Modelling Low Default Portfolios

Joran Zoutendijk

Technische Universiteit Delft

TU Delft Delft University of Technology

Challenge the future

Modelling
Low Default Portfolios

by

Joran Zoutendijk

in partial fulfillment of the requirements for the degree of

Master of Science
in Applied Mathematics

at the Delft University of Technology,
to be defended publicly on Tuesday July 31, 2018 at 13:00 AM.

Supervisor:	Dr. P. Cirillo,	TU Delft
Thesis committee:	Prof. dr. H. P. Lopuhaä,	TU Delft
	Dr. W. den Dunnen,	EY

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Preface

This master thesis contains the research done for the final part of the Applied Mathematics programme at the TU Delft. Three years ago I chose this master with the specialisation in Financial Engineering and have not regretted this choice a single moment. I have been able to combine a solid foundation in mathematics with an understanding of the intriguing world of finance. I have followed all the courses within the programme with great pleasure, in which the passionate and helpful teachers at the TU Delft played a vital role.

The topic of low default portfolios proved to be a good suggestion by my supervisor. The research allowed for a pleasant combination of mathematics, programming, and the writing of the thesis itself. I can look back upon the last eight months with great satisfaction and am pleased with the end result.

I want to acknowledge the support received from both of my supervisors, Pasquale Cirillo of the TU Delft and Wilco den Dunnen of EY. Their suggestions and comments have guided me during the progress and for this I want to show my gratitude. Also, I want to thank EY for the possibility of the thesis internship, the professional environment and helpful colleagues encouraged me to stay motivated throughout my thesis.

Joran Zoutendijk
Amsterdam, July 2018

Contents

1

Credit Risk and Low Default Portfolios

Banks fulfil an essential role in today's society. They safeguard the deposits of individuals and companies, and grant loans to fund mortgages, investments and other expenses. These activities are not without risk, which is why risk management is a vital aspect of operating a bank.

Recall for example the financial crisis of 2007 and 2008. As Holt discusses in [1], one of the main causes was the relaxed standards for mortgage loans. Many loans were granted by banks to individuals without good enough solvency, and when many could not repay their debt the banks became in trouble. This even caused one of the largest banks at that time, Lehman Brothers, to collapse in 2008.

Considering the impact banks have on the global economy, they are under the supervision of a regulator, who monitors the banks' ability to manage their risk in an adequate manner. All banks are under supervision of their 'national' regulator, for European banks this is the European Central Bank (ECB). In order to ensure worldwide financial security and fair competition national banks organised themselves in the 'Basel Committee on Banking Supervision' (BCBS). This committee creates the regulation framework all banks in the world have to follow. Since they have no official jurisdiction the official laws have to created by all the national regulators themselves. Even though national regulators regularly disagree with each other, in practice the framework of the BCBS is always implemented by them.

First, a more thorough description of credit risk will be given and thereafter the issues and regulation regarding low default portfolios.

1.1. Credit Risk

The BCBS distinguishes in the document [2] known as 'Basel II' between three main types of risk:

- **Credit risk.** *The potential that a bank borrower or counterparty will fail to meet its obligations in accordance with agreed terms.*
- **Operational risk:** *The risk of loss resulting from inadequate or failed internal processes, people and systems or from external events.*
- **Market risk:** *The risk of losses in on and off-balance-sheet positions arising from movements in market prices.*

In the financial industry the expression 'credit risk' is commonly used by practitioners. As a mathematician, one would generally assume that when a term is widely used, its definition is agreed upon by its users. In practice however, the lack of precise formulation may be cause for confusion. Normally this will mostly be a non-issue, but in the development of credit risk models it is important to precisely define which *subrisks* are included and which are not. Subrisks of credit risk include (but are not limited to) [3]:

- **Default risk.** This is usually what is meant by credit risk by laymen. Default risk is the risk of not getting your money back as a lender in the case the borrower defaults. This is why estimation of the probability of default of obligors plays a central role in credit risk management.
- **Recovery risk.** Even if a company defaults it is no guarantee that the debtholders lose all their capital. In case of default the assets of a company are liquidated and used to repay the outstanding debt. The risk that an obligor repays less of its debt than previously expected is known as recovery risk.
- **Spread risk.** The credit spread between two loans is the difference between their interest rates. If the credit market becomes more risk averse, a higher premium is required to lend capital which translates to a rise in interest rates. When interest rates increase the price of their underlying bonds go down, which causes a loss for holders of these bonds without the credit quality of the corresponding company decreasing.
- **Downgrade risk.** Most large companies have their ability to repay their debts assessed by credit rating agencies (eg. Moody's, Standard & Poor's, and Fitch). If the credit rating of a company deteriorates their risk weight increases and the lender must hold more regulatory capital.
- **Concentration risk.** When a bank lends a million dollar to one company, this is considered riskier than lending half a million to two companies. This phenomenon is known as concentration risk and urges banks to diversify their portfolio.

Credit risk can be further decomposed into more subrisks, but this is out of scope for this thesis. Also, these subrisks are not independent, but can influence each other. The focus of this research is the estimation of the probability of default when it is low, so default risk is the most relevant type of of subrisks. That is of course not to say that other types are less important.

1.1.1. Risk components and Approaches

Broadly spoken a bank can distinguish between two types of losses, ones that it can expect and ones that it can not. The first type, the expected loss (EL), are the losses incurred under normal business circumstances. These costs are covered on an ongoing basis by provisions the bank holds. Expected losses are viewed as a cost of doing business. However, losses are not deterministic but vary over time. Sometimes they are lower than expected and sometimes they are higher. In that case they are known as an unexpected loss (UL), exceeding the expected amount a bank might lose.

Expected Loss

How much a bank is expected to lose on a loan depends on three factors: The size of the loan, the probability that the borrower is not able to repay the debt, and the amount the bank recovers even if the borrower defaults. Suppose the maturity of a loan is fixed at one year. In credit risk modelling the expected loss is written as:

$$\mathrm{EL} = \mathrm{EAD} \times \mathrm{PD} \times \mathrm{LGD}. \tag{1.1}$$

Where the acronyms stand for:

- **Exposure At Default (EAD).** The size of the loan at the time of default. Sometimes, as it is in the case of bonds, this is a known quantity. Often times it is however a random quantity, as in the case of credit cards for example. The EAD is shown to increase with the PD since the need for funds increases in bad times. This is called 'wrong way risk' and needs to be taken into account in the development of models.
- **Probability of Default (PD):** The probability that the counterparty defaults. The estimation of this quantity is the focus of this thesis. It is arguably the most important of the risk components as it is the only parameter banks may estimate under one of the internal rating based approaches discusses later.
- **Loss Given Default (LGD):** Given in a percentage. Often times obligors are not able to repay their whole debts, but still a part of it. After liquidating the assets of a counterparty the bank may recover a part of the loan. It is equivalent to the Recovery Rate (RR) through $\mathrm{LGD} = 1 - \mathrm{RR}$. The LGD may also be correlated with the PD as shown in [4].

Unexpected Loss

If losses are always as expected, banks would have no trouble managing their credit risk. More interesting to consider is the unexpected loss, how much capital should a bank hold in order to protect itself against unforeseen losses? One extreme is a bank holding so much capital that it is protected against all its obligors defaulting. This is unreasonable of course, as it is too expensive to hold so much capital that otherwise could have been invested. The other extreme is a bank holding no capital to protect against unexpected losses at all. A very dangerous practice as it is a known fact these losses can occur, causing much stress on the solvency of the bank when they do. It is therefore important for banks and their supervisors to find a balance between these two extremes. Also note that this can be a cause of friction between these two parties, as banks wish to hold as little capital as possible while supervisors prefer them to be more on the safe side.

The BCBS has determined that banks should hold a certain amount of Capital Requirements (CR) to mitigate unexpected losses. The CR is very simply computed as a percentage of the Risk Weighted Assets (RWA). Normally this is 8%, but it can be slightly adjusted by the regulator to make it more or less conservative. The real objective is therefore to determine the RWA of a portfolio. There are two main approaches, the Standardised (STA) and Internal Rating Based (IRB) approach. Within the last one two subapproaches can be identified.

Risk weight table for PSEs

Option 1: Based on external rating of sovereign

External rating of the sovereign	AAA to AA–	A+ to A–	BBB+ to BBB–	BB+ to B–	Below B–	Unrated
Risk weight under Option 1	20%	50%	100%	100%	150%	100%

Figure 1.1: Risk weight table of exposure to non-central government public sector entities [5]

Standardised Approach

The STA approach is as its name suggests the most simple one. In this approach the RWA is determined as the weighted sum of items on and off the balance sheet. The banks may use external ratings to determine in which weight should be applied to the loan. The weights themselves are provided by the regulator, see figure 1.1 for an example. Historically, the STA approach was used by smaller banks who did not have the resources to create (and validate) their own models. Larger banks prefer to use the IRB approaches as the Capital Requirements are usually lower under these.

Internal Rating Based Approaches

The second method to compute the RWA of a portfolio of obligors is the IRB approach. The main underlying formula for this approach will be derived first. Two subapproaches, foundation (F-IRB) and advanced (A-IRB) will be discussed thereafter. An important concept in the IRB approach is the risk measure known as the Value-at-Risk (VaR):

Definition 1 - Value-at-Risk.

For a given confidence level $\alpha \in (0,1)$ the Value-at-Risk (VaR) is that value for which the probability of the loss exceeding it is $1-\alpha$. If $F(x)$ denotes the loss distribution, the VaR is defined as:

$$\mathrm{VaR}(\alpha) = \inf\{x : F(x) \geq \alpha\}$$

The Basel Committee decided that the capital banks hold must cover the 99.9% VaR of a certain loss distribution. This means that only in the worst 0.1% of scenarios losses occur for which banks are not prepared. For a graphical representation of the relation between the EL, UL and VaR, see figure 1.2.

The VaR depends on the chosen loss distribution. The BCBS used a model known in the literature under a few different names, but in this thesis referred to as the 'Vasicek model', which is discussed in

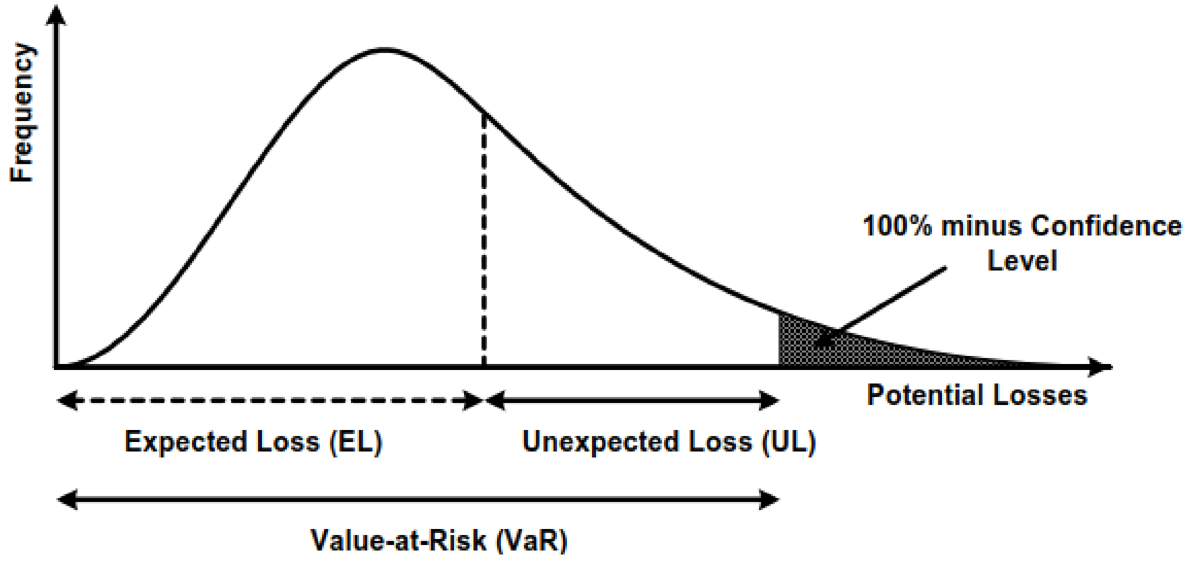

Figure 1.2: Structure of the loss distribution [6]

detail in chapter 3. For now it is enough to know that for a given individual PD estimate p, correlation value ρ, and factor Y ('state of the economy'), an unconditional PD is computed:

$$\mathrm{PD} = \Phi\left(\frac{\Phi^{-1}(p) + \sqrt{\rho}\Phi^{-1}(Y)}{\sqrt{1-\rho}}\right).$$

Recall the definition of the expected loss in equation 1.1. The unexpected loss is computed in the same manner, but by plugging a quantity known as the 'Worst Case Default Rate' (WCDR) for the PD. Coming back to the 99.9% VaR, the WCDR is determined by setting the random factor Y to 0.999:

$$\mathrm{WCDR} = \Phi\left(\frac{\Phi^{-1}(p) + \sqrt{\rho}\Phi^{-1}(0.999)}{\sqrt{1-\rho}}\right).$$

Now consider a portfolio of n obligors, the 99.9% VaR of the portfolio is given by

$$\mathrm{VaR}_{0.999} = \sum_{i=1}^{n} \mathrm{EAD}_i \times \mathrm{WCDR}_i \times \mathrm{LGD}_i.$$

As seen in figure 1.2, the unexpected loss is the difference between the VaR and the EL. This is also the capital a bank is required to have to protect itself against unexpected losses. The CR is therefore given by

$$\mathrm{CR} = \sum_{i=1}^{n} \mathrm{EAD}_i \times \left(\Phi\left(\frac{\Phi^{-1}(\mathrm{PD}_i) + \sqrt{\rho}\Phi^{-1}(0.999)}{\sqrt{1-\rho}}\right) - \mathrm{PD_i}\right) \times \mathrm{LGD}_i. \tag{1.2}$$

A key assumption not mentioned yet is that the portfolio is assumed to be granular, i.e. large and well diversified. Also note that the only random component Y is assumed to be normal, so that there may be an underestimation of the tail risk. Furthermore, the scaling component and so-called maturity adjustment have been omitted. The RWA can be computed by $\mathrm{RWA} = 12.5 \times \mathrm{CR}$, since 12.5 is the reciprocal of 8%. Please refer to the Basel documentation described in the next section for more information regarding the IRB formula.

Equation 1.2 shows that in the IRB approach three quantities are used in the calculation of the Capital Requirements, the EAD, PD and LGD. The correlation coefficient ρ is also needed, but given by a deterministic function of the PD. The correlation coefficient can attain values between 0.12 and 0.24. Next, the two internal rating based approaches are discussed.

Foundation Internal Rating Based approach (F-IRB)

The Foundation Internal Rating Based Approach (F-IRB) allows banks to use their own internal models to estimate the PD. Together with the EAD and LGD which are supplied by the regulator, the Capital Requirements are then computed through the IRB formula. The CR under the standardised approach are often higher than those under the F-IRB. This means there is an incentive for banks to use the latter, because less Capital Requirements means that there is more capital available to lend and receive interest upon. Another advantage of the F-IRB approach with respect to the standardised, is that it causes banks to be more risk sensitive, leading to better credit risk management.

Advanced Internal Rating Based approach (A-IRB)

In the A-IRB, banks are not only allowed to use internal models for the estimation of the PD, but also for the EAD and LGD. The capital requirements are generally even further reduced under the A-IRB. However, there is also a lot more effort required in the development and validation of the models. Because of stricter regulation the A-IRB has been used less often in the past few years.

Not all banks are allowed to use internal models. A lot of requirements need to be met before the regulator permits the use of them and banks need to be able to prove on a regular basis that their models are sufficiently calibrated and conservative. For more details on this regulation again please refer to the Basel documentation summarized in the next section.

1.2. Low default portfolios

First a general introduction to low default portfolios will be given, thereafter a brief history of the Basel regulation regarding low default portfolios will be presented.

1.2.1. Introduction to low default portfolios

As the name suggests, a low default portfolio is a portfolio of obligors where the observed amount of defaults is low. In practice this means that not only is the probability of default of the obligors small, but the portfolio itself as well. A large mortgage portfolio of possibly millions of obligors with 'only' thousand defaults is therefore no LDP, but a portfolio of 100 large corporates with 1 observed default is.

There are several problems in the modelling of low default portfolio, but they are all caused by the lack of data. It is possible that no defaults at all are observed in a portfolio, but the true probability of default is always larger than zero, as many examples have proven. And even when some defaults are observed, there is a high chance of underestimating the probability of default. Another problem is the volatility of the observed defaults. Many years can pass without observing any defaults, but that does not mean there will be no more financial crises.

These issues are very relevant for banks, as a significant amount of their exposure comes from low default portfolios. On the one hand, they want to hold enough capital to handle a certain amount of risk. On the other hand, they do not want to hold too much capital as this cuts into their profits. The BCBS creates regulation to ensure the banks hold enough capital, the most important documents related to low default portfolios are discussed in the next section.

1.2.2. Regulation

There have been four important documents released by the BCBS:

- (June 2004), *International Convergence of Capital Measurement and Capital Standards [2]*, also known as Basel II
- (September 2005), *Validation of low-default portfolios in the Basel II Framework [7]*

- (July 2013), *Analysis of risk-weighted assets for credit risk in the banking book [8]*
- (December 2017), *Basel III: Finalising post-crisis reforms [5]*, also known as Basel IV.

Note that the exclusion of the document released in 2011, known as Basel III [9], is done on purpose since it contains no significant changes to the regulation regarding low default portfolios.

Basel II

In June 2004 the BCBS released the accords known as Basel II. It is noteworthy that the term 'low default portfolio' is not mentioned once explicitly in the documentation. However, the following passage from article 462 says:

'Where only limited data are available, or where underwriting standards or rating systems have changed, the bank must add a greater margin of conservatism in its estimate of PD.'

Articles 416 and 527(b) mention similar sentiments. Also, in article 285 the BCBS specifies a floor on the PD of 0.03% for corporate and bank exposures (not for sovereigns).

Validation of low-default portfolios in the Basel II Framework

The financial industry, encountering problems in the validation of low default portfolio models, asked the Basel Committee for directions as it was worried that these portfolios might be excluded from the IRB approach. In 2005 the BCBS published this report on how the industry should deal with low default portfolios.

In the document it is noted that there is no strict definition of a low default portfolio, but that they are a continuum between two extremes. They are not excluded from the (Foundation or Advanced) IRB approach, but the BCBS notes that *'relatively sparse data might require increased reliance on alternative data sources and data-enhancing tools for quantification and alternative techniques for validation'*.

Some of the actions banks could take to mitigate the relative lack of data proposed in the article are the following:

- Pooling data with other market participants. Even though a portfolio may have a small number of defaults for one bank, if portfolios of similar exposures are combined the resulting portfolio may be fit for regular estimation methods.
- Combining different rating categories. If rating categories such as triple, double, and single A are merged into a single rating grade, the resulting portfolio could have a more reasonable number of defaults.
- Using the upper bound of the PD estimate. If the PD estimate itself is deemed as too unreliable, an upper bound could be used in the calculation of capital. In chapter 3 a method to derive such upper bounds is described.

Analysis of risk-weighted assets for credit risk in the banking book

In 2013 the BCBS conducted a Regulatory Consistency Assessment Programme (RCAP), doing an analysis of risk-weighted assets for credit risk in the banking book. They found that there is a substantial variation between banks in RWAs for credit risk. Three quarters of this variation can be explained by the risk profile of the banks, while one quarter comes from different practices by banks and regulators.

The Basel Committee asked 32 large banks across the world to estimate risk parameters of low default portfolios specifically and concluded:

'The low-default nature of the assessed portfolios, and the consequent lack of appropriate data for risk parameter estimation, may be one of the factors leading to differences across banks, especially for banks' estimates of LGDs in the sovereign and bank asset classes.'

While the Committee notes the largest amount of variation comes from LGD estimates, there are also significant differences among PD estimates. Some examples of practices that vary between banks, which influence the difference among the PD estimates, are the combining of data sources, the length of the data series used, and (judgemental) adjustments to the PD.

In the article the BCBS presents some suggestions to reduce the variation of credit risk RWA, these include:

- **Enhanced disclosure by banks.** Banks could disclose more, and more granular, information to the regulator. This, in combination with the use of more standardised definitions would allow banks and their exposures to be more comparable and could better explain some of the RWA variation.
- **Additional guidance on aspects of the Basel framework.** The Committee notes that it could provide more guidance on for example the use of external data ('particularly for low-default portfolios').
- **Harmonisation of national implementation requirements.** The BCBS only provides a framework of regulation, which the national regulators then implement. Some of the RWA variability stems from differences in this implementation and could be reduced if regulation would be more harmonised in the various jurisdictions.
- **Constraints on IRB parameter estimates.** The Committee notes that reducing the flexibility of the IRB approaches may reduce variability in RWA calculation among banks. Regarding LDPs, it is even noted that 'creation of such benchmarks could fill a valuable niche, especially for low-default portfolios, creating reference points for supervisors and banks'.

Basel III: Finalising post-crisis reforms

Following the RCAP in 2013, the key objective of the revisions done to Basel III in december 2017 is to reduce excessive variability of risk-weighted assets. These revisions will be part of regulation as of January 1 2022.

Articles 184 and 230 suggest not much may have changed in the guidance the BCBS is willing to provide to the industry. In article 184 it is written:

'Given the difficulties in forecasting future events and the influence they will have on a particular borrower's financial condition, a bank must take a conservative view of projected information. Furthermore, where limited data are available, a bank must adopt a conservative bias to its analysis.'

However, some concrete adjustments regarding regulation on low default portfolios have been made as well. One of the more drastic changes is the output floor on RWA. The RWA computed using the IRB Approach cannot be lower than 72.5% of the RWA according to the Standardised approach. For low default portfolios an important revision is the raise of the input floor on the PD (from 0.03% to 0.05%). This holds for exposures to banks and corporates, while the treatment of sovereign exposures is unchanged from Basel II.

Future regulation

The landscape of regulation regarding credit risk is ever-changing. As it should be, because banks and risks themselves are continually developing as well. It is interesting what future events may cause drastic (or mild) changes to the modelling of credit risk and low default portfolios in particular.

One very interesting technique, which forms the core of this thesis, is the Bayesian approach. It is the authors opinion that additional guidance by the regulator regarding this approach would seem a sensible next development. The BCBS could for example provide a list of priors which it deems reasonable. Instead of imposing bounds on PD estimates themselves, the (hyper)parameters of the prior could also be limited. Further discussion of Bayesian techniques are found in chapter 2.

2

Statistical tools

The research and analysis done in this thesis rely on various statistical tools. While a certain mathematical foundation is expected of the reader, some of the more advanced techniques are discussed in this chapter. First, the Bayesian approach to probability is described. Thereafter a summary of dependence modelling through the use of copulas is given. Then a short introduction to goodness of fit testing, including some results, is presented. Finally confidence intervals and their different interpretations are discussed.

2.1. Bayesian Statistics

Many of the results of this thesis are based on the use of Bayesian techniques. This section is dedicated to providing a concise theoretical background on the use of Bayesian statistics. First a short introduction to the different interpretations in probability is presented. From there the Bayesian approach is explained and the relationship between Bayesian point estimators and decision theory is investigated. The final section is about computational techniques needed later on in this thesis.

2.1.1. Interpretations of probability

In later chapters performance of the maximum likelihood estimator and Bayesian estimators is compared. There are also instances where confidence and credible intervals are used in the same figure. These methods do not only differ in results, but also in underlying interpretation of probability. While the notion of probability is used freely in this thesis, the reader should be beware that there are fundamental differences in the approaches. Following [10], three philosophical visions on the concept of probability are presented:

1. **Classical Probability:** 'Assigning probabilities in the absence of any evidence.' This school views probability as the fraction of the number of 'success' events to the total number of events. One of the main advocates of this approach was Laplace. Large disadvantages are the need for finiteness and equiprobability.

2. **Frequentism:** 'The probability of an event is the relative frequency over time'. The approach on which methods as maximum likelihood estimation and confidence intervals are based. Famous frequentists are Venn and von Mises. A flaw of this view on probability is the so-called *reference class problem*.

3. **Subjective Probability:** 'Probability is a degree of belief'. Subjectivists like de Finetti think of probability as the fair price of a bet on the subject of interest. The fact that ones assignments of probability may not be coherent is a drawback of this perspective.

While the underlying philosophical interpretation of probability might be very different between these approaches, they can actually co-exist in practice. Later in this thesis it will be shown that in certain situations the frequentist maximum likelihood method fails and a subjective Bayesian approach offers a solution.

2.1.2. The Bayesian approach

The unpublished work of Thomas Bayes titled 'An Essay towards solving a Problem in the Doctrine of Chances' [11] was recognized post-mortem by Richard Price. In the manuscript Bayes proposed a method of updating beliefs in the face of new evidence. This method, now known as Bayesian statistics, finds its origin in the Bayes Theorem. Although discovered and named after Bayes, it was in fact Laplace who first used the formal representation of the theorem.

Theorem 1 (Bayes Theorem)**.** *Consider a sample space Ω with probability measure P, and let A and B be two events in Ω such that $P(B) \neq 0$, then:*

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}.$$

Proof. The proof easily follows from the definition of conditional probability. □

Although it is common in the literature to illustrate this theorem with an example, for the sake of brevity this custom is not followed. An example can be found in [12]. This example shows the *likelihood* is often not that helpful; why bother with the probability of observing the data when they are already known? That is where the Bayesian notion of *prior* information comes in. These prior beliefs can for example come from historical evidence or expert judgement. Using Bayes theorem these beliefs can then be updated in combination with the observed data, resulting in *posterior* beliefs.

Using the same line of reasoning as above, the move to doing inference on a parameter is made with ease. Consider a likelihood function $f(x \mid \theta)$, where x is the data and θ the parameter (vector) of interest. Before observing the data, a prior $\pi(\theta)$ is determined to represent the prior beliefs about the parameter θ. Once the experiment is done and the data is collected, Bayes theorem can be used to update the prior distribution using the likelihood function to obtain the posterior distribution. Inference is then done on the posterior distribution of θ, given by:

$$\begin{aligned}\pi(\theta \mid x) &= \frac{\pi(\theta)f(x \mid \theta)}{f(x)} \\ &= \frac{\pi(\theta)f(x \mid \theta)}{\int_\Theta \pi(\theta)f(x \mid \theta)d\theta} \\ &\propto \pi(\theta)f(x \mid \theta),\end{aligned}$$

and when written in words:

$$\text{posterior} \propto \text{prior} \times \text{likelihood}.$$

2.1.3. The prior distribution

The prior $\pi(\theta)$ can in principle be chosen as any distribution. In fact, the prior does not even have to be a proper distribution (i.e. integrating to 1). However, in order to do inference on the posterior this does have to be a proper distribution and therefore the normalizing factor has to be finite:

$$\int_\Theta \pi(\theta)f(x \mid \theta)d\theta < \infty.$$

Example 1. Consider a Poisson likelihood where for the rate parameter λ a $Gamma(\alpha, \beta)$ prior has been chosen. The posterior is then given by:

$$\begin{aligned}\pi(\lambda \mid k, \alpha, \beta) &\propto \frac{e^{-\lambda}\lambda^k}{k!}\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\lambda\beta} \\ &= \frac{e^{-\lambda(\beta+1)}\lambda^{(k+\alpha-1)}\beta^\alpha}{k!\Gamma(\alpha)} \\ &\propto e^{-\lambda(\beta+1)}\lambda^{(k+\alpha-1)},\end{aligned}$$

which is the probability density function of a $Gamma(k + \alpha, \beta + 1)$ distribution. When the prior and the posterior distribution are in the same family of distributions, the prior is called a *conjugate prior* for that posterior. Conjugate priors are used frequently in practice as they omit the need for the numerical integration of the normalization constant.

Up to now the prior distribution has been denoted as $\pi(\theta)$, but it is certainly possible to model the prior beliefs by a distribution depending on other parameters. These parameters are then called the *hyperparameters*. If they are written as (the vector) τ the posterior simply changes to

$$\pi(\theta \mid x, \tau) \propto \pi(\theta \mid \tau) f(x \mid \theta).$$

The hyperparamaters can be chosen so the prior distribution fits certain beliefs. It is for example possible to conduct a survey under experts and from that estimate the hyperparameters. Another possibility is to use a *hyperprior*, specifying ones prior belief in the hyperparameters. This type of modelling is called *hierarchical modelling*. Denote $\tilde{\pi}(\tau)$ as the hyperprior, the posterior distribution is then obtained by integrating out the hyperparameter of the joint posterior:

$$\pi(\theta \mid x) = \frac{\int_T f(x \mid \theta)\pi(\theta \mid \tau)\tilde{\pi}(\tau)d\tau}{\int_\Theta \int_T f(x \mid \theta)\pi(\theta \mid \tau)\tilde{\pi}(\tau)d\tau d\theta}.$$

Young and Smith note in [13] this is a very effective tool, usually resulting in answers robust to model misspecification. Observe that in the above notation it is assumed the hyperprior does not depend on any parameters. Analogous to the 'normal' prior, it is possible to model the hyperprior with a distribution depending on parameters. Then these parameters have to be estimated, or again have a (hyperhyper)prior of their own.

It is important to note that within the Bayesian school two, sometimes conflicting, approaches exist: the subjective and objective Bayesian. The first is based on the Subjective view on probability, the second on the Classical view.

Subjective Bayes

This school of Bayesians views probability as the degree of personal belief. The subjective Bayesian approach is very useful for incorporating expert judgement into statistical models. Especially in the case when there is little data available this can be extremely valuable. The prior beliefs of the experts can be translated to a distribution by means of a survey.

Another advantage of the subjective over the objective approach is it does not violate the *likelihood principle*. This states that information that comes from the data is contained in the likelihood function. Since objective Bayesian priors may depend on the form of the likelihood or the data itself, it violates this principle.

Objective Bayes

The objective Bayesian approach finds its origins at Bayes and Laplace themselves. Bayes introduced a uniform prior on the probability parameter p in a Binomial model, which Laplace justified by the 'principle of insufficient reason'. This argument claims that a priori there is no reason to prefer one value of the parameter over the other, resulting in a uniform prior over the parameter space. Objective Bayesians adopted and extended this approach, and are in favor of *non-informative priors*. These priors are chosen with respect to some sense of objectivity in order to reduce the influence the personal beliefs of the researcher have on the posterior distribution.

Conjugate priors Conjugate priors are arguably the most used priors in the (financial) industry. As the size of data sets and complexity of models continues to grow it is often not computationally feasible to employ non conjugate priors. Since these priors depend on the likelihood function and not on the beliefs of someone it can be argued the conjugate prior has some sense of objectivity.

Uniform priors As previously mentioned, an example of a prior employed by objective Bayesians is the uniform one. While there is in principle nothing wrong with using a uniform prior, it is actually wrong to call it objective or non-informative as it represents the *subjective* prior belief that all values of the parameter are equiprobable. A better example of an objective prior is Jeffreys prior.

Jeffreys prior In [14] Jeffreys introduced what is now known as 'Jeffreys prior'. It is a prior distribution shown to be invariant under reparametrization and therefore a good candidate for an objective prior. The distribution depends on the Fisher Information Matrix.

Definition 2 - Fisher Information Matrix [15].

Suppose X is a (continuous) random variable with parameter vector $\theta = \theta_1, \dots, \theta_d$, given by probability density function $f(X \mid \theta)$. The Fisher Information $I(\theta)$ is then given by

$$I_{i,j}(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta_i}\log f(k \mid \theta)\right)\left(\frac{\partial}{\partial\theta_j}\log f(k \mid \theta)\right)\right],$$

and under appropriate conditions:

$$= -\mathbb{E}\left[\frac{\partial^2}{\partial\theta_i\theta_j}\log f(X \mid \theta)\right].$$

Now for a given likelihood function, Jeffreys prior is defined as the square root of the determinant of the Fisher Information matrix:

$$\pi(\theta) = \sqrt{\det I_{i,j}(\theta)}.$$

Note that a large difference with subjective priors is that Jeffreys prior cannot be determined without knowing the likelihood first. This is is a violation of the likelihood principle which says that all information of the data should be contained in the likelihood function. In the literature it is often claimed that Jeffreys prior is data dependent, but this is not entirely true. The prior is namely the same no matter what data is observed. So even though the likelihood is needed and therefore knowledge about the distribution of data is required, Jeffreys prior can be determined before observing the actual data. For further discussion on the subject of Jeffreys and data-dependent priors, a reference to the dissertation [16] of Darnieder is made. This discussion is concluded with noting that he agrees 'Jeffreys prior is not explicitly data-dependent as the methods of empirical Bayes are'.

2.1.4. Decision Theory

In order to pave the way for statistical inference using a Bayesian approach, the concept of decision theory will be introduced first. This framework allows for both a frequentist and a Bayesian approach. It will be shown that both these approaches rely on decision rules dependent on *risk functions* in order to do statistical inference. The difference between both approaches is how this risk function is defined and what properties it should satisfy.

By introducing Bayesian inference on the basis of decision theory all concepts will be developed in a formal manner. The approach of Young and Smith in chapters 2 and 3 in [13] is followed. Starting from the formulation of the components used, the notions of decision rules, risk functions and the Bayes principle will be explained. For the sake of clarity, all notation will be given for the continuous case. The discrete case is analogous by replacing integrals by summations and probability density functions by probability mass functions.

Formulation of components

In this section a formal description of a (statistical) decision problem is given. A decision problem is characterised by the following components:

1. **A parameter space** Θ**,** representing the space of parameter(s) on which inference is done. The probability parameter p of a Binomial distribution can attain all real values in the unit interval, so $\Theta = [0, 1]$

2. **A sample space** $\mathcal{X}$**,** the set of values the data can take on. The support of the Binomial distribution are all non-negative integers up to (a given) n, so $\mathcal{X} = \{0, 1, \dots, n\}$. In this case an observation $x = (x_1, x_2, \dots, x_N)$ is made, where $x \in \mathbb{N}^N$

3. **A probability distribution** $\{F(x), x \in \mathcal{X}, \theta \in \Theta\}$ The assumed (parametric) distribution of the data. In the example the Binomial distribution.
4. **An action space** $\mathcal{A}$**,** consisting of all possible actions the conductor of the experiment can take. In this case choosing an estimate $\hat{p}$, so $\mathcal{A} = \Theta = [0, 1]$
5. **A loss function** $L : \Theta \times \mathcal{A} \to \mathbb{R}$**,** describing the loss $L(\theta, a)$ that occurs when choosing an action $a \in \mathcal{A}$ when the true state of nature is $\theta \in \Theta$. Possible loss functions are squared error loss, absolute error loss and zero-one loss:
   - $L_1(\theta, a) = (\theta - a)^2$
   - $L_2(\theta, a) = |\theta - a|$
   - $L_3(\theta, a) = 1 - \delta(\theta - a)$

   Where $\delta(x)$ denotes the Dirac delta function [17]. Later in this section the relation between these loss functions and point estimators is studied.
6. **A set** $\mathcal{D}$ **of decision rules,** mapping an observation $x \in \mathcal{X}$ to an action $a \in \mathcal{A}$. Thus, $\mathcal{D}$ consists of functions $d : \mathcal{X} \to \mathcal{A}$. In the parameter estimation example, a decision rule could be:

$$d(x) = d(x_1, x_2, \dots, x_N) = \frac{1}{N \cdot n} \sum_{i=1}^{n} x_i.$$

The core principle of decision theory is that decision rules should be compared according to their risk function $R(\theta, d)$, where:

$$\begin{aligned} R(\theta, d) &= \mathbb{E}\left[L(\theta, d(X))\right] \\ &= \int_{\mathcal{X}} L(\theta, d(X)) f(x \mid \theta) dx. \end{aligned}$$

Ideally, a decision rule $d \in \mathcal{D}$ is used so that the risk function $R(\theta, d)$ is the smallest for all $\theta \in \Theta$. In practice however, this is almost never possible.

Bayes Risk

It is not surprising that, in addition to the six components used in the frequentist approach, the Bayesian setting requires the definition of another element:

7. **A prior** $\pi(\theta)$ describing the prior beliefs about the parameter θ.

