Parameter Uncertainty in Credit Risk Portfolio Models

Public version

Parameter Uncertainty in Credit Risk Portfolio Models

Master Thesis
Author: Chuck Yuen Liu
Program: Business Mathematics & Informatics
Supervisor VU University Amsterdam: Prof. Aad van der Vaart
Second supervisor VU University Amsterdam: Prof. Rob van der Mei
Supervisor Rabobank: Dr. Ir. Erik Winands

VU University Amsterdam
Faculty of Exact Sciences
De Boelelaan 1081a
1081 HV Amsterdam
The Netherlands

Rabobank Nederland
Group Risk Management
Risk Model Validation & Methodologies
Utrecht
The Netherlands

Management Summary

Credit risk on portfolio level is estimated using a statistical model with the 'probability of default' or PD as the main parameter. In this context parameter uncertainty concerns the inability to make an accurate estimation of the required capital due to insufficient data. It is shown that the calculated required capital for an assumed risk level is in fact underestimated. However parameter uncertainty is only (seriously) present in small portfolios (<200 obligors). If taken into account parameter uncertainty for these small portfolios, the additional capital to be held may be higher if one can make a reliable estimation of the economic state in previous years. Not making the assumption of correct estimation of the economic state, additional capital could be great as well. The correct estimation of the economic state poses a challenge, as this is not straightforward and the economic state variable contains parameter uncertainty as well.

The numbers in this report are only an indication of a possible impact, as the results are dependent on a number of strong assumptions. The capital not incorporating parameter uncertainty only uses default rate data and an economic state variable, but in practice more information is available and used to calculate the capital requirement. That is we ignore the fact that in practice PDs are estimated via a mathematical model in combination with expert judgments. While in this report we assume that the mean of historical default rates are directly used as an estimate. Finally in practice conservatism is included in determining the capital to be held. The conservatism takes into account the uncertainty in the estimate of the required capital and includes parameter uncertainty. Further research to take into account this conservatism and the different PD models in banks, can be conducted to actually determine whether more capital is required or not.

CONTENTS

1 INTRODUCTION†

In the past two decades the awareness towards risk management increased from the institutional as well as from the regulatory perspective. The Basel Committee on Banking Supervision (BCBS), based at the Bank for International Settlements (BIS), was established by the central banks of the Group of Ten countries (G10) and formulates recommendations on supervisory standards and guidelines on banking laws and regulations. Although the Basel Committee doesn't hold any official authority, due to its broad composition of national governments and central banks, recommendations are followed and enforced by law by practically all individual regulating authorities over the world.

In 1988 it introduced a capital risk measurement system known as the Basel Capital Accord or Basel I, which focuses primarily on credit risk and required banks to hold a minimum of 8% capital of its risk weighted assets (enforced by the end of 1992) to cover unexpected losses. The second Capital Accord called Basel II made further advancement in risk measurement, it was introduced in 2004 and put into force by 2008. The more comprehensive accord ensured capital allocation to be more risk sensitive, separating measurement in credit, market and operational risk and enhanced disclosure requirements allowing market participants to asses capital adequacy of banks.

With the introduction of Basel II, financial institutions are given the choice to follow either the standardized approach or the Internal Ratings-Based (IRB) Approach to credit risk measurement. The IRB approach encourages to think actively about risk management by demanding banks to develop their own statistical models for the calculation of the risk weighted assets. These models are built and based on knowledge within banks, however approval is needed from the relevant regulator. As a consequence methodologies need to be substantiated by banks. The choice for IRB is beneficial as the minimum capital requirement may be less.

From the bank's perspective adequate credit risk assessment is essential as the goal of credit risk management is to correctly measure credit risk such that one is able to maximize the risk-adjusted return on capital (RAROC) while maintaining credit and other risk exposures within acceptable parameters [1] . The risk-adjusted return on capital is influenced by the amount that the bank sets asides to cover credit risk at portfolio level that is the total losses due to the fraction of the obligors that defaults. The total allocation is determined by the risk appetite of the bank and its risk assessment of their portfolio.

Another motivation for banks lies in the ratings they receive from external agencies like Standard & Poor's (S&P), Moody's, Fitch and the Dominion Bond Rating Service (DBRS). The higher the rating, the safer a bank is perceived by investors, consumers and regulators which results in lower funding costs. On top credit ratings affect the faith of one bank in the other. Lower rated banks are perceived as having a higher default risk and therefore financial transactions with the relevant bank might be limited and/or at a higher cost.

† This chapter is based on the BIS website [1] and The Dutch Central Bank [2].

These developments have stimulated the risk awareness and measurement within the banking industry. A closer specification of risk is possible and a desired risk appetite can be further quantified. Depending on their risk appetite banks can hold the minimum capital requirement or more. The more capital is set aside to cover losses, the less capital is available for profitable investments. However the less capital is available for losses, the higher the probability of insolvency and ultimately bankruptcy.

In the light of accurate risk measurement, this paper will focus on credit risk. The total capital requirements for banks was €77 billion in the first quarter of 2010 for the whole Dutch banking industry, while 88% of this sum amounted to credit risk [2].

In this thesis on portfolio credit risk, the Asymptotic Single Risk Factor (ASRF) model plays a central role. It is prescribed by regulators for banks that adopt their own internal risk models to quantify capital requirements for credit risk. In practice the parameters of the ASRF model are estimated from historic data, point estimates are used for inference. Implicitly one assumes that these estimated parameters are true, that is the estimated parameters equal the true parameters. Making this assumption and thus neglecting parameter uncertainty causes erroneous estimates of the true value that is at risk [14].

A distinction must be made in the type of parameter uncertainty, one uncertainty concerns sampling error or noise due to limited data. For any consistent estimator in combination with infinite data, the consistent estimator equals the true parameter that one seeks. However when data becomes finite, a consistent estimate is in general not equal to the true parameter. The gap between the consistent estimate and the true parameter is a consequence of sampling error or noise. The other type of parameter uncertainty relates to the evolution of parameters over time, i.e. the population characteristics today has changed compared to the population characteristics yesterday.

The goal is investigate parameter uncertainty due to sampling error. This is done by assessing the gap between the estimated value at risk and the true value that is at risk for a bank. Additionally the goal is to estimate the value at risk that incorporates parameter uncertainty.

1.1 RESEARCH QUESTION

We formulate the problem around parameter uncertainty into formal research questions.

Main research question

- **How can parameter uncertainty in credit risk be dealt with?**

Sub questions

I. What is known in the literature about parameter uncertainty?
II. What is the impact of parameter uncertainty on the ASRF model?
III. How can parameter uncertainty be incorporated?

1.2 OUTLINE THESIS

In the next chapter We shed some light on the upcoming of risk management within banks and the resulting accords that have been introduced with the main focus on credit risk. The credit

risk model used in the advanced approach is based on the ASRF model, which will be derived in chapter 3. Capitalizing on the ASRF model we show results relevant for usage in quantifying parameter uncertainty. Chapter 4 gives an overview of relevant research conducted in this thesis, answering research question I. Chapter 5 highlights the theory on Bayesian statistics and its differences with the more familiar frequentist statistics. In chapter 6 the methodologies to investigate parameter uncertainty are explained, one conditioning the information on the economic state of the available data and one not capitalizing on this information. After which in chapter 7 the results are displayed for artificial generated data according to the ASRF model, answering research question I and II. The last paragraph of chapter 7 is devoted to the application in real portfolio data. Finally the conclusions in chapter 8 summarizes the findings in this thesis.

2 THE BASEL ACCORDS†

When the Organization of Arab Petroleum Exporting Countries (OAPEC) announced an oil embargo by the end of 1973, crude oil importing prices increased dramatically for West-European countries. The already unfavorable trade balance produced unprecedented deficits. The financing of the deficits was primarily managed by Euro currency markets with a key role for commercial banks. The embargo increased the speculation on the foreign exchange rate resulting in sharp fluctuations including the exchange rate between the Deutsche Mark and US dollar.

Some banks benefited from the resulting speculations on the foreign exchange rates whereas others incurred heavy losses. In the early Wednesday morning on the 26st of June in 1974, German regulators closed the Herstatt Bank, one of Germany's largest private banks primarily active in the foreign exchange market. Reportedly annual turnover in foreign transactions by the Herstatt Bank was over 60 billion Deutsche Mark a year. It is known now that this relatively large turnover was mainly a churning of losses accompanied by bookkeeping misconduct.

Many banks sold before and also on that 26th of June European currencies such as the Deutsche Mark to the Herstatt bank in exchange for delivery of US dollars on the 26th of June (spot contracts) or later. However having paid their leg of the transaction, they never received the other leg in dollars. This particular risk on failure in settlements of contracts was from then known as Herstatt or settlement risk.

The failure led to concerns of financial institutions' assessments of the inherent risks of their businesses along with German policy makers losing some of their confidence on the reliance of the market-orientated discipline. Even though Herstatt bank was (internationally) not a major player, foreign exchange markets were disrupted. International currencies went to certain big banks mostly located in New York while the same banks delayed and avoided foreign exchange deals with any suspicious European or Asian banks.

These developments and the desire to level the playing field internationally has led to the establishment of the Basel Committee in the same year Herstatt failed, with the main responsibility to improve the banking supervision worldwide. In 1988 the first capital accords were accepted by all member states with implementation in the industry by the end of 1992.

2.1 BASEL I

The first accord specified the required capital to cover the unexpected losses to be 8% of its risk weighted assets (RWA) and exposures, where capital is distinguished into two tiers. The Tier 1 or core capital consists of equity capital and disclosed reserves while other capital, the supplementary capital or Tier 2 capital is considered capital of lower quality. It included undisclosed reserves, revaluation reserves, general provisions, hybrid capital instruments and subordinated debt. Tier 2 capital was restricted to 100% of the tier 1 capital, so that the higher

† The opening is based on Goodhart [3], Kaufman [4] and Becker [5]. For more information on Basel I see BCBS [6], for Basel II see BCBS [7,8] and for Basel III see BCBS [9,10].

quality tier 1 capital would be at least 50% of the core capital or at least 4% of the risk weighted assets.

Furthermore deductions of the core capital were applied consisting of goodwill and investments in subsidiaries engaged in banking and financial activities which are not consolidated in national systems.

The assets and off balance sheet exposures are weighted according to broad categories of relative riskiness , with 5 different weights of 0%, 10%, 20%, 50% or 100%.

For example:

- 0% weight applies to cash, claims on national governments in national currencies.
- 0%, 10%, 20% or 50% weights (determined by the national regulator) for claims on banks incorporated in the OECD and loans guaranteed by OECD incorporated banks.
- 50% weight to loans fully secured by mortgage on residential property that is or will be occupied by the borrower or that is rented.
- 100% weight to claims on the private sector and claims on commercial companies owned by the public sector.

