoAgain = 1;
       wLower = 1;
   end
   if (f(end) > 0.01)&(wUpper ~= 20)
       DoAgain = 1;
       wUpper = wUpper*2;
   end
   if DoAgain
     Inc = (wUpper - wLower)/20;
```

```
    wVec = wLower:Inc:wUpper;
    f = SysPDefBayesPareto2(wVec,xi,xpars,nt,st,gammaP);
  end
  PosVec = find(f >= 0.01);
  if PosVec(1) ~= 1
     PosVec = [(PosVec(1) - 1) PosVec];
  end
  if PosVec(end) ~= length(f)
     PosVec = [PosVec (PosVec(end) + 1)];
  end
  wVec = wVec(PosVec);
  f = f(PosVec);
  if length(PosVec) < 10
     Inc = (wVec(end) - wVec(1))/20;
     wVec = wVec(1):Inc:wVec(end);
     f = SysPDefBayesPareto2(wVec,xi,xpars,nt,st,gammaP);
  end
  v = SysPDefBayesh(wVec,f);
  w = exp(v);
  wLower = wVec(1) - 0.2*(wVec(end) - wVec(1));
  if wLower < 1
      wLower = 1;
  end
  wUpper = wVec(end) + 0.2*(wVec(end) - wVec(1));
  if wUpper > 5
      wUpper = 5;
  end
  if Count > 0
     P(Count) = w;
  end

  % Draw a value of xt
  for CountT = 1:T
    xLower = xAllLower(CountT);
    xUpper = xAllUpper(CountT);
    Inc = (xUpper - xLower)/20;
    xVec = xLower:Inc:xUpper;
    f = SysPDefBayesPareto4(xVec,CountT,xi,w,xpars,tau,nt,st);
    Changed = 0;
    Lastf = f(end);
    PosMax = find(f == max(f));
    PosMax = PosMax(1);
    xMax = xVec(PosMax);
    while f(1) > 0.01
        Changed = 1;
        xLower = xLower - 3;
        f = SysPDefBayesPareto4([xLower xMax],CountT,xi,w,xpars,tau,nt,st);
    end
    while Lastf > 0.01
        Changed = 1;
        xUpper = xUpper + 3;
        f = SysPDefBayesPareto4([xMax xUpper],CountT,xi,w,xpars,tau,nt,st);
        Lastf = f(end);
    end
    if Changed
      Inc = (xUpper - xLower)/20;
      xVec = xLower:Inc:xUpper;
      f = SysPDefBayesPareto4(xVec,CountT,xi,w,xpars,tau,nt,st);
    end
    PosVec = find(f >= 0.01);
```

```
      if PosVec(1) ~= 1
        PosVec = [(PosVec(1) - 1) PosVec];
      end
      if PosVec(end) ~= length(f)
        PosVec = [PosVec (PosVec(end) + 1)];
      end
      xVec = xVec(PosVec);
      f = f(PosVec);
      if length(PosVec) < 10
        Inc = (xVec(end) - xVec(1))/20;
        xVec = xVec(1):Inc:xVec(end);
        f = SysPDefBayesPareto4(xVec,CountT,xi,w,xpars,tau,nt,st);
      end
      x = SysPDefBayesh(xVec,f);
      xpars(CountT) = x;
      xLower = xVec(1) - 0.2*(xVec(end) - xVec(1));
      xUpper = xVec(end) + 0.2*(xVec(end) - xVec(1));
      xAllLower(CountT) = xLower;
      xAllUpper(CountT) = xUpper;
      if Count > 0
        XPARS(Count,CountT) = x;
      end
    end