It is now possible to define the Bayes risk of a decision rule d and a prior $\pi(\theta)$:

$$r(\pi, d) = \int_{\Theta} R(\theta, d) \pi(\theta) d\theta.$$

A Bayes decision rule is a decision rule $\hat{d}$ that minimizes the Bayes risk over all $d \in \mathcal{D}$ for a given prior $\pi(\cdot)$:

$$\hat{d} = \arg\min_{d \in \mathcal{D}} r(\pi, d).$$

The *Bayes principle* prescribes the use of a Bayes decision rule. In order to find such a rule, it helps to rewrite the Bayes risk as follows:

$$\begin{aligned} r(\pi, d) &= \int_{\Theta} R(\theta, d) \pi(\theta) d\theta \\ &= \int_{\Theta} \int_{\mathcal{X}} L(\theta, d(x)) f(x \mid \theta) dx \; \pi(\theta) d\theta \\ &= \int_{\Theta} \int_{\mathcal{X}} L(\theta, d(x)) f(x \mid \theta) \pi(\theta) dx \; d\theta, \end{aligned}$$

by definition of conditional probabilities:

$$= \int_{\Theta} \int_{\chi} L(\theta, d(x)) f(x) \pi(\theta \mid x) dx \, d\theta,$$

because the integrand is non-negative:

$$= \int_{\chi} \int_{\Theta} L(\theta, d(x)) f(x) \pi(\theta \mid x) d\theta \, dx$$

$$= \int_{\chi} f(x) \left[\int_{\Theta} L(\theta, d(x)) \pi(\theta \mid x) d\theta \right] dx.$$

The integral in brackets $\int_{\Theta} L(\theta, d(x))\pi(\theta \mid x)d\theta$ is the weighted average of the loss function with respect to the posterior distribution, called the *expected posterior loss*. Notice that in order to find the Bayes decision rule $\hat{d}$ it is sufficient to find such a rule that minimizes the expected posterior loss for a given x. Young and Smith note this illustrates an intuitively natural attribute of the Bayesian procedure; based on a certain observed x it is only necessary to take into account the losses that occur from the decision $d(x)$. It is not necessary to think about the losses that *might have* taken place, but did not.

Posterior point estimators

In this section the three loss functions L_1, L_2 and L_3 described earlier are shown to correspond to three well known point estimators. Assume the observed data is $X = x$, for each loss function the Bayes rule $\hat{d} = \hat{d}(x)$ minimizes

$$r(d, \pi) = \int_{\Theta} L(\theta, d(x)) \pi(\theta \mid x) d\theta.$$

1. **Squared error loss: posterior mean.** $L_1(\theta, d) = (\theta - d)^2$.

   $$r(d, \pi) = \int_{\Theta} (\theta - d)^2 \pi(\theta \mid x) d\theta.$$

   Differentiating with respect to d and setting equal to zero gives

   $$\int_{\Theta} (\theta - d) \pi(\theta \mid x) d\theta = 0.$$

   Since the posterior density integrates to 1 this is equivalent to

   $$d = \int_{\Theta} \theta \pi(\theta \mid x) d\theta,$$

   the posterior mean of θ.

2. **Absolute error loss: posterior median.** $L_1(\theta, d) = |\theta - d|$.

   $$r(d, \pi) = \int_{\Theta} |\theta - d| \, \pi(\theta \mid x) d\theta$$
   $$= \int_{-\infty}^{d} (d - \theta) \pi(\theta \mid x) d\theta + \int_{d}^{\infty} (\theta - d) \pi(\theta \mid x) d\theta.$$

   Differentiating with respect to d and setting equal to zero gives

   $$\int_{-\infty}^{d} \pi(\theta \mid x) d\theta - \int_{d}^{\infty} \pi(\theta \mid x) d\theta = 0.$$

   Since the posterior integrates to 1, $\hat{d}$ is the solution of

   $$\int_{-\infty}^{d} \pi(\theta \mid x) d\theta = \frac{1}{2},$$

   the posterior median of θ.

3. **Zero-one loss: posterior mode.** $L_1(\theta, d) = 1 - \delta(\theta - d)$.

$$\begin{aligned} r(d,\pi) &= \int_\Theta (1 - \delta(\theta - d))\,\pi(\theta \mid x)d\theta \\ &= 1 - \int_\Theta \delta(\theta - d)\pi(\theta \mid x)d\theta \\ &= 1 - \pi(d \mid x). \end{aligned}$$

   It is now easily observed that the Bayes risk is minimized for that value of d for which the posterior is maximized. So, $\hat{d}$ is the posterior mode.

2.1.5. Computational techniques

The algorithms and notation in this section are based on chapter 3.7 in [13]. Suppose there are d parameters under consideration and write $\theta = (\theta_1, \theta_2, \dots, \theta_d) \in \Theta \subseteq \mathbb{R}^d$. In order to do inference on the posterior distribution

$$\pi(\theta \mid X = x) = \frac{f(x \mid \theta)\pi(\theta)}{\int_\Theta f(x \mid \theta)\pi(\theta)d\theta},$$

the normalising constant has to be computed. In many (high-dimensional) cases this is a non-trivial integral which is hard to compute, even using numerical computation techniques. To avoid this issue an algorithm called the 'Gibbs Sampler' could be employed.

Gibbs sampler

Instead of computing the integral, samples are drawn from

$$\pi(\theta \mid X = x) \propto \pi(\theta) f(x \mid \theta),$$

and quantities of interest are estimated through the simulated sample. Algorithm 1 provides a way to sample from the multidimensional posterior distribution by sampling from a one-dimensional conditional distribution. Note that there are two problems with the Gibbs sampler:

1. The sampling distribution is proven to converge to the posterior but it takes some time before convergence is reached.

2. Consecutive observations θ^n, θ^{n+1} are not independent.

The first issue is resolved by allowing for a burn-in period: the first $M-1$ sampled vectors are discarded, where M is large enough to allow for convergence. A solution for the second issue is thinning, only saving every k'th observation: $\theta^1, \dots,\ \theta^{1+k}, \dots,\ \theta^{1+2k}, \dots$. However, [18] argues it is seldom necessary and therefore not used in this thesis.

Algorithm 1 Gibbs sampler

Choose initial vector $\theta^0 = \left(\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_d^{(0)}\right)$
for n = 1, 2, ..., N **do**
  for i = 1, 2, ..., d **do**
    Generate a new value $\theta_i^{(n)}$, conditional on
    $\theta_1^{(n)}, \dots, \theta_{i-1}^{(n)}, \theta_{i+1}^{(n-1)}, \dots, \theta_d^{(n-1)}$ and $X = x$
  end for
end for
return $(\theta^M, \dots, \theta^N)$

Metropolis-Hastings

While the Gibbs sampler provides a way to sample from the posterior distribution, it does not describe how to draw samples from the one-dimensional conditional distribution. This is where the Metropolis-Hastings sampler comes in.

To simplify notation denote the distribution to sample from as $f(x)$. A so-called proposal distribution $q(x, y)$ must be specified. A usual choice for $q(x, y)$ is $\frac{1}{h}g(\frac{y-x}{h})$ where $g(\frac{y-x}{h})$ is a density symmetric around 0. Let U be a standard uniform random variable. The M-H algorithm is described in Algorithm 2

Algorithm 2 Metropolis-Hastings

```
Choose initial value x^(0)
for n = 1, 2, ..., N do
    Given X^(n) = x^(n-1), draw a value y from Y from the density q(y | x), set α = min( f(y)q(x, y) / f(x)q(y, x) )
    and draw a value u from U

    if u ≤ α then
        X^(n+1) = y
    else
        X^(n+1) = X^(n)
    end if
end for
```

When one step of the Metropolis-Hastings is used to simulate a $\theta_i^{(n)}$ conditional on the other parameters the algorithm is described as 'Metropolis within Gibbs'.

2.2. Copulas

2.2.1. Introduction

In the financial industry copulas are widely used in for example portfolio management. In a multivariate setting, they allow the user to separately model the marginals and their dependence structure. The alternative is fitting a multivariate distribution, but it can prove to be difficult to find the right one. By having the freedom to choose (different) distributions for all the margins and a *copula* for the dependence structure it is possible to find a better fit for the data, especially in high dimensional problems. This is done by changing the marginals to a uniform distribution through the probability integral transform:

Theorem 2 (Probability Integral Transform)**.** *Let X be a random variable with cumulative distribution function $F_X(x)$. Then $U = F_X(X)$ is a standard uniform distributed random variable.*

Proof.

$$\begin{aligned} F_U(u) &= P(U \leq u) \\ &= P(F_X(X) \leq u) \\ &= P(X \leq F_X^{-1}(u)) \\ &= F_X(F_X^{-1}(u)) \\ &= u, \end{aligned}$$

which is the cumulative distribution function of a standard uniform random variable and therefore completes the proof □

After the marginal densities have been transformed to standard uniform random variables, the dependence structure can be modelled. Consider the multivariate setting where $(X_1, X_2, \dots, X_n)$ is a random vector with continuous marginal cumulative distribution functions given by $(F_1, F_2, \dots, F_n)$ and $(U_1, U_2, \dots, U_n)$ is a vector of standard uniform random variables. Then:

$$\begin{aligned}(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) &= P\left(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\right) \\ &= P\left(U_1 \leq F_1(x_1), U_1 \leq F_2(x_2), \dots, U_n \leq F_n(x_n)\right) \\ &= C\left(F_1(x_1), F_2(x_2), \dots, F_n(x_n)\right),\end{aligned}$$

where the copula C satisfies the definition coming from [19]:

Definition 3 - Copula.

An n-dimensional copula is a function C with domain $[0,1]^n$ such that

1. *C is grounded and n-increasing,*
2. *C has margins C_k, $k = 1, 2, \dots, n$, which satisfy $C_k(u) = u$, $\forall u \in [0,1]$*

2.2.2. Theory

If there is a theorem that could be named the 'Fundamental Theorem of Copulas' it has to be what Sklar calls in [20] the 'Extension Theorem for Copulas'. It proves the claim that joint distributions can be modelled by their marginals and a copula C. Formulation of the following theorem follows [19]:

Theorem 3 (Sklar's Theorem)**.** *Let H be an n-dimensional distribution function with margins $F_1, F_2, \dots, F_n$. Then there exists an n-copula C such that $\forall x \in \bar{\mathbb{R}}^n$*

$$H(x_1, x_2, \dots, x_n) = C\left(F_1(x_1), F_2(x_2), \dots, F_n(x_n)\right).$$

If $F_1, F_2, \dots, F_n$ are all continuous, then C is unique; otherwise C is uniquely determined on the Cartesian product of the ranges of the marginal distribution functions. Conversely, if C is an n-copula and $F_1, F_2, \dots, F_n$ are distribution functions, then the function H defined above is an n-dimensional distribution function with margins $F_1, F_2, \dots, F_n$.

Proof. The proof is found in [20]: □

Every copula can be bounded from above and below by the so called 'Fréchet-Hoeffding bounds'.

Theorem 4 (Frćhet–Hoeffding bounds)**.** *[21] Let C be an n-copula, then $\forall \underline{u} \in [0,1]^n$ the following holds:*

$$W^n(\underline{u}) \leq C(\underline{u}) \leq M^n(\underline{u}),$$

where:

$$\begin{aligned}W^n(\underline{u}) &= \max\left(1 - n + \sum_{i=1}^{n}, 0\right), \\ M^n(\underline{u}) &= \min\left(u_1, u_2, \dots, u_n\right).\end{aligned}$$

2.2.3. Examples

In this section several copulas will be examined. Every copula will be an n-copula defined on the unit cube $[0,1]^n$ unless otherwise mentioned.

Gaussian Copula

Suppose Σ is a n-dimensional correlation matrix and define Φ_Σ^n as the joint distribution function of the multivariate standard normal. Furthermore, let Φ and Φ^{-1} be the univariate distribution and quantile function respectively. Then the Gaussian copula is defined by

$$C_\Sigma^{GA}(\underline{u}) = \Phi_R^n\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \dots, \Phi^{-1}(u_n)\right).$$

A well known drawback of gaussianity in general is the underestimation of tail risk, in the univariate as well as the multivariate setting. A copula which is more fitting when the estimation of tail risk is of interest is the t-copula.

t-Copula

In the same way the Gaussian copula relies on the normal distribution, the t-copula relies on the t-distribution. Suppose $t^n_{\nu,\Sigma}$ is the n-variate distribution function of a t-distribution with mean the zero vector, shape matrix Σ and degrees of freedom ν. Also, let t_ν be the distribution function of the univariate t-distribution with degrees of freedom ν. Then the t-copula is given as follows:

$$C^t_{\nu,\Sigma}(\underline{u}) = t^n_{\nu,\Sigma}\left(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2), \dots, t_\nu^{-1}(u_n)\right).$$

Archimedean Copulas

The Gaussian and t-Copula are derived from elliptical distributions. Drawbacks are the lack of a close form and the restriction to radial symmetry. As often observed in financial data, big losses tend to have stronger dependence than big profits. Archimedean copulas offer solutions to both of these problem, they have a closed form representation and some allow for asymmetrical modelling. However, a disadvantage of Archimedean copulas is the more difficult extension to dimensions higher than 2. If a copula C satisfies 2.1 in the following theorem [19] it is an Archimedean copula:

Theorem 5. *Suppose $\phi : [0,1] \to [0,\infty)$ is continuous, strictly decreasing function such that $phi(1) = 0$ and denote $\phi^{(-1)}$ as its generalized inverse. Let $C : [0,1]^2 \to [0,1]$ be a function given by*

$$C(u,v) = \phi^{(-1)}\left(\phi(u) + \phi(v)\right). \tag{2.1}$$

Then C is a copula if and only if ϕ is convex.

Proof. The proof is given in [22]. □

The attention is now restricted to bivariate copulas, which is justified by both the fact that higher dimensional ones are not used in this thesis and because it is easier to graphically represent them. Two examples of bivariate Archimedean copulas are the following:

- **Franks Copula:** the only copula that is satisfies the radial symmetry. It is given by

$$C^F_\theta(u,v) = -\frac{1}{\theta}\log\left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right), \quad \theta \in \mathbb{R}\setminus\{0\}.$$

- **Claytons Copula:** this dependence structure allows for asymmetrical modelling:

$$C^{Cl}_\theta(u,v) = \max\left([u^{-\theta} + v^{-\theta} - 1]^{-\frac{1}{\theta}}, 0\right), \quad \theta \in [-1,\infty)\setminus\{0\}.$$

In order to graphically represent the aforementioned copulas a choice of parameters has to be made. It would be thoughtless to just manually pick some parameters for which the plots look nice, because then they cannot be compared with one another. That is why the concept of *Kendall's tau* [19] is introduced:

Definition 4 - Kendall's tau.
Consider the random vector (X,Y) and an independent copy of that vector $(\hat{X},\hat{Y})$. Kendall's tau is then given by

$$\tau(X,Y) = P\left([X-\hat{X}][Y-\hat{Y}] > 0\right) - P\left([X-\hat{X}][Y-\hat{Y}] < 0\right).$$

This measure of dependence is preferable over the correlation measure when dealing with non-elliptical distributions. Instead of measuring linear dependence, Kendall's tau measures monotone dependence. For the previously described copulas, where the Gaussian and t-copula are elliptical copulas, Kendall's tau is given in relation to its parameters:

- **Elliptical copulas:** Suppose X is an elliptically distributed random variable with shape matrix Σ, its correlation coefficients given by $\Sigma_{ij} = R_{ij}$, then:

$$\tau^E_{R_{ij}} = \frac{2}{\pi}\sin^{-1}(R_{ij}).$$

- **Franks copula:**

$$\tau_\theta^F = 1 - \frac{4}{\theta}\left(1 - \frac{1}{\theta}\int_0^\theta \frac{t}{e^t - 1}dt\right).$$

- **Claytons copula:**

$$\tau_\theta^C = \frac{\theta}{\theta + 2}.$$

In figure 2.1 the four copulas described in this section are presented. As for the choice of parameters; Kendall's tau has been chosen as 0.5. Assuming a symmetric correlation matrix for the elliptical copulas this results in $R_{12} = R_{21} = \frac{1}{2}\sqrt{2}$. Additionally, the degrees of freedom parameter ν of the t-copula has been set to 4. Regarding the Archimedean copulas, Franks copula has been plotted for $\theta \approx 5.74$ and Claytons copula for $\theta = 2$.

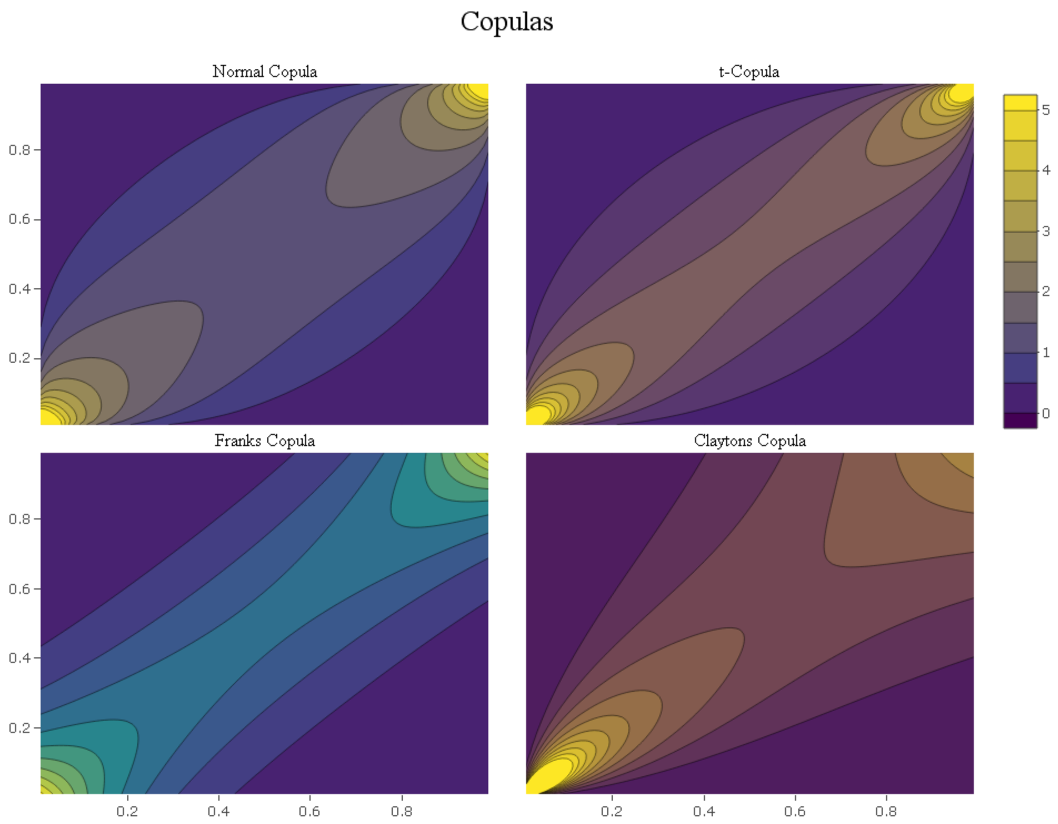

Figure 2.1: Normal, t-, Franks and Claytons copula for $\tau = 0.5$

The figure shows the radial symmetry of both the elliptical copulas and Franks copula. Furthermore it can be seen that the t-copula allows for the modelling of more extreme risks, in contrast with the normal and Franks copula. Claytons copula does not display radial symmetry, but instead puts more mass on the lower tail of the bivariate distribution than on the upper tail.

2.2.4. Remarks

Copulas allow for the separate modelling of marginals and dependence and are therefore widely used in practice. They seem to be the holy grail of multivariate dependence modelling, but could they be too good to be true? Mikosch argues in [23] this is the case.

Year	1	2	3	4	5	6	7	8	9	10
Number of defaults	0	0	1	0	1	2	0	4	0	1

Table 2.1: Default observations per year in a mock portfolio of $n = 10$ obligors.

One of his points of criticism is there is no practical reason for choosing the transformation to the unit interval. The transformation $-\log(1 - F(X))$ yields standard exponential random variables and one could fit a dependence structure on those as well. Secondly, different choices of marginals could lead to different choices of copulas, so the claimed separation of marginal and dependence modelling may not be so robust after all. Related to this, there is higher statistical uncertainty when estimating all the marginals and the copula than in estimating one multivariate distribution. More remarks on the disadvantages of copulas can be found in the article of Mikosch.

There are certainly advantages in using copulas to model the dependence structure, and their usefulness is not automatically disregarded. But it is important to be aware of their drawbacks as well. In the proceedings of this research a thoughtful consideration will be made if a multivariate distribution may or may not be a better way to model dependence than the copula approach.

2.3. Goodness of Fit testing

In order to test if a model can be used to study a given data set, it is wise to check if this data satisfies the assumptions of the model. One can argue that one of the most important assumptions to test is whether the data is really distributed as the model specifies. This can be verified by conducting a goodness of fit test with null- and alternative hypothesis given by:

- H_0 : The sample belongs to the assumed distribution
- H_1 : The sample does not belonged to the assumed distribution

When testing a hypothesis two types of errors can be made:

- Type I error: Rejecting the null hypothesis when it is true
- Type II error: Not rejecting the null hypothesis when it is false

Thus, a type I error occurs when the sample comes from the assumed distribution but the test says it does not. This error is controlled by the significance level α of the test. So if $\alpha = 0.05$ at most 5% of the samples will be rejected when tested against their true distribution. A type II error arises when the sample does not come from the assumed distribution, but the test does not reject it. The lower the type II error, the higher the so called *power* of the test. Notice the interaction between these two types of errors: when the significance level is lowered, type I errors decrease but type II errors increase.

First, the exact multinomial test is described and its drawback discussed. An approximation to this test, the χ^2-test is presented thereafter. Another test, based on the maximum of the distribution is visited later in this thesis.

2.3.1. Exact multinomial test

The exact multinomial test is explained by means of an example. Suppose a distribution for the amount of defaults in a portfolio with probability mass function $f(k \mid p)$ is proposed, where p is the probability of default parameter. Furthermore, consider the following data of a portfolio with $n = 10$ obligors per year:

First, the sample has to be divided into bins of number of observed defaults. Since the maximum amount of defaults in the sample is lower than the size of the portfolio, a 'rest' bin containing the number of defaults > 4 is defined. This is needed since under the null hypothesis the probability of observing more than 4 defaults is larger than zero, even though no observations in that bin are made.

Number of defaults, k	0	1	2	3	4	>4
Amount of observations in sample	5	3	1	0	1	0
Probability of observing k defaults under H_0	0.389	0.385	0.171	0.045	0.008	0.001

Table 2.2: Number of observations per default amount k, with probabilities of observing k defaults under H_0.

Also, since no value for the probability parameter p has been specified, a maximum likelihood estimation has been performed which resulted in $p = 0.09$. Table 2.2 is equivalent to the following partitioning into bins:
The test exact multinomial test is based on the multinomial distribution. A multinomial random variable is the result of throwing n independent balls into m urns, each with a given probability of attraction p_i [24]. The probability mass function of a Mult$(p_1, \ldots, p_m, n)$ is given by

$$f_{MN}(K_1 = k_1, \ldots, K_m = k_m \mid p_1, \ldots, p_m, n) = \frac{n!}{k_1!k_2!\ldots k_m!} p_1^{k_1} p_2^{k_2} \ldots p_m^{k_m}.$$

Now an exact multinomial test can be conducted to test if the data comes from a Mult$(p_1, \ldots, p_6, n = 10)$ distribution (since there are 6 default bins and 10 years of data in the example), with probabilities p_i found in table 2.2. Following the approach of [25], a multinomial test is conducted as follows:

1. For all possible outcomes $(k_1, \ldots, k_m)$, compute the probabilities $f(k_1, \ldots, k_m \mid p_1, \ldots, p_m, n)$.
2. Order these probabilities from smallest to largest.
3. Add together all these increasing probabilities until the cumulative probability first exceeds α.
4. Reject the null hypothesis if the sample outcome is among the outcomes used in calculating the cumulative probability in the previous step.

Although this seems a fairly simple goodness of fit test, the first step in the procedure is extremely difficult in practice. The amount of possible outcomes grows exponentially, with an order of $\frac{n^m}{m}$. In this example that is still doable because the portfolio only consists of 10 obligors. However, in a more standard low default portfolio of 100 obligors $\frac{100^6}{6}$ probabilities need to be evaluated, which is unfeasible. An approximation to the exact multinomial test is considered next.

2.3.2. χ^2-test

In practice the χ^2-test is heavily used in goodness of fit testing of categorical data, but has its applications in discrete and continuous data as well. In these cases the sample space of the data must be must be partitioned into a finite collection of disjoint bins, just as in the example is done.

The test compares the amount of observations in each category to the expected amount under the null hypothesis. Define O_i and E_i as, respectively, the amount of observations and the expected amount under the null hypothesis in category i, where the categories are the number of times a certain amount of defaults is observed, see table 2.3.

Number of defaults, k	0	1	2	3	4	>4
Amount of observations in sample	5	3	1	0	1	0
Expected amount of defaults k under H_0	3.89	3.85	1.71	0.45	0.08	0.01

Table 2.3: Number of observations per default amount k, with expected amount of defaults under H_0.

The test statistic is given below:

$$T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(d),$$

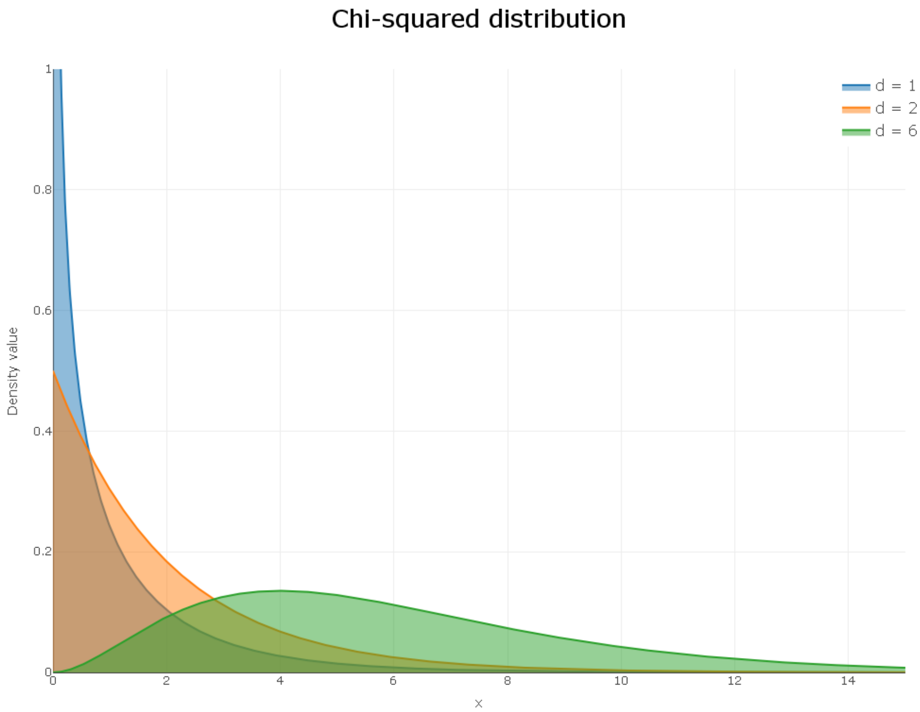

Figure 2.2: Probability density function of the χ^2 distribution for various degrees of freedom d

where $\chi^2(d)$ is a χ^2-distribution with degrees of freedom parameter d. Suppose $Z_1, Z_2, \ldots, Z_d \overset{iid}{\sim} N(0,1)$, then it is known that:

$$\sum_{i=1}^{d} Z_i^2 \sim \chi^2(d).$$

So a χ^2 distribution with degrees of freedom parameter d is the sum of the squares of d independently distributed standard normal variables. Let X be a $\chi^2(d)$ distributed random variable, the probability density function of X is given by

$$f(x) = \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}.$$

The mean and variance of the χ^2 distribution are given by $\mathbb{E}[X] = d$ and $\mathrm{Var}(X) = 2d$.

Under the null-hypothesis the test statistic is χ^2 distributed with degrees of freedom d equal to $k - c$ where k is the amount of bins and c the amount of estimated parameters, see [26]. The null hypothesis is rejected if the p-value of the sample (probability under the null hypothesis of obtaining a more extreme result than the sample) is less than a certain significance level α, or if the the test statistic T is larger than the critical value c_α.

The corresponding test statistic has value $T = 12.12$ and the degrees of freedom of the χ^2-distribution has value $6-2 = 4$ since the size parameter of the binomial has been fixed and the probability parameter has been estimated. The p-value of this sample is the probability of observing a more extreme result, so if $F(x)$ is the distribution function of a $\chi^2(4)$ random variable, the p-value of this sample is:

$$1 - F(12.12) = 0.016$$

A usual choice, which will be followed in this thesis, for the level of α is 0.05 and for this significance level the null-hypothesis that the sample comes from a Binomial distribution is rejected.

Although the χ^2-test is popular under practitioners, it will be shown that it is not suited for goodness of fit testing in the setting of low default portfolios. These findings are supported by the literature, as [25] summarises some heuristics on the minimum amount of expected frequencies. Some authors suggest a value of at least 10 or 20 for the expected frequencies, while the 'most widely used' Conan's rule is all expected frequencies higher than 1, and 80% higher than 5. In a general low default portfolio, the zero default bin will have a very high expected frequency, but even observing one default is often unlikely.

2.3.3. An alternative: the maximum test

In an effort to improve goodness of fit testing when dealing with a small sample, a test based on the maximum is investigated. In [24] results about the maximum of an equiprobable Multinomial distribution are derived. The assumption of equal probabilities however is hard to justify in low default samples. One could merge certain default bins, in table 2.2, 2 up to > 4 have together a probability of 0.225, which is not close enough to 0.385. Generally, no approximation of equiprobability may exist. Therefore a general method based on sampling is proposed, able to deal with any distribution under H_0.

The maximum test compares the distribution of the maximum under the null hypothesis with the maximum of the sample. The advantage of this test is the sensitivity to more extreme values, in contrast with the χ^2-test. However, a drawback is that the maximum is the only value under consideration. This may cause the test to not reject the null hypothesis because the sample maximum is close to the maximum under H_0, while the body of the distribution is very different from the body under the null hypothesis.

The procedure for conducting the maximum test is explained through the same example. Suppose again a test for the Binomial distribution is carried out. First, use maximum likelihood estimation to compute the probability parameter p, resulting in $p = 0.09$. Then, draw 1000 samples from a Bin$(0.09, 1000)$ distribution and save each sample maximum. Look up the empirical 95% quantile of these maxima and reject the null hypothesis if the maximum of the data is larger than (or equal to) this quantile. The simulation resulted in a value of 4 for the empirical quantile of the maximum. The maximum of the data was also 4, so the null hypothesis is rejected.

2.3.4. Comparison of the χ^2 and maximum test

The performances of the χ^2 and maximum test are evaluated in case of simulated data where the underlying distribution is known. Consider a portfolio of n obligors (the, perhaps unrealistic, assumption is made that n is constant) and a certain amount of defaults k_t for every year $t \in \{1, \dots, T\}$. The following procedure has followed:

Let $n = 100$, $T \in \{10, 100\}$ and define $B = 100$ as the sample size. Assume a distribution and a certain set of parameters. Distributions under consideration are Poisson, Binomial, Vasicek and CR+. Draw B samples from that distribution and for each of those samples perform the χ^2-test and the maximum test against the same four distributions.

This is then done 10 times for $T = 10, 100$ and both tests, so 40 times the above procedure in total. The mean and standard deviations of the number of rejected samples are presented in tables 2.4, 2.5, 2.6 and 2.7 in the appendix and a summary is given by the following points:

- It is not possible to distinguish between Binomial and Poisson, which is coherent with the Poisson approximation of the Binomial when p is small.
- The χ^2-test has less type I errors when sampling from Binomial or Poisson. Especially when the PD is low the maximum test performs poorly for type I errors.

- However, when the true distribution is Vasicek or CR+, the χ^2 test performs badly concerning type I errors. This is especially true when the variance is higher. The maxtest in this case has quite acceptable type I errors.
- Both tests make enormous type II errors when the underlying distribution is Binomial or Poisson. The tests do not reject the Vasicek and CR+ distributions even when $T = 100$. This is probably due to the fact that the Vasicek and CR+ are extensions of the Binomial and Poisson and are equal to these distribution if the correlation/volatility parameter is set to 0. Even though the χ^2-test should penalize the extra parameter estimated, this is not sufficient to reduce type II errors.
- When sampling from a Vasicek or CR+ distribution, type II errors are low in the two tests. The maximum test is still doing a bit better than the χ^2.

From a regulatory point of view, it is most important to minimize type I errors when the underlying distribution is Vasicek or CR+. The maximum test is the most consistent in rejecting the Binomial and Poisson in this case, with the added bonus of also having much less type II errors. It is therefore preferable over the χ^2-test.

Table 2.4: ($T = 10$, χ^2-test) Mean and standard deviation of # rejected samples. $B = 100$ is the sample size.

Distribution	Parameters		Binomial		Poisson		Vasicek		CR+	
			mean	sd	mean	sd	mean	sd	mean	sd
Binomial	$p = 0.01$		8,0	2,1	5,3	2,2	3,3	2,2	3,8	1,8
Binomial	$p = 0.003$		7,5	3,0	5,4	2,3	4,8	2,1	5,3	2,1
Poisson	$\lambda = 1$		6,9	1,5	4,3	1,5	3,2	1,8	4,9	1,7
Poisson	$\lambda = 0.3$		8,8	3,8	6,0	3,6	5,9	3,3	6,4	3,3
Vasicek	$p = 0.01$	$\rho = 0.18$	52,4	4,6	48,6	5,4	14,0	4,2	10,3	3,9
Vasicek	$p = 0.01$	$\rho = 0.8$	85,0	3,3	80,1	3,2	76,1	3,8	71,3	3,9
Vasicek	$p = 0.003$	$\rho = 0.18$	32,5	4,1	25,0	3,6	17,5	2,7	15,0	2,5
Vasicek	$p = 0.003$	$\rho = 0.8$	90,1	3,7	85,3	2,9	85,0	2,9	82,8	3,6
CR+	$\mu = 0.01$	$\sigma = 1$	37,7	6,0	31,2	6,5	6,7	2,0	5,8	1,9
CR+	$\mu = 0.01$	$\sigma = 2$	69,5	3,6	65,4	4,8	32,1	3,8	25,2	3,8
CR+	$\mu = 0.003$	$\sigma = 1$	65,2	6,3	55,9	5,0	47,3	5,2	39,5	5,4
CR+	$\mu = 0.003$	$\sigma = 2$	85,8	2,7	79,8	2,8	78,2	2,9	73,3	3,2

Table 2.5: ($T = 10$, Maxtest) Mean and standard deviation of # rejected samples. $B = 100$ is the sample size.

Distribution	Parameters		Binomial		Poisson		Vasicek		CR+	
			mean	sd	mean	sd	mean	sd	mean	sd
Binomial	$p = 0.01$		9,7	3,0	8,8	3,3	0,0	0,0	0,0	0,0
Binomial	$p = 0.003$		25,4	3,9	24,1	2,8	4,2	1,8	15,9	2,3
Poisson	$\lambda = 1$		8,7	3,9	8,1	3,6	0,0	0,0	0,0	0,0
Poisson	$\lambda = 0.3$		27,5	4,0	26,1	4,5	4,9	3,0	15,7	3,6
Vasicek	$p = 0.01$	$\rho = 0.18$	56,5	6,2	54,8	6,3	0,3	0,5	1,2	1,4
Vasicek	$p = 0.01$	$\rho = 0.8$	93,6	2,2	94,7	2,0	38,6	4,5	44,7	3,3
Vasicek	$p = 0.003$	$\rho = 0.18$	57,6	5,6	57,1	6,0	10,7	2,1	24,9	5,3
Vasicek	$p = 0.003$	$\rho = 0.8$	97,7	0,8	97,8	1,6	71,5	5,7	76,7	5,9
CR+	$\mu = 0.01$	$\sigma = 1$	42,0	5,6	41,0	5,1	0,1	0,3	0,5	0,7
CR+	$\mu = 0.01$	$\sigma = 2$	77,0	3,4	77,4	3,3	2,8	0,9	5,6	1,7
CR+	$\mu = 0.003$	$\sigma = 1$	82,9	4,0	82,7	4,2	25,2	4,2	39,1	4,7
CR+	$\mu = 0.003$	$\sigma = 2$	95,6	1,7	94,9	2,0	56,3	3,5	63,5	3,3

Table 2.6: ($T = 100$, χ^2-test) Mean and standard deviation of # rejected samples. $B = 100$ is the sample size.