2.2 BASEL II

The revised framework, Basel II, made significant changes and consisted of three pillars. Pillar 1 consisted of the minimum capital requirement to cover unexpected losses distinguishing credit, market and operational risk, calculating the capital requirements each separately. Pillar 2 covers the supervisory review process including residual risks not covered in the first pillar. The last pillar treats market discipline and disclosure requirements allowing market participants to gauge the bank's capital adequacy.

Our focus lies with credit risk and consists of two different approaches the first a simple and the second a more sophisticated approach:

1. The Standardized approach (SA)
2. The Internal Ratings-based approach (Foundation-IRB or Advanced-IRB)

The simplest method, the standardized approach is based on risk weights dependent on external credit risk ratings issued by rating agencies. This allowed a more risk weighted approach as weights were determined not only by type but also by their credibility. For example the risk weights for claims on sovereigns and corporate companies are stated in the tables below.

Credit assesment	AAA to AA-	A+ to A-	BBB+ to BBB-	BB+ to B-	Below B-	Unrated
Risk Weight	0%	20%	50%	100%	150%	100%

Table 1 Risk Weights for sovereigns and central banks

Credit assesment	AAA to AA-	A+ to A-	BBB+ to BB-	Below BB-	Unrated
Risk Weight	20%	50%	100%	150%	100%

Table 2 Risk Weights for corporate companies

Another difference is the weight of 150% in the standardized approach, where the maximum weight in Basel I is 100%. The weights are further increased for assets with higher risk such as securitization tranches rated between BB+ and BB- at 350% risk weight. The ratings follow Standard & Poor's notations, but ratings of other agencies satisfying certain criteria are allowed as well. A minimum capital ratio of 8% of the risk weighted assets (RWA) is required by the Basel Committee for banks to keep aside. The capital requirement K for credit can thus be determined by:

$$K\ (in\ terms\ of\ Exposure\ at\ default) = 8\% * RWA$$

$$Risk\ weighted\ assets\ (RWA) = \sum_{all\ portfolios} risk\ weight_{portfolio} * portfolio$$

Table 3 Standardized approach

In the Internal Ratings Based approach, the capital requirement is calculated using the ASRF model for five broad classes or portfolios of assets: a) corporate, b) sovereign, c) bank, d) retail and e) equity and are each further subdivided in sub-classes. The parameters PD, LGD and EAD (and also M the effective maturity) in the ASRF model are allowed to be estimated by internal developed models subject to some minimum standards while in the Foundation Internal Ratings Based approach only PD is estimated internally using regulator's estimates for the other parameters. The derivation and explanation of the parameter will be treated in chapter 3.

Furthermore in the IRB approach for credit risk a scaling factor of 1.06 is used to broadly maintain the aggregate level of minimum capital requirements. while it also provides incentives to adopt the more advanced IRB approach. For the classes corporate, sovereign and bank exposures the following formulas hold:

$$K\ (in\ terms\ of\ Exposure\ at\ default) = LGD * (Value\ at\ Risk\ - expected\ loss)$$
$$Scaled\ capital\ requirement\ K^{scaled} = K * 1.06$$

$$Risk\ weighted\ assets\ (RWA) = K^{scaled} * 12.5 * EAD$$

Table 4 IRB approach

The expected losses are subtracted from the 'Value at Risk' as the capital requirement is meant to cover the unexpected losses. The calculation of the Value at Risk, the expected and unexpected losses will also be treated in chapter 3. For the classes corporate, sovereign, bank and retail exposures not in default or hedged is calculated by PD times LGD times EAD. As one can observe the total capital requirement is proportional to EAD and LGD while these are calculated separately. Finally the RWA can be calculated to compare the scaled capital requirement with the capital requirement in the standardized approach.

The Value at Risk (VaR) is one well known measure to asses risk and is specified with a confidence parameter $\alpha \in (0,1)$ parameter (typically low) and for a specific time horizon.

$$VaR_{1-\alpha}(Total\ Loss\ of\ portfolio) \stackrel{\text{def}}{=} q_{1-\alpha}(Total\ Loss\ of\ portfolio)$$

Where $q_{1-\alpha}(.)$ denotes the quantile function which returns the $(1-\alpha)^{th}$ quantile.

The quantile function is defined as $q_p(X) = \inf\{x \in \mathbb{R}: F(x) \geq p\}$, where X is the random variable having cumulative distribution $F(x)$.

Now we get $VaR_{1-\alpha}(Total\ Loss\ of\ portfolio) = \inf\{v \in \mathbb{R}: F(v) \geq 1-\alpha\} =$
$\inf\{v \in \mathbb{R}: \mathbb{P}(Total\ Loss\ of\ portfolio \leq\ v) \geq 1-\alpha\}$

Definition Value at Risk(VaR)

$$VaR_{1-\alpha}(Total\ Loss\ of\ portfolio) = \inf\{v \in \mathbb{R}: \mathbb{P}(Total\ Loss\ of\ portfolio \leq\ v) \geq 1-\alpha\}^*$$

Thus the VaR is defined as the $(1-\alpha)^{th}$ quantile of the total portfolio loss distribution.

For the VaR the Dutch regulator has set the time horizon at one year with $\alpha = 0.1\%$. In practice the term Regulatory Capital (RC) is used and is defined to be the capital requirement set out in table 3 or 4 depending on the choice of approach.

Next to RC banks calculate the Economic Capital (EC), this is the amount the bank is willing to set aside determined by the bank's individual risk appetite when there is no regulation. In practice EC is always equal or larger than RC, this provides an additional loss buffer on top of the amount for RC and will also determine for a large part the credit ratings assigned by external rating agencies.

2.3 Parameter uncertainty in Basel II

The estimates of VaR is highly sensitive to deviations from estimates of the probability of default (PD) . In the graph below we plotted the VaR measure against different PD, it is clear that an erroneous estimate of the PD has huge impact on the VaR, especially for portfolios with smaller PDs.

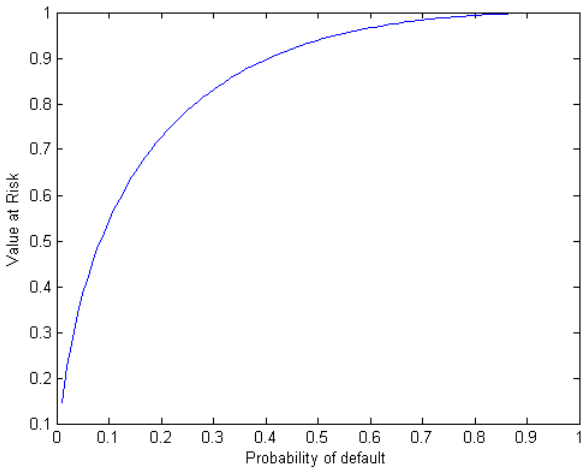


Figure 1 VaR with asset correlation at 0.20

* This is identical to the definition of Value at Risk in [11].

We zoom further in on a portfolio with true PD=0.03 but is estimated ranging from 0.01 to 0.05. Observe that the VaR as a function of PD is concave so that the derivative is decreasing, i.e. deviations from the true VaR that is larger for underestimation than for overestimation of PD.

The deviations are large for the PD estimated incorrectly on the down side banks can be undercapitalized by 50%, which would serious endanger banks in down turn economies. While estimates on the up side would mean that banks would be holding more capital than needed up to 33%.

Probability of default	0.01	0.02	0.03	0.04	0.05
Value at Risk	-50%	-22%	base	18%	33%

Table 5 Impact of wrong estimates of the PD (0.2 correlation)

The impact of lower asset correlations is relatively small, for example with the same true PD=0.03 but changing the asset correlation to 0.1 adds only 2-8% extra to the range of 0.01-0.05 for the PD. Thus impact of underestimation and overestimation of the VaR is mainly driven by the uncertainty in estimating the PD. .

Probability of default	0.01	0.02	0.03	0.04	0.05
Value at Risk	-55%	-25%	base	22%	41%

Table 6 Impact of wrong estimates of the PD (0.1 correlation)

2.4 TOWARDS BASEL III

As the full commitment of all member states of the G20 to fully adopt the Basel II agreements was set in 2011, the development of its successor named Basel III continued. In July 2009 the measurement of risks related to securitization and trading book exposures were agreed and in December 2010 the committee released Basel III. In the new capital accord higher levels of capital requirements are set and a new global liquidity framework is introduced. The members agreed to start implementation of Basel III in 1 January 2013 in phases, with the final phase implemented by 2019.

The phasing refers for example to the minimum common equity requirement in Basel II being only 2% but gradually increasing to 3.5% by January 2013, 4.0% in January 2014 and finally 4.5% in January 2015. The common equity is part of the Tier 1 capital requirement, where the Tier 1 must increase from 4% to 6% in the same period. New is the capital conservation buffer set at a minimum of 2.5% to be kept aside, which is to absorb losses during periods of stress. During these periods of stress banks are allowed to draw from this buffer, but when capital gets closer to the minimum regulatory capital constraints, the tighter the regulation will be on the respective bank. New is also a countercyclical buffer which is implemented to protect the banks from periods of excess credit growth, a requirement set by national regulators in the range of 0-2.5% of common equity or other fully loss absorbing capital.

For the new global liquidity framework the liquidity coverage ratio (LCR) will be introduced in January 2015 and the net stable funding ratio will move to a minimum standard by January 2018.

Banks that pose a potential threat to the stability of the financial sector, so called system banks will be required to hold additional capital. The exact details are not known yet, but this and the above mentioned increasing capital requirements have a positive effect for banks to absorb unexpected losses while the capital requirement based on the ASRF model to absorb unexpected losses in Basel II will remain in the Basel III framework.

3 CREDIT PORTFOLIO MANAGEMENT*

One important characteristic on the choice of credit risk models for regulatory supervisors was that credit risk models should be portfolio invariant. That is the capital required for any given load should only depend on the risk of that loan and must not depend on the portfolio it is added to [3]. The desired characteristic turns out to have a strong influence on the structure of the portfolio mode, it is essentially shown [19] that the ASRF models are portfolio invariant.

Consequently for researching the impact of parameter uncertainty of credit risk at the portfolio level, the ASRF model will be used.

In the ASRF model credit risk is determined by a set of parameters or risk components, which are assumed independent of each other. Each parameter is estimated using an internal model specific for the portfolio.

$PD \in (0,1)$,the probability of default.
$\rho \in [0,1]$, the correlation between obligors have equal pair wise correlation.
$LGD \in [0,1]$, the loss given default denoted in a fraction of the loss (sometimes the recovery rate is used, which is one minus the LGD).
$EAD \in \mathbb{R}^+$, the exposure given default denoted in money units.

The correlation parameter has been chosen by the Basel Committee to be a deterministic constant or a deterministic function of the PD. For example in residential mortgage exposures (retail) the correlation parameter is set at 0.15. For the classes corporate, sovereign and bank, correlation is determined by a function mapping it between 12% and 24% [7].

$$\rho = 0.12 * w + 0.24 * (1 - w)$$
$$where\ w = \frac{1 - e^{-50PD}}{1 - e^{-50}}$$

Some adjustments are permitted such as for small and medium sized companies (the corporate class) where the correlation is adjusted downwards depending on the annual sales.