    % Pause and display
     if etime(clock,t0) > 30
       t0 = clock;
       disp(strcat(num2str(round(1000*(Count + NBurn)/(NReps + ...
           NBurn))/10),'%'))
         pause(1)
     end

end

%Plot a histogram of the marginal posterior density of the draws of the
%Probability of Default.
%hist(Draws(:,1));
Draws = [P XPARS];
```

List of References

Altman, E. (2006). Default recovery rates and lgd in credit risk modeling and practice: an updated review of the literature and empirical evidence. *New York University, Stern School of Business*.

BCBS (2004). *International convergence of capital measurement and capital standards: A revised framework*. Bank for International Settlements.

BCBS (2005). Validation of low-default portfolios in the basel ii framework. *Basel Committee Newsletter No. 6 (September 2005)*.

BCBS (2011). *Basel III: A global regulatory framework for more resilient banks and banking systems*. Bank for International Settlements.

Benjamin, N., Cathcart, A. and Ryan, K. (2006). Low default portfolios: A proposal for conservative estimation of default probabilities. *Financial Services Authority*.

Berning, T.L. (2010). *Improved estimation procedures for a positive extreme value index*. Ph.D. thesis, Stellenbosch: University of Stellenbosch.

Bernstein, P.L. (1996). *Against the gods: The remarkable story of risk*. Wiley, New York.

Bickel, D.R. and Frühwirth, R. (2006). On a fast, robust estimator of the mode: Comparisons to other robust estimators with applications. *Computational Statistics & Data Analysis*, vol. 50, no. 12, pp. 3500–3530.

BIPRU (2011). *Prudential sourcebook for Banks, Building Societies and Investment Firms*. Prudential Regulatory Authority.

Blochwitz, S., Hohl, S., Tasche, D. and Wehn, C.S. (2004). Validating default probabilities on short time series. *Capital & Market Risk Insights, Federal Reserve Bank of Chicago*.

Bluhm, C. (2015). Credit risk modeling. [ONLINE] Available from: `http://www.christian-bluhm.net/Personal/Credit_Risk_Modeling.html`. [Accessed on 05/22/2015].

Bluhm, C., Overbeck, L. and Wagner, C. (2010). *An Introduction to Credit Risk Modeling*. Chapman and Hall/CRC, Florida.

Chang, Y.-P. and Yu, C.-T. (2014). Bayesian confidence intervals for probability of default and asset correlation of portfolio credit risk. *Computational Statistics*, vol. 29, no. 1-2, pp. 331–361.

Clifford, T., Sebesteyn, K. and Marianski, A. (2013). Low default portfolio modelling: Probability of default calibration conundrum. Credit Scoring and Credit Control XIII Conference.

Crouhy, M., Galai, D. and Mark, R. (2000). *Risk Management*, vol. 1. McGraw-Hill, New York.

Crouhy, M., Galai, D. and Mark, R. (2001). Prototype risk rating system. *Journal of banking & finance*, vol. 25, no. 1, pp. 47–95.

Dahlin, F. and Storkitt, S. (2014). Estimation of loss given default for low default portfolios. *Royal Institute of Technology*.

Dwyer, D.W. (2006). The distribution of defaults and bayesian model validation. *Moody's KMV*.

Erlenmaier, U. (2011). The shadow rating approach: Experience from banking practice. In: *The Basel II Risk Parameters*, pp. 37–74. Springer.

Fernandes, G. and Rocha, C.A. (2011). Low default modelling: A comparison of techniques based on a real brazilian corporate portfolio. *Credit Scoring and Credit Control XII*, pp. 24–26.

Forrest, A. (2005). Likelihood approaches to low default portfolios. *Dunfermline Building Society. Version*, vol. 1, no. 14, p. 9.

Frunza, M. (2013). Are default probability models relevant for low default portfolios? *Available at SSRN 2282675*.

Genest, B. and Brie, L. (2013). Basel ii irb risk weight functions. *Global Research and Analytics Dpt. of Chappuis Halder and Cie*.

Gupta, A.K. and Nadarajah, S. (2004). *Handbook of beta distribution and its applications*. CRC Press, USA.

Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pp. 1029–1054.