Distribution	Parameters		Binomial		Poisson		Vasicek		CR+	
			mean	sd	mean	sd	mean	sd	mean	sd
Binomial	$p = 0.01$		8,8	3,0	6,6	3,8	3,7	2,3	5,4	3,2
Binomial	$p = 0.003$		6,8	1,8	3,6	1,2	2,1	1,7	2,5	1,9
Poisson	$\lambda = 1$		11,4	2,1	8,3	2,6	4,7	1,7	6,4	2,2
Poisson	$\lambda = 0.3$		7,9	3,4	4,4	2,0	2,7	2,0	3,2	2,2
Vasicek	$p = 0.01$	$\rho = 0.18$	99,8	0,4	99,7	0,7	22,0	3,6	30,1	4,1
Vasicek	$p = 0.01$	$\rho = 0.8$	100,0	0,0	99,9	0,3	85,0	2,9	88,7	2,7
Vasicek	$p = 0.003$	$\rho = 0.18$	83,9	4,7	80,4	3,7	9,5	3,8	12,0	4,5
Vasicek	$p = 0.003$	$\rho = 0.8$	92,9	3,5	92,4	3,3	73,9	6,1	68,3	5,3
CR+	$\mu = 0.01$	$\sigma = 1$	97,5	1,6	96,4	2,1	8,1	3,9	9,3	2,9
CR+	$\mu = 0.01$	$\sigma = 2$	100,0	0,0	100,0	0,0	23,8	5,8	24,4	6,0
CR+	$\mu = 0.003$	$\sigma = 1$	99,1	0,9	98,7	1,3	27,1	3,5	18,1	4,0
CR+	$\mu = 0.003$	$\sigma = 2$	98,5	1,1	98,2	1,0	73,0	4,3	59,7	6,3

Table 2.7: ($T = 100$, Maxtest) Mean and standard deviation of # rejected samples. $B = 100$ is the sample size.

Distribution	Parameters		Binomial		Poisson		Vasicek		CR+	
			mean	sd	mean	sd	mean	sd	mean	sd
Binomial	$p = 0.01$		12,7	3,3	11,0	2,5	4,4	1,3	4,0	1,2
Binomial	$p = 0.003$		22,9	5,8	21,3	5,4	8,9	3,1	8,0	3,0
Poisson	$\lambda = 1$		16,1	4,6	13,3	3,5	6,0	2,7	5,1	1,6
Poisson	$\lambda = 0.3$		23,1	3,0	23,1	3,2	7,3	2,8	7,6	2,1
Vasicek	$p = 0.01$	$\rho = 0.18$	98,9	1,2	98,8	1,1	4,4	1,2	12,5	4,2
Vasicek	$p = 0.01$	$\rho = 0.8$	100,0	0,0	100,0	0,0	22,5	3,2	0,0	0,0
Vasicek	$p = 0.003$	$\rho = 0.18$	90,0	3,2	89,4	3,8	1,4	1,2	8,7	2,8
Vasicek	$p = 0.003$	$\rho = 0.8$	99,9	0,3	99,7	0,7	16,9	3,1	4,4	2,8
CR+	$\mu = 0.01$	$\sigma = 1$	92,1	2,5	90,6	2,7	0,9	1,0	3,7	2,4
CR+	$\mu = 0.01$	$\sigma = 2$	99,9	0,3	99,9	0,3	0,0	0,0	2,4	1,6
CR+	$\mu = 0.003$	$\sigma = 1$	99,5	0,7	99,4	0,7	0,1	0,3	0,6	0,7
CR+	$\mu = 0.003$	$\sigma = 2$	99,5	0,7	99,4	0,7	3,6	1,6	1,1	1,0

2.4. Confidence Intervals

Point estimators are useful in the setting of credit risk modelling, especially since the IRB formula requires them as input for the computation of required capital. From a statistical perspective however, only considering point estimators if often a bad practice since they give no indication of how reliable a certain estimate is. Confidence intervals give a range of plausible values and could therefore be used to asses the reliability of point estimators.

Note that confidence intervals are a frequentist tool and recall from earlier in this chapter that the maximum likelihood estimator belongs to the frequentist approach as well. These methods are fundamentally different from the Bayesian ones. Bayesians prefer to speak about 'credible intervals'. The estimation of both these intervals and the conceptual differences between these approaches are discussed here.

Confidence Intervals

Suppose the confidence level is $\gamma = 0.95$, a confidence interval will contain the true value of the parameter in 95% of the cases if one were to draw an infinite amount of samples. Note that this does not mean the interval contains the true parameter with a probability of 95%. In the frequentist setting there is a true parameter, which either lies in the interval or not so there is no notion of probability needed.

Confidence intervals around the MLE can be constructed through its asymptotic convergence. Recall the definition of Fisher information from before, then the following theorem gives the limit distribution of the maximum likelihood estimator.

Theorem 6 (Asymptotic normality of the MLE [15])**.** *Suppose X is a (continuous) random variable with parameter θ, given by probability density function $f(X \mid \theta)$. Let $\hat{\theta}_{MLE}$ denote the maximum likelihood estimator and $I(\theta)$ the Fisher Information of θ. Suppose n observations have been made, then:*

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{d} N\left(0, \frac{1}{I(\theta)}\right).$$

Proof. See [15] for a sketch of the proof. □

In order to extract confidence intervals from this result, one issue still needs to be resolved. Note that in theorem 6 the Fisher Information of the *true* parameter θ is used. However, this parameter is unknown and so its Fisher Information as well. Rice notes in [15] that the MLE $\hat{\theta}_{MLE}$ can be used as a substitute for the true parameter. Writing $z_{\alpha/2}$ as the $\frac{\alpha}{2}$-quantile of a standard normal distribution, the confidence interval with significance level α is given by

$$\left(\hat{\theta}_{MLE} - z_{\alpha/2}\frac{1}{\sqrt{nI(\hat{\theta}_{MLE})}}, \quad \hat{\theta}_{MLE} + z_{\alpha/2}\frac{1}{\sqrt{nI(\hat{\theta}_{MLE})}}\right).$$

2.4.1. Credible Intervals

In the Bayesian approach credible intervals have a different meaning than their frequentist counterpart. After having specified a priori knowledge in the form of a prior and updating the likelihood the posterior distribution is derived. This distribution assigns probability to the true parameter p. Deriving a credible interval of confidence level $\gamma = 0.95$ then comes down to selecting an interval (or union of intervals) to which the parameter p belongs to with a probability of 95%. Two possible choices for a Bayesian credible interval are [27]:

- **Highest Posterior Density (HPD):** The narrowest interval, assuming a unimodal distribution this interval will contain the mode of the posterior distribution. The HPD is given by the set $\{\theta : \pi(\theta \mid x) \geq k\}$ with k such that the following holds:

$$\gamma = \int_{\{\theta:\pi(\theta|x)\geq k\}} \pi(\theta \mid x) d\theta.$$

- **Equal Tailed (ET):** An interval chosen so that the probability below the interval is equal to the probability above it. This interval will contain the median of the posterior distribution. The ET interval is given by (θ_L, θ_U) with:

$$\theta_L = \inf\left\{\theta : \int_{-\infty}^{\theta} \pi(\theta \mid x) d\theta \geq \frac{\alpha}{2}\right\},$$
$$\theta_U = \inf\left\{\theta : \int_{-\infty}^{\theta} \pi(\theta \mid x) d\theta \geq 1 - \frac{\alpha}{2}\right\}.$$

In this thesis the choice for the Bayesian credible interval is the Equal Tailed interval. The ET interval is robust to outliers and has no numerical complications in the case of a bimodal posterior distributions as the HPD has.

3
State of the Art

The content of this chapter serves two purposes. The first is to present a general overview of the literature written on the subject of low default portfolios. The second is to describe the starting point from which the research in this thesis is done. First, two factor models under the names of Merton and Vasicek are are discussed, as especially the last one is frequently used in the literature. With knowledge of these models a selection of articles published on low default portfolios is presented. Two of these articles are thoroughly investigated, as their approaches are used in the new analysis done in this thesis.

3.1. Factor models

The Vasicek model is often used by practitioners in both the academic world and the financial industry. The corresponding distribution can be seen as an extension of the Binomial distribution, but with correlated defaults. In this chapter the model and its pros and cons will be thoroughly investigated. It will be shown that the Vasicek model relies on many of the same assumptions as Merton's model, which is why the latter is presented first.

3.1.1. Merton's model

In 1973 Robert C. Merton published an article under the title 'On the pricing of corporate debt: the risk structure of interest rates' [28]. This paper assumed a geometric Brownian motion for a companies asset value, and derived under certain assumptions a result for credit risk modelling analogous to the Black and Scholes theorem for option pricing. In 1997 Myron Scholes and Merton received the Nobel prize for their work (Fisher Black unfortunately had passed away in 1995). It is therefore for good reason this result is sometimes named the 'Black, Scholes and Merton' theorem.

Model structure

Merton assumes in his paper that the value of a company at time t is given by V_t, which is comprised of the sum of its equity S_t and debt B_t:

$$V_t = S_t + B_t. \tag{3.1}$$

Under the Modigliani-Miller theorem [29] the value of the company does not depend on the proportion of equity (or liabilities) to value. In the model the debt is modelled as a zero-coupon bond expiring in T with face value B. A company defaults when it cannot repay its debt at maturity T. This is of course a heavily simplified representation of the real world, since a company can also default before the final maturity. Other assumptions, such as frictionless markets, are found in the original document of Merton in A1 to A8.

The event of default is given by $\{V_T < B\}$, and so the probability of default by $P(V_T < B)$. In order to compute this probability it is necessary to know more about the dynamics of V_t. Merton therefore assumes V_t to follow a geometric Brownian motion, which translates to satisfying this stochastic differential equation:

$$dV_t = \mu V_t dt + \sigma V_t dW_t.$$

Where W_t is a standard Brownian motion, μ is the trend of the asset value and σ its volatility parameter. Suppose Z is a standard normally distributed random variable, by making use of Itô's lemma it is found that:

$$\begin{aligned} V_t &= V_0 \exp\left\{(\mu - \frac{1}{2}\sigma^2)T + \sigma W_t\right\} \\ &= V_0 \exp\left\{(\mu - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}Z\right\}, \end{aligned} \tag{3.2}$$

and thus:

$$\begin{aligned} \log V_T &= \log V_0 + (\mu - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}Z \\ &\sim N\left(\log V_0 + (\mu - \frac{1}{2}\sigma)T, \sigma^2 T\right). \end{aligned}$$

The probability of default can now be written as follows:

$$\begin{aligned} P\left(V_T < B\right) &= P\left(\log V_T < \log B\right) \\ &= P\left(\sigma\sqrt{T}Z + \log V_0 + (\mu - \frac{1}{2}\sigma)T < \log B\right) \\ &= P\left(Z < \frac{\log\frac{B}{V_0} - (\mu - \frac{1}{2}\sigma)T}{\sigma\sqrt{T}}\right) \\ &= \Phi\left(\frac{\log\frac{B}{V_0} - (\mu - \frac{1}{2}\sigma)T}{\sigma\sqrt{T}}\right). \end{aligned}$$

Relation to option pricing

As mentioned before, Merton proved the result for credit risk modelling what Black and Scholes showed for option pricing. This can be seen by further examining the event of default. It is also important to note that when a default happens debtors get paid before shareholders do (the latter case can not happen under the assumptions of the model).

At time T there are two possible situations, $V_T < B$ when a company has defaulted and $V_T \geq B$ when it has not. Since debt holders get the remaining value of the company whereas shareholders do not get paid in the case of default it can be written using equation 3.1:

$$S_T = \begin{cases} V_T - B & \text{if } V_T \geq B \\ 0 & \text{if } V_T < B \end{cases} = \max\left(V_T - B, 0\right),$$

and similarly:

$$B_T = \begin{cases} B & \text{if } V_T \geq B \\ V_T & \text{if } V_T < B \end{cases} = B - \max(B - V_T, 0).$$

So the value of equity and debt are equivalent to the payoff of a European call and put respectively. As V_T is assumed to follow a Geometric Brownian Motion, the relationship to the Black and Scholes theorem is clear.

3.1.2. The Vasicek model

In this section the portfolio model [30] developed by Vasicek in 2002 is described. It is derived from Merton's model and extended with a Gaussian correlation factor. This model is known under as the 'CreditMetrics^{TM}/ KMV One-Factor Model' [31], the 'one-factor probit model' [32] and 'Vasicek's model' [33]. From now on the model is referred to as the 'Vasicek model' (not to be confused by Vasicek's interest rate model!).

Individual probability of default

Just as in Merton's model, it is assumed company i defaults if it cannot repay its obligations of value B_i at maturity. In the model the loans of each company share the same time of maturity T. Denote the value of company i as V_t^i, which is also assumed to follow a Geometric Brownian Motion. Introducing dependence on the specific company and using previous result:

$$V_t^i = V_0^i \exp\left\{(\mu_i - \frac{1}{2}\sigma_i^2)T + \sigma_i\sqrt{T}Z_i\right\}.$$

Now let p_i be the *individual* probability of default of obligor i. Following the derivation in Merton's model it is found that:

$$p_i = P(V_T^i < B_i) = \Phi\left(c_i\right),$$

where:

$$c_i = \frac{\log\frac{B}{V_0^i} - (\mu_i - \frac{1}{2}\sigma_i)T}{\sigma_i\sqrt{T}}.$$

Probability of dependent defaults

The main difference of the Vasicek model in comparison to Merton's model is the introduction of a correlation effect. This is done through the random variable Z_i. This factor is still assumed to be a standard normal random variable, but such that there is positive correlation between two different factors. Therefore, write:

$$Z_i^\rho = \sqrt{\rho}Y + \sqrt{1-\rho}\,\xi_i.$$

Where $Y \sim N(0,1)$ independent of $\xi_i \overset{iid}{\sim} N(0,1)$. Y is known as the *systemic factor* and includes all the effects that the companies in the portfolio have in common. These effects could be macroeconomic, industrial or country-specific for example. The *asset correlation* ρ is the only quantity in which dependencies between obligors are captured. A higher value of ρ means a stronger influence of the common factor Y and thus a higher dependence between obligors. Remark that asset correlation is different from *default correlation*, the former being larger than the latter.

The factor ξ_i is transformed through the quantity $\sqrt{1-\rho}$ so that Z_i^ρ admits its standard normal form. ξ_i is called the *idiosyncratic* factor and stands for the company specific risk. The mean, variance and covariance of Z_i^ρ are easily verified to be:

$$\begin{aligned}\mathbb{E}\left[Z_i^\rho\right] &= 0,\\ \text{Var}\left(Z_i^\rho\right) &= 1,\\ \text{Cov}\left(Z_i^\rho, Z_j^\rho\right) &= \rho \quad \text{for } i \neq j.\end{aligned}$$

Now write p_i^ρ for the probability of default of company i in the Vasicek model, where p_i stands for the individual probability of default and ρ for the asset correlation. p_i^ρ is again given by the event when the asset value $V_T^{\rho,i}$ of the company is smaller than its debt B. By applying the same arguments as in Merton's model the *conditional* probability of default can therefore be written:

$$\begin{aligned}[p_i^\rho \mid Y = y] &= P\left(V_T^{\rho,i} < B\right)\\ &= P\left(Z_i^\rho < c_i\right),\end{aligned}$$

note that the individual probability $p_i = \Phi(c_i)$ and therefore $c_i = \Phi^{-1}(p_i)$,

$$\begin{aligned}&= P\left(Z_i^\rho < \Phi^{-1}(p_i)\right)\\ &= P\left(\sqrt{\rho}\, y + \sqrt{1-\rho}\, \xi_i < \Phi^{-1}(p_i)\right)\\ &= P\left(\xi_i < \frac{\Phi^{-1}(p_i)c_i - \sqrt{\rho}\, y}{\sqrt{1-\rho}}\right)\\ &= \Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho}\, y}{\sqrt{1-\rho}}\right).\end{aligned}$$

Remark that this is the IRB formula from chapter 1. Also note that in fact the underlying structure of the model is just an equicorrelated multivariate normal distribution with standard marginals. In the copula setting this would be the same as assuming standard normal marginals and modelling the dependence structure by a Gaussian copula. Also, in the limiting case where $\rho = 0$, p_i^ρ is equal to the individual probability of default p_i.

Default correlation

The parameter ρ is known as the asset correlation. A remark has to be made that under the Vasicek model this is not the same as default correlation. Consider two Bernoulli random variables X_1 and X_2 with success probability p, and let success be the event of default:

$$X_i = \begin{cases} 1 & \text{if } Z_i^\rho < c_i \\ 0 & \text{if } Z_i^\rho \geq c_i \end{cases} \qquad i = 1, 2.$$

Also,

$$\mathbb{E}\left[X_1, X_2\right] = P(Z_1^\rho < c_1, Z_2^\rho < c_2) = \underline{\Phi}(c_1, c_2, \rho),$$

where $c_1 = c_2 = p$ and $\underline{\Phi}(\cdot)$ denotes the bivariate normal distribution. The default correlation ρ^D can now be written as:

$$\rho^D = \frac{\text{Cov}(X_1, X_2)}{\sigma(X_1)\sigma(X_2)} = \frac{\underline{\Phi}(c_1, c_2, \rho) - p^2}{p(1-p)}.$$

Note that this also shows that the default correlation and probability of default are not independent, a result used later in chapter 5.

3.1.3. Remarks

Some assumptions made in the Vasicek model may not be completely warranted. It is of course possible for a company to default at any other time than some fixed maturity T. Moreover, the normality assumption as been shown not to hold in many cases as it severely underestimates (mutual) tail probabilities. The assumption that all dependencies are captured in a single factor influencing the asset correlation may also not be justified.

Even though these assumptions might seem too restrictive, the resulting model does have clear advantages. It is quite intuitive so that its use is not restricted to experts in the industry, see the IRB formula. Also, Vasicek derives in [30] a loss distribution for large portfolios.

The normality assumption for the common factor Y can be easily be lifted in favor of an other distribution. Normality for the idiosyncratic component ξ_i can also be swapped for a different distribution, but then numerical methods may be needed to compute the conditional probability, whereas the lack of these is one of the advantages of the Vasicek model.

3.2. Review of the literature

In this section the literature concerning low default portfolios will be investigated. Although there is only a limited number of papers published on the subject, it would take up too much space to review all of them. Therefore a conscious selection of influential and diverse articles has been made, so that they form a good representation of the state of the art. The first five publications are presented in chronological order, the last two are described more extensively because it contains an approach also incorporated in this thesis.

- Hanson, Schuermann (2005): Confidence Interval for Probabilities of Default [34]
- Kiefer (2006): Default Estimation for Low-Default Portfolios [35]
- Van der Burgt (2008): Calibrating Low-Default Portfolios, using the Cumulative Accuracy Profile [36]
- Fernandes, Rocha (2011): Low default modelling - a comparison of techniques based on a real Brazilian corporate portfolio [37]
- Orth (2012): Default Probability Estimation in Small Samples – With an Application to Sovereign Bonds [38]
- Pluto, Tasche (2005): Estimating Probabilities of Default for Low Default Portfolios [32]
- Tasche (2012): Bayesian Estimation of Probabilities of Default for Low Default Portfolios [39]

3.2.1. Hanson, Schuermann (2005): Confidence Interval for Probabilities of Default

In this publication Hanson and Schuermann concerned themselves with the estimation of confidence intervals in the context of low default portfolios. They made use of rating transition matrices in their procedure as the one of figure 3.1). Two main approached are considered: the cohort and duration approach. In the former, only the begin and end state of the portfolio was considered. In the latter also movements of obligors in between is taken into consideration. Two methods of confidence interval estimation are proposed, an analytical and a empirical or *bootstrap* method.

2016 One-Year Corporate Transition Rates By Region (%)

From/to	AAA	AA	A	BBB	BB	B	CCC/C	D	NR
Global									
AAA	81.25	12.50	0.00	0.00	0.00	0.00	0.00	0.00	6.25
AA	0.00	90.11	6.50	0.00	0.00	0.00	0.00	0.00	3.39
A	0.00	0.82	91.02	4.01	0.00	0.00	0.00	0.00	4.15
BBB	0.00	0.00	2.62	87.70	3.23	0.17	0.06	0.00	6.23
BB	0.00	0.00	0.00	3.12	80.37	6.07	0.23	0.47	9.74
B	0.00	0.00	0.00	0.00	3.84	74.00	5.55	3.68	12.92
CCC/C	0.00	0.00	0.00	0.00	0.99	14.36	40.59	32.67	11.39

Figure 3.1: Example of a credit migration matrix, credit scores are given in increasing riskiness. 'D' stands for 'Default' and 'NR' stands for 'No Rating' [40]

Analytical confidence intervals

Under the assumption that the amount of defaults is Binomially distributed, the Wald confidence interval for a given estimate of the probability of default $\hat{PD}$ is given by

$$CI_W = \hat{PD} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{PD}\,(1-\hat{PD})}{N}},$$

where $z_{\frac{\alpha}{2}}$ is the $1-\frac{\alpha}{2}$ quantile of the standard normal distribution and N is the number of companies in the portfolio. Another analytical confidence interval is obtained by setting the amount of observed defaults as k and defining

$$\tilde{PD} = \frac{k + z^2_{\frac{\alpha}{4}}}{N + z_{\frac{\alpha}{2}}},$$

the so-called Agresti and Coull confidence interval is then given by

$$CI_{AC} = \tilde{PD} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\tilde{PD}\,(1-\tilde{PD})}{N + z_{\frac{\alpha}{2}}}}.$$

The third analytical confidence interval, the Clopper-Pearson interval, is not based on asymptotics as the first two but instead derived from the Binomial distribution. The endpoints of the interval are given as solution to the following two equations:

$$\begin{cases} \sum_{i=k}^{N} \binom{N}{i} PD^i (1-PD)^{N-i} & = \frac{\alpha}{2} \\ \sum_{i=0}^{k} \binom{N}{i} PD^i (1-PD)^{N-i} & = \frac{\alpha}{2} \end{cases}$$

Bootstrapped confidence intervals

Instead of relying on analytical assumptions to derive confidence intervals, Hanson and Schuermann also considered a non-parametric approach. By resampling the rating histories of the firms the entire migration matrix is sampled and from the last column the resampled amount of defaults is obtained. Then the quantiles of the empirical distribution of defaults is used to obtain bootstrapped confidence intervals.

3.2.2. Kiefer (2006): Default Estimation for Low-Default Portfolios

In this article Kiefer examines the Bayesian approach in the setting of low default portfolios. He considers a Binomial likelihood for the amount of defaults and puts a prior on the probability parameter p. This prior $\pi(p)$ is assumed to follow a Beta(α, β) distribution truncated to the interval $[a, b] \subseteq [0, 1]$:

$$\pi(p \mid \alpha, \beta, a, b) = \frac{\Gamma(\alpha+\beta)}{(b-a)\Gamma(\alpha)\Gamma(\beta)} \left(\frac{a-p}{a-b}\right)^{\alpha-1} \left(\frac{p-b}{a-b}\right)^{\beta-1}$$

The four parameters are estimated through the use of expert judgement. Kiefer mentioned in his paper it is usually preferred to ask experts about probability of seeing data and using the predictive distribution to work backwards to inference about the parameter, as experts may find it difficult to answer questions about the parameter directly. However, the expert he consulted had no problem of thinking about quantiles of probabilities, which may be more common in this specific field. Kiefer had the expert define the minimum, maximum, the quartiles including the median, the 90% quantile and the mean of the parameter p. He set a and b to the minimum and maximum and determined α and β using relative squared loss.

Combining the Binomial likelihood and the truncated Beta prior, a posterior distribution of p is constructed. A summary statistic, or point estimator, is required in for example the Basel capital formula. Kiefer proposes the posterior mean or mode as two possible estimators and concludes they may be favoured over the maximum likelihood estimator.

3.2.3. Van der Burgt (2008): Calibrating Low-Default Portfolios, using the Cumulative Accuracy Profile

Van der Burgt proposes in this article the use of the Cumulative Accuracy Profile (CAP) to model low default portfolios. The CAP curve, or power curve, is normally used to assess the discriminative power of a model. It is closely related to the Receiver Operating Characteristic (ROC) [41]. In the context of default modelling, the CAP curve is constructed by sorting debtors from worse to better credit rating and then plotting the cumulative percentage of defaults against the cumulative percentage of obligors. A perfect credit rating system predicts all defaults happening in the worst grades, while a random system has no discriminatory power at all, see figure 3.2 for reference.

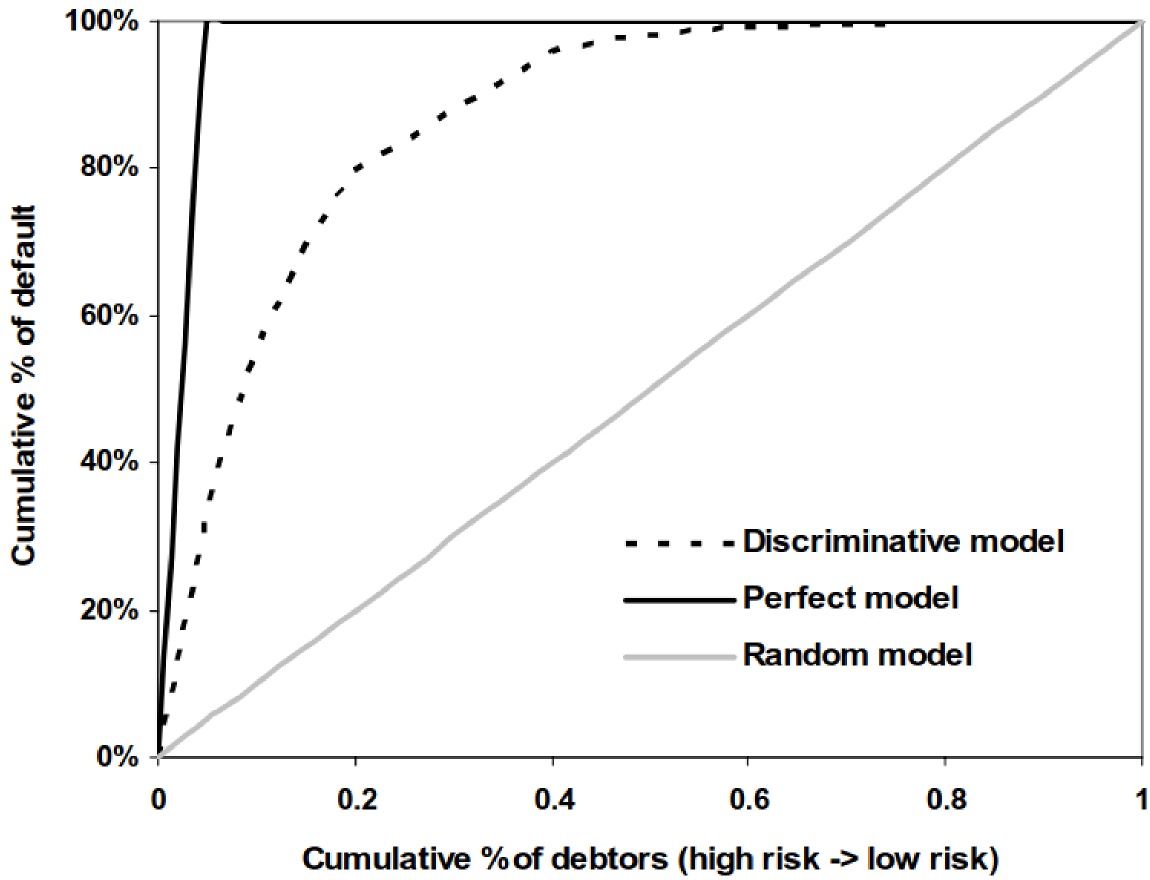

Figure 3.2: Example of Cumulative Accuracy Profile

[36]

Assuming a model has at least some discriminatory power, the CAP curve has a concave shape. The author then proposes to model the CAP curve by the following equation:

$$y(x) = \frac{1 - e^{-kx}}{1 - e^{-k}}$$

Where the parameter k is the so-called concavity parameter. This function is then calibrated to data from a rating agency such as S&P, minimizing the Root-Mean-Square error using the concavity parameter k to assure the best fit. After fitting the function to the data and finding an optimal value for k the probability of default in rating class R is obtained by the following result from [42]:

$$PD(R) = \hat{D}\frac{dy}{dx} = \frac{k\hat{D}}{1 - e^{-k}} e^{-kx_r},$$

where $\hat{D}$ is the average observed default rate in the whole portfolio (number of defaults/number of obligors), and x_r is some averaged mid-point to represent the cumulative percentage of counterparties in rating class R.

Van der Burgt then proceeds to apply the model first to artificial data and then to a real portfolio of sovereign defaults. He concludes the article by noting some flaws in the approach. The model is for example very sensitive to the rating class in which the defaults are observed, as the ordering of the obligors is what drives the CAP curve. The method also fails to produce nonzero estimates for a portfolio where no counterparties have defaulted, a property which is highly desirable.

3.2.4. Fernandes, Rocha (2011): Low default modelling - a comparison of techniques based on a real Brazilian corporate portfolio

Fernandes and Rocha considered in their article various techniques for the modelling of low default portfolios. A short summary of each technique is presented and afterwards the results are discussed.

Methods

Classical Logistic Regression Logistic regression is a model belonging to the family of Generalized Linear Models (GLM), see [43] for more information. The response variable $Y \in \{0,1\}$ is a dichotomous random variable modelled as a non-linear function of linear risk drivers. Denote Y_i as the observation of default event, attaining a value of 1 if company i defaulted and 0 otherwise. Write $\mathrm{x_i}$ as a vector of explanatory variables and β as a vector of weights, then:

$$P(Y_i \mid \mathrm{X_i} = \mathrm{x_i}) = \frac{1}{1 + e^{-\mathrm{x_i}\beta}}.$$

A maximum likelihood method is then conducted to find the optimal vector of weights $\hat{\beta}$.

Limited logistic regression This form of regression is the same as the classical logistic regression above, apart from the introduction of an extra parameter ω setting an upper bound on the probability:

$$P(Y_i \mid \mathrm{X_i} = \mathrm{x_i}) = \omega \frac{1}{1 + e^{-\mathrm{x_i}\beta}}.$$

Again maximum likelihood is used to find the vector of weights β and upper bound ω.

Bayesian logistic regression Consider again the classic logistic regression. Instead of the frequentist approach using the maximum likelihood method to find an estimate for β, Fernandes and Rocha also examined putting a Bayesian prior on the parameters and doing inference on its corresponding posterior. They claimed their prior was 'uninformative' and chose it to be $N(0, 10^6)$. Remark however that this is not true, since this is actually a subjective prior imposing the same probability on all values.

Artificial oversampling via SMOTE To deal with the problem of low data, a Synthetic Minority Oversampling Technique is a way to artificially create more data. Two data points $\mathrm{x_1}, \mathrm{x_2} \in \mathbb{R}^n$ are chosen randomly, and then a new observation is created by sampling uniformly from the hypercube defined by $\mathrm{x_1}$ and $\mathrm{x_2}$. After inflating the amount of observations a regular form of analysis can be applied to the new data set. This method of course introduces a bias in the estimation of PD. Fernandes and Rocha therefore propose a bias-correcting method as well.

Summary of results

The aforementioned methods are applied to a real Brazilian portfolio consisting of corporate defaults. Using a bootstrap method to reduce variability, two performance measures are used to test the effectiveness of the proposed approaches: The Gini coefficient and Kolmogorov test statistic. According to the first measure the limited logistic regression was the best model, while according to the second this was the classical logistic regression method.

3.2.5. Orth (2012): Default Probability Estimation in Small Samples – With an Application to Sovereign Bonds

In this publication Orth proposes the use of empirical Bayes in estimating small default probabilities. The resulting estimator is capable of multi-period predictions. First a standard estimator used in the industry will be described and afterwards the empirical Bayesian extension. Finally the resulting estimator is applied to a portfolio of sovereign bonds.

Standard Estimator

All the obligors belonging to rating class r at time $t = 1, \dots, T$ form a *cohort*. The number of obligors in that cohort at the start of the observation period is denoted by $N^r_{t,1}$ and the number that are still in the portfolio (not censored or defaulted) at time $t+s$ by $N^r_{t,s}$. Furthermore, $D_{t,s}$ are the number of defaults in period $t+s$ whereas $L^r_{t,s}$ are the number of censored obligors in that time-period. Now let λ^r_s be the marginal default rate, i.e. the probability of a company with rating r defaulting after s periods conditional on surviving $s-1$ periods. Denote by Y_{it} the lifetime of obligor i starting in time t and R_{it} as the rating of that company. The default rate is then given by the following probability:

$$\lambda^r_s = P(Y_{it} = s \mid Y_{it} > s-1, R_{it} = r).$$

Orth then notes that a standard estimator of this default rate used often in the industry is:

$$\hat{\lambda}^r_s = \frac{\sum_{t=1}^{T} D^r_{t,s}}{\sum_{t=1}^{T} N^r_{t,s} - \frac{L^r_{t,s}}{2}}.$$

Since the instantaneous default rate are often not of interest, but instead the cumulative default rates, Orth writes:

$$PD^r_s = P(Y_{it} \leq s \mid R_{it} = r),$$

with standard estimator:

$$\hat{PD}^r_s = 1 - \prod_{j=1}^{s} (1 - \hat{\lambda}^r_j).$$

Empirical Bayes extension

In his paper, the author also considers multiple portfolios but for convenience this, and dependence on t are omitted. Orth assumes the likelihood of the amount of defaults to admit a Binomial distribution:

$$D^r_s \sim \text{Bin}(N^r_s, \lambda^r_s).$$

Now, just as in the regular Bayesian setting, a prior is defined on the parameter λ^r_s. Since this default rate parameter lies in the unit interval, Orth has chosen to model this default rate using the flexible Beta distribution:

$$\lambda^r_s \sim \text{Beta}(\alpha^r_s, \beta^r_s).$$

Now instead of using expert judgement such as Kiefer does in [35], the empirical Bayesian approach uses data to estimate the hyperparameters. This is done by using a different, but comparable data set than the one under consideration. In this case, the portfolio consists of sovereign bonds of different rating grades. By using data of corporate bonds rated by the same rating agency the parameters of the priors are estimated. Then, for each rating grade a Beta posterior is constructed and the mean is used as a point estimator.

3.2.6. Pluto, Tasche (2005): Estimating Probabilities of Default for Low Default Portfolios

In this article, Pluto and Tasche introduce the concept of *most prudent estimation*. They estimate the PD by upper confidence bounds, making use of external credit rating grades. Suppose three rating grades, decreasing in credit quality, exist: A, B and C. Define the PD of these rating grades and the number of obligors in each rating as p_A, p_B, p_C and n_A, n_B, n_C respectively. Since the credit quality of obligors in rating grade A is better than those in B, which in turn is better than the credit quality of borrowers in rating grade C, the following assumption of inequality is reasonable:

$$p_A \leq p_B \leq p_C.$$

Pluto and Tasche consider both the case of dependent and independent defaults. They also distinguish between the case of no defaults and some defaults but both are captured by letting 'some' defaults include zero defaults. First the concept of most prudent estimation is applied to the single-period (year) case, after which also a multi-period extension is considered.

Independent defaults

Denote X as the random variable representing the amount of defaults and suppose k defaults are observed in the portfolio. Under the assumption of independence and equiprobability the amount of defaults follows a binomial distribution. The cumulative binomial distribution is given by

$$P\left(X \leq k\right) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}.$$

The most prudent principle is used to produce a conservative estimate for p_A. The principle postulates that companies in rating grade A have the same credit quality as those in rating B and C and therefore:

$$p_A = p_B = p_C.$$

So that p_C forms an upper bound on the probability of default in rating grade A. When the equation above is true the rating grades have equal credit quality and therefore the whole portfolio can be seen as homogeneous in terms of risk.