We will research the impact of the parameter uncertainty by PD only, making the assumption that ρ, LGD and EAD are known and deterministic.

3.1 THE ASRF MODEL

The total loss distribution of the credit portfolio is modeled using the Vasicek model. A portfolio is considered which consists of n identical obligors (from a credit risk perspective), each with an exposure of 1/n (the total exposure equals one). The asymptotic case is considered in which $n \to \infty$, so that the portfolio is perfectly fine-grained or perfectly granular.

* The ASRF model is based on a paper of Vasicek [11]. This chapter gives careful derivation and extends the results.

At time 0 all obligors' assets are equal in value after which each individual obligor's asset evolves in its own way. At a later time point the value of each obligor is evaluated. An obligor is in default at this later time point whenever the value of its assets is below a certain threshold at the time of the evaluation. Thus the value of its assets $V_{i,t}$ is a direct measure of the credit condition of each obligor i at time t. $V_{i,t}$ changes continuously over time t and satisfies the following stochastic differential equation (SDE) for i=1,…,n and t≥0:

$$dV_{i,t} = uV_{i,t}\,\mathrm{dt} + \sigma V_{i,t}\,dW_{i,t}$$
$$Where\ u, \sigma \in \mathbb{R}^{+}\ and\ W_{i,t}\ a\ Brownian\ motion$$

The SDE has an explicit solution given by:

$$V_{i,t} = V_{i,t-\Delta} e^{\left(u - \frac{\sigma^2}{2}\right)\Delta + \sigma(W_{i,t} - W_{i,t-\Delta})}$$

Which is known as the Geometric Brownian Motion with $V_{i,t-\Delta}$ the starting value of the process and constants u (called the drift), σ which scales the volatility of the Brownian motion and $\Delta \in \mathbb{R}^{+}$ over which the value of the assets of each obligor is reevaluated. The constants $V_{i,t-\Delta}$, u, σ and Δ are considered to be fixed and identical for all n obligors.

Each obligor's asset value is determined by each obligor's individual Brownian motion. From the properties of Brownian motion it follows that each $W_{i,t}$ is $Normal(0,t)$ distributed. The Brownian motions between the obligors are assumed to be jointly normal and correlated with linear correlation coefficient ρ.

The dependence between the obligors can be modeled as follows:

$$W_{i,1} = \sqrt{\rho}M + \sqrt{1-\rho}Z_i$$
$$Where\ M \sim N(0,1)\ and\ Z_i \sim N(0,1)$$
$$M, Z_1, .., Z_n\ independent$$

Furthermore we have set $\Delta = 1$, that is we take an evaluation period of 1 year time following the Basel II framework. As a consequence the Brownian motion processes $W_{i,t}$ turn into a $W_{i,1}$ variable which follow a standard normal distribution for each individual obligor i.
M can be interpreted as the common factor and models the systematic risk, while Z_i is the idiosyncratic factor modeling the unsystematic risk of each individual obligor.

We fill this in the original explicit solution of the SDE of the asset process:

$$V_{i,t} = V_{i,t-1} e^{(u - \frac{\sigma^2}{2}) + \sigma(\sqrt{\rho}M + \sqrt{1-\rho}Z_i)}$$

The log is often taken to get the log asset value:

$$\ln(V_{i,t}) = \ln(V_{i,t-1}) + u - \frac{\sigma^2}{2} + \sigma(\sqrt{\rho}M + \sqrt{1-\rho}Z_i)$$

From this we infer that the log asset value is normally distributed or that the asset value is log normally distributed.
Now there is a default in one year time if the asset value of obligor i goes below the barrier $D_i = D\ for\ all\ i$ (all obligors are identical, so are their default points):

$$\rightarrow \ln(V_{i,t}) = \ln(V_{i,t-1}) + u - \frac{\sigma^2}{2} + \sigma(\sqrt{\rho}M + \sqrt{1-\rho}Z_i) < \ln(D)$$

$$\sqrt{\rho}M + \sqrt{1-\rho}Z_i < \frac{\ln(D)\ - ln(V_{i,t-1}) - u + \frac{\sigma^2}{2}}{\sigma}$$

3.1.1 PROBABILITY OF DEFAULT

Now we want to find the probability of default, PD from one year time.

$$PD = \mathbb{P}(Default\ of\ obligor\ i) = \mathbb{P}\left(\sqrt{\rho}M + \sqrt{1-\rho}Z_i < \frac{\ln(D)\ - ln(V_{i,t-1}) - u + \frac{\sigma^2}{2}}{\sigma}\right)$$

$$= \Phi\left(\frac{\ln(D)\ - ln(V_{i,t-1}) - u + \frac{\sigma^2}{2}}{\sigma}\right)$$

Since $\sqrt{\rho}M + \sqrt{1-\rho}Z_i \sim Normal(0,1)$ and where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. From this we infer that:

$$\Phi^{-1}(\mathrm{PD}) = \frac{\ln(D)\ - ln(V_{i,t-1}) - u + \frac{\sigma^2}{2}}{\sigma}$$

$\Phi^{-1}(\mathrm{PD})$ is the (standardized) default point and PD is the unconditional one year probability of default.

3.1.2 CONDITIONAL PROBABILITY OF DEFAULT

Often the PD is conditioned on the common factor M_t, as we will show the conditional PD is a function of the unconditional PD and the correlation coefficient ρ. In that case the constants u and σ are not needed. It can be rewritten as follows:

$$\mathbb{P}\left(\sqrt{\rho}M + \sqrt{1-\rho}Z_i < \frac{\ln(D)\ - ln(V_{i,t-1}) - u + \frac{\sigma^2}{2}}{\sigma} | M\right) = \mathbb{P}\left(\sqrt{\rho}M + \sqrt{1-\rho}Z_i < \Phi^{-1}(PD)|\mathrm{M}\right)$$

$$= \mathbb{P}\left(Z_i < \frac{\Phi^{-1}(PD) - \sqrt{\rho}M}{\sqrt{1-\rho}} |\mathrm{M}\right) = \Phi\left(\frac{\Phi^{-1}(PD) - \sqrt{\rho}M}{\sqrt{1-\rho}}\right)$$

It follows that the PD conditioned on the common factor is given by:

$$\mathbb{P}(Default\ of\ obligor\ i|M=m)=\Phi(\frac{\Phi^{-1}(PD)-\sqrt{\rho}m}{\sqrt{1-\rho}})\coloneqq \mathrm{p(m)}$$

Conditioned on the common factor, the loss on obligor i denote L_i, is Bernoulli distributed with parameter p(m). One can interpret the common factor as the state of the economy, given a high common factor or given a high state of economy, the conditional probability of default is lower than in a lower state of economy.
Notice we have just given the conditional probability of default as a function of the common factor and the unconditional probability of default which can be interpreted as the long term default rate throughout the whole economic cycle. The probability of default conditioned on the common factor is also known as the point in time PD (PIT PD) while the unconditional probability of default is also known as the through-the-cycle PD (TTC PD). The plot below shows the conditional PD varying the common factor using a correlation 0.2.

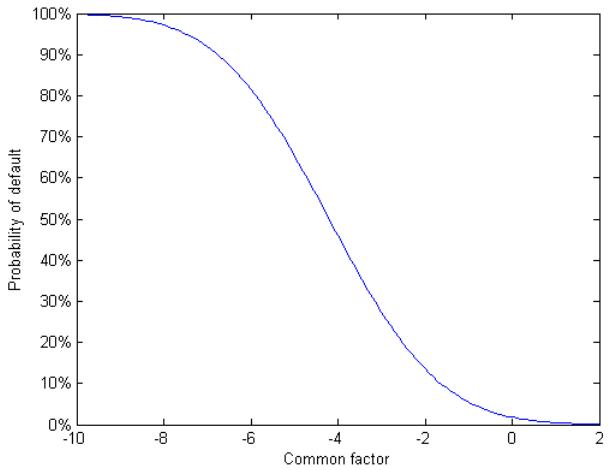


Figure 2. The conditional PD as a function of the PD=0.03 and the common factor

The formula can be seen the mapping of the economic state and the through the cycle PD towards a point in time PD. This gives the observed default rate depending on the prevailing economic state at a particular point in time. This formula will be used to switch between the conditional probability of default and the unconditional probability of default.

3.1.3 THE EXPECTED LOSS

The **expected loss** (EL) of obligor i assuming a deterministic LGD and EAD and independence between them is given by:

$$\mathbb{E}[Loss\ of\ obligor\ i]=PD*LGD*EAD$$

For non independence see Kupiec (2008).

3.1.4 THE TOTAL LOSS DISTRIBUTION

The total loss, denote TL, is the sum of all the individual obligor's losses.

$$TL = \frac{1}{n}\sum_{i=1}^{n} L_i$$

Conditioning on the common factor, the $L_i|M$ are independent and identically distributed.

$$with\ L_i|M \coloneqq \begin{cases} 1 & with\ \mathrm{p}(M) \\ 0 & with\ 1-\mathrm{p}(M) \end{cases}$$

So that the sum of the individual obligor's loss, $nTL|M = \sum_{i=1}^{n} L_i|M$ is a sum of n independent $Bernoulli(\mathrm{p}(M))$ variables which is $Binomial\big(n, \mathrm{p}(M)\big)$ distributed. As $n \to \infty$, $nTL|M$ is normally distributed by the Demoivre-Laplace limit theorem with the same mean and variance , see chapter 5.4.1 of [13].

Thus the **conditional distribution of the total loss** is $TL|M \sim Normal(\mathrm{p}(M), \frac{\mathrm{p}(M)(1-\mathrm{p}(M))}{n})$ and the variance tends to 0 as $n \to \infty$. For large n, the distribution is given by $\mathbb{P}(TL < x|M) \approx 0\ if\ x < \mathrm{p}(M) - \varepsilon$ and $\mathbb{P}(TL < x|M) \approx 1\ if\ x > \mathrm{p}(M) + \varepsilon$. And as $n \to \infty$, the conditional total loss equals its expectation, $\mathbb{P}(TL \approx \mathrm{p}(M)|M) = 1$.

This result is compatible with proposition 1 from [19], in which the distribution of the total loss TL given the common factor $M = m$ converges (almost surely) to the conditional expectation of the loss $E[L|m] = \Phi\left(\frac{\Phi^{-1}(PD) - \sqrt{\rho}m}{\sqrt{1-\rho}}\right)$ as $n \to \infty$. The proposition holds under two general conditions which are met in the ASRF model. According to Gordy, this result implies that in the limit one only needs $E[L|M]$ to answer questions about the unconditional distribution of the total loss in the asymptotic case.