Hansen, L.P., Heaton, J. and Yaron, A. (1996). Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics*, vol. 14, no. 3, pp. 262–280.

Hull, J.C. (2006). Defining copulas. *Risk Magazine Ltd. London*, vol. 19, no. 10, pp. 62–64.

Hull, J.C. (2012). *Options, futures, and other derivatives*. Pearson/Prentice Hall, New Jersey.

Huynh, H.T., Soumare, I. *et al.* (2011). *Stochastic simulation and applications in finance with MATLAB programs*. John Wiley & Sons, West Sussex, England.

Jarvis, J. (2012). The crisis of credit visualized. [ONLINE] Available from: `http://crisisofcredit.com/`. [Accessed on 05/24/2015].

Jarvis, J. (2015). The european debt crisis visualized. [ONLINE] Available from: `http://jonathanjarvis.com/european-debt-crisis`. [Accessed on 05/24/2015].

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 186, pp. 453–461. The Royal Society.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

Kennedy, K., Mac Namee, B. and Delany, S.J. (2012). Using semi-supervised classifiers for credit scoring. *Journal of the Operational Research Society*, vol. 64, no. 4, pp. 513–529.

Kiefer, N.M. (2009). Default estimation for low-default portfolios. *Journal of Empirical Finance*, vol. 16, no. 1, pp. 164–173.

Kiefer, N.M. (2010). Default estimation and expert information. *Journal of Business & Economic Statistics*, vol. 28, no. 2, pp. 320–328.

Kiefer, N.M. (2011). Default estimation, correlated defaults, and expert information. *Journal of Applied Econometrics*, vol. 26, no. 2, pp. 173–192.

Kruger, M. (2015). Validation and monitoring of pd models for low default portfolios using proc mcmc. *SAS Global Forum 2015*.

McNeil, A.J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools: Concepts, Techniques and Tools*. Princeton university press, New Jersey.

Meissner, G. (2009). *Credit derivatives: application, pricing, and risk management*. John Wiley & Sons, Oxford.

Merton, R.C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, vol. 29, no. 2, pp. 449–470.

Moody's (2015). About moody's ratings: How to get rated. `https://www.moodys.com/Pages/amr002001.aspx`. (Visited on 06/15/2015).

Orth, W. (2011). Default probability estimation in small samples-with an application to sovereign bonds. *Discussion Papers in Statistics and Econometrics, University of Cologne*, , no. 5.

Pluto, K. and Tasche, D. (2010). Internal-ratings-based approach. *Encyclopedia of Quantitative Finance*.

Pluto, K. and Tasche, D. (2011). Estimating probabilities of default for low default portfolios. pp. 75–101.

Rice, J. (2007). *Mathematical statistics and data analysis*. Cengage Learning, Belmont.

Rizzo, M.L. (2007). *Statistical computing with R*. CRC Press, Florida.

Romano, C. (2002). Applying copula function to risk management. In: *Capitalia, Italy. [ONLINE] Available from:http://www.icer.it/workshop/Romano.pdf*.

Schuermann, T. (2004). What do we know about loss given default? *Wharton Financial Institutions Center Working Paper*.

Standard and Poor's (2014). 2013 annual global corporate default study and rating transitions.

Tasche, D. (2013). Bayesian estimation of probabilities of default for low default portfolios. *Journal of Risk Management in Financial Institutions*, vol. 6, no. 3, pp. 302–326.

Van Der Burgt, M.J. (2008). Calibrating low-default portfolios, using the cumulative accuracy profile. *Journal of Risk Model Validation*, vol. 1, no. 4, pp. 17–33.

Van Gestel, T. and Baesens, B. (2009). *Credit risk management: basic concepts: financial risk components, rating analysis, models, economic and regulatory capital*. Oxford University Press, Oxford.

Vasicek, O. (2002). The distribution of loan portfolio value. *Risk*, vol. 15, no. 12, pp. 160–162.

Wackerly, D., Mendenhall, W. and Scheaffer, R. (2007). *Mathematical statistics with applications*. Cengage Learning, Belmont.

Wilde, T. and Jackson, L. (2006). Credit portfolio risk low-default portfolios without simulation. *Risk Magazine, London*, vol. 19, no. 8, p. 60.