The most prudent estimator for p_A is then defined using a confidence interval for the probability of default. This confidence interval consists of all values not rejected at significance level α. A value is rejected when its p-value (probability of obtaining the same, or more extreme result than the sample under the null hypothesis) is smaller than α. The confidence level γ of a confidence interval is related to the significance level α through $1-\gamma = \alpha$. The confidence interval thus exists of all $p \in [0,1]$ for which holds:

$$P(X \leq k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} > \alpha = 1-\gamma. \tag{3.3}$$

The most prudent estimate is the smallest value of p so that equation 3.3 holds. Because of continuity the most prudent estimate is the solution of the equation:

$$\sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} = 1-\gamma.$$

According to the most prudent principle, the estimate for p_A should be based on all the obligors in the portfolio: n_A, n_B and n_C. Suppose k_A, k_B and k_C defaults are observed in each rating grade and set $k_{A+} = k_A + k_B + k_C$. The most prudent estimator for p_A is the solution of

$$\sum_{i=0}^{k_{A+}} \binom{n_A+n_B+n_C}{i} {p_A}^i (1-p_A)^{n_A+n_B+n_C-i} = 1-\gamma,$$

solved for p_A for a given confidence level γ.

Now, the most prudent estimate for p_B is analogous to the one of p_A, so set $k_{B+} = k_B + k_C$. However, the assumption that $p_A = p_B$ can not be made, since p_A is not an upper bound for p_B and this would violate the principle of most prudent estimation. Thus, the assumption is made that $p_B = p_C$ and rating grades B and C can be viewed as a homogeneous rating class. Therefore the most prudent estimator for p_B (for given confidence level γ) is the solution of

$$1-\gamma = \sum_{i=0}^{k_{B+}} \binom{n_B+n_C}{i} {p_B}^i (1-p_B)^{n_B+n_C-i}.$$

Now since p_C does not have an upper bound only observations in grade C are used for the estimation of p_C. Therefore, the most prudent estimator of p_C is the solution of

$$1-\gamma = \sum_{i=0}^{k_C} \binom{n_C}{i} {p_C}^i (1-p_C)^{n_C-i}.$$

Dependent defaults

Consider the same three rating grades as before and again suppose there are k defaults observed in the portfolio. Since defaults are now not assumed to be independent they are no longer Binomially distributed. For the modelling of dependence Pluto and Tasche propose the Vasicek model.

Recall that this model is based on given individual probabilities of default p_i and a one factor Gaussian correlation parameter ρ. This resulted in a conditional probability $\left[p_i^\rho \mid Y = y\right]$. This quantity is denoted by Pluto and Tasche as

$$G(p,\rho,y) = \Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho}\, y}{\sqrt{1-\rho}}\right). \tag{3.4}$$

Subsequently, this conditional probability is used as the probability parameter p in the Binomial distribution, giving the following probability mass function

$$P(X = k \mid Y = y) = \binom{n}{i} G(p,\rho,y)^i (1 - G(p,\rho,y))^{n-i}.$$

In order to obtain the *unconditional* probability of default, the normal factor Y needs to be integrated out. By denoting $\phi(y)$ as the probability density function of a standard normal random variable the unconditional probability mass function is

$$P(X = k) = \int_{-\infty}^{\infty} \phi(y) \binom{n}{k} G(p,\rho,y)^k (1 - G(p,\rho,y))^{n-k} dy,$$

and the cumulative distribution function is given by

$$\begin{aligned} P\left(X \le k\right) &= \sum_{i=0}^{k} P\left(X = i\right) \\ &= \int_{-\infty}^{\infty} \phi(y) \sum_{i=0}^{k} \binom{n}{i} G(p,\rho,y)^i (1 - G(p,\rho,y))^{n-i} dy, \end{aligned}$$

with the function G defined as in equation 3.4. Now, analogous to the independent case, the most prudent estimators for p_A, p_B and p_C are given as the solutions for the equations:

$$1-\gamma=\int_{-\infty}^{\infty}\phi(y)\sum_{i=0}^{k_{A+}}\binom{n_A+n_B+n_C}{i}G(p_A,\rho,y)^i(1-G(p_A,\rho,y))^{n_A+n_B+n_C-i}dy,$$

$$1-\gamma=\int_{-\infty}^{\infty}\phi(y)\sum_{i=0}^{k_{B+}}\binom{n_B+n_C}{i}G(p_B,\rho,y)^i(1-G(p_B,\rho,y))^{n_B+n_C-i}dy,$$

$$1-\gamma=\int_{-\infty}^{\infty}\phi(y)\sum_{i=0}^{k_C}\binom{n_C}{i}G(p_C,\rho,y)^i(1-G(p_C,\rho,y))^{n_C-i}dy.$$

Multi-period case

So far all results are derived under the assumption that companies were observed in one year. Consider now the case where data consisting of multiple years of observations is available. An approach could be to assume independence between years of observations, pool all this data per rating grade and proceed as in the single period case. However, some valuable relations in the data may be ignored this way.

Pluto and Tasche lifted the single period assumption and proposed a model which not only introduced *cross-sectional* asset correlation as in the original Vasicek model, but now extended the model to also include *inter-temporal* asset correlation. The framework of the model remains the same; there are three rating grades A, B and C with n_A, n_B and n_C obligors in each grade. At the end of the observation period $T > 1$ the amounts of default k_A, k_B and k_C are observed. The authors assume a company can only default once and only obligors present at the start of the observation period are considered.

Instead of only considering one point of maturity T as before, a company can now default at the end of each time period $t = 1, 2, \dots, T$. Still adhering to the underlying structure of Merton's model, company i defaults in time t when its asset value $V_{i,t}^{\rho}$ is below a certain level at that time. Pluto and Tasche model the intertemporal effect through the common factor and write for the now time-dependent asset value:

$$V_{i,t}^{\rho}=\sqrt{\rho}S_t+\sqrt{1-\rho}\;\xi_{i,t}.$$

Where the idiosyncratic component $\xi_{i,t} \overset{iid}{\sim} N(0,1)$ for all i and t. S_t on the other hand is no longer assumed to be time-independent. Let $\mathrm{S} = (S_1, S_2, \dots, S_T)$ denote the vector of common factors. The key assumption of the multi-period extension is:

$$\mathrm{S}\sim N(0,\Sigma),$$

where the entries of the correlation matrix are given by $\Sigma_{s,t}^{\theta} = \theta^{|s-t|}$ for some $\theta \in (0,1)$ and $s,t \in \{1,\dots,T\}$. Or in matrix form:

$$\Sigma_{\theta}=\begin{pmatrix}1 & \theta & \theta^2 & \cdots & \theta^{T-1}\\ \theta & 1 & \theta & & \vdots\\ \theta^2 & \theta & \ddots & & \vdots\\ \vdots & & & \ddots & \theta\\ \theta^{T-1} & & \dots & \theta & 1\end{pmatrix}\in(0,1]^{T\times T}. \tag{3.5}$$

Obligor i has defaulted when $V_{i,t}^{\rho}$ has at least been under the critical debt value once at the end of a time period $t = 1, \dots, T$. Recall that this critical debt value was equal to $\Phi^{-1}(p)$ under the Vasicek model, where p was the individual probability of default. The conditional probability of default of obligor i is therefore given by

$$\begin{aligned}P\,(\text{Obligor i defaults}\mid \mathrm{S}) &= P\left(\min_{t=1,\dots,T}V_{i,t}^{\rho}<\Phi^{-1}(p)\mid \mathrm{S}\right)\\ &=1-P\left(V_{i,1}^{\rho}>\Phi^{-1}(p),\dots V_{i,T}^{\rho}>\Phi^{-1}(p)\mid \mathrm{S}\right),\end{aligned}$$

conditional on S these probabilities are independent:

$$
\begin{aligned}
&= 1 - \prod_{t=1}^{T} P\left(V_{i,t}^{\rho} > \Phi^{-1}(p) \mid \mathrm{S}\right) \\
&= 1 - \prod_{t=1}^{T} P\left(\sqrt{\rho}S_t + \sqrt{1-\rho}\,\xi_{i,t} > \Phi^{-1}(p) \mid \mathrm{S}\right) \\
&= 1 - \prod_{t=1}^{T} P\left(\xi_{i,t} > \frac{\Phi^{-1}(p) - \sqrt{\rho}S_t}{\sqrt{1-\rho}} \mid \mathrm{S}\right) \\
&= 1 - \prod_{t=1}^{T} (1 - \Phi\left(G(p,\rho,S_t)\right)) \\
&= \pi(p,\rho,\mathrm{S}).
\end{aligned}
$$

The absence of dependence on the specific obligor is because the probability of default is assumed to be the same for all obligors in the portfolio. Once the underlying structure is defined, the authors apply the most prudent estimation principle to the multi-period model. Suppose again that X is the random variable representing the amount of defaults in the portfolio. Recall that just as in the single-period model, for a chosen confidence level γ and observed amount of defaults k the following equation needs to be solved for some p:

$$
\begin{aligned}
1 - \gamma &= P(X \le k \mid \mathrm{S}) \\
&= \sum_{l=0}^{k} P(X = l \mid \mathrm{S}) \\
&= \sum_{l=0}^{k} \binom{n}{l} \left(\pi(p,\rho,\mathrm{S})\right)^{l} \left(1 - \pi(p,\rho,\mathrm{S})\right)^{n-l}.
\end{aligned} \tag{3.6}
$$

Up to this point only conditional probabilities were considered. In order to use the multi-period model the unconditional default distribution needs to be derived. Denote the multivariate normal distribution with zero mean vector and covariance matrix Σ^{θ} as in equation 3.5 as $\phi_{\Sigma^{\theta}}(S_1,\dots,S_T)$. Then by writing out the unconditional cumulative distribution function of the amount of defaults X, equation 3.6 becomes:

$$
\begin{aligned}
&1 - \gamma = P(X \le k) = \\
&\int \dots \int \phi_{\Sigma^{\theta}}(s_1,\dots,s_T) \sum_{l=0}^{k} \binom{n}{l} \left(\pi(p,\rho,\mathrm{S})\right)^{l} \left(1 - \pi(p,\rho,\mathrm{S})\right)^{n-l} d(s_1,\dots,s_T).
\end{aligned}
$$

For large or even moderate levels of T it is not feasible to solve the above equation by means of numerical integration. Instead the authors suggest the use of Monte Carlo simulation. By setting the appropriate upper bounds $p_A = p_B = p_C$, $p_B = p_C$ and p_C the most prudent estimates for p_A, p_B and p_C are found respectively, using the corresponding number of obligors and defaults.

Concluding remarks

In this article Pluto and Tasche introduced a model for low default data according to the most prudent principle. It takes into account the number of obligors and amount of defaults for each rating grade in the portfolio and produces PD estimates for each of those grades, even when no defaults are observed. The estimates for the probability of default turn out to be fairly conservative, which can be seen as an advantage since risks are overestimated rather than underestimated. However, this will result in higher capital requirements for banks which in their turn cut into the profit.

A disadvantage of the most prudent approach is it may lead to non monotone (with respect to the rating grades) PD estimates when there are many defaults in high rating grades and less in lower grades. Another open issue is the choice of the confidence level, as this is of large impact on the PD estimates. This makes the model extremely sensitive to subjective influences. The authors note that high confidence levels may be overly conservative and therefore argue for levels that are moderate.

3.2.7. Tasche (2012): Bayesian Estimation of Probabilities of Default for Low Default Portfolios

In this section the publication of Tasche [39] from 2012 is followed. After his original article with Katja Pluto, Dirk Tasche revisited in this publication the most prudent estimation approach. The method was criticised because in the multi-period version three parameters had to be pre-defined: the cross-sectional correlation ρ, time-correlation θ and confidence level γ. While estimation methods for the first two parameters were supplied, the confidence level remained an open issue. Tasche proposes the Bayesian approach in this article to forego the need of choosing a confidence level.

First, the derivation of the conservative prior is discussed, as Tasche showed it has a special relation to the most prudent estimator. Second, estimators are derived when applied to the same three models considered in the article of Pluto and Tasche from 2005. An extension of the model and its results are discussed in chapter 5.

Estimators

Three estimators are considered in the article:

- Most prudent estimator
- Conservative Bayesian estimator
- Uniform Bayesian estimator

The Bayesian estimators correspond to the mean of the posterior distribution with one of the two prior distributions. See chapter 2 for more information on Bayesian statistics.

Conservative prior Tasche shows in [39] the relationship between the Most Prudent Estimator in the Binomial model and what he calls the 'conservative prior'. Consider a portfolio of n obligors and k defaults observed in a single time period. Recall that under assumption of independence the amount of defaults X is Binomially distributed and the most prudent estimator is for a given confidence level γ given as the solution of

$$1-\gamma = P(X \leq k) = \sum_{i=0}^{k} \binom{n}{i} p^i p^{n-i}.$$

Tasche notes the relation between the cumulative distribution functions of a Binomial and Beta variable. Let p b the probability parameter of a Binomial variable X and $Y \sim \mathrm{Beta}(k+1, n-k)$, then:

$$P(X \leq k) = 1 - P(Y \leq p).$$

Writing $F_Y^{-1}(\cdot)$ as the quantile function of Y, it is easily seen that the most prudent estimator $\hat{p}_{mpe}$ is obtained as the γ quantile of the $\mathrm{Beta}(k+1, n-k)$ distribution:

$$\hat{p}_{mpe} = F_Y^{-1}(\gamma).$$

A key finding in the article is the proposal of a certain prior distribution that relates the posterior to the most prudent estimator. Recall that a $\mathrm{Beta}(\alpha, \beta)$ prior in combination with a Binomial likelihood results in a $\mathrm{Beta}(\alpha + k, \beta + n - k)$ posterior for the probability parameter p. So by setting $\alpha = 1$ and $\beta = 0$

the previously seen Beta distribution is obtained. Now set the prior as the probability density function of a Beta distribution with those parameters and ignore constants not depending on p. The prior for p then becomes:

$$\pi(p) = \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$
$$\propto \frac{1}{1-p}.$$

This prior is called the *conservative* prior by Tasche. The corresponding posterior is a $\mathrm{Beta}(k+1, n-k)$ distribution. Since credible intervals are in a Bayesian approach defined as quantiles of the distribution, it is easily seen that the upper Bayesian credible bound of confidence level γ corresponds precisely to the most prudent estimator. Although these two estimators are related, they are not the same because the conservative estimator is not defined as an upper quantile but as the mean of the posterior.

Uniform prior In order to assess the level of prudentiality of the conservative prior, Tasche also considered the uniform distribution on the interval $(0, u)$ as a prior. Generally, two parameters representing the bounds of the interval have to be specified for the uniform distribution. Since a lower bound of 0 is very plausible in the situation of a low default probability only an upper bound is to be chosen. The prior is given by

$$\pi(p) = \frac{1}{u}, \qquad 0 < u \leq 1.$$

The larger this upper bound, the more conservative the resulting estimator will be. When $u = 1$ the uniform distribution can be seen as a $\mathrm{Beta}(1,1)$ distribution, since:

$$\frac{1}{B(1,1)} p^{1-1}(1-p)^{1-1} = 1.$$

Independent defaults, Binomial model

Consider again the independent default model described in the previous article, dealing with a homogeneous portfolio of n obligors, each having the same probability of default p. Under this model the amount of defaults has a Binomial distribution.

Most Prudent Estimator The most prudent estimator $\hat{p}_{MPE}$ is for confidence level γ given as the solution of

$$1 - \gamma = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}.$$

And as previously seen this can be written as:

$$\hat{p}_{mpe} = F_Y^{-1}(\gamma), \qquad \text{where } Y \sim \mathrm{Beta}(k+1, n-k).$$

Conservative Bayesian estimator A useful result is the formula of the mean of the Beta distribution. Suppose $Y \sim \mathrm{Beta}(\alpha, \beta)$, the mean of Y is then given by:

$$\mathbb{E}[Y] = \frac{\alpha}{\alpha+\beta}.$$

Repeating earlier results, the conservative prior is a special case of the $\mathrm{Beta}(1,0)$ distribution and together with a Binomial likelihood generates a $\mathrm{Beta}(k+1, n-k)$ posterior. Therefore, the conservative Bayesian estimator for the Binomial model is given by:

$$\hat{p}_{CBE} = \frac{k+1}{n+1}.$$

Uniform Bayesian estimator Tasche derives in [39] the posterior distribution using the uniform prior and shows it to be:

$$\pi(p \mid k) = \begin{cases} \frac{b_{k+1,n-k+1}(p)}{P(Y \le u)}, & \text{for } 0 < p < u, \\ 0, & \text{for } u \le p < 1. \end{cases}$$

With $Y \sim \text{Beta}(k+1, n-k+1)$ and $b_{k+1,n-k+1}$ its corresponding density function. Taking the expectation of this posterior, Tasche shows that the uniform Bayesian estimator for the Binomial model is given by

$$\hat{p}_{UBE} = \frac{(k+1)P(Y_{k+2,n-k+1} \le u)}{(n+2)P(Y_{k+1,n-k+1} \le u)}.$$

Where $Y_{\alpha,\beta} \sim \text{Beta}(\alpha, \beta)$. By making use of the fact that the Beta distribution is defined only on the unit interval, in the special case where $u = 1$ the UBE simplifies to:

$$\hat{p}_{CBE} = \frac{k+1}{n+2}.$$

Dependent defaults, Vasicek model

Most of the structure of the independent case will be retained, but the strict assumption of independence will be lifted. To model dependence, Tasche proposes the same 'Vasicek model' as in his paper of 2005, which is discussed in detail in chapter 2.

Consider again a homogeneous portfolio of n obligors, each having the same individual probability of default p. Also assume that the dependence of defaults is exclusively given by the asset correlation ρ. Furthermore, retain the intertemporal independence assumption of before. Recall that under this framework the probability of observing k defaults is given by

$$P(X = k) = \int_{-\infty}^{\infty} \phi(y) \binom{n}{k} G(p,\rho,y)^k (1 - G(p,\rho,y))^{n-k} dy,$$
$$G(p,\rho,y) = \Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho}\, y}{\sqrt{1-\rho}}\right).$$

Most prudent estimator The computation of the most prudent estimator $\hat{p}_{mpe}$ goes analogous to the Binomial case, only a numerical routine must be applied as no analytical solution is known. The most prudent estimator is given as the solution of

$$1 - \gamma = \int_{-\infty}^{\infty} \phi(y) \sum_{i=0}^{k} \binom{n}{i} G(p,\rho,y)^i (1 - G(p,\rho,y))^{n-i} dy.$$

Conservative and uniform Bayesian estimators Again, the summary statistic applied to the posterior distribution is derived from its squared error loss, corresponding to the posterior mean. However, because the Binomial likelihood has been replaced by the likelihood of the Vasicek model, analytical solutions do not longer exist. Therefore the posterior mean needs to be numerically evaluated as well. Denote the prior (either conservative or uniform) as $\pi(p)$, the posterior mean is then given by

$$\hat{p} = \frac{\int_0^1 p\, \pi(p) f(k \mid p,\rho) dp}{\int_0^1 \pi(p) f(k \mid p,\rho) dp},$$

where $f(k \mid p,\rho)$ is the likelihood of the Vasicek model. Note that no method for estimating the asset correlation ρ is specified. Tasche uses pre-defined values for ρ in the interval $[0.12, 0.24]$, the same range as used in the IRB formula.

Multi-period observations

Recall the multi-period extension of the Vasicek model of Pluto and Tasche described in the previous article. Assume each time period $t = 1, 2, \ldots, T$ has length of one year. Let X_t denote the random amount of defaults observed in year t and k_t the actual observed amount of defaults. The amount of obligors in the portfolio may also vary under this model, write therefore n_t as the amount of obligors in the portfolio in year t.

In this paper, Tasche revisits this model. Using the previously defined notation, the (unconditional) probability density function of the multi-period Vasicek model is given by

$$P\left(X_1 = k_1, \ldots, X_T = k_T \mid p, \rho, \theta\right) = \tag{3.7}$$
$$\int \ldots \int \phi_{\Sigma_\theta}(s_1, \ldots, s_T) \prod_{t=1}^{T} \binom{n_t}{k_t} G(p, \rho, S_t)^{k_t} (1 - G(p, \rho, S_t)^{n_t - k_t} ds_1, \ldots, ds_T.$$

Maximum likelihood estimator The maximum likelihood estimator maximises the likelihood given in equation 3.7. Given the amount of defaults $k_1, \ldots, k_T$ in the portfolio, the maximum likelihood estimates are given by

$$\left(\hat{p}_{MLE}, \hat{\rho}_{MLE}, \hat{\theta}_{MLE}\right) = \underset{p,\rho,\theta}{\arg\max} \int \ldots \int \phi_{\Sigma_\theta}(s_1, \ldots, s_T) \prod_{t=1}^{T} \binom{n_t}{k_t} G(p, \rho, S_t)^{k_t} (1 - G(p, \rho, S_t)^{n_t - k_t} ds_1, \ldots, ds_T.$$

There is obviously no analytical solution for the maximum likelihood estimator in the multi-period model. The function to be maximised involves a T-dimensional integral so applying the logarithm and maximising the log-likelihood neither gives analytical results. Therefore a Monte Carlo method has been applied by Tasche, where instead of integrating out the systemic factors $S_1, \ldots, S_T$ they are simulated a large number of times.

Following notation by Tasche, a sample of n independent realisation $(s_1^{(1)}, \ldots s_T^{(1)}, \ldots, s_1^{(n)}, \ldots s_T^{(n)})$ of the joint normally distributed systemic factors $S_1, \ldots, S_T$ are drawn. Then the likelihood function is approximated by

$$P\left(X_1 = k_1, \ldots, X_T = k_T \mid p, \rho, \theta\right) \approx \frac{1}{n} \sum_{i=1}^{n} \prod_{t=1}^{T} \binom{n_t}{k_t} G(p, \rho, s_t^{(i)})^{k_t} (1 - G(p, \rho, s_t^{(i)})^{n_t - k_t}.$$

There is one essential issue however, which is not addressed by Tasche in his paper. Not only is the Monte Carlo simulation needed to estimate the value of the likelihood, but in order to find the maximum of the likelihood function an optimisation routine is also used.

These routines, such as the 'BFGS' method found in the *optim* package in **R**, generally rely on some gradient search algorithm. However, if a naive approach is used and for every iteration in the optimisation routine a new sample is drawn from the multivariate normal with correlation matrix Σ_θ, the algorithm easily can get stuck because the function it tries to optimise is different in each iteration.

A solution would be to draw n samples and keep these the same during the optimisation routine. However, the objective is also to provide a maximum likelihood estimate for θ. The distribution from which is sampled depends on θ, so if only at the start of the optimisation a sample would be drawn, the algorithm has no way of updating θ to find the value for which the likelihood is maximised.

Therefore the influence of θ and the sample on the likelihood need to be separated from each other, so that the algorithm does not get stuck because of random variations of the target function but still has a way to evaluate it for different values of θ. The proposed solution makes use of the Cholesky decomposition [44]:

Definition 5 - Cholesky decomposition.

Suppose A is a Hermitian, positive definite matrix. Then the following expression

$$A = LL^T,$$

where L is a real-valued lower triangular matrix with positive entries and L^T its transpose, is known as the Cholesky decomposition.

If the matrix A in definition 5 is the variance-covariance matrix of a multivariate normal distribution and z is a vector of uncorrelated standard normal samples, then the vector Az is a vector of samples from the multivariate normal distribution with variance-covariance matrix A.

So in order to find the maximum likelihood estimate for θ, n samples of size T are drawn from a standard normal distribution, which are fixed for the whole optimisation routine. Then in each iteration of the algorithm, the Cholesky decomposition of Σ_θ is computed for the changing value of θ and used to produce a correlated sample.

There are several problems with estimating the parameters of the model using this technique. The first is mentioned before, but when zero defaults are observed the maximum likelihood estimator for p is $\hat{p}_{MLE} = 0$, while the true value of the PD is known to be positive, albeit small. Secondly, for larger problems such as Example 2 in the article of Tasche, the computation is slow, and still gets stuck often due to the very small values of the likelihood. Using the package *BBoptim* instead of *optim* seems to improve performance, but convergence is still slow and not always guaranteed.

The third issue is the number of observations with respect to the amount of parameters to be estimated. The sample $k_1, \dots, k_T$ is only one draw from the distribution given in equation 3.7, so the estimators will not be very robust. A possible solution to this problem could be to follow the approach by Tasche, which is pre-defining the correlation values.

Most prudent estimator In addition to the notation in the previous section, denote $X = X_1 + \dots + X_T$ as the random total number of defaults in the time period from $t = 1$ to $t = T$ and $k = k_1 + \dots + k_T$ as the actually observed total number of defaults in that time period. The most prudent estimator is, for a given confidence level γ, defined as the solution of

$$\begin{aligned} 1 - \gamma &= P(X \leq k) \\ &\approx \int \dots \int \phi_{\Sigma_\theta}(s_1, \dots, s_T) \exp\{-I(s_1, \dots, s_T)\} \\ &\sum_{j=0}^{k} \frac{I(s_1, \dots, s_T)^j}{j!} ds_1, \dots, ds_T, \end{aligned}$$

where

$$I(s_1, \dots, s_T) = \sum_{t=1}^{T} n_t G(p, \rho, s_t).$$

Here the Poisson approximation of the Binomial distribution is used, see chapter 4 for details. The approximation holds because the sum of Poisson distributed random variables is itself Poisson distributed with its intensity equal to the sum of the intensities.

Conservative and uniform Bayesian estimators Similar to the single period cases of both independent and dependent defaults, prior information on the parameters can be incorporated by the use of Bayesian techniques. Write the likelihood function in equation 3.7 as $f(\underline{k} \mid p, \rho, \theta)$. Write for the conservative or uniform prior $\pi(p)$, the mean of the posterior distribution is then:

$$\hat{p} = \frac{\int_0^1 p\ \pi(p) f(k \mid p, \rho, \theta) dp}{\int_0^1 \pi(p) f(k \mid p, \rho, \theta) dp}.$$

Following this approach both a pre-defined value for ρ and θ must be specified. For ρ again values in the interval $[0.12, 0.24]$ are proposed, and for θ Tasche has chosen a value of 0.6.

Concluding remarks

The paper of Pluto and Tasche from 2005 was criticised due to the fact that three values needed to be specified, ρ, θ and γ. Especially the last one proved difficult, as estimates are always very subjective. In this paper, Tasche showed a relation between the most prudent estimate and the conservative prior. Employing Bayesian techniques, the confidence level γ was no longer needed.

Still, there remain some open issues. The first problem is that in the multi-period model numerical results are no longer obtainable by numerical integration, so a sampling technique must be used. The second issue is that the approach still requires pre-specified values for the asset and time correlation parameters ρ and θ. Misspecification of one or both of these parameters may lead to underestimating the true probability of default. Both of these issues are addressed in the extension of the Vasicek model in chapter 5.

4

An extension: Credit Risk+

In 1997 Credit Suisse developed the CreditRisk+ [45] model, which is a portfolio approach to credit risk modelling. The main application of the CR+ model is for quantitative portfolio management. The model assumes the individual probability of default and its standard deviation of obligors to be known, and uses these quantities do derive the loss distribution and the related expected loss. Since the interest of this thesis is the estimation of default probability, some (slight) changes will be made to the CR+ approach. The model is based on the notion of *default rate*, which is actually the expected number of defaults in the portfolio. First, the default rate is assumed to be fixed, it can be shown that under some more assumptions, the number of defaults in a portfolio is Poisson distributed. Thereafter correlation between defaults is introduced on the basis of a common factor.

4.1. Fixed rate of default

Consider a portfolio consisting of N obligors, and define the probability of default of obligor i as p_i. Note that these probabilities do not have to be equal, but are fixed over a single period time horizon. Define the probability generating function of obligor i as:

$$G_i(z) = \sum_{n=0}^{\infty} P(\text{n defaults for i}) z^n$$

and since an individual company can only default or not,

$$\begin{aligned} &= \sum_{n=0}^{1} P(\text{n defaults for i}) z^n \\ &= (1 - p_i) z^0 + p_i z \\ &= 1 + p_i(z - 1), \end{aligned}$$

which of course is the probability generating function of a Bernoulli random variable. Now define $G(z)$ as the probability generating function of the whole portfolio, because of independence the following holds:

$$\begin{aligned} G(z) &= \prod_{i=1}^{N} G_i(z) \\ \log(G(z)) &= \sum_{i=1}^{N} \log(1 + p_i(z - 1)), \end{aligned}$$

since default probabilities are small, and $\log(1+x) \to x$ as $x \to 0$:

$$\log(G(z)) \approx \sum_{i=1}^{N} p_i(z-1)$$
$$G(z) = \exp\left(\sum_{i=1}^{N} p_i(z-1)\right) = e^{\lambda(z-1)}, \qquad \lambda = \sum_{i=1}^{N} p_i.$$

This is the probability generating function of a Poisson random variable. And so it can be concluded from the assumption that individual default probabilities tend to zero, the number of defaults in the portfolio is Poisson distributed with parameter λ. Now let $X \sim \text{Pois}(\lambda)$, then the first two moments of X are given by $\mathbb{E}[X] = \text{Var}(X) = \lambda$. However, historical evidence has shown the standard deviation of the number of defaults in portfolios to be larger than $\sqrt{\lambda}$. This leads to the consideration of uncertain rates of default discussed in the next section.

4.2. Uncertain rate of default

The extension of the model from a fixed rate of default to default rate uncertainty is done through the concept of sector analysis. Since the number of defaults may vary largely from year to year, one can view these as realisations of a random variable with an expected value equal to the average default rate over several years. This means that default probabilities *themselves* are random over time. Credit Suisse makes three summarising comments on this situation:

- The probability of default is not fixed over time, this holds also for obligors having the same credit quality.
- The variability of these probabilities stems from a (small) number of common economic factors
- Default rates are still rare events, so the Poisson approximation still holds

In the CR+ model the common economic factors influencing the portfolio are addressed by dividing the obligors into k different sectors: $S_1, \ldots, S_k$. Examples of these sectors could be macroeconomic, country- or industry related quantities. In this thesis it is assumed that all obligors in the portfolio are influenced by one and the same sector. Credit Suisse has shown in their original document how to derive the model for multiple sectors. In practice this means that analytical results no longer hold and one has to resort to numerical computation techniques.

A sector is driven by a single underlying factor, which influences the expected rate of defaults of that sector. The factor is modelled as a random variable X_k with mean μ_k and standard deviation σ_k. The expected amount of defaults in sector k is simply the sum of the individual probabilities

$$\mu_k = \sum_{i \in S_k} p_i.$$

The variability of defaults in each sector is only driven by the random variable X_k, so the probability of default for each *individual* obligor X_i will be modelled as a random variable proportional to X_k. Credit Suisse therefore makes the following assumption:

$$X_i = p_i \frac{X_k}{\mu_k}.$$

Now it is not hard to derive the relationship between the individual standard deviations and the one of the sector:

$$\begin{aligned}\sum_{i\in S_k}\sigma(X_i) &= \sum_{i\in S_k}\sigma\left(p_i\frac{X_k}{\mu_k}\right)\\ &= \frac{\sigma_k}{\mu_k}\sum_{i\in S_k}p_i\\ &= \frac{\sigma_k}{\mu_k}\mu_k\\ &= \sigma_k.\end{aligned}$$

So, both the mean and the standard deviation of the sector default rate are estimated as the sums over the means and standard deviations of the individual obligors belonging to that sector. In order to obtain the distribution of defaults under the variable default rate framework the probability generating function is again considered. Because of independence between sectors:

$$\begin{aligned}G(z) &= \sum_{n=0}^{\infty}P(\text{n defaults in the portfolio})z^n\\ &= \prod_{i=1}^{k}G_k(z).\end{aligned}$$

Therefore attention can be focussed on the probability generating function $G_k(z)$ of a single sector. Conditional on the common factor X_k the pgf is already known to be

$$G_k(z)|\,(X_k = x) = e^{x(z-1)}.$$

In order to obtain the unconditional probability the probability density function of X_k is set as some density $f_k(x)$. Then:

$$\begin{aligned}G_k(z) &= \sum_{n=0}^{\infty}P(\text{n defaults in }S_k)z^n\\ &= \sum_{n=0}^{\infty}z^n\int_0^{\infty}P(\text{n defaults in }S_k,\,X_k = x)dx\\ &= \sum_{n=0}^{\infty}z^n\int_0^{\infty}P(\text{n defaults in }S_k|X_k = x)f_k(x)dx\\ &= \int_0^{\infty}e^{x(z-1)}f_k(x)dx.\end{aligned}$$

The last step in obtaining the unconditional distribution of defaults is to specify the distribution of X_k. The CreditRisk+ model assumes this to be the Gamma distribution with mean μ_k and standard deviation σ_k. Before proceeding several properties of the Gamma distribution and its mixture with the Poisson are discussed.

4.2.1. Poisson Gamma mixture

The random variable X_k representing the default rate of sector k is assumed to be Gamma distributed. The two main reasons for this choice are the flexibility of the distribution and its analytic tractability when used as a mixture for the Poisson distribution. Before proving the last claim some properties of the Gamma are discussed.

Gamma distribution

The Gamma distribution is a continuous distribution on the domain $[0, \infty)$. There are several parametrizations of the Gamma distribution. For now, the choice of Credit Suisse of a shape parameter α and scale parameter β is followed. This leads to the probability density function:

$$f_G(x \mid \alpha,\ \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} e^{\frac{-x}{\beta}} x^{\alpha-1}, \tag{4.1}$$

where $\Gamma(z)$ denotes the gamma function (also defined for complex values), given by

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx.$$

A result quite often used in this thesis is the following identity, using partial integration:

$$\begin{aligned}
\Gamma(z+1) &= \int_0^\infty x^z e^{-x} \\
&= [-x^z e^{-x}]_0^\infty - \int_0^\infty -z x^{z-1} e^{-x} dx \\
&= 0 + \int_0^\infty z x^z e^{-x} dx \\
&= z\Gamma(z).
\end{aligned}$$

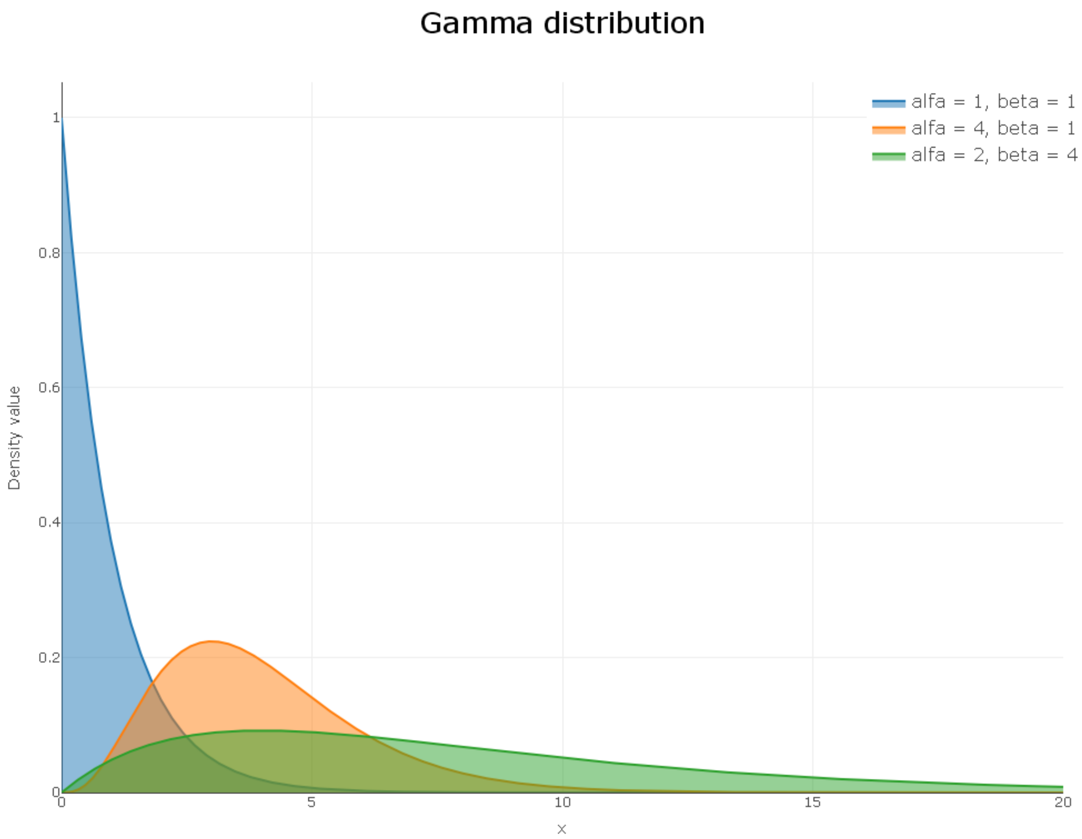

Figure 4.1: The probability density functions of three parametrizations of the Gamma distribution.