The **unconditional distribution of the total loss** $\mathbb{P}(TL < x) = \mathbb{E}[\mathbb{P}(TL < x|M)] \approx \mathbb{E}\big[1_{x > \mathrm{p}(M)}\big] = \mathbb{E}\big[1_{M > \mathrm{p}^{-1}(x)}\big] = 1 - \Phi\big(\mathrm{p}^{-1}(x)\big) = \Phi\big(-\mathrm{p}^{-1}(x)\big)$.

The inverse function is given by: $\mathrm{p}^{-1}(x) = \frac{-(\Phi^{-1}(\mathrm{x})\sqrt{1-\rho} - \Phi^{-1}(\mathrm{PD}))}{\sqrt{\rho}}$, so that $\mathbb{P}(TL < x) = \Phi\big(-\mathrm{p}^{-1}(x)\big) = \Phi\left(\frac{\Phi^{-1}(\mathrm{x})\sqrt{1-\rho} - \Phi^{-1}(\mathrm{PD})}{\sqrt{\rho}}\right)$.

3.2 VALUE AT RISK

Now the portfolio $VaR_{1-\alpha}$ is calculated, it is defined as $q_{1-\alpha}(Total\ Loss\ of\ portfolio)$. Since the total loss distribution is continuous, $\inf\{v \in \mathbb{R}: \mathbb{P}(Total\ Loss\ of\ portfolio \leq v) \geq 1-\alpha\}$ can be replaced with $\{v \in \mathbb{R}: \mathbb{P}(Total\ Loss\ of\ portfolio \leq v) = 1-\alpha\}$. Thus we can find $VaR_{1-\alpha}$ from the following equation: $\mathbb{P}(Total\ Loss\ of\ portfolio \leq VaR_{1-\alpha}) = 1-\alpha$.

We derive $VaR_{1-\alpha}$, it follows from the equation:

$$\mathbb{P}(TL \leq VaR_{1-\alpha}) = 1 - \alpha$$
$$\Phi\left(-PD^{-1}(VaR_{1-\alpha})\right) = 1 - \alpha$$
$$\Phi\left(PD^{-1}(VaR_{1-\alpha})\right) = \alpha$$
$$VaR_{1-\alpha} = PD\left(\Phi^{-1}(\alpha)\right)$$

So that the **one year VaR at $(\mathbf{1-\alpha})$-level** $VaR_{1-\alpha} = \Phi(\frac{\Phi^{-1}(PD)-\sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}})$
This can be interpreted as the long term VaR, long term meaning the required VaR throughout the economic cycle. This homogeneous result holds also for a heterogeneous portfolio under the granularity assumption which follows from results in [19].

Up till now we have assumed that PD is deterministic. One way to quantify parameter uncertainty around PD is by replacing the deterministic PD by a random distributed PD in the final expression of the VaR. Consequently the one year VaR becomes a distribution too. Of this distribution we could look at the **expected one year VaR at $(\mathbf{1-\alpha})$-level**: $\mathbb{E}[VaR_{1-\alpha}|PD] = \mathbb{E}\left[\Phi\left(\frac{\Phi^{-1}(PD)-\sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right)|PD\right]$, where PD is a random variable.

However simply replacing the deterministic PD with a distributional PD in the last step is not valid, as the the Var is a non-linear function of PD and the steps to derive the VaR would be different with a distributional PD. In the next section the bias in the estimation of the VaR is exposed when in the estimation of the VaR the PD is simply replaced by an unbiased estimator for the PD.

3.2.1 The unbiased estimated VaR is biased downwards

In practice the VaR is estimated by estimating the parameter PD. Let the 'true VaR' $VaR_{1-\alpha}^{\text{true}}$ be the VaR that a bank is interested in and let the 'true PD' be the unobservable parameter PD that calculates the true VaR. The true PD is estimated by an unbiased estimator, resulting in an 'estimated VaR' $\text{VaR}_{1-\alpha}^{\text{estimated}}$.

For example let the true PD=0.03 and the data at hand are five observed default rates $(y_1, y_2, \ldots., y_5) = (0.02,0.04,0.03,0.01,0.06)$ generated by the true PD. One unbiased estimator of the true PD is the average of the observed default rates $\frac{y_1+y_2+\cdots+y_5}{5}$ so that the unbiased estimate is 0.032. The $\text{VaR}_{1-\alpha}^{\text{estimated}}$ is then calculated using the derived formula $VaR_{1-\alpha} = \Phi(\frac{\Phi^{-1}(PD)-\sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}})$ and filling in the unbiased estimate for PD.

Let X be the sampling distribution of an unbiased estimator of the true parameter PD, then the expected estimated VaR based on unbiased estimates x, i.e. $\text{VaR}(X) = \text{VaR}_{1-\alpha}^{\text{estimated}}$, will be strictly smaller than the true VaR.

We can see the VaR as a function of X, we write $\text{VaR}(X) = \Phi(\frac{\Phi^{-1}(X)-\sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}})$, we will show that $\mathbb{E}[VaR(X)] < VaR(\mathbb{E}[X])$ or $\mathbb{E}[\text{VaR}_{1-\alpha}^{\text{estimated}}] < VaR_{1-\alpha}^{\text{true}}$. Where $\mathbb{E}[X]$ is the expectation of the unbiased estimator and is by definition the true parameter PD, hence $VaR(\mathbb{E}[X])$ is the true VaR. And $\text{VaR}_{1-\alpha}^{\text{estimated}}$ is the VaR based on the unbiased estimator of PD. The term $\mathbb{E}[VaR(X)]$ is the expected VaRs based on the unbiased estimates.

A Taylor expansion is utilized to expose the tendency for $\text{VaR}_{1-\alpha}^{\text{estimated}}$ to underestimate.

We approximate $VaR(.) := f(.)$ around $\mathbb{E}[X] = \mu_X$.

$$f(X) = f(\mu_X) + f^{(1)}(\mu_X)(X-\mu_X) + \frac{f^{(2)}(\mu_X)}{2!}(X-\mu_X)^2 + .. + \frac{f^{(k)}(\mu_X)}{k!}(X-\mu_X)^k + R_k(X)$$

Where $R_k(X)$ is the remainder term given by $R_k(X) = \frac{f^{(k+1)}(\varepsilon)}{(k+1)!}(X-\mu_X)^{k+1}$ with ε in between X and μ_X. $X \in (0,1)$ so that $\varepsilon > 0$.

Let's look at k=1:

$$f(x) = f(\mu_X) + f^{(1)}(\mu_X)(X-\mu_X) + \frac{f^{(2)}(\varepsilon)}{2!}(X-\mu_X)^2$$

Now we take the expectation on both sides and

$$\mathbb{E}[VaR(X)] = VaR(\mu_X) + VaR^{(1)}(\mu_X)\mathbb{E}[(X-\mu_X)] + \frac{VaR^{(2)}(\varepsilon)}{2!}\mathbb{E}[(X-\mu_X)^2]$$

And we see that the expected:

$$\mathbb{E}[\text{VaR}_{1-\alpha}^{\text{estimated}}] = \text{VaR}_{1-\alpha}^{\text{true}} + \frac{VaR^{(2)}(\varepsilon)}{2!}\mathbb{E}[(X-\mu_X)^2]$$

or

$$\text{VaR}_{1-\alpha}^{\text{true}} = \mathbb{E}[\text{VaR}_{1-\alpha}^{\text{estimated}}] - \frac{VaR^{(2)}(\varepsilon)}{2!}\mathbb{E}[(X-\mu_X)^2].$$

Since $VaR(.)$ is a concave function (see figure 1 and formal proof below), the second derivative $VaR^{(2)}(.)$ is negative, we have shown that $\text{VaR}_{1-\alpha}^{\text{estimated}}$ tends to be smaller than the $\text{VaR}_{1-\alpha}^{\text{true}}$ and in expectation there is a bias of size $-\frac{VaR^{(2)}(\varepsilon)}{2!}\mathbb{E}[(X-\mu_X)^2]$ proportional to the variance of the estimator. The more data we have at hand the less variance an estimator for the true PD carries, as data increases the bias goes to zero and in expectation $\text{VaR}_{1-\alpha}^{\text{estimated}}$ converges to $\text{VaR}_{1-\alpha}^{\text{true}}$.

We will proof this formally by calculating the second derivative showing that it is negative.

Proof: $VaR^{(2)} < 0$

Write for convenience $f(.)$ for $VaR(.)$.

$$f(y) = \Phi\left(\frac{\Phi^{-1}(y)-\sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right)$$

$$f'(y) = \phi\left(\frac{\Phi^{-1}(y) - \sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right)\frac{1}{\sqrt{1-\rho}}\frac{1}{\phi(\Phi^{-1}(y))} = \frac{1}{\sqrt{1-\rho}}e^{-\frac{1}{2}\left(\frac{\Phi^{-1}(y)-\sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right)^2}e^{\frac{1}{2}\left(\Phi^{-1}(y)\right)^2}$$

$$= \frac{1}{\sqrt{1-\rho}}e^{-\frac{1}{2}\left(\frac{\Phi^{-1}(y)-\sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right)^2+\frac{1}{2}\left(\Phi^{-1}(y)\right)^2} = \frac{1}{\sqrt{1-\rho}}e^{-\frac{1}{2}\frac{1}{1-\rho}\left(\Phi^{-1}(y)-\sqrt{\rho}\Phi^{-1}(\alpha)\right)^2+\frac{1}{2}\left(\Phi^{-1}(y)\right)^2}$$

$$Define\ \mathrm{Z} := -\frac{1}{2}\frac{1}{1-\rho}\left(\Phi^{-1}(y) - \sqrt{\rho}\Phi^{-1}(\alpha)\right)^2 + \frac{1}{2}\left(\Phi^{-1}(y)\right)^2$$

$$f''(y) = \frac{1}{\sqrt{1-\rho}}e^{\mathrm{Z}}\frac{\mathrm{d}}{\mathrm{d}y}(\mathrm{Z})$$

$$\frac{\mathrm{d}}{\mathrm{d}y}(\mathrm{Z}) = -\frac{1}{1-\rho}\left(\Phi^{-1}(y) - \sqrt{\rho}\Phi^{-1}(\alpha)\right)\frac{d}{dy}\left(\Phi^{-1}(y)\right) + \frac{d}{dy}\left(\Phi^{-1}(y)\right)$$

$$f''(y) < 0\ if\ \frac{1}{1-\rho}\left(\Phi^{-1}(y) - \sqrt{\rho}\Phi^{-1}(\alpha)\right) < 1$$

$$f''(y) < 0\ if\ \Phi^{-1}(y) > \sqrt{\rho}\Phi^{-1}(\alpha)$$

In practice the term $\Phi^{-1}(\varepsilon)$ is larger than $\sqrt{\rho}\Phi^{-1}(\alpha)$, since α is very small (for regulatory capital $\alpha = 0.01\%$). The probability that ε is smaller than 0.01% is negligible. Further more $\rho\epsilon(0,1)$ and derivative of the inverse cumulative normal distribution is positive and given by $\frac{d}{dy}\left(\Phi^{-1}(y)\right) = \frac{1}{\phi(\Phi^{-1}(y))}$ can be found by combining $\frac{d}{dy}(\Phi(\Phi^{-1}(y))) = \phi(\Phi^{-1}(y))\frac{d}{dy}\left(\Phi^{-1}(y)\right)$ and $\frac{d}{dy}(\Phi(\Phi^{-1}(y))) = \frac{d}{dy}(\mathrm{y}) = 1$. So that $f''(\varepsilon) < 0$, the function $VaR(.)$ is concave.