The flexibility of the Gamma is shown in figure 4.1. The mean and variance of the Gamma distribution are $\mathbb{E}[X_k] = \alpha_k\beta_k$, $\mathrm{Var}(X_k) = \alpha_k\beta_k^2$. In order to match the moments μ_k and σ_k are related to α_k and β_k through

$$\mu = \alpha_k\beta_k \text{ and } \sigma^2 = \alpha_k\beta_k^2,$$

or equivalently

$$\alpha_k = \frac{\mu^2}{\sigma^2} \text{ and } \beta_k = \frac{\sigma^2}{\mu}.$$

Mixture of Poisson and Gamma: Negative Binomial

The 'classical' definition of the probability density function $f_{NB}(k \mid r,\ p)$ of the negative binomial distribution is given by

$$f_{NB}(k \mid r,\ p) = \binom{r+k-1}{r} p^k(1-p)^r,$$

and gives the probability of observing k successes before r failures have occurred in a Bernoulli experiment with probability p. In this definition r is constricted to the positive integers, but using the following extension r can attain any positive real value:

$$f_{NB}(k \mid r,\ p) = \frac{\Gamma(r+k)}{k!\Gamma(r)} p^k(1-p)^r.$$

Denote $f_P(k \mid \lambda)$ as the density function of a Poisson random variable with parameter λ and $f_G(x \mid \alpha,\ \beta)$ as the density of a Gamma random variable with parameters $\alpha,\ \beta$. Now assume λ to be Gamma distributed, the unconditional distribution is derived as follows:

$$\begin{aligned}
P(X = k \mid \alpha,\ \beta) &= \int_0^\infty f_P(k \mid \lambda) f_G(\lambda \mid \alpha, \beta) d\lambda \\
&= \frac{1}{k!\beta^\alpha\Gamma(\alpha)} \int_0^\infty \lambda^{k+\alpha-1} e^{-\lambda(1+\frac{1}{\beta})} d\lambda \\
&= \frac{1}{k!\beta^\alpha\Gamma(\alpha)} \int_0^\infty \left(\frac{u}{1+\frac{1}{\beta}}\right)^{k+\alpha-1} e^{-u}(1+\frac{1}{\beta})^{-1} du \\
&= \frac{1}{k!\beta^\alpha\Gamma(\alpha)} \frac{\Gamma(\alpha+k)}{\left(1+\frac{1}{\beta}\right)^{\alpha+k}} \\
&= \frac{\Gamma(\alpha+k)}{k!\beta^\alpha\Gamma(\alpha)} \left(\frac{\beta}{1+\beta}\right)^{\alpha+k} \\
&= \frac{\Gamma(\alpha+k)}{k!\Gamma(\alpha)} \left(\frac{\beta}{1+\beta}\right)^{k} \left(\frac{1}{1+\beta}\right)^{\alpha},
\end{aligned}$$

which is the density of a Negative Binomial $NB\left(\alpha, \frac{\beta}{1+\beta}\right)$. If the parametrisation in terms of mean and standard deviation is followed, the density function is given by:

$$f_{NB}(k \mid \mu,\ \sigma) = \frac{\Gamma\left(\frac{\mu^2}{\sigma^2}+k\right)}{k!\Gamma\left(\frac{\mu^2}{\sigma^2}\right)} \left(\frac{\sigma^2}{\mu+\sigma^2}\right)^k \left(\frac{\mu}{\mu+\sigma^2}\right)^{\frac{\mu^2}{\sigma^2}}. \tag{4.2}$$

Now consider a Negative Binomial random variable $X \sim NB\left(\frac{\mu^2}{\sigma^2}, \frac{\mu}{\mu+\sigma^2}\right)$. The mean and variance of X are given by $\mathbb{E}[X] = \mu$, $\text{Var}(X) = \mu + \sigma^2$. This result should not be surprising, as it is actually a general property of mixed distributions.

Up to this point the CR+ model is described in its original form, omitting the derivations of loss distributions since this is not in scope of the research. It is interesting to remark however that Credit Suisse assumes the average individual probability of defaults and their standard deviations to be known, or estimated using the credit rating of the obligor to be more specifically. They assume these quantities to be the input of the model, whereas the output is the distribution of defaults.

Since the objects under investigation in this thesis are the individual PDs these can not be assumed to be known. Luckily, the amount of defaults in the portfolio *is* known and therefore the probability mass function of the Negative Binomial, $f_{NB}(k \mid \mu, \sigma)$, can be viewed as a likelihood. The estimation of the parameters of this likelihood is part of the next chapter.

4.3. Multi-period extension

In the original model of Credit Suisse no method of incorporating intertemporal dependence is mentioned. Generally, as is illustrated in the sovereign default data set in chapter 6, there is quite some time dependence found in the data. Therefore a multi-period extension of the Credit Risk+ model is considered here. A similar approach to the Vasicek model from chapter 3 is followed. Instead of using a multivariate normal distribution for the factor, a Gaussian copula is utilised with the same correlation matrix:

$$\Sigma_\theta = \begin{pmatrix} 1 & \theta & \theta^2 & \cdots & \theta^{T-1} \\ \theta & 1 & \theta & & \vdots \\ \theta^2 & \theta & \ddots & & \vdots \\ \vdots & & & \ddots & \theta \\ \theta^{T-1} & & \dots & \theta & 1 \end{pmatrix} \in (0,1]^{T \times T}.$$

Suppose there are k_t defaults observed in each time period $t = 1, 2, \dots, T$. Denote $F_{NB}(k \mid \mu, \sigma)$ as the cumulative mass function of the Negative Binomial distribution and write

$$u_t = F_{NB}(k_t \mid \mu, \sigma) \qquad t = 1, \dots, T.$$

Now, one could try to write the likelihood function of the multi-period CR+ model as the density of a T-dimensional Gaussian copula with correlation matrix Σ_θ:

$$f(k_1, \dots, k_T \mid \mu, \sigma, \theta) = c^{\Sigma_\theta}(u_1, \dots, u_T).$$

However, using copulas to model dependence between marginals when these are discrete brings quite some problems with it, see [46] for details. One of these issues, as described in the article, is that the copula density is no longer well-defined. An alternative method to define this density is found through the 'inclusion-exclusion' method, making use of the distribution function of the copula:

$$f(k_1, \dots, k_T \mid \mu, \sigma, \theta) = \sum_{j_1=0}^{1} \dots \sum_{j_T=0}^{1} (-1)^{j_1+\dots+j_T} \cdot C^{\Sigma_\theta}(F_{NB}(k_1 - j_1 \mid \mu, \sigma), \dots, F_{NB}(k_T - j_T \mid \mu, \sigma)).$$

The likelihood function for the multi-period extension of the CR+ model given above is now well-defined. There is however, a major drawback to this approach, namely the amount of computations needed. The copula distribution needs to be calculated a total of 2^T times. For small to moderate values of T this is feasible, but for 20 years of data the calculation of the likelihood requires more than a million computations of the copula distribution.

Even though this model seems not suited for portfolios of longer time-periods, there may be a way to significantly reduce the amount of computations. Since the portfolio under consideration is classified as low default, many of the observations include zero defaults and are therefore the same. This also means that many of the values of the distribution of the copula are the same, and only need to be computed once. Further research to the reduction of computations is therefore recommended. Performance of the multi-period extension is discussed in chapter 6.

5

Estimating the probability of default

In this chapter the single time-period Vasicek and Credit Risk+ models are investigated. First, estimators for the Vasicek model from chapter 3 are implemented and compared. Some estimators are proposed in the literature, but new ones are introduced as well. Second, new techniques used in the Vasicek model are also applied to the Credit Risk+ model from chapter 4, where some quite remarkable results are obtained. The multi-period extension of these models and their estimators are applied to a real data set consisting of sovereign defaults in chapter 6.

5.1. Vasicek model

Recall the Vasicek model described in chapter 3. Various estimators for the probability of default will be compared for different levels of correlation. The case $\rho = 0$ is equal to the independent, Binomial model. Moreover, $\rho = 0.12$ is denoted as 'low correlation' and $\rho = 0.24$ as 'high correlation'. Finally, a bivariate prior on both the probability parameter p and correlation ρ is considered.

5.1.1. Estimators

The article [39] of Tasche from 2012 showed promising results regarding the modelling of low default portfolios using Bayesian techniques. In order to more thoroughly investigate the use of priors in the Vasicek model, and compare them to other estimators, six different estimators are proposed:

- Maximum likelihood estimator
- Most prudent estimator
- Bayesian estimators with a
  - Conservative prior
  - Uniform prior
  - Pareto prior
  - Triangular prior

Maximum Likelihood Estimator

The MLE works especially well for large problems when the amount of observations n can be assumed to tend to infinity. It can for example be shown that the estimator is consistent, i.e. $\hat{\theta}_{MLE} \xrightarrow{P} \theta$. However, this is of course under the assumption that the parametrisation of the problem to the likelihood function is fully correct, which is never completely the case. Furthermore, analytical solutions only exist in very simple problems. For real life applications one therefore has to resort to numerical techniques, which may be unstable especially when the dimension of the parameter vector is high. Finally, it will be shown that the MLE is not conservative at all while this is one of the most important conditions under the Basel framework.

Most Prudent Estimator

Following both [32] and [39] the Most Prudent Estimator (MPE) finds it origin in confidence intervals and hypothesis testing. Suppose the null hypothesis $H_0 : p \geq p_0$ is tested against the alternative $H_1 : p < p_0$. When the null hypothesis can be rejected at significance level α $(= 1 - \gamma)$, an upper bound p_0 for p is found. Let X denote the random amount of defaults, depending on the probability parameter p and k the observed amount of defaults. The Most Prudent Estimator $\hat{p}_{MLE}$ is given by the solution of

$$1 - \gamma = P(X \leq k). \tag{5.1}$$

The main appeal of the MPE is the level of conservativeness. Especially for what constitutes in hypothesis testing as 'normal' levels of γ such as 0.95 or 0.99 the Most Prudent Estimator produces results that are conservative. However, one of the main points of criticism was that the estimator is *too* conservative, leading to high amounts of required capital by banks. Another disadvantage of this approach is that the confidence level γ must be prespecified, adding a factor of subjectiveness in the model.

Bayesian Estimators

The advantages and disadvantages of the Bayesian approach are extensively discussed in chapter 2. In this section the four Bayesian priors and their resulting estimators are investigated. A prior and a likelihood together form a posterior distribution, but in order to come to a point estimate a choice for the summary statistic has to be made. As seen previously, different type of loss functions give rise to different posterior estimators. Here the choice for squared error loss and therefore the posterior mean has been made. In chapter 3 it is shown that in the Binomial model the posterior mean for some priors is known analytically. In the dependent case when $\rho > 0$ the posterior mean is computed by numerical integration:

$$\hat{p} = \frac{\int_0^1 p \; \pi(p) f(k \mid p, \rho) dp}{\int_0^1 \pi(p) f(k \mid p, \rho) dp}.$$

Conservative prior The conservative prior is proposed by Tasche in [39], who showed its relation to the most prudent estimator. It is actually a Beta distribution with shape parameters $\alpha = 1$ and $\beta = 0$, given by

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

In figure 5.1 it can be seen that the prior has most of its mass closer to 1, thus putting more prior probability on a higher values of p.

Uniform prior To compare the results of the conservative prior, Tasche also proposed a uniform prior on the interval $[0, u]$, see figure 5.2. Generally, two parameters representing the bounds of the interval have to be specified for the uniform distribution. Since a lower bound of 0 is very plausible in the situation of a low default probability only an upper bound is to be chosen. The prior is given by

$$\pi(p) = \frac{1}{u}, \qquad 0 < u \leq 1.$$

Note that, in contrast with the conservative prior, when $u < 1$ the uniform prior puts zero prior probability on larger values of p. This causes the posterior to be more sensitive to misspecification of the prior. Some consider the uniform distribution as non-informative, but imposing equal probability on the domain is arguably as subjective as the other priors under consideration. Still, the uniform prior might be useful in cases where an expert has not much prior information except for the range in which the PD might lie.

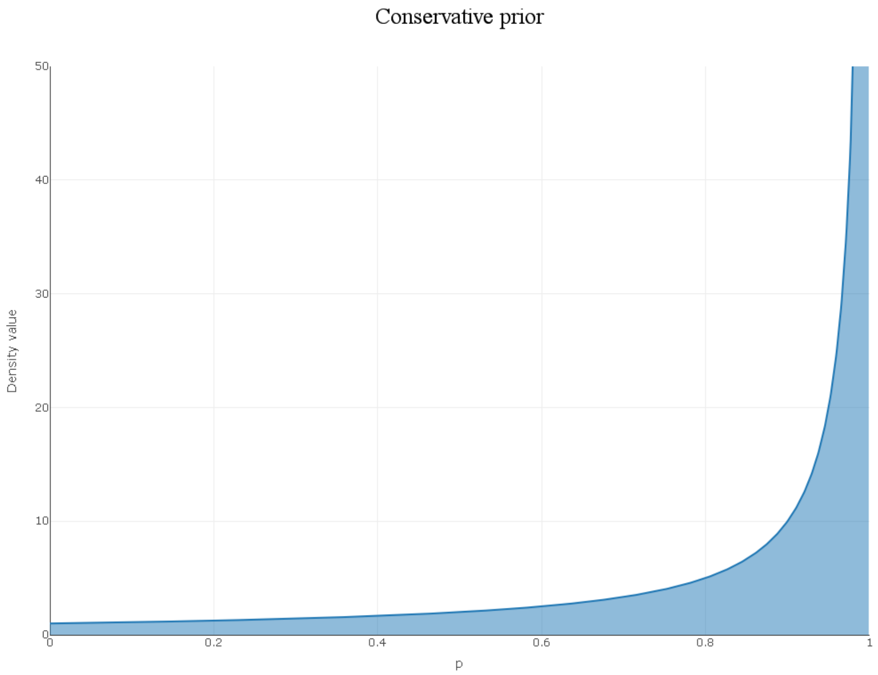

Figure 5.1: The conservative prior proposed by Tasche

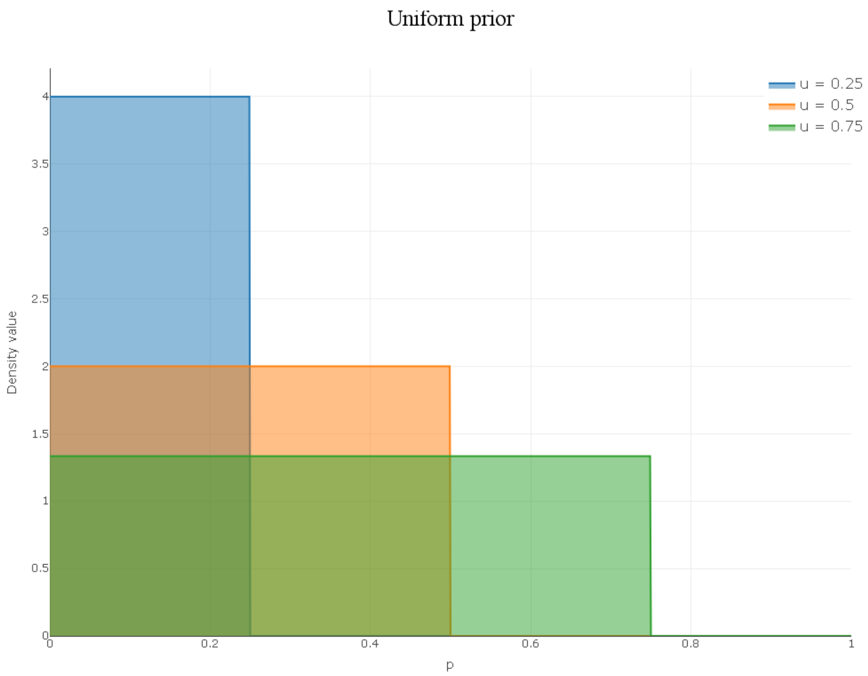

Figure 5.2: The uniform prior for various values of upper bound u

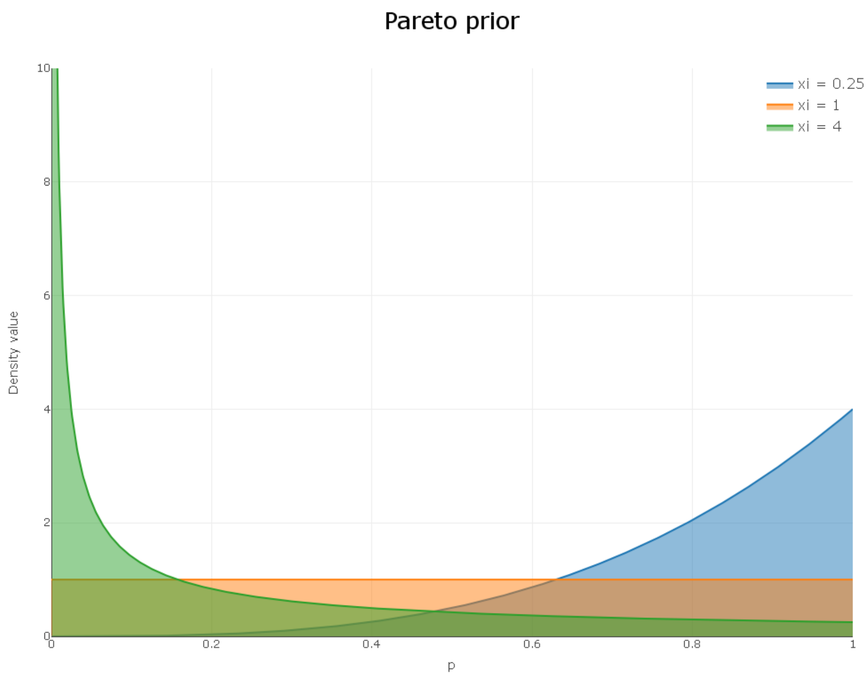

Figure 5.3: The Pareto prior for various values of parameter ξ

Pareto prior In his master thesis [47] Venter also considers the Bayesian approach in the modelling of low default portfolios. There he introduces a prior based on a transformation of a Pareto distribution. The probability density function of the Pareto distribution is given by:

$$f(\psi \mid \alpha, x_m) = \begin{cases} \frac{\alpha x_m^\alpha}{\psi^{\alpha+1}} & \text{for } \psi \geq x_m, \\ 0 & \text{for } \psi < x_m. \end{cases}$$

Venter chooses x_m to be 1 and sets $\alpha = \frac{1}{\xi}$, this then simplifies to the following density function:

$$f_\Psi(\psi \mid \xi) = \frac{1}{\xi}\psi^{-\frac{1}{\xi}-1}, \quad \psi \geq 1.$$

The domain of this density function is $[1, \infty)$. Since a prior for the probability parameter p is desired, the change of variable technique is applied. The variable Ψ is transformed to $P = g(\Psi) = \frac{1}{\Psi} = g^{-1}(\Psi)$. Thus the prior for p becomes:

$$\begin{aligned} \pi_P(p) &= \left|\frac{d}{dp}(g^{-1}(p))\right| \cdot f_\Psi(g^{-1}(p) \mid \xi) \\ &= \frac{1}{p^2}\frac{1}{\xi}\left(\frac{1}{p}\right)^{-\frac{1}{\xi}+1} \\ &= \frac{1}{\xi}p^{\frac{1}{\xi}-1}. \end{aligned}$$

See figure 5.3 for a plot of the Pareto prior. This prior is very flexible in comparison with the other priors seen so far. For higher values of ξ the distribution becomes more conservative, while lower values of ξ allow for (generally speaking) more realistic prior knowledge. Also, there is a direct relationship between the Pareto prior with parameter ξ and the $\text{Beta}(\frac{1}{\xi}, 1)$ distribution:

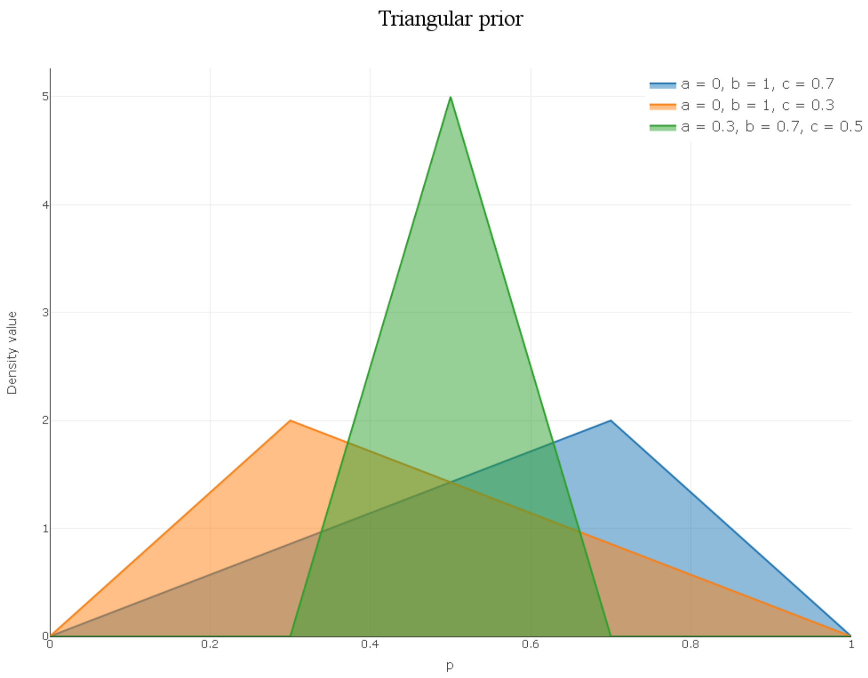

Figure 5.4: The triangular prior for various values of the parameters a, b and c

$$\frac{1}{\xi} p^{\frac{1}{\xi}-1} \propto \frac{1}{B(\frac{1}{\xi}, 1)} p^{\frac{1}{\xi}-1} (1-p)^{1-1}.$$

Triangular prior The triangular distribution $T(a, b, c)$ is defined by three parameters: a lower bound a, upper bound b and mode c. The prior resulting from this distribution is given by:

$$\pi(p) = \begin{cases} 0, & \text{for } p < a, \\ \frac{2(p-a)}{(b-a)(c-a)} & \text{for } a \leq p < c, \\ \frac{2}{(b-a)} & \text{for } p = c, \\ \frac{2(b-p)}{(b-a)(b-c)} & \text{for } c < p \leq b, \\ 0 & \text{for } b < p. \end{cases}$$

For possible representations of the triangular prior, see figure 5.4. A special case of the triangular distribution arises as the average of two independent standard uniform variables U_1 and U_2 [48]:

$$\frac{U_1 + U_2}{2} \sim T\left(0, 1, \frac{1}{2}\right).$$

Which is a symmetrical triangular distribution on the unit interval. By definition the domain of the triangular prior for the parameter p is confined to the unit interval, but just as in the uniform case smaller bounds may be chosen. The triangular distribution is quite often used as a prior when incorporating expert judgement, since an expert only needs to specify a range and a mode which are more instinctive than the estimation of quantiles.

Performance metrics

Much attention is put on the special case of zero defaults, but all other possible amount of defaults can of course not be ignored. It would be unfeasible to compare the six estimators for every default $k = 0, 1, \ldots, n$ and therefore some performance metric could be of help. An important property of an

Table 5.1: Probability of underestimating the default for a true PD of $p = 0.55$

k	$P(X = k)$	$P(X \leq k)$	$\hat{p}$
0	0.1	0.1	0
1	0.15	0.25	0.2
2	0.25	**0.5**	0.4
3	0.25	0.75	**0.6**
4	0.15	0.9	0.8
5	0.1	1	1

estimator is, in view of the Basel regulation, being conservative and therefore one of the two criteria on which an estimator will be judged. While being conservative is important, high PD estimates mean a higher required capital for banks which lead to increased costs. Therefore, accuracy is the other property on which estimators will be compared.

Conservativeness The choice in this thesis has been made to measure the level of conservativeness of an estimator based on the *Probability of Underestimating the true Probability of Default* (PU_{PD}). Given a true parameter p the Binomial probability mass function gives the probability of observing the amount of defaults $k = 0, 1, \dots, n$. For each value of k the estimators under consideration give a monotone increasing PD estimate. For the lower values of k they might underestimate the true PD while for others they will overestimate it, of course depending on the estimator. The cumulative probability of the largest value of k for which an estimator underestimates the true PD is defined as the PU_{PD}.

An example is provided in table 5.1. The table gives the possible defaults $k = 0, 1, \dots, 5$ in the first column. The probability of observing each amount of default given the true value of the PD is given in the second column, while the cumulative probability is written in the third. The final column shows the estimate of the PD according to the estimator of which the PU_{PD} is computed. Now all one has to do is to look up the largest value of k for which $\hat{p}$ is lower than the true value $p = 0.55$, which gives a PD_{PU} of 0.5.

Accuracy The accuracy of an estimator is measured as the *(relative) Mean Squared Error* (MSE). For each value of k the squared difference between the estimator and the true value of the PD is taken and multiplied by the probability of observing k defaults, or in formula:

$$MSE = \frac{1}{n} \sum_{k=0}^{n} P(X = k) \left(\hat{p} - p\right)^2 .$$

The ideal estimator would give the perfect prediction of p for every value of k. This is surely unattainable in practice, but some estimators do perform better than others in terms of accuracy.

5.1.2. Independent defaults

The aforementioned estimators are applied to the Binomial model. Point estimates and confidence intervals can be computed for any amount of observed defaults k. The special case of no default is of particular interest, since the conventional method of maximum likelihood fails indisputably. See figure 5.5 for the performance of the six estimators when no defaults are observed.

As expected, the maximum likelihood estimator is equal to zero and when no defaults are observed it is also impossible to compute the corresponding confidence interval. The most prudent estimator is shown in the literature to be a very conservative estimator, but has a PD estimate comparable to the Bayesian estimators, with the exception of the Pareto estimator. The Pareto estimator has the lowest PD of the Bayesian estimators. Interestingly, the triangular estimator is even more conservative than the most prudent estimator. Note also that the conservative and uniform point estimates are very close, although the conservative has a wider confidence interval. In order to investigate this behaviour, the

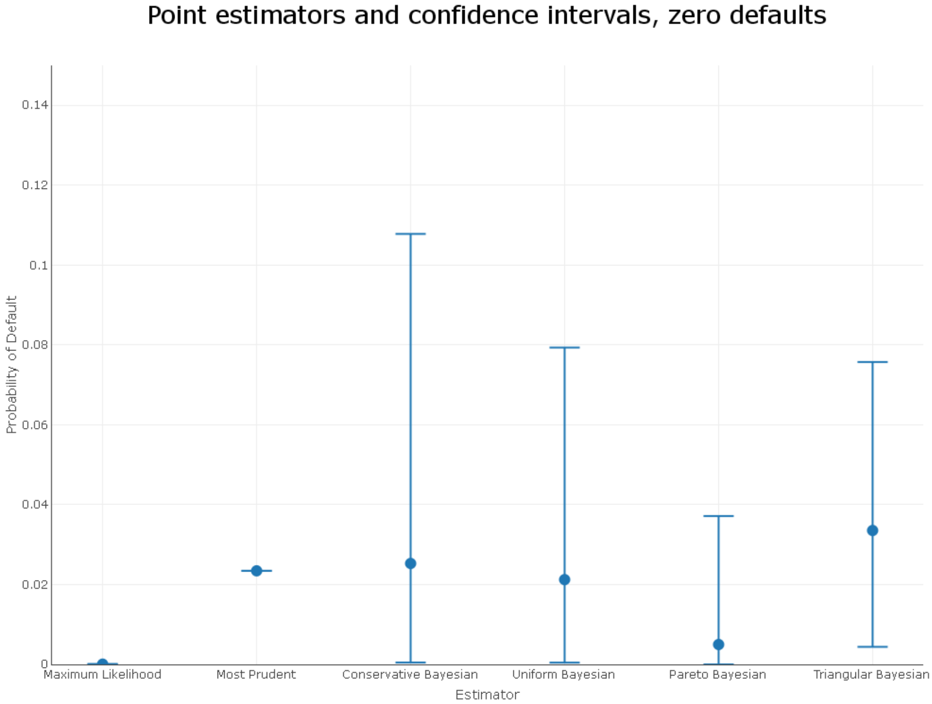

Figure 5.5: Estimators for the Binomial model with their parameters given by: MLE(), MPE($\gamma = 0.75$), CBE(), UBE($u = 0.1$), PBE($\xi = 4$), TBE($a = 0, b = 0.1, c = 0.05$)

four priors and the likelihood function (for $k = 0$) have been plotted in respectively figure 5.6 and figure 5.7 for $0 \leq p \leq 0.1$.

First please note that the y-axes for the priors are very different. However, they can be compared independent of scale since only the shape of the prior matters. After all, the product of the prior and likelihood gets normalised to form a posterior distribution.

Notice the likelihood function has almost all its mass located on those p smaller than 0.05. Since the likelihood for values of p larger than 0.05 is so small, it does not really matter what value the prior attains there; multiplying by something practically zero is still zero (unless the prior probability tends to infinity there). Since the triangular prior is the most conservative on the same interval the likelihood still has mass, it makes sense that the triangular estimator has the highest PD estimate for $k = 0$. Also, by looking at the plot of the conservative and uniform prior for $0 \leq p \leq 0.1$ it is not surprising their posterior estimates are practically identical.

In the Bayesian setting usually a prior is specified before an experiment is conducted and afterwards used in combination with the data (likelihood) to form a posterior. This example has shown how important it is to review the prior and likelihood after analysis is done. One would expect a priori that the conservative prior produces conservative estimates as it puts the most probability on higher values of p. Upon closer inspection however, for values where the prior actually has an impact (lower than 0.05) the conservative prior is approximately uniform. Please let this be a warning for the reader!

In figure 5.8 the performance in terms of both conservativeness and accuracy of the six estimators is shown for a true PD value of $p = 0.01$. Results for other values of the true PD are similar when it is small.

Note that the estimators only attain three different levels of the PU_{PD}. This is not a coincidence, as the distribution of k is discrete and therefore estimators have the same probability of underestimating the

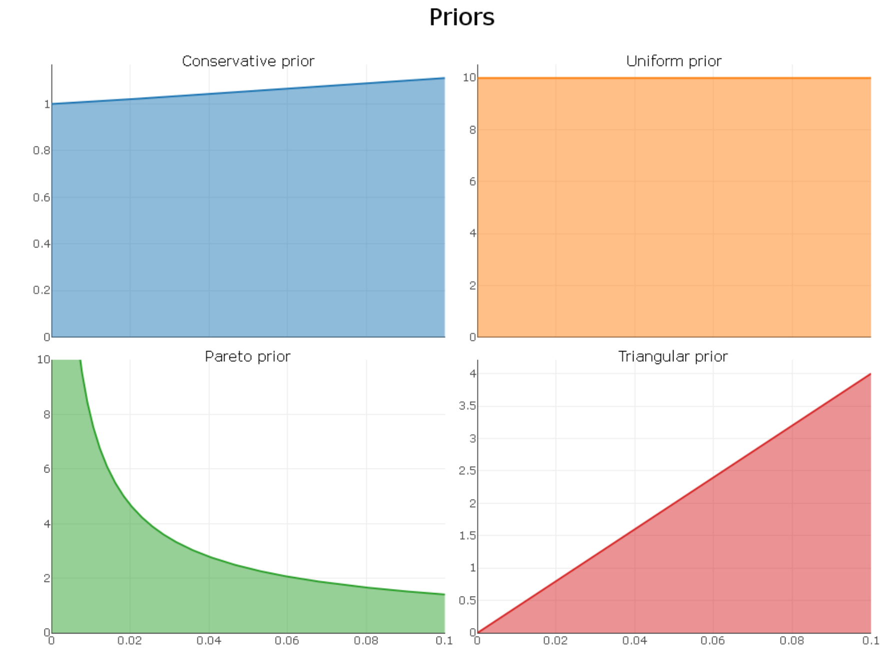

Figure 5.6: The four Bayesian priors with their parameters given by: CBE(), UBE($u = 0.1$), PBE($\xi = 4$), TBE($a = 0, b = 0.1, c = 0.05$)

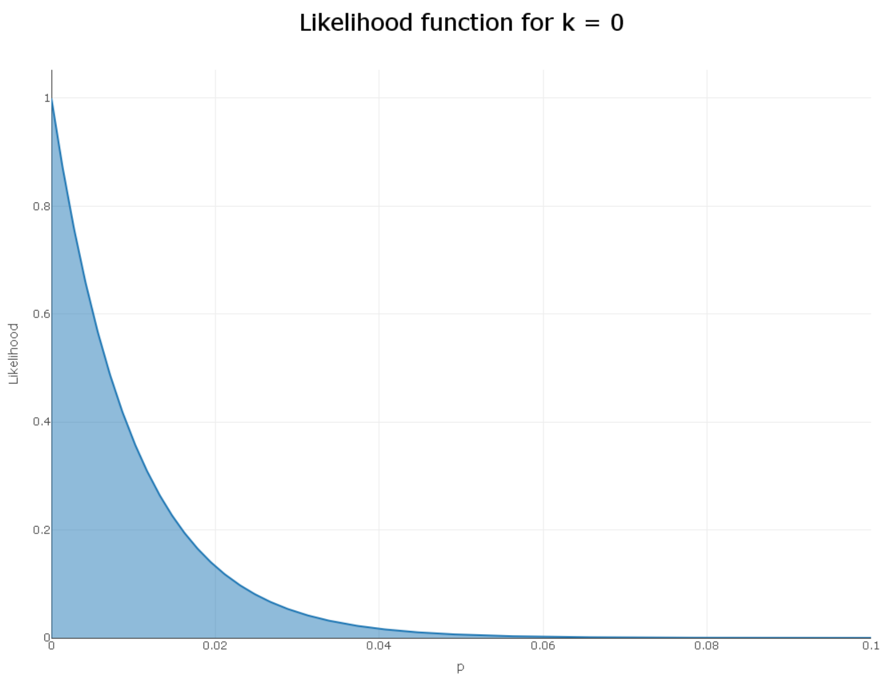

Figure 5.7: Likelihood function of a Binomial distribution for 0 observed defaults, plotted for $0 \leq p \leq 0.1$

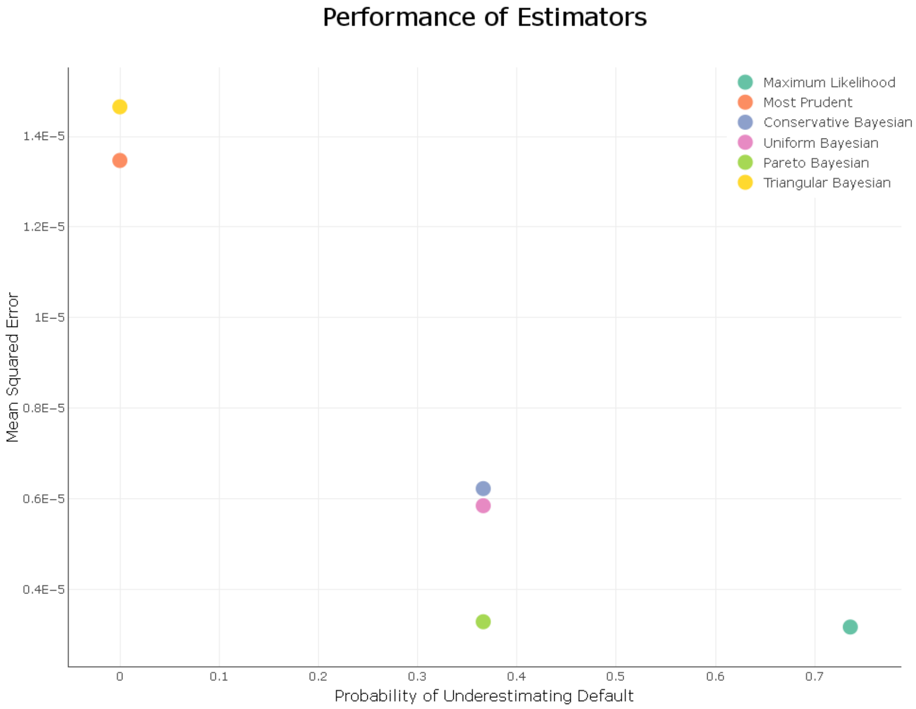

Figure 5.8: Performance of the six estimators for the Binomial model in terms of conservativeness and accuracy when the true value of the PD is $p = 0.01$

PD. There is also a clear relation between the level of conservativeness and the accuracy; the more conservative an estimator is, the lower its respective accuracy.

The most prudent and triangular estimators are the most conservative, just as seen in figure 5.5. The most prudent estimator would be slightly preferred based on this plot as it is more accurate than the triangular. The maximum likelihood estimator is the least conservative, as earlier results would suggest. An advantage of this estimator is that it is the most accurate of ones under consideration. Three estimators have a moderate level of conservativeness, the conservative, uniform and Pareto. Of those three the Pareto estimator has the highest accuracy, almost the same as the maximum likelihood. Based on figure 5.8 the Pareto prior and its corresponding posterior mean would be a very reasonable estimator.

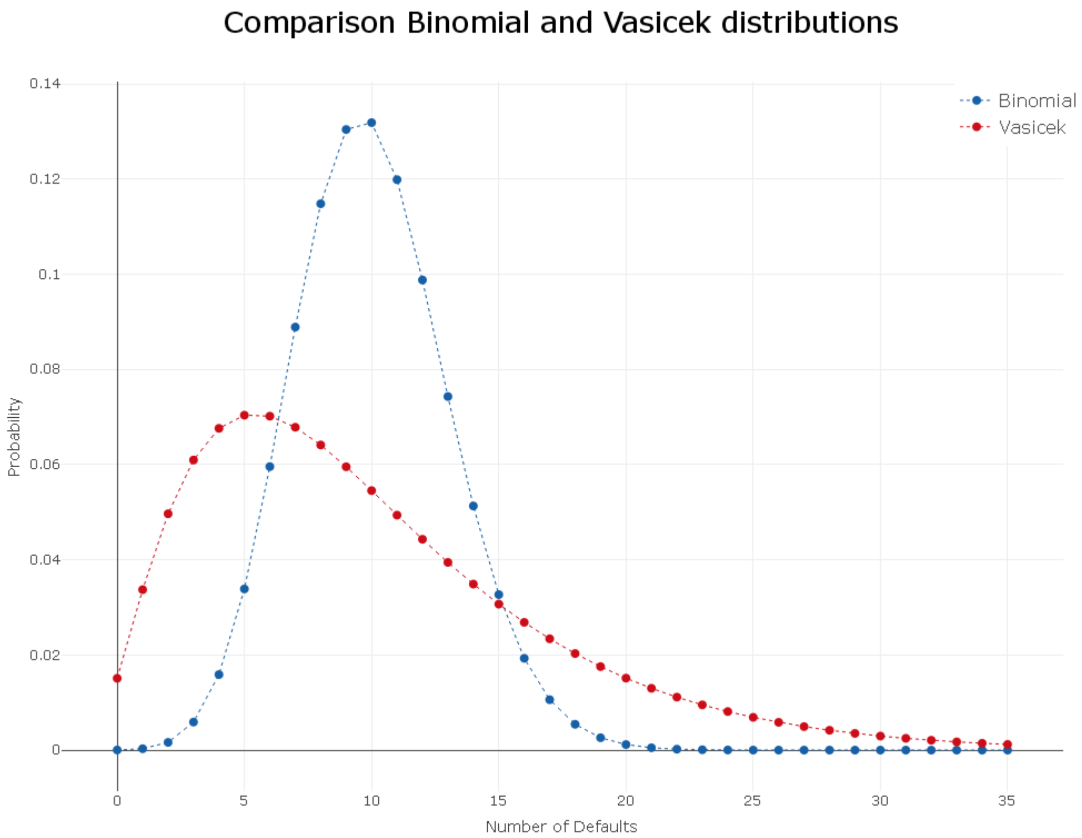

Figure 5.9: Comparison of the Binomial and Vasicek distribution. Probability density functions have been plotted for parameters $p = 0.1$ (both models) and $\rho = 0.12$ (Vasicek).

5.1.3. Dependent defaults

Now that the case of independent defaults has been carefully examined and the results discussed, the dependent case is investigated. Consider again a homogeneous portfolio of n obligors, each having the same individual probability of default p. Also assume that the dependence of defaults is exclusively given by the asset correlation ρ. Furthermore, retain the intertemporal independence assumption of before. Recall that under this framework the probability of observing k defaults is given by

$$P(X = k) = \int_{-\infty}^{\infty} \phi(y) \binom{n}{k} G(p, \rho, y)^k (1 - G(p, \rho, y))^{n-k} dy,$$

$$G(p, \rho, y) = \Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho}\, y}{\sqrt{1 - \rho}}\right).$$

In figure 5.9 a comparison between the independent Binomial, and dependent Vasicek, models is shown. The Vasicek distribution assigns more probability to the more extreme (lower and higher) number of defaults.

Since only one homogeneous portfolio in one time period is considered, only one observation is available. It would therefore be unfeasible to produce a robust estimate for both the individual PD p and the asset correlation ρ. Hence the asset correlation ρ is assumed to be known for know, later in this section this assumption will be relieved.

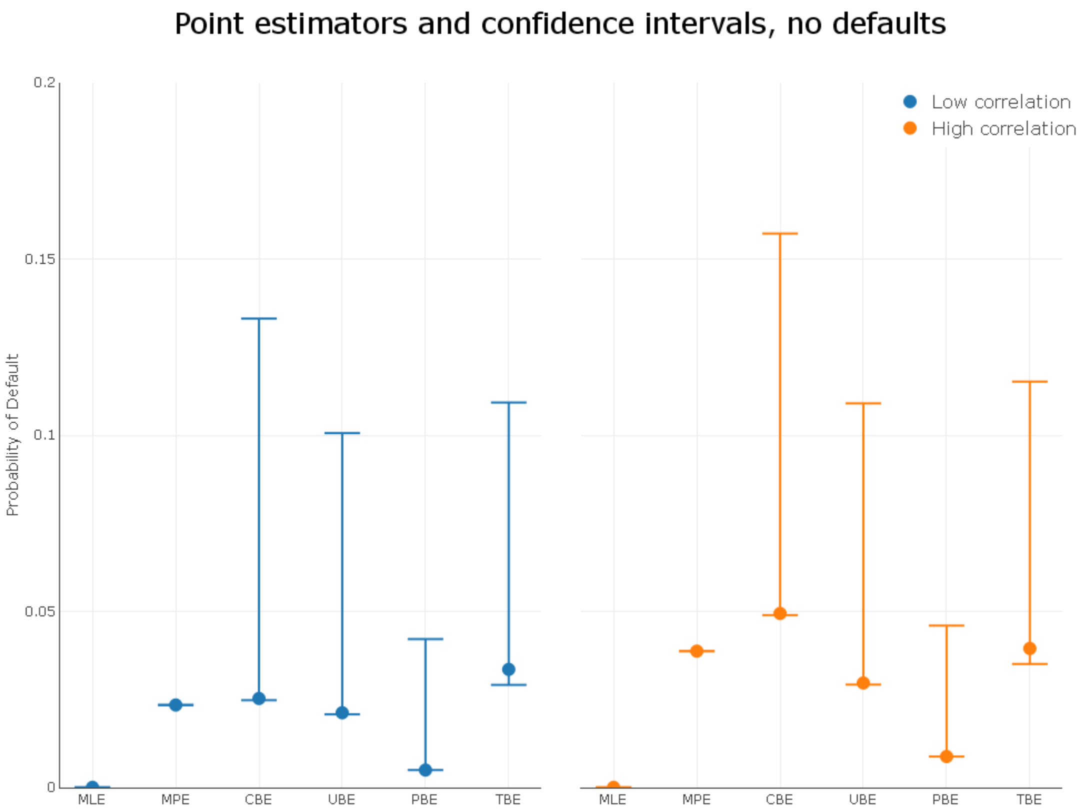

Figure 5.10: Estimators for the Vasicek model with their parameters given by: MLE(), MPE($\gamma = 0.75$), CBE(), UBE($u = 0.1$), PBE($\xi = 4$), TBE($a = 0, b = 0.1, c = 0.05$)

Maximum Likelihood Estimator

Contrary to the Binomial model, the likelihood function in the Vasicek model has no analytical solution. Therefore an optimization routine is applied where inside each iteration the integral is numerically evaluated. This, as all numerical computations and producing of plots, is done with the statistical software **R** [49]. The MLE of the of Vasicek model is

$$\hat{p}_{MLE} = \arg\max_{p} \int_{-\infty}^{\infty} \phi(y) \binom{n}{k} G(p,\rho,y)^k (1 - G(p,\rho,y))^{n-k} dy.$$

Results

In order to compare the performance of the estimators an asset correlation level ρ needs to be specified. The lower and upper bound for ρ specified by the BCBS are 0.12 and 0.24 respectively. Although the Basel Committee does not claim these are reasonable input parameters for other models, these levels seem satisfactory for testing purposes. From now on $\rho = 0.12$ is known as the 'low correlation' case and $\rho = 0.24$ as the 'high correlation'.

In figure 5.10 point estimators and their confidence intervals are shown for the no default case. The maximum likelihood estimator again fails to produce a nonzero estimate, which is as expected. The most prudent estimator is more conservative in the high correlation case than in the low correlation one. Considering the Bayesian estimators, the triangular Bayesian estimator is just as in the Binomial model the most conservative for a low correlation level. In the high correlation case the conservative Bayesian estimator increases its conservativeness drastically. The Pareto Bayesian estimator has again the lowest PD estimates of the Bayesian estimators and its confidence interval is not overlapping with the conservative estimator.

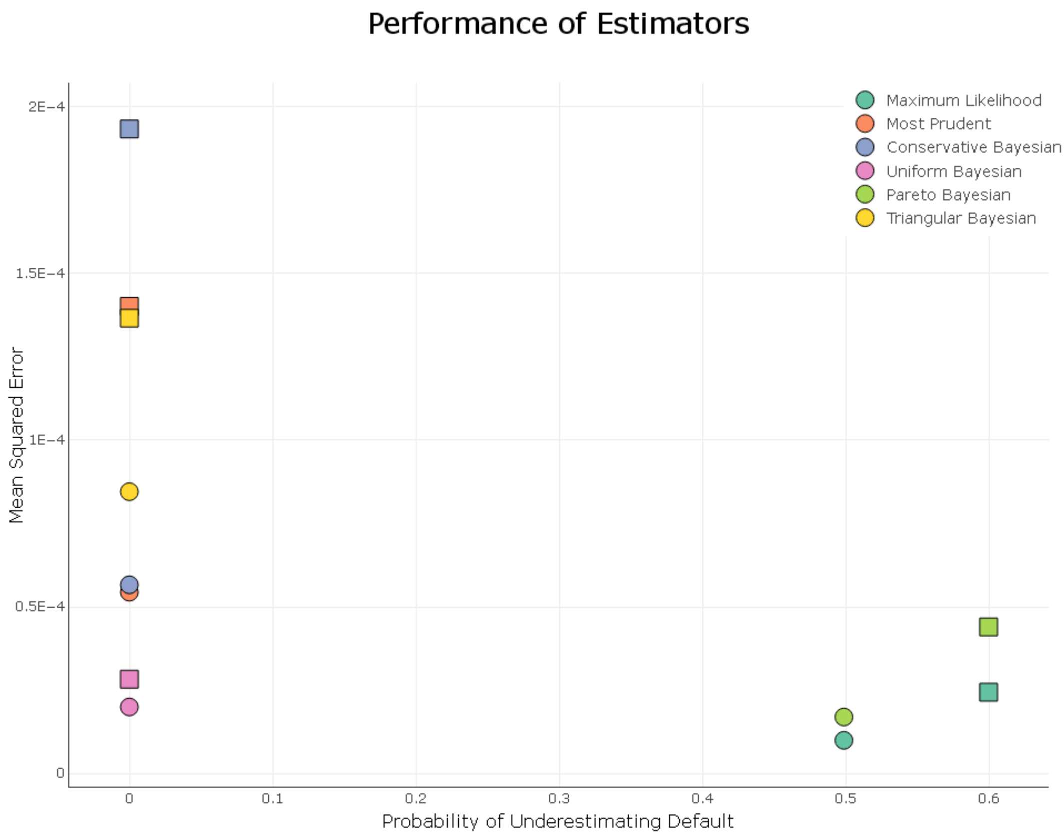

Figure 5.11: Performance of the six estimators for the Vasicek model in terms of conservativeness and accuracy when the true value of the PD is $p = 0.01$. Circles are for the low correlation model, squares for the high correlation.

While the zero default case gives a reasonable indication of the level of conservativeness of the estimators, for a more complete view the same technique as in the Binomial case is deployed. The accuracy of each estimator, measured by the relative mean squared error, is plotted against the conservativeness, measured by the probability of underestimating the true PD. The results are shown in figure 5.11.

For all estimators, results are better in the low correlation case than in the high correlation. This can be seen by the fact that for $\rho = 0.24$ each estimator has a higher MSE and equal or higher PU_{PD} than for $\rho = 0.12$. The maximum likelihood and Pareto estimators are again the least conservative, where the maximum likelihood outperforms the Pareto estimator in terms of accuracy. The other estimators all have zero PU_{PD} for a true PD of 0.01. A special remark has to be made on the performance of the uniform Bayesian estimator because it is both conservative and has an MSE comparable to the maximum likelihood estimator.

5.1.4. Bivariate prior

As figure 5.10 has shown, the difference in PD estimates for different values of ρ is quite large. Therefore there is a large risk of misspecifying this value, possibly leading to the underestimation of the true probability of default. Therefore, instead of assuming ρ to have a fixed value, it is now proposed to incorporate prior knowledge about the asset correlation by employing the Bayesian technique. Hopefully, this will reduce the variability of estimators and provides a more robust way to estimate the PD in case of low defaults.

In order to model prior information on ρ a uniform distribution on $[0, 1]$ is chosen. Another possible choice would be a uniform prior on $[0.12, 0.24]$, similar to the earlier fixed values of ρ. However,

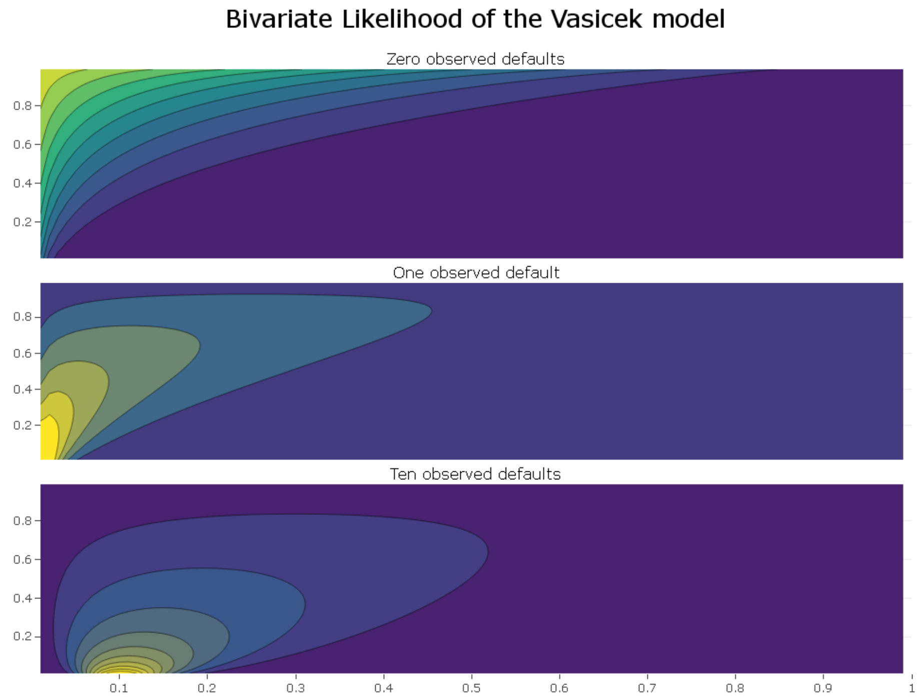

Figure 5.12: The bivariate likelihood of the Vasicek model for $k = 0, 1, 10$. The plots do not share the same scale.

contrary to the probability parameter for which prior information is available (since the portfolio under consideration is known to be low-default), no viable information for ρ is known. Remark that no claim of objectivity is made; other priors with the unit interval as domain are possible candidates as well.

Now, two priors have to be specified; one for the probability p and one for the correlation ρ. Together, these priors can be viewed as a *bivariate* prior. For clarity's sake, just two priors for the probability parameter p are chosen. For low values of p it is seen that the conservative and uniform priors are similar, so only the uniform one is used in this case. Also, the Pareto prior still seems able to capture the most realistic distribution of p and therefore it is also incorporated in the analysis.

One essential component is omitted up to now, the dependence structure between the correlation ρ and the individual PD p. First independence is assumed between the correlation and PD, thereafter a Clayton copula is utilised. Suppose the univariate priors for p and ρ are specified together with their dependence structure. If the bivariate prior is denoted as $\pi(p, \rho)$, the bivariate posterior distribution is given by

$$\pi(p, \rho \mid k) = \frac{\pi(p, \rho) f(k \mid p, \rho)}{\int_0^1 \int_0^1 \pi(p, \rho) f(k \mid p, \rho) dp d\rho}. \tag{5.2}$$

Remark the likelihood function is now a function of both p and ρ and thus also bivariate. Figure 5.12 shows the behaviour of the likelihood for various values of k. For the zero default case the likelihood has a very different shape than for non-zero defaults. Note that the scale in the figure is not the same for each subplot, this is not a problem however since the posterior is normalised so only the shape of the likelihood matters.

Independent priors

In the independent case, the bivariate prior is simply the product of the two univariate priors. For the prior on ρ a uniform distribution on $[0, 1]$ is used, while for p a uniform and Pareto prior is chosen. The bivariate independent uniform (IU) and independent Pareto (IP) priors are then written as

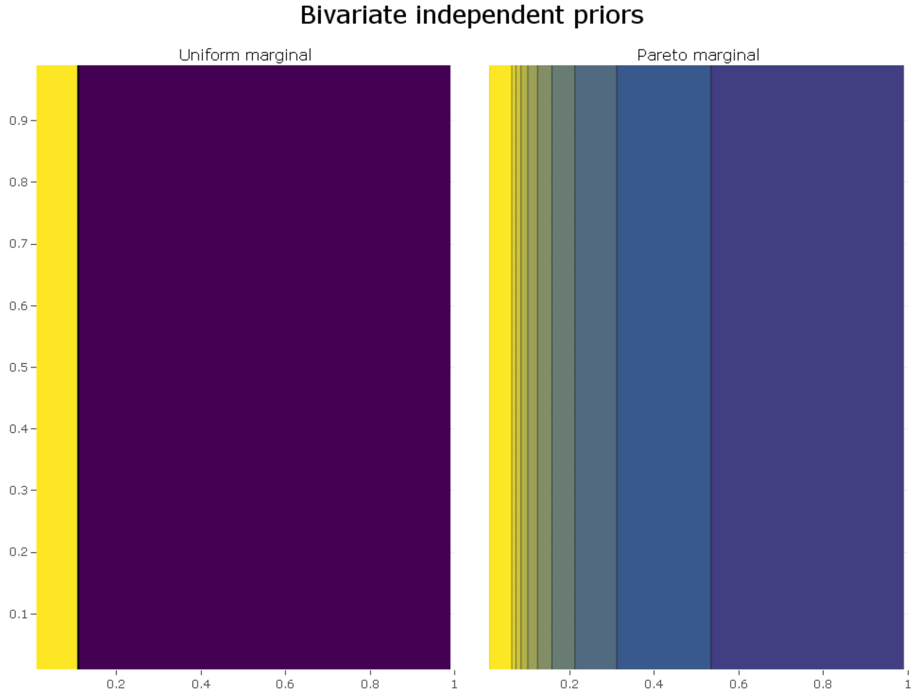

Figure 5.13: Bivariate independent priors, uniform marginal with parameter $u = 0.1$ and Pareto marginal with parameter $\xi = 4$.

$$\pi_{IU}(p, \rho \mid u) = \frac{1}{u} 1_{[0,u]}(p) 1_{[0,1]}(\rho),$$
$$\pi_{IP}(p, \rho \mid \xi) = \frac{1}{\xi} p^{\frac{1}{\xi}-1} 1_{[0,1]}(\rho).$$

Dependent priors

The assumption of independence between the probability of default and the asset correlation of the obligors is a quite strong one, even in the IRB formula this assumption is not done, see chapter 1. One would expect in portfolios where defaults are uncommon the variability of defaults to be low as well. Similarly, in higher default portfolios there is assumed to be more correlation. For testing purposes these relationships are assumed to be asymmetrical, just as the real world is seldom symmetrical. Recall Claytons copula described in chapter 2, given by

$$C_\theta^{Cl}(u, v) = \max\left([u^{-\theta} + v^{-\theta} - 1]^{-\frac{1}{\theta}}, 0\right), \quad \theta \in [-1, \infty) \setminus \{0\}.$$

The parameter value of $\theta = 0.5$ is chosen arbitrarily to ensure a moderate level of dependence. Since copulas rely on the probability integral transform to ensure uniform marginals, the cumulative distribution functions of univariate priors are needed. Write therefore the cdf of the uniform and Pareto prior as

$$\Pi_{UBE}(p \mid u) = \frac{1}{u} p 1_{[0,u]}(p)$$
$$\Pi_{PBE}(p \mid \xi) = p^{\frac{1}{\xi}}.$$

Then the dependent bivariate prior with uniform (DU) and Pareto (DP) marginals are given by

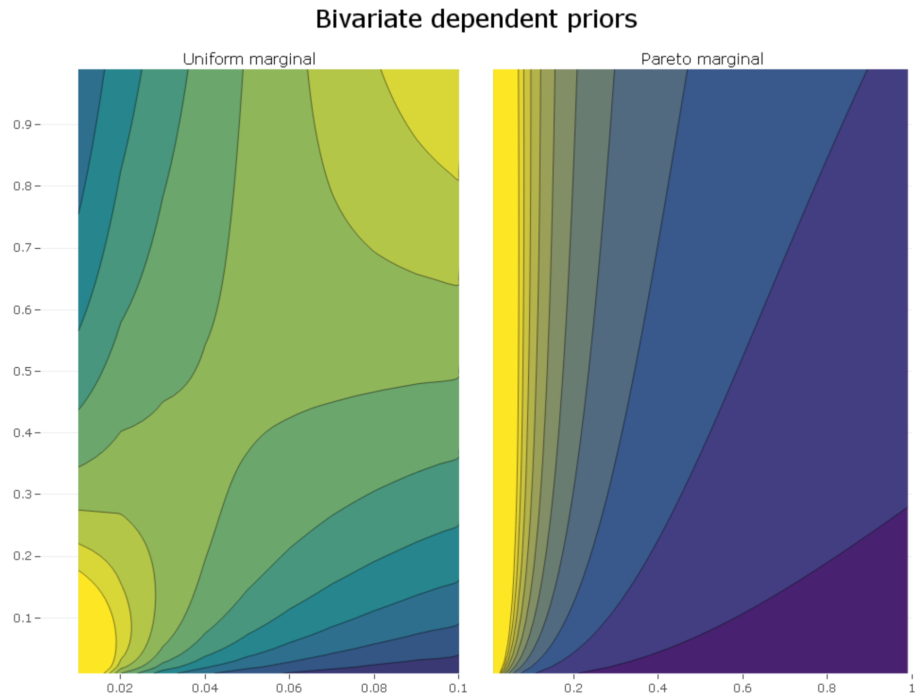

Figure 5.14: Bivariate independent priors, uniform marginal with parameter $u = 0.1$ and Pareto marginal with parameter $\xi = 4$. Note the uniform marginal is only shown for $0 \leq p \leq 0.1$ as it is zero elsewhere.

$$\pi_{DU}(p, \rho \mid u) = C_\theta^{Cl}(\Pi_{UBE}(p \mid u), \Pi_{UBE}(\rho \mid u = 1))$$
$$\pi_{DP}(p, \rho \mid \xi) = C_\theta^{Cl}(\Pi_{PBE}(p \mid \xi), \Pi_{UBE}(\rho \mid u = 1)).$$

Now the bivariate prior is combined with the likelihood function of the Vasicek model using equation 5.2, resulting in a posterior distribution. In order to analyse the performance of the bivariate prior, point estimators for the probability parameter p need to be defined. This is done in the same fashion as in the univariate case, by computing the mean of the posterior distribution. In order to do this, the posterior distribution first needs to be marginalised with respect to p. Therefore, the *posterior marginal* of p is defined as:

$$\pi(p \mid k) = \int_0^1 \pi(p, \rho \mid k) d\rho.$$

And the posterior mean is then equal to

$$\hat{p} = \int_0^1 p \; \pi(p \mid k) dp.$$

For illustration purposes the posterior distribution for zero observed defaults using the bivariate dependent uniform prior and its marginal are shown in figure 5.15.

The posterior means and the Bayesian 95% credible intervals of the four marginalised posterior distributions are presented in figure 5.16. The Pareto estimators have a much larger confidence interval, explained by the fact that the uniform priors only put mass on $p \leq 0.1$. The dependent estimators have higher PD estimates than their independent counterparts and the Pareto estimates higher than the uniform ones. Remark also that the estimators lie between the low and high level of correlation,

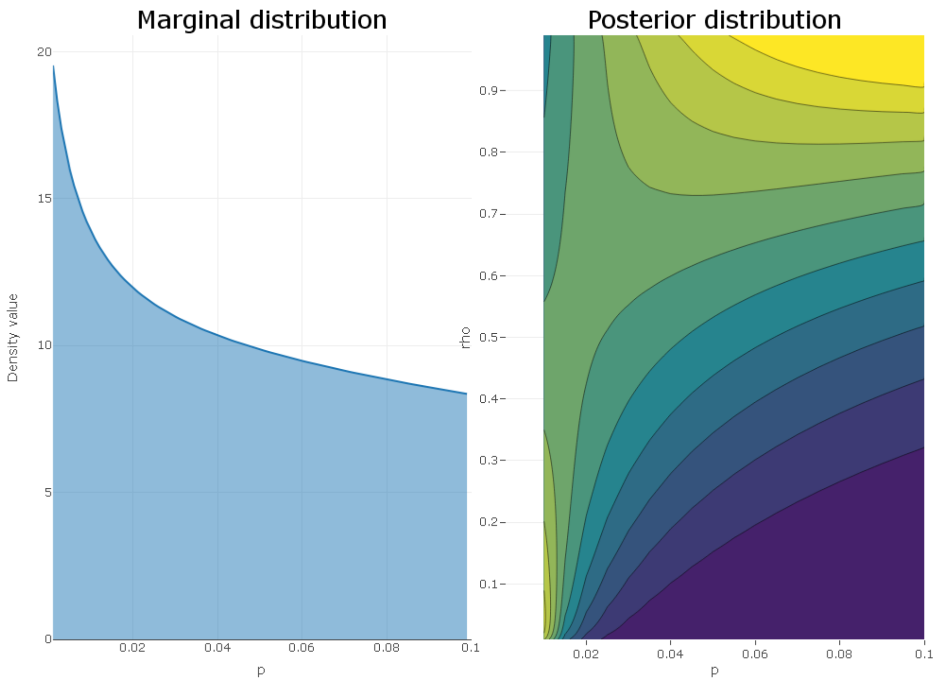

Figure 5.15: Posterior distribution given the dependent uniform prior with parameter $u = 0.1$ and the marginal distribution of p.

making them preferable over the fixed level correlation estimators due to their robustness. The choice of which (bivariate) prior to use is up to the modeller, if no information are available it makes sense to use the bivariate uniform prior. However, if there are experts with more prior knowledge these can be utilised to obtain better informed prior distributions possibly making use of the (dependent) Pareto prior.

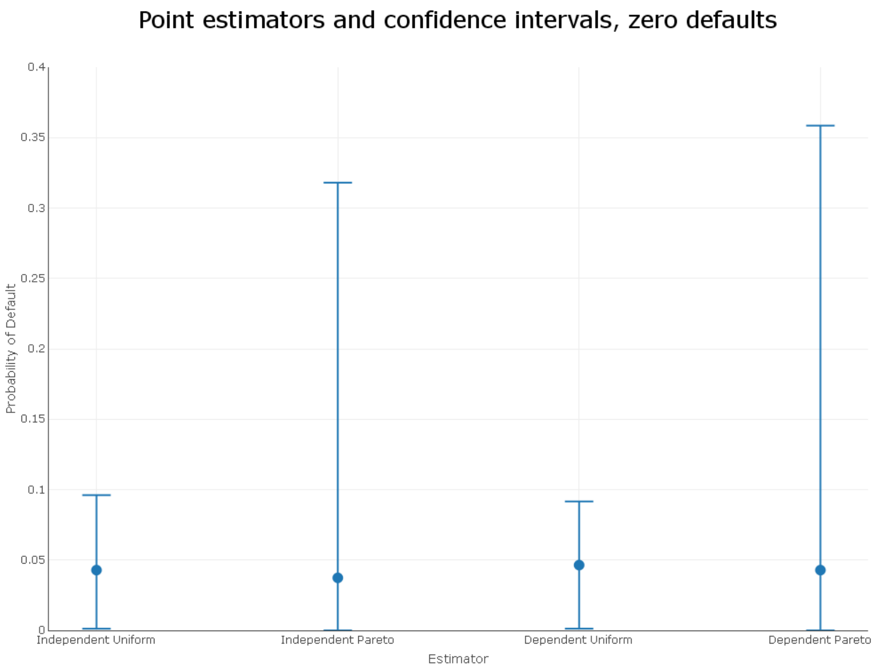

Figure 5.16: Estimators and confidence intervals for the Vasicek model with Bivariate prior. Point estimators are respectively 4.28%, 3.73%, 4.64% and 4.28%.

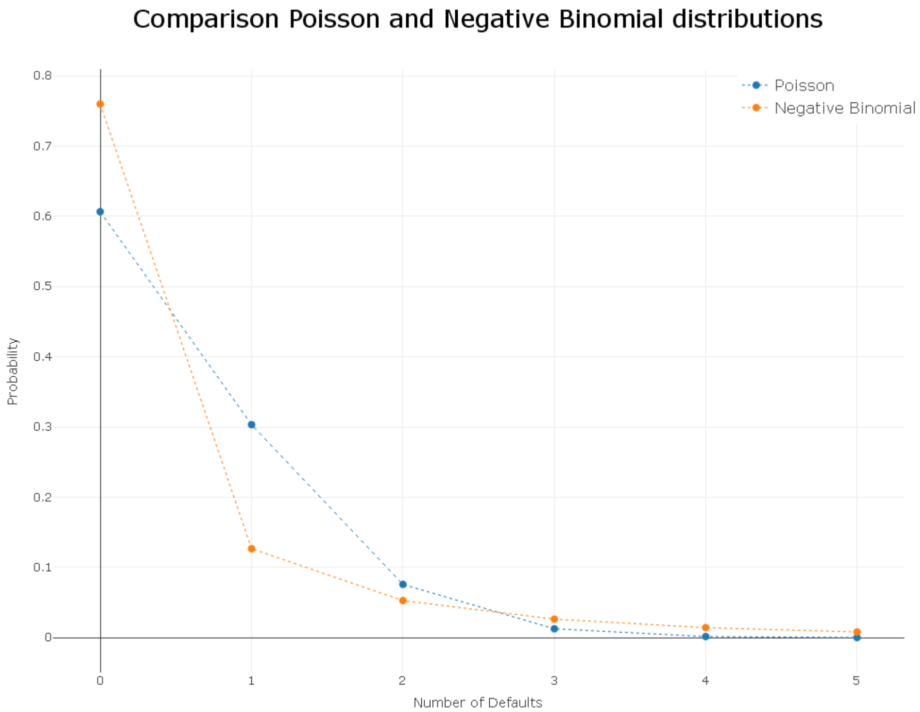

Figure 5.17: The probability mass functions of a Poisson distribution with $\lambda = 0.5$ and a Negative Binomial distribution with $\mu = 0.5$ and $\sigma = 1$.

5.2. Credit Risk+ model

Recall the CR+ model described in chapter 4, where instead of the probability of default p a default rate μ is considered. In the independent case this means that defaults are Poisson distributed instead of Binomial, and in the dependent case the Negative Binomial replaces the Vasicek distribution. For the sake of brevity, only the dependent CR+ model is considered, so remember the density of the Negative Binomial distribution:

$$f_{NB}(k \mid \mu, \sigma) = \frac{\Gamma\left(\frac{\mu^2}{\sigma^2} + k\right)}{k!\Gamma\left(\frac{\mu^2}{\sigma^2}\right)} \left(\frac{\sigma^2}{\mu + \sigma^2}\right)^k \left(\frac{\mu}{\mu + \sigma^2}\right)^{\frac{\mu^2}{\sigma^2}}.$$

Recall that the underlying reason for abandoning the Poisson distribution was the unrealistic implicit assumption between the mean and the variance. Figure 5.17 shows that the Negative Binomial distribution is more heavily skewed toward the left, assigning more probability to the zero default case than the Poisson distribution does.

5.2.1. Bayesian extension of CreditRisk+

The same problem setting as in the Vasicek model is encountered: A bivariate likelihood of which one marginal represents the probability of default and the other represents its volatility.

Similar to the analysis done on the Vasicek model, a bivariate prior is put on the parameters of the likelihood. Since different parametrisations are used for the Negative Binomial distribution, it is important to note that the choice of parametrisation matters in the Bayesian context, as the same prior on different parametrisations of the likelihood will generally lead to different posteriors. The exception to this rule is Jeffreys prior, which will be seen later in this chapter.

Since a prior should display the knowledge one has before observing the data (from a subjective Bayesian standpoint), it is more intuitive to put a prior on the (μ, σ) parametrisation than the (α, β) one.

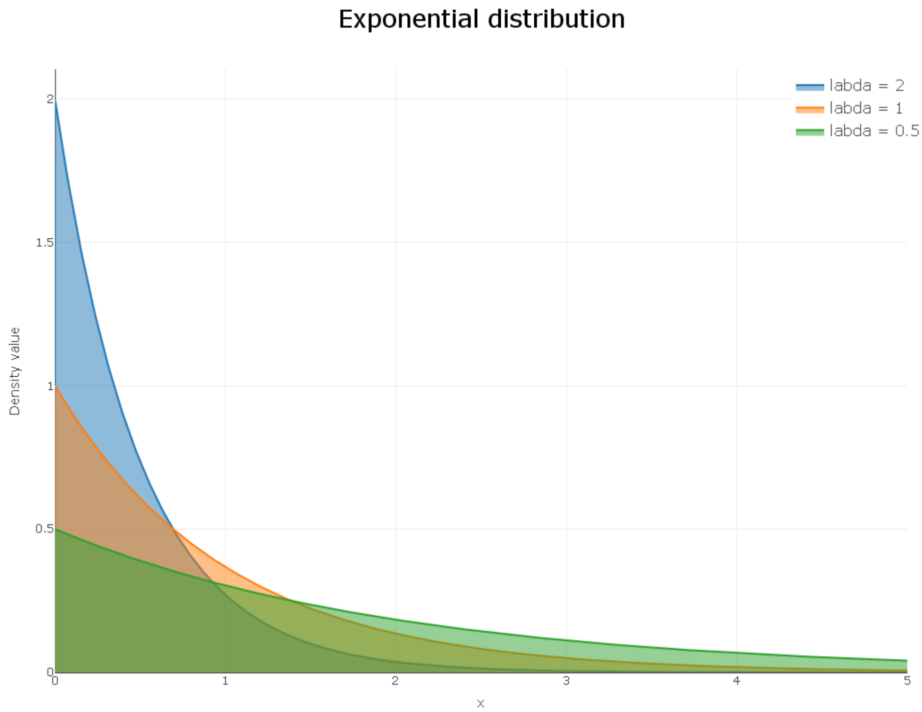

Figure 5.18: Probability density function of the exponential distribution for parameters $\lambda = 0.5, 1, 2$.

The CR+ model is derived from assumptions about (μ, σ) and these have a real-world interpretation. Naturally, the domain of the (continuous) prior is $\mathbb{R}^+ \times \mathbb{R}^+$.