3.2.2 INCORPORATING PARAMETER UNCERTAINTY IN PD

[14] shows, using results from [19], that the VaR or the quantile of the total loss distribution assuming a random distributed PD and ρ can be found under some regularity conditions by looking at the quantile of $E[L|M, PD, \rho]$ defined by: $E[L|M, PD, \rho] = \Phi\left(\frac{\Phi^{-1}(PD)-\sqrt{\rho}M}{\sqrt{1-\rho}}\right)$, where M, PD and ρ are random distributed variables.

Formally we can write:

$$As\ n \to \infty, q_{1-\alpha}(TL) \to q_{1-\alpha}(E[L|M, PD, \rho])$$

Where $q_{1-\alpha}(TL)$ is the quantile of the total loss distribution, which is the objective of the VaR measure. Unlike Tarashev we will assume that the correlation ρ is known and deterministic. The above result still holds and ρ can be viewed as a degenerate distribution.

Observe that Tarashev's result is compatible with the earlier derivation of the VaR, assuming that the parameters PD and ρ are constants:

$$q_{1-\alpha}(E[L|M, PD, \rho]) = q_{1-\alpha}\left(\Phi\left(\frac{\Phi^{-1}(PD)-\sqrt{\rho}M}{\sqrt{1-\rho}}\right)\right) = \Phi\left(\frac{\Phi^{-1}(PD)-\sqrt{\rho}q_{1-\alpha}(M)}{\sqrt{1-\rho}}\right) =$$

$$\Phi\left(\frac{\Phi^{-1}(PD)-\sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right).$$

Thus we can see Tarashev's results as a more generalized formula for the calculation of the VaR. In chapter 7 it will be clear that the underestimation problem, as shown in the previous section, vanishes with the use of this alternative formula.

We would like to compare this estimator with other estimators for the true VaR.

3.2.3 *DIFFERENT ESTIMATORS*

In general there is no way to find the true VaR since the true PD is unknown and can only be estimated based on the data at hand.

The data is summarized in a matrix Δ, n rows representing the number of obligors in the portfolio and T columns representing the number of years of data. An element of the matrix is denoted by $\Delta_{i,t}$, which is the element in row i and column t for i = 1,..n and t = 1,...T. It is an indicator variable attaining the value one if there is an observed default at year t of obligor i.

$$\Delta = \begin{pmatrix} \Delta_{1,1} & \cdots & \Delta_{1,T} \\ \vdots & \ddots & \vdots \\ \Delta_{n,1} & \cdots & \Delta_{n,T} \end{pmatrix}$$

Next to the defaults, a corresponding vector of the common factor is possibly known. An entry m_t denotes the common factor in year t for t = 1,..,T.

$$M = \begin{pmatrix} m_1 \\ \vdots \\ m_T \end{pmatrix}'$$

So next to the defaults $\Delta_{1,t}, \dots, \Delta_{n,t}$ a corresponding common factor m_t is known for each year t.

The observed defaults are converted into observed default frequencies (ODF), a vector of T years.

$$\delta = \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_T \end{pmatrix}'$$

Where $\delta_t = \frac{1}{n}\sum_{i=1}^{n} \Delta_{i,t}$ is the observed default frequency or rate in year t. If the common factor is known, one can use the relationship between the conditional probability of default $p(m_t)$, common factor m_t and unconditional probability of default PD to convert conditional ODFs into unconditional ODFs.

$$p(m_t) = \mathbb{P}(Default\ of\ obligor\ i | M = m_t) = \Phi\left(\frac{\Phi^{-1}(PD) - \sqrt{\rho} m_t}{\sqrt{1-\rho}}\right)$$

This formula gives the conditional probability of default and can be rewritten to give the unconditional PD as a function of the conditional probability of default and the common factor.

$$PD = \Phi\left(\Phi^{-1}\big(p(m_t)\big)\sqrt{1-\rho} + \sqrt{\rho} m_t\right)$$

The observed default frequency δ_t which is generated under the latent common factor m_t is converted into an unconditional observed default frequency δ_t^* using the above relationship when the common factor is known.

$$\delta^* = \begin{pmatrix} \delta_1^* \\ \vdots \\ \delta_T^* \end{pmatrix}'$$

Where $\delta_t^* = \Phi(\Phi^{-1}(\delta_t)\sqrt{1-\rho} + \sqrt{\rho}m_t))$ for each year t.

Since additional information in the form of the common factor is used, the estimation of the unconditional PD or the TTC PD is more accurate based on the unconditional observed default frequencies. The unconditional default frequencies are of interest since in the ASRF model the TTC PD is of interest. In practice in the banking industry the probability of default is estimated using a logit regression model with defaults data and obligor specific characteristics. These models varies over banks and the type of portfolio. We will purely look at the uncertainty due to PD uncertainty and for this purpose compare different measures for VaR.

The VaR that is of interest and which is attempted to be estimated but is unknown is called the true VaR. It is based on the true value of PD.

$\text{VaR}_{1-\alpha}^{\text{true}} = \Phi(\frac{\Phi^{-1}(PD_{true})-\sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}})$, where PD_{true} is the true but unknown population parameter.

The true VaR is estimated with $\text{VaR}_{1-\alpha}^{\text{estimated}}$ based on an unbiased estimator of the PD. In this thesis the unbiased estimator is chosen to be the sample average of the observed default rates (=ODFs) and is called $PD_{\text{estimated}}$

$$PD_{\text{estimated}} = \begin{cases} \frac{1}{T}\sum_{t=1}^{T}\delta_t \;\; if\ common\ factor\ is\ unknown \\ \frac{1}{T}\sum_{t=1}^{T}\delta_t^* \;\; if\ common\ factor\ is\ known \end{cases}$$

and

$$\text{VaR}_{1-\alpha}^{\text{estimated}} = \Phi(\frac{\Phi^{-1}(PD_{\text{estimated}})-\sqrt{\rho}\Phi^{-1}(\alpha)}{\sqrt{1-\rho}})$$

The parameter uncertainty incorporating measure will be denoted as $\text{VaR}_{1-\alpha}^{\text{parameter uncertainty}}$ or $\text{VaR}_{1-\alpha}^{\text{pu}}$. In this thesis, PD is found in a different way from Tarashev. It will be clear in chapter 6 how the distribution of PD using the data, is derived. .

$\text{VaR}_{1-\alpha}^{\text{pu}} = q_{1-\alpha}\left(\Phi\left(\frac{\Phi^{-1}(PD)-\sqrt{\rho}M}{\sqrt{1-\rho}}\right)\right)$, where PD and M are random variables.

And the distribution of PD is a function of $\delta\ or\ \delta^*$ and M.

3.2.4 FORMULA IN TERMS OF MONEY UNITS AND HETEROGENEOUS PORTFOLIOS

The results can easily be translated from the fractions of losses to actually losses in money units. Since we have assumed that LGD and EAD are deterministic, the VaR in terms of money units is simply the VaR measure multiplied with LGD and EAD. Until now all results were derived for a homogenous portfolio, by [19] the same result holds for heterogonous portfolios in the

asymptotic case. In the literature review a number of papers have investigated the non-asymptotic case and it is shown that for small portfolios the works approximately as well.

4 LITERATURE OVERVIEW*

Tarashev (2009) uses a Bayesian inference approach, measuring the credit VaR and allowing uncertainty on the PD and ρ. It is pointed out that in the ASRF model the assumption of a single risk factor is violated since next to the common factor we have estimation risk factors in the form of PD and rho. Thus the usual formula for VaR, is incorrect and an alternative formula called the 'correct VaR' is derived and proven to converge to the true VaR. The formula of the correct VaR equals to the earlier defined $\mathrm{VaR}_{1-\alpha}^{\mathrm{pu}}$, but the posterior parameter PD in the formula is derived differently in this thesis (see chapter 6) from Tarashev.

Investors who are interested in the credit VaR are assumed to have monthly asset values available which they use to estimate the correlation coefficient while using the same asset values to determine the default rate at the end of the year. The estimators for both parameters are assumed to be delivered by minimum variance unbiased estimators, the noise in the point estimates are considered to attain the Cramer-Rao lower bounds. In combination with an uniform prior on the (0.1) interval for the PDand a beta distribution for the correlation estimator given the observed correlation between assets, the posterior distribution is calculated. The PD posterior is constructed less transparent. Tarashev states three criteria the posterior must satisfy and finishes with the statement that the implied posterior distribution of the 'default point' Φ^{-1}(PD) is normal. This in turn would imply that the posterior of PD is uniformly distributed.

The results are presented in the form of add-ons which are the differences in percentages between the correct and an estimated VaR (called naïve VaR in the paper) varying the numbers of obligors, n=50, 200, 1000 and the number of years of data T=5,10 years. For a fair comparison along cross-sections and time, one single dataset is used. Results indicate that the impact of PD uncertainty is much larger than correlation uncertainty. The difference between the naïve and correct VaR cannot neglected: in the base case of regulatory capital, with alfa at 0.1% ,true PD=0.01, n=200 and T=10 the difference is 27%. While neglecting the correlation uncertainty results in 24.9% difference in the base case.

Making use of the special choice of the distribution of the estimators, bounded by the Cramer Rao lower bounds, Tarashev derives approximate expressions for the correct VaR incorporating either the PD or correlation uncertainty only. This can be rewarding if time is limited and a first approximation is only needed.

Tarashev points out that the ASRF model assuming e.g. homogeneity of exposures and time invariant risk parameters can effect the observed correct VaR. While also pointing out that the parameterization of the distributions of the posterior are convenient but unrealistic in practice. He states that it is likely that these assumptions depress the correct VaR and proposes to treat these VaRs as lower bounds by investors.