Exponential priors

First note that the priors used in the Vasicek model can no longer be applied due to the half infinite domain now under consideration. Having demonstrated the advantages of specifying a prior for the correlation/volatility parameter because of the robustness, only the bivariate case is studied. Also, since defaults are expected to be rare the prior should decay quite rapidly, which is why the exponential distribution is determined to be a good candidate for the marginals. Finally, for investigation purposes both independent and dependent priors are considered. These requirements lead to the consideration of the following priors:

- Independent Exponential
- Dependent Exponential (Marshall and Olkin)
- Dependent Exponential (Gumbel type I)

Since all these three bivariate distributions are derived from the univariate exponential, a concise description of that distribution will be given first. Suppose $X \sim \text{Exp}(\lambda)$, its density is given by

$$f(x \mid \lambda) = \lambda e^{-\lambda x}.$$

See figure 5.18 for a graphical representation of the probability density function. The first two moments are given by $\mathbb{E}[X] = \frac{1}{\lambda}$ and $\text{Var}(X) = \frac{1}{\lambda^2}$.
The reason no copulas are used in the dependence modelling is because of the remarks made earlier in chapter 2, since there exist nice analytical distributions with the desired properties copulas are not needed and therefore not incorporated. Now the three bivariate extensions of the exponential distribution will be investigated.

Independent exponential prior A simple first choice is the independent exponential prior: the product of two exponential distributions which may have different parameters:

$$\pi(\mu,\ \sigma \mid \lambda_1,\ \lambda_2) = \pi(\mu \mid \lambda_1)\pi(\sigma \mid\ \lambda_2)$$
$$= \lambda_1 e^{-\lambda_1\mu}\lambda_2 e^{-\lambda_2\sigma}.$$

Marshall and Olkin exponential [50] Imposing the memoryless property of the univariate exponential distribution as a condition for the bivariate case, formulated as:

$$S(s_1+t, s_2+t) = S(s_1,s_2)S(t,t) \qquad s_1,s_2,t \geq 0.$$

With $S(s,t)$ defined as the survival function:

$$S(s,t) = P(X_1 > s, X_2 > t).$$

It is then shown that the survival function for the bivariate Exponential $E_{MO}\,(\lambda_1,\lambda_2,\eta)$ is given by:

$$S(x_1,x_2 \mid \lambda_1,\lambda_2,\eta) = \exp\{-\lambda_1 x_1 - \lambda_2 x_2 - \eta \max(x_1,x_2)\},$$

and since

$$\frac{\partial^2 S(x_1,x_2 \mid \lambda_1,\lambda_2,\eta)}{\partial x_1 \partial x_2} = f(x_1,x_2 \mid \lambda_1,\lambda_2,\eta)$$

the probability density function (now used as a prior) is given by

$$\pi(\mu,\sigma \mid \lambda_1,\lambda_2,\eta) = \begin{cases} \lambda_2\,(\lambda_1+\eta)\exp\left(-(\lambda_1+\eta)\mu - \lambda_2\sigma\right), & \text{for } \mu \geq \sigma \\ \lambda_1\,(\lambda_2+\eta)\exp\left(-(\lambda_2+\eta)\sigma - \lambda_1\mu\right), & \text{for } \mu < \sigma. \end{cases} \tag{5.3}$$

Gumbel type I [51]
In the book of Balakrishnan and Lai the Gumbel Type I bivariate Exponential distribution is described. The density function (again, now seen as a prior) is given by

$$\pi(\mu,\sigma \mid \eta) = [(1+\theta\mu)(1+\eta\sigma) - \eta]\exp\left(-(\mu+\sigma+\eta\mu\sigma\right).$$

See figures 5.19, 5.20 and 5.21 for plots of the three exponential bivariate priors. It can be seen that the Marshall Olkin prior puts more mass on values of mutual extremes of μ and σ, while the Gumbel prior has a negative correlation coefficient.
Combining the three priors with the data (likelihood) through Bayes' formula produces for each prior a posterior. The likelihood function for zero observed defaults and the resulting three posteriors are shown in figure 5.22. Recall that the bivariate posterior is given by

$$\pi(\mu,\sigma \mid k) = \frac{f(k \mid \mu,\sigma)\pi(\mu,\sigma)}{\int_0^\infty \int_0^\infty f(k \mid \mu,\sigma)\pi(\mu,\sigma)d\mu d\sigma},$$

the marginalised posterior distribution for μ by

$$\pi(\mu \mid k) = \int_0^\sigma \pi(\mu,\sigma \mid k)d\sigma,$$

and the posterior mean of that distribution by

$$\hat{\mu} = \int_0^\infty \mu\ \pi(\mu \mid k)d\mu.$$

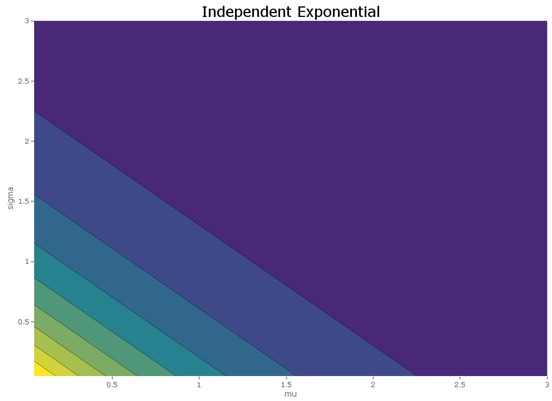

Figure 5.19: Independent exponential prior with parameters $\lambda_1 = \lambda_2 = 1$.

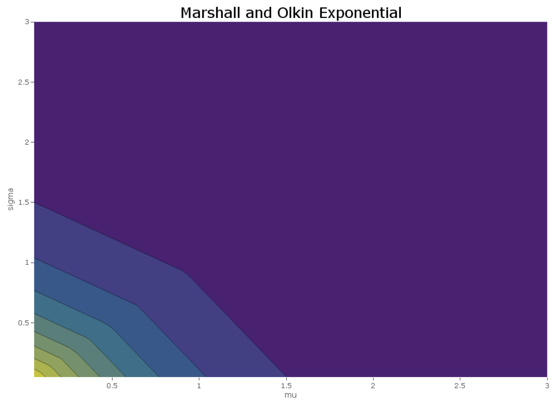

Figure 5.20: Marshall and Olkin prior with parameters $\lambda_1 = \lambda_2 = 1$, $\eta = 0.5$.

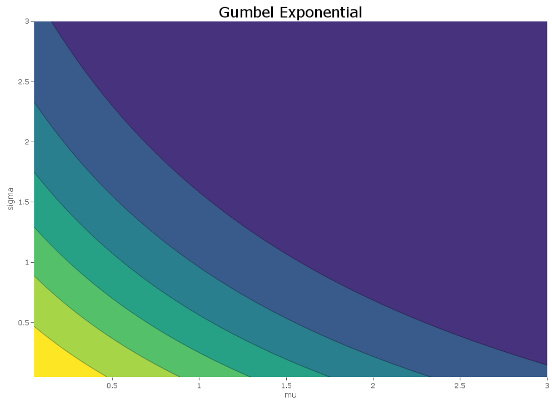

Figure 5.21: Gumbel type I prior with parameter $\eta = 0.5$

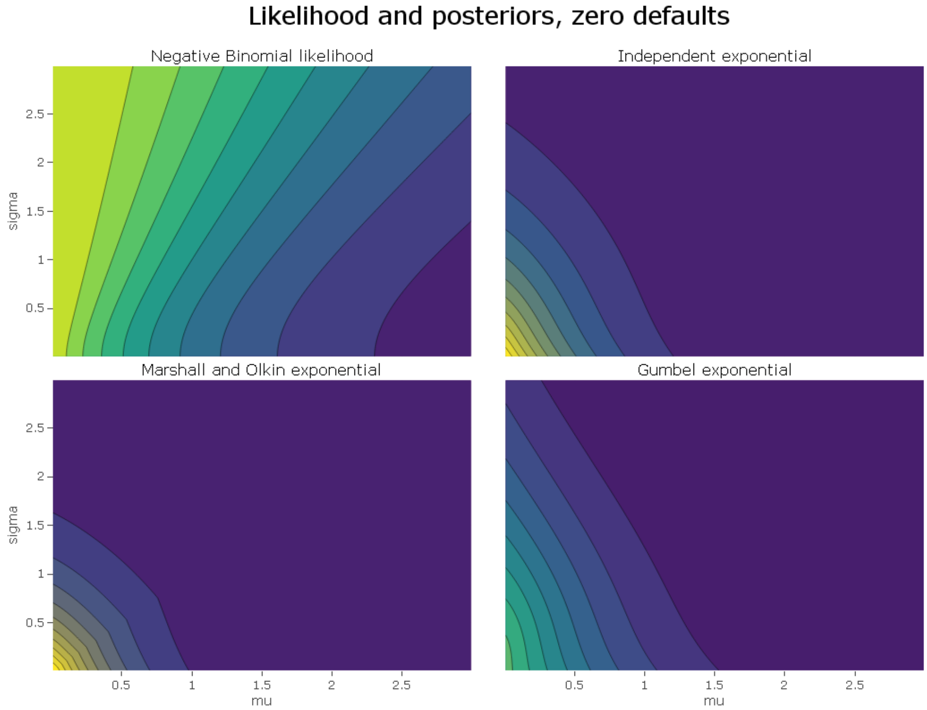

Figure 5.22: Likelihood function and posterior distributions for three bivariate exponential distributions.

An interesting result is found when investigating the posterior means. An intuitive condition which an estimator should definitely satisfy is that of monotonicity: When more defaults are observed, the corresponding estimate of the probability of default cannot decrease. However, the means of the marginalised posteriors with the three aforementioned priors seem to violate this principle as shown in figure 5.23.

A reason for this behaviour might be the non-existence of the first moments. Therefore the posterior medians are presented in figure 5.24. The median of the marginalised posterior distribution for μ is given by:

$$\hat{\mu} = \inf\left\{q : \int_0^q \pi(\mu \mid k) d\mu \geq 0.5\right\}.$$

As can be seen, the posterior median also displays the same strange non-monotonicity. Acknowledging that making errors is only human, it is quite possible that there are errors in the numerical integration code. Therefore an additional method, based on sampling, to compute the posterior mean and median is utilised. The 'Metropolis within Gibbs' algorithm described in chapter 2 is employed to obtain samples $(\mu_1, \sigma_1), \ldots, (\mu_N, \sigma_N)$ from the bivariate posterior distribution. From these samples the posterior means and medians are determined.

Figure 5.25 shows for the independent exponential prior that the numerical and sampling routine produce almost identical results. For the other priors the sampling routine also aligns with the numerical integration method. Now it can be said with sufficient confidence the non-monotonicity of the posterior mean and median are not caused by implementation errors but by some mathematical phenomenon. The underlying cause has not been identified yet but could be the starting point of future research. Seeing as these subjective priors display an undesirable property, an objective approach is tried in the next section.

Jeffreys prior

Recalling the description in chapter 2, Jeffreys prior is given by

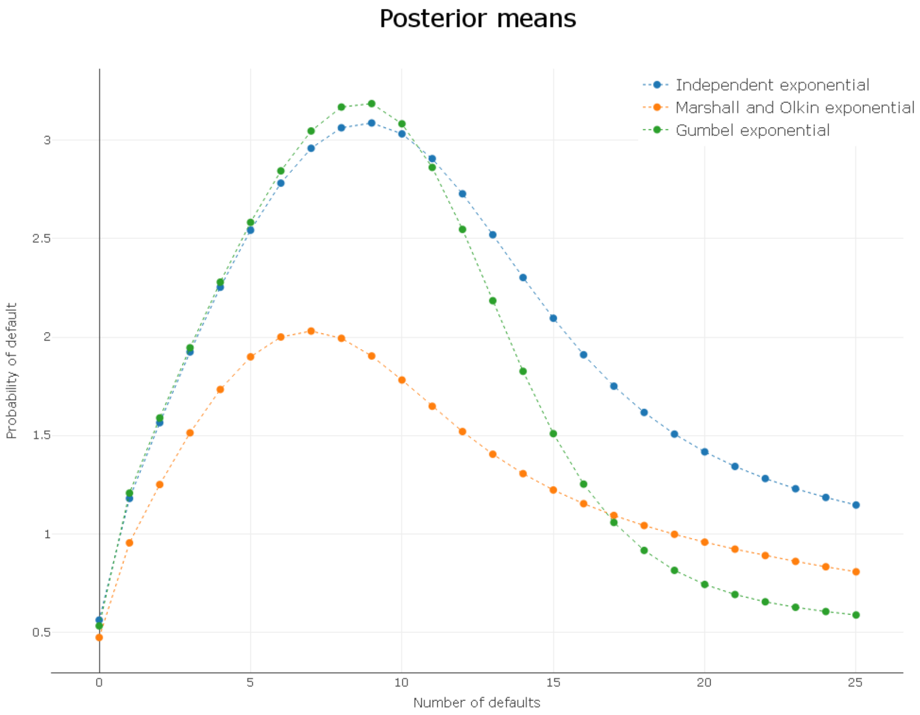

Figure 5.23: Means of the marginalised posterior distribution of μ, using the three exponential priors under consideration. Results are obtained by numerical integration.

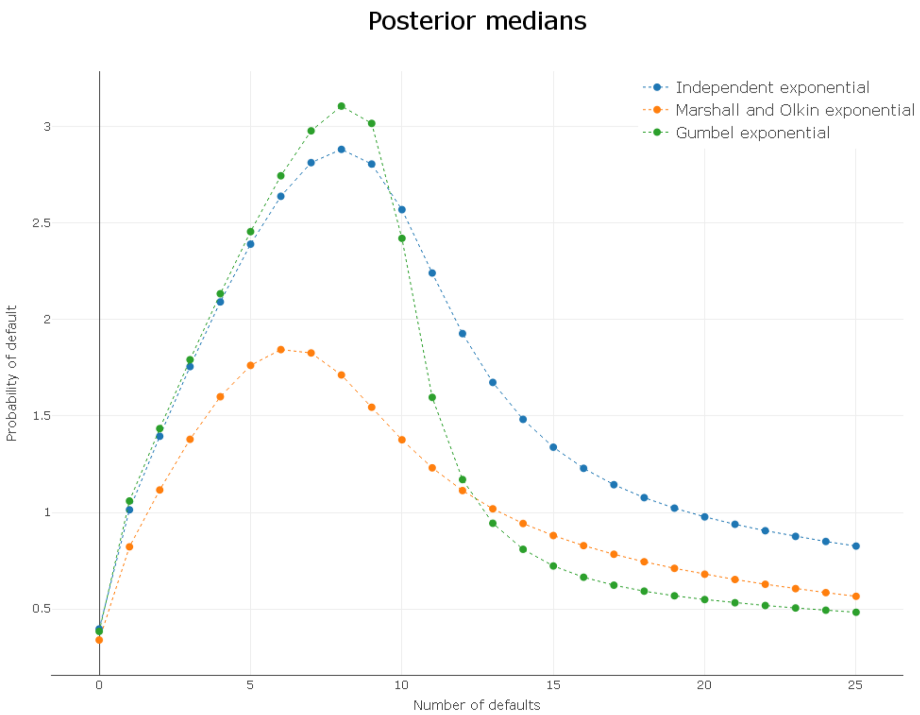

Figure 5.24: Medians of the marginalised posterior distribution of μ, using the three exponential priors under consideration. Results are obtained by numerical solving and integration.

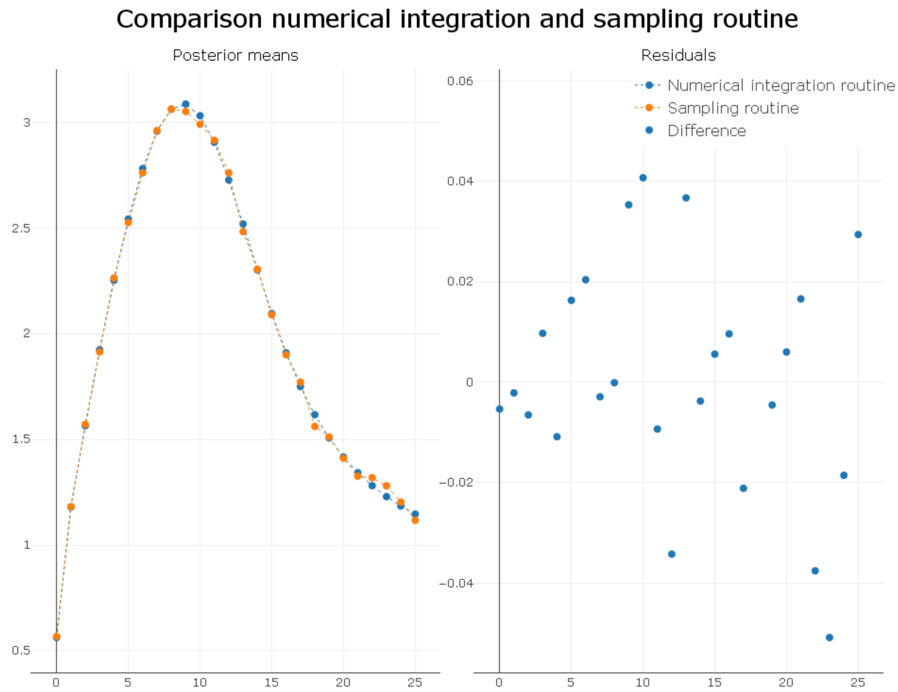

Figure 5.25: Comparison for the posterior mean of the numerical integration routine and the Metropolis within Gibbs sampling method. 50.000 samples are drawn from the bivariate posterior, discarding the first 10.000 as 'burn-in' samples.

$$\pi(\mu,\ \sigma) = \sqrt{\det I(\mu,\ \sigma)},$$

where $\det I(\mu,\ \sigma)$ is the determinant of the Fisher information matrix

$$I_{i,j}(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta_i}\log f(k \mid \theta)\right)\left(\frac{\partial}{\partial\theta_j}\log f(k \mid \theta)\right)\right],$$

with $\theta_1 = \mu$ and $\theta_2 = \sigma$. The likelihood under consideration is the Negative Binomial with the (μ, σ) parametrization given in equation 4.2. The logarithm of the likelihood is given by

$$\begin{aligned}\log f(k \mid \mu,\ \sigma) =& \log\Gamma\left(\frac{\mu^2}{\sigma^2} + k\right) - \log(k!) - \log\Gamma\left(\frac{\mu^2}{\sigma^2}\right) + \\ & k\log\left(\frac{\sigma^2}{\mu}\right) - \log\left(\frac{\mu+\sigma^2}{\mu}\right) - \frac{\mu^2}{\sigma^2}\log\left(1 + \frac{\mu^2}{\sigma^2}\right).\end{aligned}$$

The partial derivatives can also be found analytically, where $\psi(x)$ is written for the derivative of the logarithm of the gamma function:

$$\begin{aligned}\frac{\partial}{\partial\mu}\log f(k \mid \mu,\ \sigma) =& \frac{2\mu}{\sigma^2}\left(\psi\left(\frac{\mu^2}{\sigma^2} + k\right) - \psi\left(\frac{\mu^2}{\sigma^2}\right) - \log\left(1 + \frac{\sigma^2}{\mu}\right)\right) - \\ & \frac{\sigma^2}{\mu^2}k\left(\frac{\mu}{\sigma^2} - \frac{\mu}{\mu+\sigma^2}\right) + \frac{\mu}{\mu+\sigma^2},\end{aligned}$$

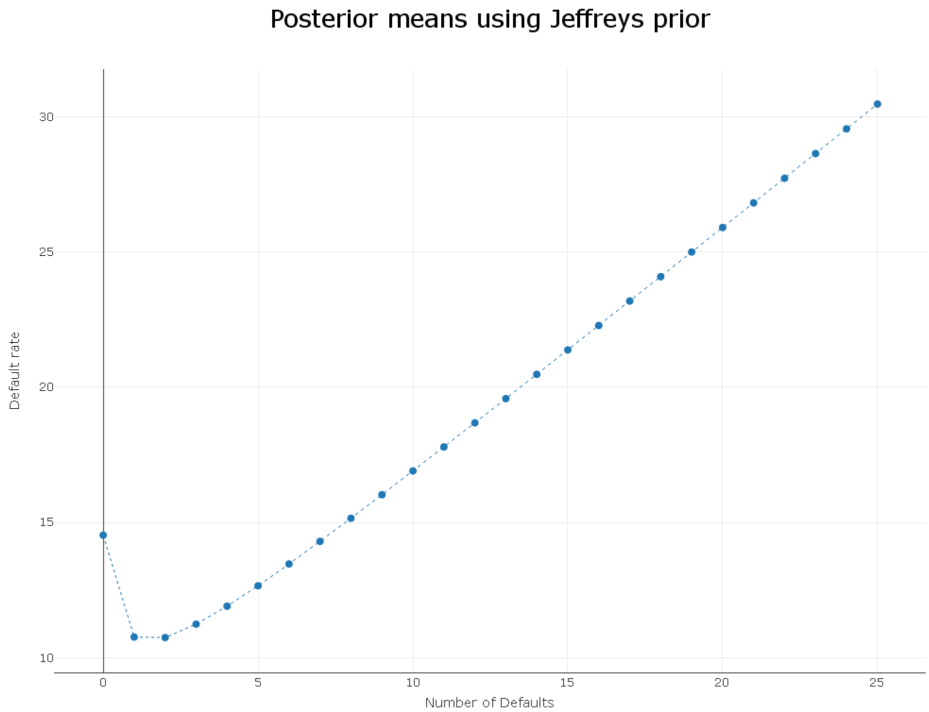

Figure 5.26: Means of the marginalised posterior distribution of μ, using Jeffreys prior. Results are obtained by numerical integration.

$$\frac{\partial}{\partial \sigma} \log f(k \mid \mu,\, \sigma) = -\frac{2\mu^2}{\sigma^3}\left(\psi\left(\frac{\mu^2}{\sigma^2}+k\right)-\psi\left(\frac{\mu^2}{\sigma^2}\right)-\log\left(1+\frac{\sigma^2}{\mu}\right)\right)+ k\left(\frac{2}{\sigma}-\frac{2\sigma}{\mu+\sigma^2}\right)-\frac{2\mu^2}{\sigma(\mu+\sigma^2)}.$$

Using these partial derivatives and the law of the unconscious statistician, the Fisher information matrix is approximated by

$$I_{i,j}(\theta) \approx \sum_{k=0}^{N} f(k \mid \theta)\left(\frac{\partial}{\partial \theta_i} \log f(k \mid \theta)\right)\left(\frac{\partial}{\partial \theta_j} \log f(k \mid \theta)\right).$$

where a value of $N = 50$ is found to be sufficiently large.

The posterior means for various values of k are computed using numerical integration and the result found in figure 5.26. The value of these estimates seems way too high, especially for the lower amounts of default. Unfortunately a verification using the Gibbs sampler was not feasible, as the variance between estimators was enormous and some of the estimates for μ were over 1500. The reason for this behaviour could be that the distribution is heavy-tailed. A so-called Maximum to Sum (MS) plot as found in [52] is a method to verify the existence of methods.

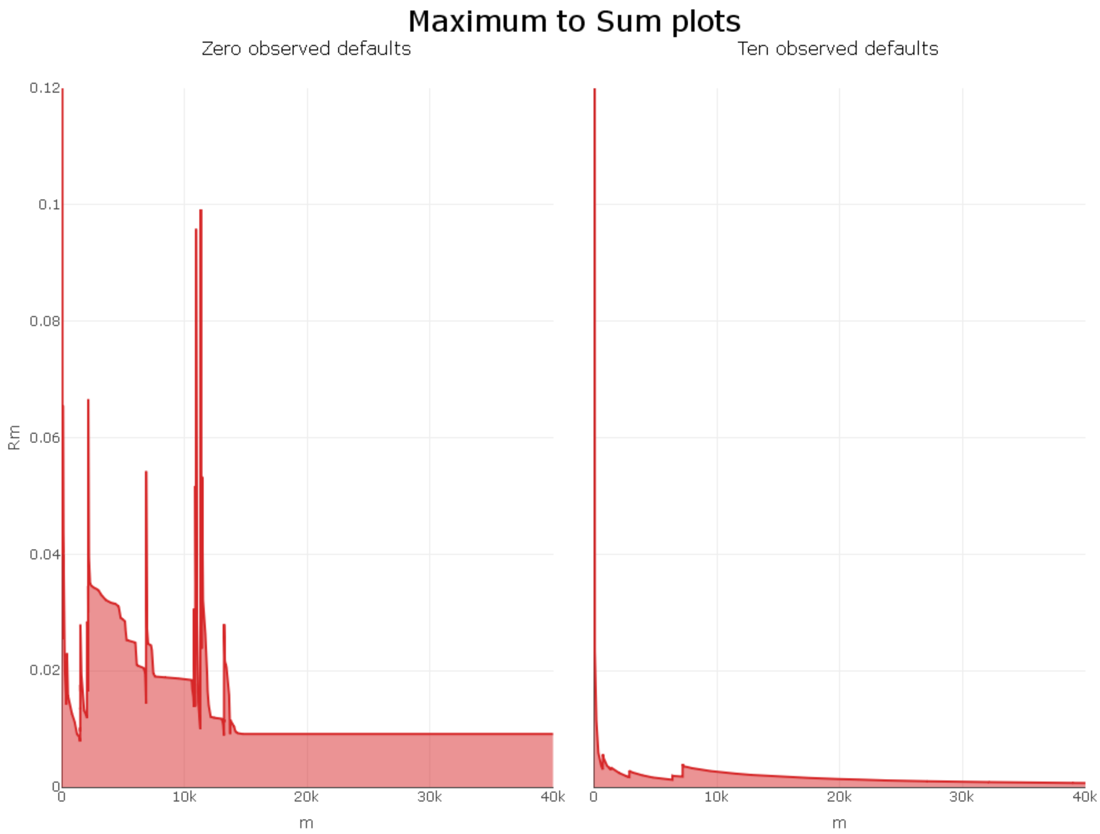

Figure 5.27: Maximum to Sum plot for the first moment. The sample comes from the posterior distribution using a Jeffreys prior and zero observed defaults.

Definition 6 - Maximum to Sum plot [52].
Consider a sample $\{X_1, X_2, \dots, X_n\}$, *define:*

$$S_m^p = \sum_{i=1}^{m} X_i^p,$$
$$H_m^p = \max\left(X_1, \dots, X_m\right).$$

Then the following graph is called a Maximum to Sum plot:

$$m \longmapsto \frac{H_m^p}{S_m^p} = R_m^p.$$

If R_m^p converges to 0 as m tends to infinity, the MS plot suggests that the p'th moment is finite. Figure 5.27 shows this may not be the case for $k = 0$, but for higher values of k it seems the mean is finite.

As the mean produces too high and unreliable estimates, the posterior median seems to be a promising alternative. The results plotted in figure 5.28 are more in the line of expectation as they show almost perfect linear behaviour. In the same plot a verification of results using the Metropolis within Gibbs sampler is plotted, which shows that the computation of the posterior median is correct.

Comparison of estimators
The CR+ model has been derived and reparametrized to the (μ, σ) representation. First three bivariate exponential priors have been put on these parameters and it is shown that posterior estimates for the

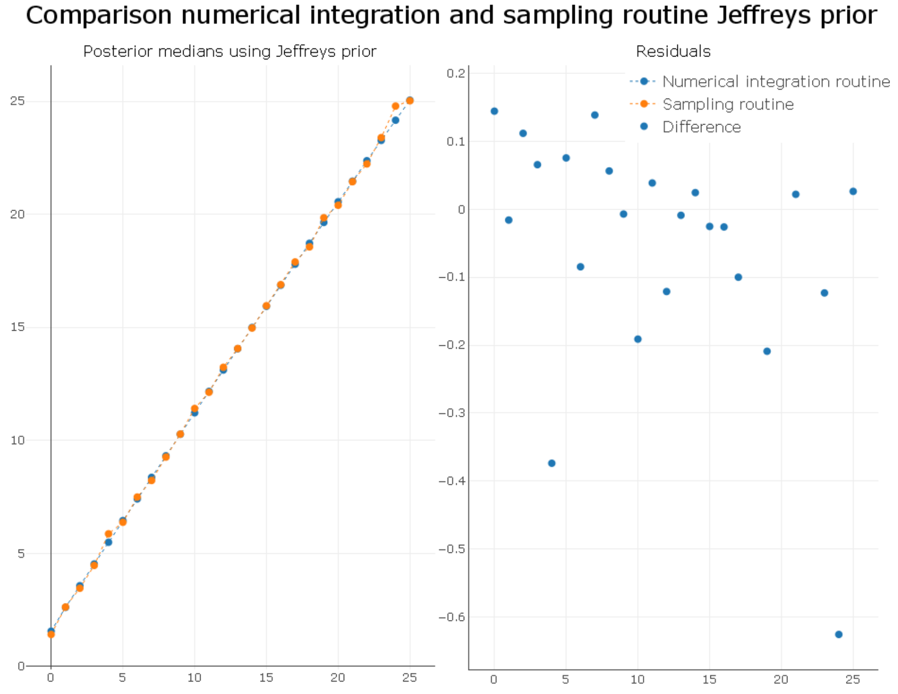

Figure 5.28: Comparison for the posterior median of the numerical integration routine and the Metropolis within Gibbs sampling method. 50.000 samples are drawn from the bivariate posterior, discarding the first 10.000 as 'burn-in' samples.

default rate parameter μ are non-monotone. Therefore an objective prior is considered, Jeffreys prior. Numerical issues arise when computing the posterior mean for μ, but the median is not subject to those. In order to asses the level of conservatism in the default rate estimates resulting from these priors, figure 5.29 shows posterior estimates for μ and their credible intervals for the case of no defaults.

One possible point of criticism has been ignored up to now, namely the exclusion of portfolio size in the model. Suppose zero defaults are observed, then under the CR+ model the default rate is the same in a portfolio of 100 or 10.000 obligors. But dependence on the portfolio size is introduced by converting default rates to default probabilities by dividing by the amount of obligors.

The default rate estimates of the Jeffreys priors is quite a bit higher than the ones of the exponential priors. For zero defaults the posterior for the default rate possibly has fat tails, which explains why the estimator is more prudent than the others. Jeffreys prior might be viewed as too conservative by practitioners, as it expects almost two defaults when none are observed. However, since it is both conservative and objective the regulator might find it an attractive prior.

In contrast with Jeffreys prior, the exponential priors allow for the incorporation of subjective prior knowledge. The parameters in this chapter have only been chosen as a way of demonstration. In real applications, experts or other techniques could be used to calibrate the parameters. A large advantage of this is that the estimates become more risk-sensitive, but banks will have to show that their calibration technique is robust and conservative.

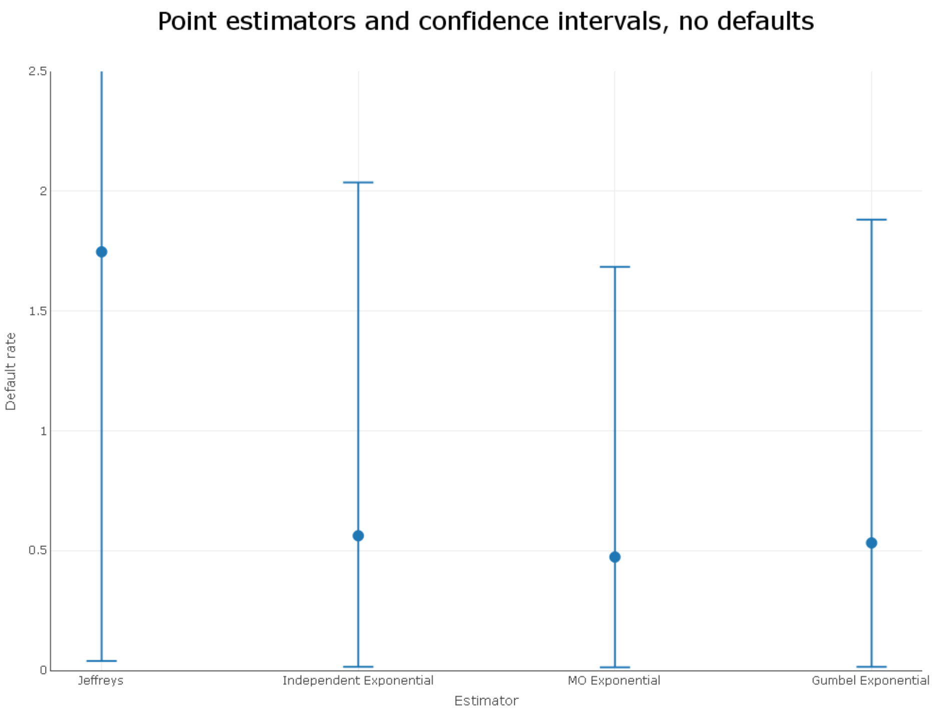

Figure 5.29: Point estimates and confidence intervals for the four priors under the CR+ model. The upper credible bound of Jeffreys prior is 22867.

6

An application to a real sovereign default portfolio

So far model performance has only been evaluated by mock data. It is therefore interesting to consider how the models perform when they are applied to a real data set. The data used consists of sovereign defaults occurring in the time period 1960-2016. The data is aggregated by the Bank of Canada (BoC) and published in [53], a review is given by them in the corresponding report [54]. First, a description of the data will be given, thereafter the multi-period extension of the Vasicek and Credit Risk+ models will be applied to the data.

6.1. Description of data

Sovereign, or government, defaults are an interesting case of defaults. They are generally even more rare and impactful than corporate defaults, so particularly suited as a low default portfolio. Historically, defaults occurred in emerging-market economies, but some advanced-economy defaults have been observed during the financial crisis as well.

The BoC mentions that aggregating and comparing sovereign defauls is quite challenging, since *'there is no single internationally recognized definition of what constitutes a sovereign default'*. In other situations, defaults could be defined as the failure to meet a contractual obligation to pay interest or principal in full on the due date. However, note that governments have for example the ability to create money. When this is done in order to meet debt payments that otherwise could not have been paid, it could constitute as a default as well. For more details please refer to [54].

In total, the database consists of 57 years of data, from 1960 to 2016. After removal of missing data, 4117 defaults in 145 different countries have been observed. The Bank of Canada uses the classification [55] of the International Monetary Fund (IMF) of countries to distinguish four country groups:

- Advanced Economies
- Emerging/frontier markets
- Heavily Indebted Poor Countries (HIPC)
- Other developing countries

As 4117 default observations in a portfolio is too much to classify it as 'low default', only a sub-portfolio will be considered. A natural choice is to restrict the portfolio to only obligors classified as Advanced Economies. In the 57 years of data, only 14 defaults have been observed among 5 countries labelled as Advanced Economy. See table 6.1 for details.

Three different forms of lending are used by the Advanced Economy countries that defaulted:

Country	Year	Total	IMF	Other official creditors	FC bonds
Cyprus	2013	1698	-	-	1698
Greece	1960	369	0	-	369
Greece	1961	378	0	-	378
Greece	1962	387	0	-	387
Greece	1963	301	0	-	301
Greece	1964	308	0	-	308
Greece	1965	314	0	-	314
Greece	2012	312420	0	0	312420
Greece	2013	212819	0	212819	0
Greece	2015	2222	2222	0	0
Ireland	2013	88290	-	88290	-
Portugal	2013	52712	-	52712	-
Puerto Rico	2015	1103	-	-	1103
Puerto Rico	2016	25904	-	-	25904

Table 6.1: Sovereign defaults of countries classified as 'Advanced Economies' in the years 1960-2016. The debt is given in millions of US dollars.

- **IMF lending.** When a sovereign misses a membership quota or loan granted by the IMF, this is considered as default. The IMF has written off some loans given to countries under the Multilateral Debt Relief Initiative (MDRI).
- **Other official creditors.** These include official creditors not belonging to the other categories described in [54], as for example national export credit agencies (ECA).
- **Foreign currency (FC) bonds.** These are bonds denominated in a, for the defaulting country, foreign currency.

In total, the IMF classifies 39 countries as Advanced Economies. This number may have varied over the 57 years under consideration, but for computation purposes deemed as constant. The portfolio therefore consists of $n = 39$ obligors, for a time period of 57 years.

6.2. Vasicek model

The multi-period extension of the Vasicek model described in chapter 3 will be applied to the sovereign default data set. First, the maximum likelihood estimator will be tried. The most prudent estimator is shown to be less accurate than certain Bayesian estimators and therefore not taken into account. The focus will be on the estimation using Bayesian techniques.

6.2.1. Maximum likelihood estimator

The maximum likelihood estimation is done using the Monte Carlo technique presented in chapter 3, making use of the Cholesky decomposition to seperate the influence of θ and the sample. To provide an estimate of the variability of the MLE, a Monte Carlo routine of 25.000 samples (of dimension $T = 57$) is done 8 times. The mean and standard errors of the estimates are given in table 6.2.

	mean	standard error
p	0.0370	0.0031
ρ	0.5618	0.0176
θ	0.8689	0.0105

Table 6.2: Results of the maximum likelihood estimator using 8 runs of a Monte Carlo simulation

Even though the Vasicek model is an extended model incorporating asset and time correlation, one would assume that the estimate for p is close to the maximum likelihood estimator of the Binomial model which it extends. However, the estimate for that model is

$$\hat{p} = \frac{\sum_{t=1}^{T} k_t}{\sum_{t=1}^{T} n_t} = \frac{14}{2223} = 0.0063,$$

five times smaller than the MLE estimate of the multi-period Vasicek model. Note also the relatively high value of the standard error of the estimate of p. There may be a problem in the implementation of the estimator, but together with the earlier issues found with maximum likelihood estimation it seems a better choice to further investigate the alternatives.

6.2.2. Bayesian estimators

Now a multivariate prior distribution will be specified for the parameters p, ρ and θ. Earlier in chapter 5 it is shown a uniform prior on the unit interval for the asset correlation ρ produced good results, therefore this prior is used in the multi-period setting as well. For the intertemporal correlation parameter θ the same prior might be even better suited, as a value of $\theta = 0.3$ is proposed in [56], $\theta = 0.6$ is used in [39] and the MLE in the previous section set $\theta = 0.86$.

For the probability parameter p two different priors are proposed, the uniform prior on the interval $[0, u]$ with $u = 0.1$, and the Pareto prior with parameter $\xi = 4$.

Numerical difficulties were already faced when imposing a bivariate prior. Now that the prior is a function on three dimensions, for the computation of the normalisation constant a three-dimensional integral needs to be evaluated. To find the a value of the posterior distribution for p another two-dimensional integral needs to be evaluated. Doing this with numerical integration may be possible with extensive effort, but implementing a sampling method is given preference.

Recall therefore the computational technique known as the Metropolis within Gibbs sampler presented in chapter 2 and used in the verification of numerical integration results of the Vasicek model. The proposal distribution $q(x, y)$, the same for all three parameters, is chosen to be a $\text{Beta}(\frac{1}{1-x}, 2)$, so that a trial value $Y = y$ is drawn from a Beta distribution with its mode equal to the previous value x.

Independent priors

When a multivariate distribution has independent marginals, it is equal to the product of the marginals, so that:

$$
\begin{aligned}
\pi_{indU}(p, \rho, \theta \mid u) &= \pi(p \mid u)\pi(\rho)\pi(\theta) \\
&= \frac{1}{u} 1_{[0,u]}(p)\ 1_{[0,1]}(\rho)\ 1_{[0,1]}(\theta), \\
\pi_{indP}(p, \rho, \theta \mid \xi) &= \pi(p \mid \xi)\pi(\rho)\pi(\theta) \\
&= \frac{1}{\xi} p^{\frac{1}{\xi}-1}\ 1_{[0,1]}(\rho)\ 1_{[0,1]}(\theta).
\end{aligned}
$$

Point estimators of the marginalised posterior are shown in figure 6.1, with exact values of the estimators in table 6.3. The median is chosen as a point estimator because the mean of the marginalised posterior for p seemed very conservative. Upon closer inspection the tail of this distribution seems quite heavy, see figure 6.2 for a sample.

Median value	uniform	MAD	Pareto	MAD
p	0.0476	0.0156	0.0077	0.0055
ρ	0.5926	0.0925	-	-
θ	0.9182	0.0349	-	-

Table 6.3: Median values and their median absolute deviations of the sample drawn from the posterior distribution with Pareto or uniform prior for p.

As expected, for the probability of default p, the Pareto prior corresponds to a less conservative estimate than the uniform prior. The first has an estimate of 0.0077, a bit higher than the naive estimate of 0.0063. It can be said that the uniform prior performs badly, its estimate is about 8 times higher than the naive estimate and thus too conservative. The Pareto prior is therefore definitely preferred, as it produces estimates just a bit more prudent than the MLE but also works in the case of zero defaults.

The estimates for ρ and θ are not of particular practical interest as they are not needed as input in the computation of capital requirement, but they are briefly discussed since they are sampled together with

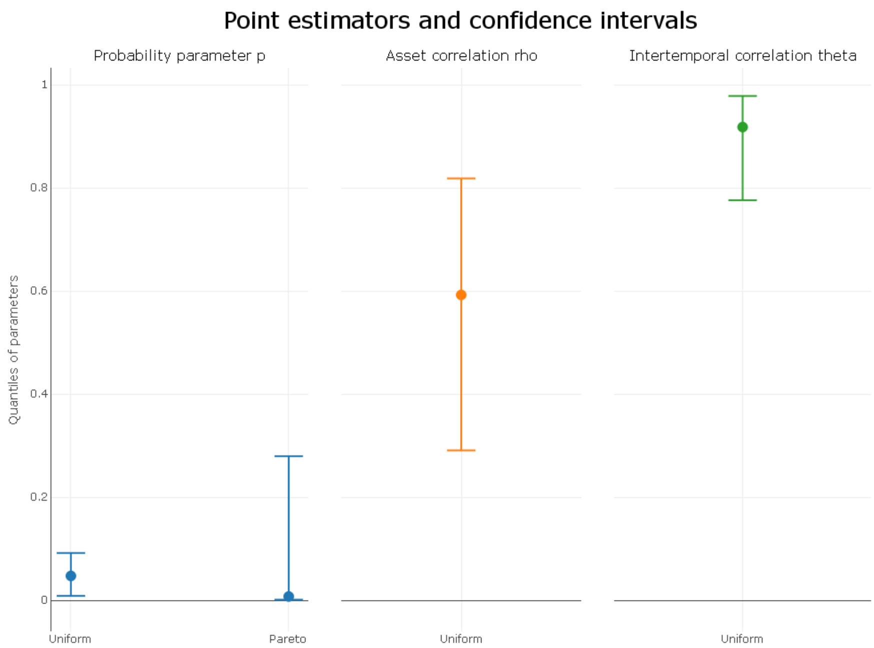

Figure 6.1: Point estimates (medians) and confidence intervals for p, ρ and θ of their marginal distributions, in the case of independence between priors

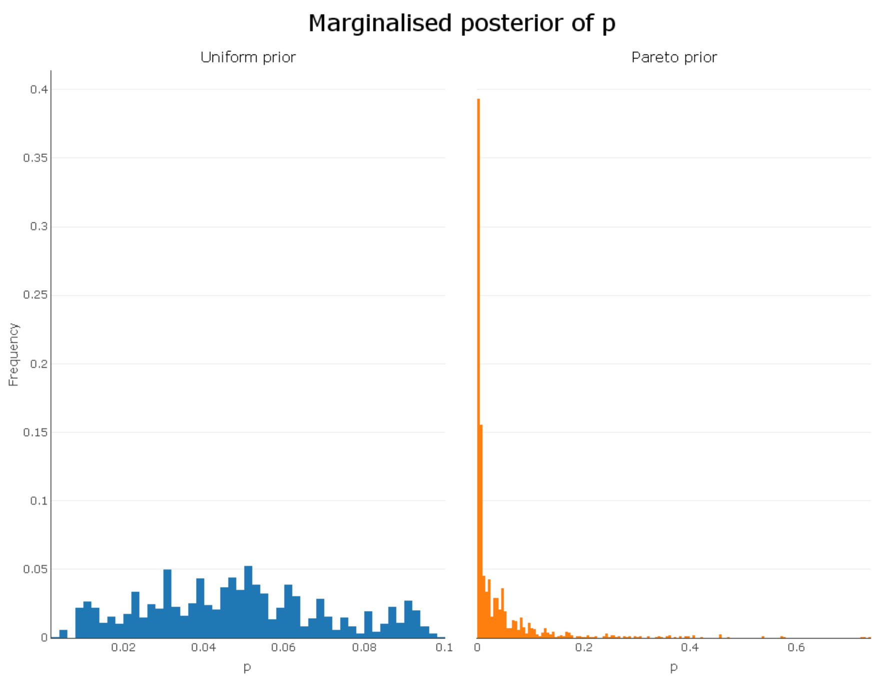

Figure 6.2: Samples of size 5000 of the marginalised posterior distribution for the probability parameter p, in the case of independent uniform and Pareto prior.

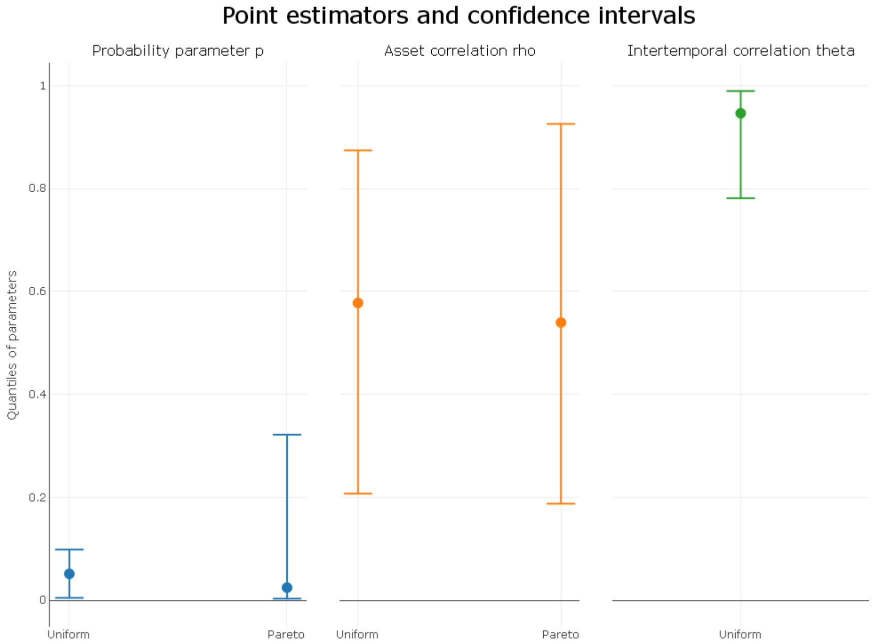

Figure 6.3: Point estimates (medians) and confidence intervals for p, ρ and θ of their marginal distributions, in the case of dependence between p and ρ.

p. The point estimates of both correlation parameters are close to their maximum likelihood estimates, so the priors do not too heavily influence the estimates. The uniform distribution therefore seems justified. Note also that the credible interval for θ is more narrow than the one of ρ. This is explained by the fact that the defaults are concentrated in two time periods ($1960-1965$ and $2013-2016$), so that there is more 'information' available for θ and therefore its credible interval more narrow.

Dependent prior

Following the approach in the bivariate single period setting, dependence between p and ρ is introduced through a Clayton copula with the dependence parameter equal to 0.5. The time correlation θ has been chosen independent of the other two parameters, as no clear relation came to mind. The two dependent priors are given by

$$\pi_{DU}(p,\rho,\theta \mid u) = C^{Cl}(\Pi_{UBE}(p \mid u), \Pi_{UBE}(\rho \mid u=1))\pi(\theta),$$
$$\pi_{DP}(p,\rho,\theta \mid \xi) = C^{Cl}(\Pi_{PBE}(p \mid \xi), \Pi_{UBE}(\rho \mid u=1))\pi(\theta),$$

with the Clayton copula given as in chapter 2.

When prior information on the dependence between p or ρ and θ is available, this can of course be easily introduced by choosing a suitable copula. A different one than the Clayton must be chosen however, since that is a bivariate copula.

In figure 6.3 the point estimators and confidence intervals in the case of dependence can be seen. All three point estimators corresponding to the uniform prior are slightly higher than the independent case, which is as expected since introducing dependence generally causes estimates to be more conservative. The estimate for p corresponding to the Pareto prior is

	uniform	MAD	Pareto	MAD
p	0.0514	0.0251	0.0245	0.0191
ρ	0.5774	0.1109	-	-
θ	0.9459	0.0215	-	-

Table 6.4: Median values and their median absolute deviations of the sample drawn from the posterior distribution with Pareto or uniform prior for p.

Year	2009	2010	2011	2012	2013	2014	2015	2016
Number of defaults	0	0	0	1	4	0	2	1

Table 6.5: Number of defaults in the last 8 years of the data set.

much more conservative than in the independent case. Where the Pareto estimate previously gave a nice result, it is now actually too conservative as it is about four times larger than the naive estimate. Three solutions are therefore proposed, in decreasing impact on the PD estimate, these are:

- Choose a different value for the hyperparameter ξ of the Pareto prior. The value $\xi = 4$ has been chosen as an example value, but with the use of experts a more realistic value can be chosen for this parameter.
- Change the marginal prior for the asset correlation ρ. In this example a uniform prior is assumed, which puts quite some mass on higher values of ρ. As in the literature values between 0.12 and 0.24 are deemed reasonable, the marginal prior might be too conservative.
- Change the dependence between p and ρ. A Clayton copula with parameter 0.5 is chosen, but there is no real justification for this choice of parameter. If less dependence between p and ρ is assumed, point estimates for p will be less conservative.

From a theoretical point of view, the dependent situation is preferred over the independent one as it is a more realistic representation of the real world situation. With the chosen parameters the point estimates might be too conservative to use in the computation of capital requirements. The model has enough flexibility (three marginal prior distributions and a dependence parameter) to incorporate expert prior information. The use of a Pareto prior for p and a uniform prior for θ seems well justified, although the parameter ξ needs calibration. A different prior than the uniform $[0, 1]$ for ρ might be in order, and further investigation is recommended.

6.3. CR+ model

Now the multi-period extension of the CR+ model described in chapter 4 will be considered. A major drawback of this extension is noted up front, the likelihood requires a number of computations of order 2^T. The data set under consideration consists of 57 years of data, but $2^{57} \approx 10^{17}$, which is unfeasible to use in practice. Therefore only the last 8 years of data of the portfolio are considered, with the amount of defaults given in table 6.5.

6.3.1. Maximum likelihood estimator

Before giving the estimates produced by maximum likelihood estimation, the 'naive' estimate of the probability of default of this subportfolio is given by

$$\hat{p} = \frac{\sum_{t=1}^{T} k_t}{\sum_{t=1}^{T} n_t} = \frac{8}{312} = 0.0256,$$

which will again serve as a reference probability. The maximum likelihood estimates for μ, σ and θ are given in table 6.6. In this table the mean default rate is also converted to the probability of default by dividing by 39, the amount of obligors in the portfolio. The maximum likelihood estimate of the probability of default is 0.0132, even lower than the naive estimate. Next, Bayesian estimators are considered.

MLE	Estimate value
μ	0.5161
p	0.0132
σ	0.9231
θ	0.6330

Table 6.6: Maximum likelihood estimates for the CR+ model.

6.3.2. Bayesian estimators

Again a multivariate prior distribution will be employed, now on the parameters μ, σ and θ of the CR+ model. The same priors considered in chapter

5 will be considered for μ and σ, an independent Exponential prior and the so-called Marshall-Olkin Exponential prior with parameter $\eta = 0.5$. For the intertemporal correlation parameter θ again a uniform prior on the unit interval is chosen, independent of μ and σ.

Just as in the multi-period extension of the Vasicek model, numerical integration techniques cause some difficulties, so the Metropolis within Gibbs sampler is used in obtaining estimates. For the proposal distribution for θ the same $\mathrm{Beta}(\frac{1}{1-x}, 2)$ distribution is chosen as before, while for μ and σ an $\mathrm{Exp}(\frac{1}{x})$ distribution is used, so that the mean is equal to the previous value x.

Independent priors

Since the multivariate prior distribution for the parameters is chosen to be independent, it is simply the product of the marginal distributions. For μ and σ an Exponential distribution with parameters λ_1 and λ_2 is chosen, so that:

$$\begin{aligned}\pi_{ind}(\mu, \sigma, \theta \mid \lambda_1, \lambda_2) &= \pi(\mu \mid \lambda_1)\pi(\sigma \mid \lambda_2)\pi(\theta) \\ &= \lambda_1 e^{-\lambda_1 \mu} \lambda_2 e^{-\lambda_2 \sigma} 1_{[0,1]}(\theta).\end{aligned}$$

Now a sample is drawn from this multivariate distribution with equal parameters $\lambda_1 = \lambda_2 = 1$ and point estimators are applied to the marginals, shown in figure 6.4 with exact values presented in table 6.7. The median value 0.0119 of the probability of default is slightly lower than the maximum likelihood estimate of 0.0132, which shows it is not overly conservative. The median of σ with value 0.7119 is lower than its maximum likelihood estimate of 0.9231. This might seem counter intuitive since the prior for σ has a mean of 1, but is less surprising considering the median of the prior is approximately 0.6931. The posterior median of θ with value 0.5146 is also lower than its MLE of 0.6695. This result is as expected since the mean and median of the uniform is 0.5, bringing the maximum likelihood estimate down.

Median value	Exponential	MAD
μ	0.4627	0.2661
p	0.0119	0.0068
σ	0.7119	0.4908
θ	0.5146	0.2621

Table 6.7: Median values and their median absolute deviations of the sample drawn from the posterior distribution with an independent Exponential prior for μ and σ.

Dependent priors

Now dependence is introduced between the parameters μ and σ. For this purpose the bivariate Marshall and Olkin Exponential distribution from chapter 5 is chosen. For the prior on the time correlation parameter θ the choice is still the uniform distribution, independent of μ and σ. The dependent prior is given by

$$\begin{aligned}\pi_{dep}(\mu, \sigma, \theta \mid \lambda_1, \lambda_2, \eta) &= \pi(\mu, \sigma \mid \lambda_1, \lambda_2, \eta)\pi(\theta) \\ &= \begin{cases} \lambda_2 (\lambda_1 + \eta) \exp\left(-(\lambda_1 + \eta)\mu - \lambda_2 \sigma\right) 1_{[0,1]}(\theta), & \text{for } \mu \geq \sigma \\ \lambda_1 (\lambda_2 + \eta) \exp\left(-(\lambda_2 + \eta)\sigma - \lambda_1 \mu\right) 1_{[0,1]}(\theta), & \text{for } \mu < \sigma. \end{cases}\end{aligned}$$

Sampled median values of the marginalised posterior distributions and their 95% Bayesian credible intervals are found in figure 6.5. Exact values of the median and of the median absolute deviation are found in table 6.8. The results of using the dependent prior are not surprising. The probability of default of 0.0134 is slightly higher than the value of 0.0119, similar to the behaviour of all extensions introducing dependence seen so far. The median value of σ of 0.8911 is higher than the independent estimate of 0.7119, which could be explained by the fact that the Marshall and Olkin Exponential distribution allows for higher simultaneous values of μ and σ. Lastly, the median value of θ in the dependent case is close to the independent one, 0.5460 and 0.5146 respectively. As for both estimates the independent uniform prior is used this behaviour is as expected.

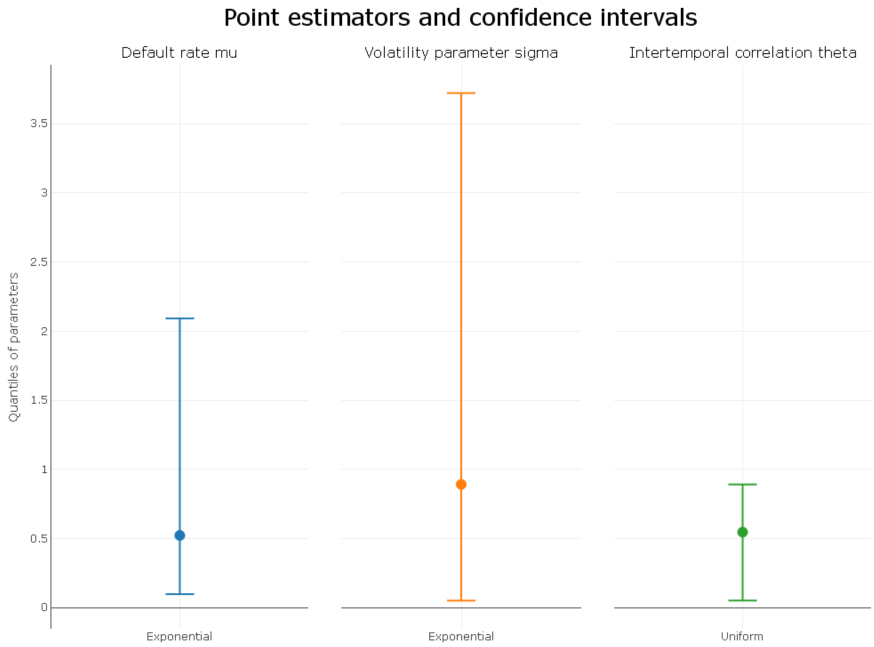

Figure 6.4: Point estimates (medians) and confidence intervals for μ, σ and θ of their marginal distributions, in the case of independence between priors

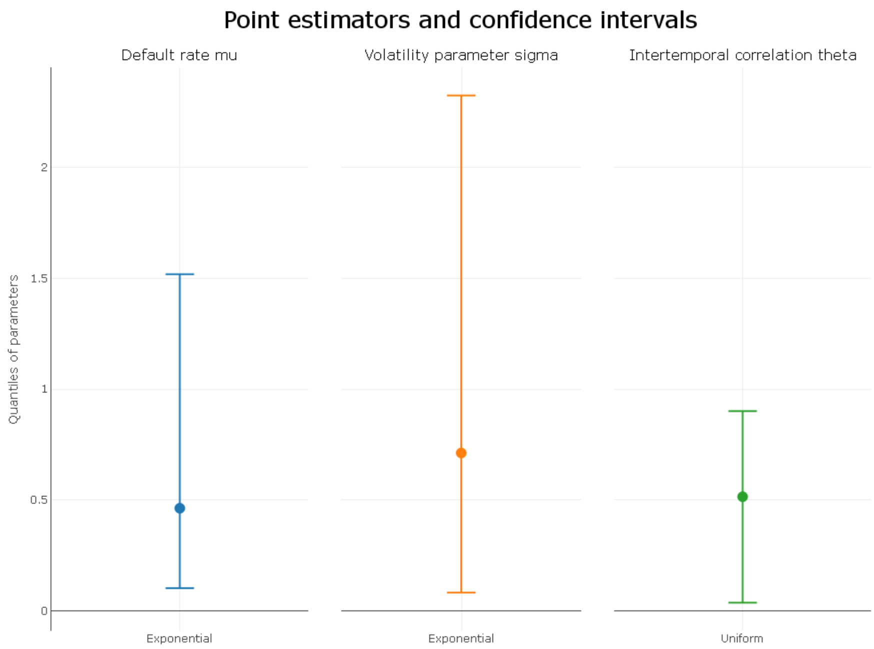

Figure 6.5: Point estimates (medians) and confidence intervals for μ, σ and θ of their marginal distributions, in the case of dependence between μ and σ.

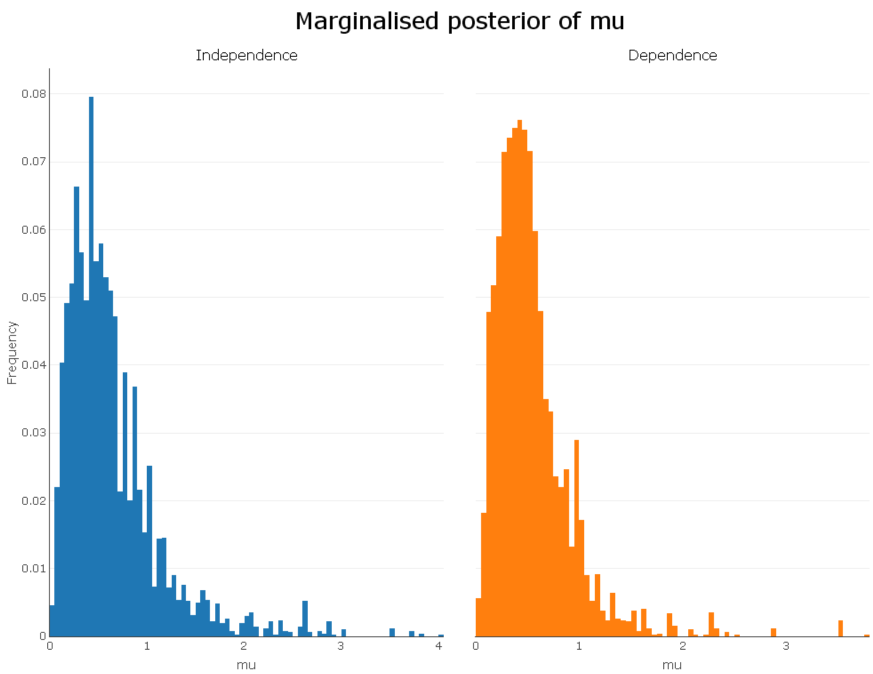

Figure 6.6: Samples of size 5000 of the marginalised posterior distribution for the probability parameter μ.

In figure 6.6 two histograms of the marginalised posterior distributions for μ, in the case of independence and dependence, are given by a sample of size 5000. Both distributions are very similar, unimodal with a right tail.

Median value	Exponential	MAD
μ	0.5221	0.3508
p	0.0134	0.0090
σ	0.8911	0.6837
θ	0.5460	0.2826

Table 6.8: Median values and their median absolute deviations of the sample drawn from the posterior distribution with a Marshall Olkin Exponential prior for μ and σ.

This chapter is concluded by some final comments on the performance of the CR+ model.

- Estimates for all three parameters in the CR+ model seem reasonable. The probability of default is in both maximum likelihood as Bayesian estimation lower than the naive estimate, caused by the introduction of dependence. This means the estimate of p is more robust, a desired property by the regulator.
- The use of a dependent prior is preferred over an independent one, as it is known for a fact such dependence exists. Estimates in both cases behave as expected.
- A major disadvantage of the model are the computational requirements, as they are of order 2^T. For moderate lengths of the time period under consideration (say $T < 10$), estimation is still feasible. For periods longer than this there may be improvement possible, discussed in chapter 4.

7

Conclusion

In this thesis two models for the estimation of the probability of default in the case of a low default portfolio are considered, the Vasicek and Credit Risk+ models. Both of these models have fundamental differences, but depend on the same type of parameters. One for the probability of default and a parameter corresponding to the correlation between obligors.

It is shown that estimators such as the maximum likelihood method may fail to provide reasonable estimates, and as an improvement Bayesian techniques are investigated. Various priors are examined, and new on the subject is the consideration of multivariate priors. Simulated results are given for the single period variant of the models, the multi-period extensions are applied to a real data set of sovereign defaults.

For the Vasicek model a prior for the probability of default, based on the Pareto distribution, seems the most promising. For the correlation parameter good performance is shown for the uniform distribution as a prior. In the bivariate case dependence is introduced by use of a Clayton copula. Furthermore, it is shown that the general form of a prior can be misleading, as it influences the likelihood only on the domain where the latter has mass.

In the Credit Risk+ model first different bivariate exponential prior distributions are investigated. Curious behaviour is shown where estimates of the probability of default actually decrease when more defaults are observed. These results are obtained both by numerical integration and with sampling techniques. As a different approach the objective Jeffreys prior is considered, but shown to be too conservative for practical use.

The multi-period extensions of the Vasicek and Credit Risk+ model are applied to a real data set consisting of sovereign defaults. Estimates in the Vasicek model are for the heuristically chosen hyperparameters too conservative, but calibration by experts might circumvent this problem. The Credit Risk+ model seems to produce reasonable estimates for the probability of default, but requires in its current form too much computational power.

The Bayesian approach is shown to be superior to other estimation methods. In this thesis parameters have been chosen heuristically, but in practice these may be calibrated based on expert judgement or comparable data sets. However, practitioners need to be careful when applying subjective methods as they need to be able to show a margin of conservativeness. Additional guidance by the regulator, specific to the Bayesian approach, is therefore much welcome and likely in the future.

All in all, the Vasicek model is preferred over the Credit Risk+ one, because of two reasons. The first is the domain of parameters, as in the Vasicek model the probability of default and correlation parameter are defined on the unit interval, prior distributions are more intuitive than those on the infinite domain of the Credit Risk+ model. The second reason is that because the Vasicek model is a factor model, the extension to a multi-period setting comes more naturally and is computationally much faster.

Some suggestions for further research are the following:

- Investigation on suitable priors for the correlation parameter in the Vasicek model. Although the uniform prior performed quite well, other priors may be able to capture prior information better.
- The use of rating classes. In this thesis these have not been used, and while rating grades have been shown to be unreliable sometimes, they are still used in the Basel framework.
- A Beta Poisson model. The Credit Risk+ model is essentially a mixing distribution of a Poisson and Gamma. But since probabilities of default are small, the Gamma distribution may be replaced with a Beta.

Bibliography

[1] J. Holt, *A summary of the primary causes of the housing bubble and the resulting credit crisis: A non-technical paper,* The Journal of Business Inquiry **8**, 120 (2009).

[2] B. C. on Banking Supervision, *International convergence of capital measurement and capital standards: A revised framework,* Bank for International Settlement (2006).

[3] M. Crouhy, D. Galai, and R. Mark, *A comparative analysis of current credit risk models,* Journal of Banking & Finance **24**, 59 (2000).

[4] Q. Meng, A. Levy, A. Kaplin, Y. Wang, and Z. Hu, *Implications of pd-lgd correlation in a portfolio setting,* Moody's Analytics (2010).

[5] B. C. on Banking Supervision, *Basel iii: Finalising post-crisis reforms,* Bank for International Settlement (2017).

[6] B. C. on Banking Supervision, *An explanatory note on the basel ii irb risk weight functions,* Bank for International Settlement (2005).

[7] B. C. on Banking Supervision, *Validation of low-default portfolios in the basel ii framework, september 2005,* Bank for International Settlement (2005).

[8] B. C. on Banking Supervision, *Analysis of risk-weighted assets for credit risk in the banking book,* Bank for International Settlement (2013).

[9] B. C. on Banking Supervision, *Basel iii: A global regulatory framework for more resilient banks and banking systems,* Bank for International Settlement (2011).

[10] A. Hájek, *Interpretations of probability,* https://plato.stanford.edu/archives/win2012/entries/probability-interpret/ (2012).

[11] M. Bayes and M. Price, *An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfrs,* Philosophical Transactions (1683-1775) , 370 (1763).

[12] T. Zheng, *Bayes' formula,* (2008).

[13] G. Young and R. Smith, *Essentials of Statistical Inference* (Cambridge University Press, New York, 2005).

[14] H. Jeffreys, *An invariant form for the prior probability in estimation problems,* Proc. R. Soc. Lond. A **186**, 453 (1946).

[15] J. Rice, *Mathematical statistics and data analysis* (Nelson Education, 2006).

[16] W. F. Darnieder, *Bayesian methods for data-dependent priors*, Ph.D. thesis, The Ohio State University (2011).

[17] H. Aratyn and C. Rasinariu, *A short course in mathematical methods with Maple* (World scientific, 2006).

[18] W. A. Link and M. J. Eaton, *On thinning of chains in mcmc,* Methods in ecology and evolution **3**, 112 (2012).

[19] P. Embrechts, F. Lindskog, and A. McNeil, *Modelling dependence with copulas,* Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich (2001).

[20] A. Sklar, *Random variables, distribution functions, and copulas: a personal look backward and forward,* Lecture notes-monograph series , 1 (1996).

[21] M. Fréchet, *Les tableaux de corrélation et les programmes linéaires,* Revue de l'Institut international de statistique , 23 (1957).

[22] R. Nelson, *An introduction to copulas, lectures notes in statistics vol. 39,* (1999).

[23] T. Mikosch, *Copulas: Tales and facts—rejoinder,* Extremes **9**, 55 (2006).

[24] M. Bonetti and Cirillo, *Exact algorithms for the multinomial counts: Maximum, minimum, range and sums of the largest order statistics,* (2017).

[25] S. D. Horn, *Goodness-of-fit tests for discrete data: A review and an application to a health impairment scale,* Biometrics **33**, 237 (1977).

[26] A. Agresti, *Categorical data analysis*, Vol. 482 (John Wiley & Sons, 2003).

[27] S. Nadarajah, M. Alizadeh, and S. Bagheri, *Bayesian and non-bayesian interval estimators for the poisson mean,* REVSTAT–Statistical Journal **13**, 245 (2015).

[28] R. C. Merton, *On the pricing of corporate debt: The risk structure of interest rates,* The Journal of finance **29**, 449 (1974).

[29] F. Modigliani and M. H. Miller, *The cost of capital, corporation finance and the theory of investment,* The American economic review **48**, 261 (1958).

[30] O. Vasicek, *The distribution of loan portfolio value,* Risk **15**, 160 (2002).

[31] B. et al, *An Introduction to credit risk modeling* (Chapman & Hall/CRC, Boca Raton, 2003).

[32] K. Pluto and D. Tasche, *Thinking positively,* Risk **18**, 72 (2005).

[33] N. Benjamin, A. Cathcart, and K. Ryan, *Low default portfolios: A proposal for conservative estimation of default probabilities,* (2006).

[34] S. Hanson and T. Schuermann, *Confidence intervals for probabilities of default,* Journal of Banking & Finance **30**, 2281 (2006).

[35] N. M. Kiefer, *Default estimation for low-default portfolios,* Journal of Empirical Finance **16**, 164 (2009).

[36] M. Van Der Burgt, *Calibrating low-default portfolios, using the cumulative accuracy profile,* Journal of Risk Model Validation **1**, 17 (2008).

[37] G. Fernandes and C. A. Rocha, *Low default modelling: A comparison of techniques based on a real brazilian corporate portfolio,* Credit Scoring and Credit Control XII , 24 (2011).

[38] W. Orth, *Default probability estimation in small samples-with an application to sovereign bonds,* (2011).

[39] D. Tasche, *Bayesian estimation of probabilities of default for low default portfolios,* Journal of Risk Management in Financial Institutions **6**, 302 (2012).

[40] D. Vazza and N. W. Kraemer, *Annual global corporate default study and rating transitions,* Standard & Poor's RatingsDirect , 1 (2016).

[41] C. D. Brown and H. T. Davis, *Receiver operating characteristics curves and related decision measures: A tutorial,* Chemometrics and Intelligent Laboratory Systems **80**, 24 (2006).

[42] E. Falkenstein, A. Boral, and L. Carty, *Riskcalc for private companies: Moody's default model,* (2000).

[43] J. Fox, *Applied regression analysis and generalized linear models* (Sage Publications, 2015).

[44] W. J. Moonan, *Linear transformation to a set of stochastically dependent normal variables,* Journal of the American Statistical Association **52**, 247 (1957).

[45] C. Suisse, *Creditrisk+: A credit risk management framework,* Credit Suisse Financial Products (1997).

[46] L. Koopman and Lucas, *Discrete copula distributions with time-varying marginals and dependence structure,* (2015).

[47] E. Venter, *Probability of Default Calibration for Low Default Portfolios: Revisiting the Bayesian Approach*, Master's thesis, University of Stellenbosch, South Africa (2003).

[48] S. Kotz and J. R. Van Dorp, *Beyond beta: other continuous families of distributions with bounded support and applications* (World Scientific, 2004).

[49] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2008), ISBN 3-900051-07-0.

[50] A. W. Marshall and I. Olkin, *A multivariate exponential distribution,* Journal of the American Statistical Association **62**, 30 (1967).

[51] N. Balakrishnan and C. Lai, *Continuous bivariate distributions. 2009,* (2009).

[52] M. Bonetti, P. Cirillo, P. M. Tanzi, and E. Trinchero, *An analysis of the number of medical malpractice claims and their amounts,* PloS one **11**, e0153362 (2016).

[53] *Database of sovereign defaults,* https://www.bankofcanada.ca/2014/02/technical-report-101/ (2017), accessed: 16-04-2018.

[54] D. T. Beers and J. Mavalwalla, *Database of sovereign defaults, 2017,* (2017).

[55] Imf, *World Economic Outlook, April 2016* (International Monetary Fund, 2016).

[56] D. Tasche and K. Pluto, *Estimating probabilities of default for low default portfolios,* (2005).