In **Tarashev & Zhu (2008)** possible specification and calibration errors in a couple measures including the regulatory VaR are analysed for the ASRF model. For the specification, the

* This section is based on the papers [12], [13], [14], [15], [16], [17], [18] and [19].

granularity assumption and multi risk factors as opposed to a single risk factor are assessed by comparing the differences in VaR between homogeneous versus heterogeneous portfolios and two risk factors versus single risk factor in the measures. For the calibration the asset correlations are investigated for estimation error and different degrees of dispersion in the average correlation coefficient in the assets. Furthermore the impact of estimation errors in PD, LGD and asset correlation is tested on a homogeneous portfolio. A small portfolio consisting of 100 assets and a large portfolio consisting of 1000 assets are used. It turns out that the impact of the assumptions is small or even negligible compared to large errors in the calibration or estimation of the parameters. The granularity effect using 1000 assets is already smaller than 1% while the PD can deviate 30% using 1000 assets and 5 years of data in its 95% confidence interval. The authors conclude that violations of the specification are virtually inconsequential especially for large portfolios. Their concern goes towards calibration in the parameters especially small sample estimation errors or rule-of-thumb values of asset correlations, these can lead to significant inaccuracies measuring for example the VaR.

A different approach is chosen in **Löffler (2001)** investigating both specification error and parameter uncertainty along PD, recovery rates (1-LGD) and correlations following the model of CreditMetrics which is similar to the ASRF model. But trading in the single common risk factor for two common risk factors. For this adopted model the general analytic result for the VaR is derived.

Acknowledging autocorrelations between the default rates and pointing out that ignoring (positive) autocorrelation underestimates standard errors [26] Löffler captures this effect in an autoregressive model of 2 lags. A non parametric bootstrap procedure to derive a distribution of the true mean default rates is constructed, drawing from the 18 year history two consecutive annual default rates after which the remaining default rates are inferred from the autoregressive model. Furthermore sampling randomness is added by assuming by using the bootstrapped default rate as the mean of a Beta distribution (with a chosen standard deviation). From this distribution is drawn to obtain of the 18 bootstrapped default rate. Averaging the 18 default rates of the bootstrap sample yields a mean default rate, which is an estimate of the true default rate. Repeating this large number of times gives the default rate estimator distribution.

For the recovery rates, it is assumed that these are uniformly distributed with mean 49.6% within a 40% interval around, based on a study by Standard & Poor's [15] and Altman and Kishore]16]. The asset correlations are estimated using a sample of 60 monthly asset returns using the original model where the correlation are dependent on the variances of the common factor and each individual normal distribution. These variances are assumed to be uncertainty by modeling both as χ^2 distributions and the final correlation distribution is found by running a Monte Carlo simulation a large number of times.

Using the distributions of all three, the VaR is calculated for many times by drawing from the respective distribution and using the analytical formula for the VaR measure. This result is a frequency distribution for this risk measure.

Löffler concludes the same as in **Tarashev & Zhu (2008)** that the additional risk of dispersion in the parameters adds little extra noise e.g. heterogeneous portfolios versus homogenous portfolios even for a small portfolio of 50 obligors. The major force in final VaR calculation lies

with the PD sampling randomness. It needs to be assessed on a case by case setting, as the eventual influence depends on the availability of data. This is also found in **Heitfield (2009),** using a 90% confidence interval for the parameter PD, the pricing of credit default obligations (CDOs) is highly sensitive to the found estimator. Finally it is pointed out that the noise can influence the interpretation of the results of backtests.

In a general framework **Gordy (2003)** shows some useful important theorems. He shows that the loss distribution in the single risk factor model is asymptotically the same as the expected loss conditioned on the stochastic common factor. Continuing on this result he shows that the quantile of the conditioned expected loss equals the quantile of the total loss distribution, so that the VaR measure can be found using only the expectation of the loss. The general framework in which he proves the results are general enough to cover all well known risk models including the ASRF model and is also general enough to cover inhomogeneous portfolios.

Looking further in the analysis of parameter uncertainty in relating fields, **Korteweg and Polson (2010)** applies bond pricing in the framework of Merton structural model. In the Bayesian framework the posterior joint distribution of the model parameters is found by using, the prices of bonds is under parameter calculated as the expectation of the bond pricing formula with the use of the parameters. This is contrasted with calculated bond prices using point estimates. It is concluded that credit spreads are much higher influenced by parameter uncertainty and the impact of different factors such liquidity, taxes and jump risk is smaller. Another application is in **Heitfield (2009)** for the pricing of Credit Default Obligations (CDOs) and inferring the rating of such instruments. It is shown that in a 90% confidence interval multiple ratings are possible for CDOs, concluding that pricing of CDOs is not robust implying that prices are uncertain.

5 BAYESIAN STATISTICS*

In statistics two main philosophical approaches are present. The frequentist approach is the most well known, probabilities are viewed as long run relative frequencies. The parameters of statistical models are assumed fixed (non-random) unknown constants, probability statements are only allowed for random quantities. Procedures are developed by looking at how they perform over all possible random samples. The probabilities don't relate to the particular random sample that was obtained.

The other is called the Bayesian approach. Since the true value of parameters of statistical models are uncertain, the parameters are modeled as random variables. Probabilities are not only interpreted as long run relative frequencies but also represents the plausibility or degree of belief of statements. This degree of belief is updated using observed data according to Bayes' theorem, named after Thomas Bayes' work in the 18th century. The theorem can be stated in terms of probabilities or probability densities, where the last one is of particular interest.

5.1 BAYES' THEOREM

For two continuous random variables X and Y the definition of the conditional probability density function X given Y is: $f_{X|Y}(x|y) = \frac{f(x,y)}{f(y)}$, where $f(x, y)$ is the joint probability distribution of X and Y and $f(y)$is the marginal probability distribution of Y. Using this definition and law of total probability given by $f(y) = \int_{x=-\infty}^{\infty} f(x, y)dx = \int_{x=-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dx$ one can see the following result:

Bayes' theorem for probability densities

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f(x)}{\int_{x=-\infty}^{\infty} f_{Y|X}(y|x)f(x)dx}$$

In the context of Bayesian statistics, X represents the unknown parameter (possible multivariate) and Y represents the data. The marginal probability density $f(x)$ embodies the *a priori* belief of the parameter's possible values while $f_{Y|X}(y|x)$ is the *likelihood* of the parameter of interest given the data. Together the posterior belief/distribution of the unknown parameter $f_{X|Y}(x|y)$ can be calculated. To emphasize X being the parameter, let $\Theta \coloneqq X$ with all θ in the sample space of Θ. The likelihood function is often written with the conditioning variable in the subscript such that $f_\Theta(y) = f_\Theta(y)$. Finally the prior density of the parameter $f(\theta)$ is often written as $\pi(\theta)$, taken all together the theorem is rewritten as:

$$f(\theta|Y) = \frac{f_\Theta(y)\pi(\theta)}{\int_{-\infty}^{\infty} f_\Theta(y)\pi(\theta)d\theta}$$

* This chapter is based on [13], [22] and [23].

5.2 THE A PRIORI CHOICE

The biggest critic on the Bayesian approach is the possibly subjective choice of prior distribution/belief $\pi(\theta)$ on the parameter. Depending on the application and the correct specification of the prior statistical inference on frequentists or Bayesian is preferred. However from the Bernstein-von-Mises theorem it follows that the posterior is independent of the prior choice once there is a large amount of samples, while the same can be said in the context of frequentist statistics.

In this context of parameter uncertainty the unknown parameter is the PD, which is limited to the interval between 0 and 1. Due to its flexible shape the Beta distribution is a popular choice on this interval. The probability density of the $Beta(\alpha, \beta)$ distribution is given by:

$$\pi(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)} 1_{[0,1]}(\theta)$$

Where $B(\alpha, \beta)$ is the Beta function which is the integral $\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$ with parameters $\alpha > 0 \; and \; \beta > 0$. The expected value and variance can be calculated to be $\mathbb{E}[\theta] = \frac{\alpha}{\alpha+\beta}$ and $var(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

If one chooses the $Beta(\alpha, \beta)$ distribution for the prior and the likelihood function $\mathrm{f}_\Theta(\mathrm{y})$ is the binomial distribution, i.e. $Y \sim binom(n, \theta)$, then the posterior distribution falls in the same probability distribution family (Beta) as the prior.

$$\mathrm{f}(\theta|\mathrm{Y}) = \frac{\mathrm{f}_\Theta(\mathrm{y})\pi(\theta)}{\int_{-\infty}^{\infty} \mathrm{f}_\Theta(\mathrm{y})\pi(\theta)\mathrm{d}\theta} = \frac{\binom{\mathrm{n}}{\mathrm{y}} \theta^y(1-\theta)^{n-y} \frac{\theta^{\alpha-1}(1-\alpha)^{\beta-1}}{B(\alpha,\beta)} 1_{[0,1]}(\theta)}{\int_{-\infty}^{\infty} \mathrm{f}_\Theta(\mathrm{y})\pi(\theta)\mathrm{d}\theta}$$
$$= \frac{\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}}{C} 1_{[0,1]}(\theta)$$

Where C is a constant. One can see that the posterior is $Beta(y + \alpha, n - y + \beta)$. A prior resulting in the same probability distribution family for the posterior, is called a conjugate prior.

Another useful distribution on the same domain of the PD is the (continuous) uniform distribution on [0,1]. It is a specific parameterization of the Beta distribution with $\alpha = \beta = 1$, this follows from the density: $\pi(\theta) = \frac{\theta^{\alpha-1}(1-\alpha)^{\beta-1}}{B(\alpha,\beta)} 1_{[0,1]}(\theta) = \frac{1}{B(1,1)} 1_{[0,1]}(\theta) = \frac{1}{\int_0^1 1 \, dt} 1_{[0,1]} = 1_{[0,1]}(\theta)$. The first moments are given by $\mathbb{E}[\theta] = 0.5$ and $var(\theta) = \frac{1}{12}$. When there is no particular reasons to give the prior of the PD more belief on certain values, then this uniform prior is a natural choice.

5.3 POINT ESTIMATION

Statistical inference such as point estimation and hypothesis testing differ among the two approaches. In the frequentist approach a particular estimator for a parameter depends on the random sample, and consequently has a probability distribution. The probability distribution of the estimator is called the sampling distribution, since it embodies the distribution of all possible random samples that could but didn't occur. After which the distribution of the

estimator is analyzed around the parameter value. The following statistical concepts are necessary for an example of point estimation comparing frequentist with Bayesian statistics.

An estimator $\hat{\theta}$ is a function of the data and the goal is to estimate the parameter θ as closely as possible. The mean squared error (MSE) measures the performance of a particular estimator, the convention in frequentist statistics is to choose the estimator with the lowest MSE which is written as $MSE(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \theta)^2\right] = var(\hat{\theta}) + (\mathbb{E}[\hat{\theta} - \theta])^2$. The estimator is called unbiased if $\mathbb{E}[\hat{\theta}] = \theta$, from which follows that $(\mathbb{E}[\hat{\theta} - \theta])^2$ is zero. An unbiased estimator seems desirable but when the variance of the estimator (first component of the MSE) is sufficiently small, the biased estimator is in the light of the MSE criterum preferable.

The Bayes risk of a estimator $\hat{\theta}_b$ is defined as $Bayes\ risk = \int \mathbb{E}\left[(\hat{\theta}_b - \theta)^2\right]\pi(\theta)d\theta$, it weights the squared error to the distribution of the parameter θ. The Bayes estimator $\hat{\theta}_b$ is the estimator that minimalises the Bayes Risk. It can be seen that the mean of the posterior distribution of θ is the Bayes estimator. One can compare the MSE of the frequentist estimator with the Bayes estimator.

Finally the Cramér Rao lower bound expresses a lower bound on the variance of any unbiased estimator. An estimator achieving this minimum variance for unbiased estimators is called an efficient estimator, which means this efficient estimator is preferred over all unbiased estimators. It is not said that an estimator exists that achieves this bound, so that the interpretation is the minimum lower bound for any unbiased estimator. H. Cramér and C.R. Rao proved that the lower bound is given by the inverse of the Fisher information I(θ):

$$var(\hat{\theta}) \geq \frac{1}{\mathrm{I}(\theta)} = \frac{1}{var(\frac{\partial \log\, \mathrm{l}_\Theta(\mathrm{y})}{\partial\theta})}$$

Where y represents the data and $\mathrm{l}_\Theta(\mathrm{y})$ repressents the likelihood function.

In the next, an example is shown comparing the two different statistics in point estimation. Assume the data $Y \sim binom(n, \theta)$ and that our interest goes to the parameter θ. In the frequentist approach the estimator is chosen to be $\hat{\theta}_f = \frac{y}{n}$, y a realization of Y. The MSE of this estimator is $MSE(\hat{\theta}_f) = var(\hat{\theta}_f) + (\mathbb{E}[\hat{\theta}_f - \theta])^2 = \frac{1}{n^2}var(y) + (0)^2 = \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n}$. Computing the Cramér Rao lower bound (CRLB) proves this estimator to be efficient and this estimator is the minimum variance unbiased estimator (MVUE).

$$\mathrm{CRLB} = \frac{1}{var(\frac{\partial \log\, \mathrm{l}_\Theta(\mathrm{y})}{\partial\theta})} = var\left(\frac{\partial \log\binom{\mathrm{n}}{\mathrm{y}}\theta^y(1-\theta)^{n-y}}{\partial\theta}\right)^{-1} = var\left(\frac{\mathrm{y}}{\theta} - \frac{\mathrm{n-y}}{1-\theta}\right)^{-1} = var\left(\frac{\mathrm{y}}{\theta(1-\theta)} - \frac{\mathrm{n}\theta}{\theta(1-\theta)}\right)^{-1} = var\left(\frac{\mathrm{y}}{\theta(1-\theta)}\right)^{-1} = \left(\frac{\mathrm{n}}{\theta(1-\theta)}\right)^{-1} \equiv MSE(\hat{\theta}_f).$$

In the Bayesian approach a non-informative uniform prior is chosen $prior = Beta(1,1)$. While the likelihood function $\mathrm{f}_\Theta(\mathrm{y})$ is specified by the binomial distribution. Thus the posterior distribution of the parameter $\theta \sim Beta(y+1, n-y+1)$ and the Bayes estimator is the mean of this posterior distribution $\hat{\theta}_b = \frac{y+1}{y+1+n-y+1} = \frac{y+1}{n+2}$, which is biased since $\mathbb{E}\left[\frac{y+1}{n+2}\right] = \frac{n\theta+1}{n+2} \neq \theta$. The MSE of this estimator is calculated.

$$MSE(\hat{\theta}_b) = var\left(\frac{y}{n+2} + \frac{1}{n+2}\right) + (\mathbb{E}\left[\frac{y}{n+2} + \frac{1}{n+2} - \theta\right])^2 = \frac{n\theta(1-\theta)}{(n+2)^2} + (\frac{n\theta}{n+2} + \frac{1}{n+2} - \theta)^2 = \frac{n\theta(1-\theta)+(1-2\theta)^2}{(n+2)^2}.$$

In the below graph $MSE(\hat{\theta}_f)$ versus $MSE(\hat{\theta}_b)$ is plotted for n=50 and 100. Depending on the true value of theta, the MSE of the one is lower than the other one. In most cases the Bayes estimator has a lower MSE, whenever theta is around 0.15 up to 0.85 so that this estimator is preferred over the frequentist's minimum variance unbiased estimator. Note that the difference gets smaller as n gets larger.

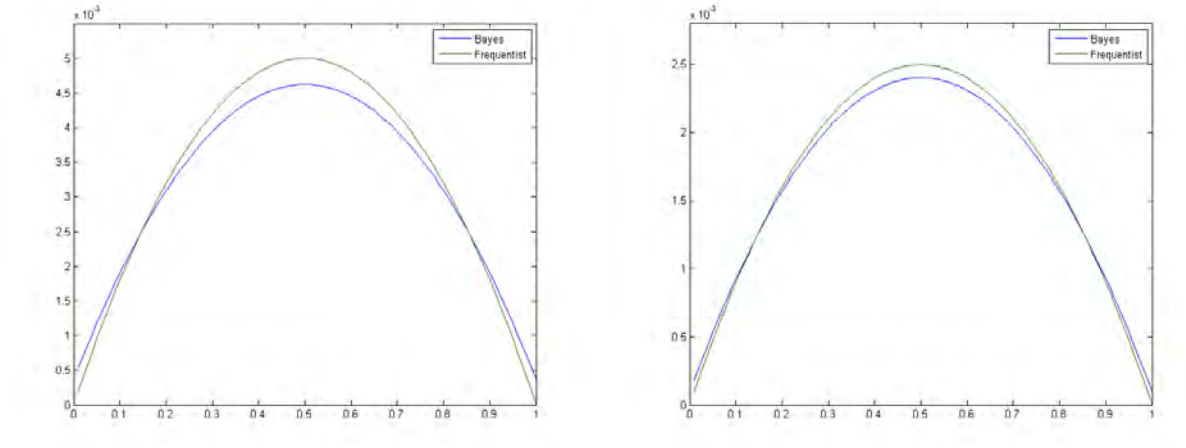


Figure 3 Comparing the MSE of an frequentist versus Bayes estimator(left:n=50 and right: n=100)

5.4 OTHER STATISTICAL INFERENCES

As in point estimation for the Bayesian approach other statistical inferences such as interval estimation and hypothesis testing is straightforward. In interval estimation the frequentist approach utilizes the sampling distribution of the parameter. The interval boundaries are random as they are dependent on the particular observed sample. The interpretation of all the $(1-\alpha)\%$ random intervals is that $(1-\alpha)\%$ of the intervals contain the true parameter. Utilizing the data one confidence interval is calculated, there is a $(1-\alpha)\%$ confidence that this is one of the intervals that contains the true parameter. A misinterpretation would be that the interval contains the true parameter with $(1-\alpha)\%$ confidence. In this approach the fixed true parameter is in the confidence interval or not. The sampling distribution of the parameter may to be approximated with the normal distribution in some occasions, as this distribution might not be known. This could interfere the correctness of the results.

In contrast the Bayesian approach results in the posterior distribution of the parameter which is conditioned on the actual sample that occurred. It embodies the belief of the true parameter. An $(1-\alpha)\%$ interval, called a Bayesian credible interval, can easily be constructed from the posterior and is interpreted as holding the true parameter with $(1-\alpha)\%$ confidence or credibility. Additionally a normal approximation is not necessary.

The same applies to hypothesis testing, again in the frequentist approach this inference is based on the sampling distribution of the parameter. For example in one sided tests, under the null

hypothesis the p-value is calculated. If this probability is low, i.e. lower than the α confidence, the null hypothesis is rejected as the observed sample has a too low probability to occur. It is not said that it couldn't occur, but the alternative hypothesis is accepted. The null hypothesis is either true or not. However the sampling distribution might not be known so that often tests with less power needs to be chosen.

The Bayesian approach for hypothesis testing relies heavily on the prior which directly influences the probabilities of the hypotheses. One these are specified the posterior probability of the null hypothesis can be calculated. In the same way as in the frequentist way for the p-value, one could use this posterior probability of the null hypothesis.

There is no ultimate choice of either the frequentist or Bayesian approach. The choice will depend the specific problem as well on the willingness to formulate a prior belief and the considerations on the applicability of both approaches. To find the posterior might be a challenge as this is in general analytically not possible. With the increasing computational power of computers, the analytical calculations can be replaced by numerical computations. In Bayesian statistics Markov Chain Monte Carlo (MCMC) techniques are specifically suitable for this purpose. One such MCMC technique is the Gibbs sampler which is utilized in the next chapter.

5.5 THE GIBBS SAMPLER*

The Gibbs sampler is utilized in chapter 6.2, the idea is set out in the general case. It may be difficult to sample directly from the joint posterior distribution $(\theta_1, \theta_2, \dots, \theta_p)$. However it may be easier to sample from the fully conditional distributions $\pi(\theta_k|\{\theta_j, j \neq k\})$. The Gibbs sampler begins with an initial guess of the realization out of $(\theta_1, \theta_2, \dots, \theta_p)$ and let this guess be $(\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})$. Then one cycle of the Gibbs sampler uses the initial guess and the knowledge on the conditional distribution:$\theta_1^{(1)}\ from\ \pi(\theta_1|\theta_j^{(0)}, j \neq 1\})$

$$\theta_2^{(1)}\ from\ \pi(\theta_2|\theta_j^{(0)}, j > 2\})$$

$$\cdot\cdot$$

$$\cdot\cdot$$

$$\theta_p^{(1)}\ from\ \pi(\theta_p|\theta_j^{(1)}, j < p\})$$

After the first iteration all conditional distributions have been sampled from, the sample of all the conditionals is utilized by going through the next cycle of the Gibbs sampler. Performing this cycle t times and as t approaches infinity, the joint distribution of $(\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_p^{(t)})$ approaches the joint distribution of $(\theta_1, \theta_2, \dots, \theta_p)$. For a large enough t, the samples starting at t and the next samples can be considered to be samples for from the joint distribution of $(\theta_1, \theta_2, \dots, \theta_p)$. In general it difficult to make comments on the convergence behavior, this convergence depends on the particular application.

* This section is based on [24], for a more extensive overview and examples see [25].

6 METHODOLOGY

Two methodologies are presented to estimate the true VaR. The first is conditioned on the knowledge of the common factor and in the second this knowledge is not used. The methodology conditioned on the common factor is utilized in chapter 7.1, 7.2, and 7.3, while the methodology not conditioned on the common factor is utilized in chapter 7.3 only. Since chapter 7.1 and 7.2 investigates the impact of uncertainty in general, many realizations or scenarios (500 in this thesis) must be generated. The methodology not conditioned on the common factor requires computationally significant more time due to large matrix operations, so that this methodology is utilized on two datasets.

The columns of Δ are assumed to be independent for simplicity, i.e. there is no dependency of the defaults over the years. This assumption is irrelevant in case the common factor is known.

6.1 METHODOLOGY CONDITIONED ON THE COMMON FACTOR

Recognize that under the Vasicek model and given the common factor for each year of the defaults data Δ, each element $\Delta_{i,t}$ is independent Bernoulli distributed with $p(m_t) = \Phi\left(\frac{\Phi^{-1}(PD)-\sqrt{\rho}m_t}{\sqrt{1-\rho}}\right)$.

The number of defaults in year t, $\sum_{i=1}^{n}\Delta_{i,t}$, is binomially distributed with parameters n (the number of obligors) and $p(m_t)$.

Thus the data is each year distributed with a different conditional PD. Eventhough we put a Beta distribution on the prior unconditional PD and have a binomial likelihood function, the posterior is not Beta distributed due to the fact the binomial defaults are parameterized by the conditional PD.

Before moving on to the implementation we state the following result theorem 7.1 from [13], with which we can calculate the probability density function of a random variable which is a function of another random variable with known probability density function.

Let X be a continuous random variable having a known probability density function f_X and $Y \coloneqq g(X)$ where $g(.)$ is a strictly increasing monotonic function that is also differentiable. Then the probability density function of Y, f_Y, is given by:

$$f_Y(y) = f_X\big(g^{-1}(y)\big)\frac{d}{dy}\big(g^{-1}(y)\big)$$

Initiation:
A distributional choice for the unconditional PD, here Beta(alfaPrior,betaPrior) distributed.

To simplify notation we write Y $\coloneqq p(m)$ and X $\coloneqq PD$.

Set the prior $\pi(.)$ and hold it numerically by obtaining the density on many different points on the interval(0,1), set $t = 1$ and continue to step 1.

Step 1.

Convert the unconditional prior density $\pi(x)$ into an conditional prior density f_Y.

We hold the density of PD, which is our prior and we are looking for the density of $p(m_t)$ which is a function of PD, $g(PD) = \Phi\left(\frac{\Phi^{-1}(PD)-\sqrt{\rho}m_t}{\sqrt{1-\rho}}\right)$. This is a strictly monotone increasing function in PD and differentiable so that we can apply theorem. The inverse is given by: $g^{-1}(p(m_t)) = \Phi(\Phi^{-1}\big(p(m_t)\big)\sqrt{1-\rho} + \sqrt{\rho}m_t))$.

$f_Y(y) = f_X\big(g^{-1}(y)\big)\frac{d}{dy}\big(g^{-1}(y)\big) = \pi\big(g^{-1}(y)\big)\frac{d}{dy}\big(g^{-1}(y)\big) = \pi\Big(\Phi\big(\Phi^{-1}(y)\sqrt{1-\rho} + \sqrt{\rho}m_t\big)\Big)\frac{d}{dy}\Big(\Phi\big(\Phi^{-1}(y)\sqrt{1-\rho} + \sqrt{\rho}m_t\big)\Big)$.

Where
$\frac{d}{dy}\Big(\Phi\big(\Phi^{-1}(y)\sqrt{1-\rho} + \sqrt{\rho}m_t\big)\Big) = \phi\big(\Phi^{-1}(y)\sqrt{1-\rho} + \sqrt{\rho}m_t\big)\sqrt{1-\rho}\frac{d}{dy}\big(\Phi^{-1}(y)\big)$.
and $\frac{d}{dy}\big(\Phi^{-1}(y)\big) = \frac{1}{\phi(\Phi^{-1}(y))}$ can be found as before by combining $\frac{d}{dy}(\Phi(\Phi^{-1}(y))) = \phi(\Phi^{-1}(y))\frac{d}{dy}\big(\Phi^{-1}(y)\big)$ and $\frac{d}{dy}(\Phi(\Phi^{-1}(y))) = \frac{d}{dy}(\mathrm{y}) = 1$.

So that we work with:

$f_Y(y) = \pi(\Phi\big(\Phi^{-1}(y)\sqrt{1-\rho} + \sqrt{\rho}m_t\big))\phi\big(\Phi^{-1}(y)\sqrt{1-\rho} + \sqrt{\rho}m_t\big)\sqrt{1-\rho}\frac{d}{dy}\big(\Phi^{-1}(y)\big)$.

The common factor data is available so that m_t is a constant.

Step 2.

The conditional prior density is updated into the conditional posterior density using the the number of defaults δ_t in year t, $f_Y(y) \to f_{Y|\delta_t}(y)$.

$f_{Y|\delta_t}(y) = \frac{\mathrm{f}(\delta_t|y)f_Y(y)}{\int_{-\infty}^{\infty}\mathrm{f}(\delta_t|y)f_Y(y)\mathrm{d}y}$. Where $\mathrm{f}(\delta_t|y) = \binom{n}{\delta_t}y^{\delta_t}(1-y)^{n-\delta_t}$ and $f_Y(y)$ is from step 1

Step 3.

Convert the conditional posterior $f_{Y|\delta_t}$ back to an unconditional posterior $f_{X|\delta_t}$.

We hold the density of the conditional PD. We are looking for the density of the unconditional PD which is a function of the conditional $p(m_t)$ and the common factor m_t, $g\big(p(m_t)\big) = \Phi(\Phi^{-1}\big(p(m_t)\big)\sqrt{1-\rho} + \sqrt{\rho}m_t))$. This is a strictly monotonic increasing function in $p(m_t)$, differentiable and m_t is a given constant. Again we can apply the theorem and use $g^{-1}(PD) = \Phi\left(\frac{\Phi^{-1}(PD)-\sqrt{\rho}m_t}{\sqrt{1-\rho}}\right)$.

$$f_{X|\delta_t}(x) = f_{Y|\delta_t}\left(\Phi\left(\frac{\Phi^{-1}(x)-\sqrt{\rho}m_t}{\sqrt{1-\rho}}\right)\right)\frac{d}{dx}\left(\Phi\left(\frac{\Phi^{-1}(x)-\sqrt{\rho}m_t}{\sqrt{1-\rho}}\right)\right) =$$

$$f_{Y|\delta_1}\left(\Phi\left(\frac{\Phi^{-1}(x)-\sqrt{\rho}m_t}{\sqrt{1-\rho}}\right)\right)\phi\left(\frac{\Phi^{-1}(x)-\sqrt{\rho}m_t}{\sqrt{1-\rho}}\right)\frac{1}{\sqrt{1-\rho}}\frac{d}{dx}\left(\Phi^{-1}(x)\right).$$ Where $f_{Y|\delta_t}$ is from step 2.

Step 4.

While t<T, let the unconditional prior density be posterior density, $f_{X|\delta_t} \coloneqq \pi(x)$, t=t+1 and repeat steps 1-4.

Step 5.

Obtain a sample out of the posterior density.

Since the cumulative distribution can be approximated very accurate, we will use the Inverse Transformation Method.

The inverse cumulative distribution is approximated from the obtained (numerical) density. We draw an uniform. A realization of the distribution of interest is obtained by filling in the uniform in the inverse cumulative distribution.

Step 6.

Using the sample of the unconditional PDs, we can calculate the different VaRs.

<u>Implementation considerations</u>

The steps 1-3 are implemented in opposite sequence, since the third and second step are dependent on the previous steps.

We define the following variables in Matlab:

Step 3: $$f_{Y|\delta_1}\left(\underbrace{\Phi\left(\frac{\Phi^{-1}(x)-\sqrt{\rho}m_t}{\sqrt{1-\rho}}\right)}_{a\_step3}\right)\overbrace{\phi\left(\frac{\Phi^{-1}(x)-\sqrt{\rho}m_t}{\sqrt{1-\rho}}\right)\frac{1}{\sqrt{1-\rho}}}^{b\_step3}\underbrace{\frac{d}{dx}\left(\Phi^{-1}(x)\right)}_{c\_step3}.$$

Step 2: $$f_{Y|\delta_t}(y) = \underbrace{\frac{\overbrace{\mathrm{f}(\delta_t|y)}^{a\_step2}\overbrace{f_Y(y)}^{b\_step2}}{\int_{-\infty}^{\infty}\mathrm{f}(\delta_t|y)f_Y(y)\mathrm{d}y}}_{int\_step2}.$$

Step 1: $$f_Y(y) = \underbrace{\pi\left(\Phi\left(\Phi^{-1}(y)\sqrt{1-\rho}+\sqrt{\rho}m_t\right)\right)}_{a\_step1}\overbrace{\phi\left(\Phi^{-1}(y)\sqrt{1-\rho}+\sqrt{\rho}m_t\right)\sqrt{1-\rho}}^{b\_step1}\underbrace{\frac{d}{dy}\left(\Phi^{-1}(y)\right)}_{c\_step1}.$$

f($\delta_t|y$) is the binomial distribution with parameters n and success rate y.

The densities will be represented in numerical vectors. At the moment a value of the density must be calculated which is not in the vector, a cubic spline interpolation technique will be used to approximate the density at that particular point.

For step 5 we use the Inverse Transformation Method and get a (sufficient large) sample of the unconditional posterior PD. With this sample we can calculate the measures of the posterior distribution, such as the standard deviation, the expectation and the $(1-\alpha)^{th}$ quantile along with a plot of the distribution.

Using this sample we can calculate in step 6 the two VaR measures and compare them as will be shown in chapter 7.

6.2 Methodology not conditioned on the common factor

The methodology 'not conditioned on the common factor' directly finds a sample of the posterior distribution of PD after observing the defaults of all obligors over all the years. The PD is directly the TTC PD of interest. Unlike the previous chapter there is no explicit updating with Bayes' theorem to the posterior density based on the likelihood function and prior of the PD. Another difference lies in the fact the prior cannot be chosen and as the name explains the possibly known common factor is not utilized.

We have at our hands the matrix Δ, the observation of the defaults of n obligors over T time periods. The common factor vector M. The assumption is that $\Delta_{i,t}$ over i are dependent through the common factor m_t and are independent of the time t.

Let the prior $PD \sim uniform(0,1)$ and $\theta = \Phi^{-1}(PD)$ so that the default points $\theta \sim N(0,1)$.
$M_t \sim N(0,1)$, independent across t and elements of the vector T by 1 vector M.
Furthermore for each t, M_t is assumed independent of PD and consequently M_t is independent of θ.

The data $\Delta_{1,t}, \dots, \Delta_{n,t}$ given θ and M_t are for each t independent Bernoulli variables, $\mathbb{P}(\Delta_{i,t}= 1|\theta, M_t) = \Phi\left(\frac{\theta-\sqrt{\rho}M_t}{\sqrt{1-\rho}}\right) = \Phi(\alpha\theta + \beta M_t)$, where $\alpha = \frac{1}{\sqrt{1-\rho}}$ and $\beta = \frac{-\sqrt{\rho}}{\sqrt{1-\rho}}$. We introduce a latent variable $U_{i,t}$ such that given (θ, M_t), $U_{i,t} \sim N(\alpha\theta + \beta M_t, 1)$ again independent across (i,t).
We get $\mathbb{P}(U_{i,t} > 0|\theta, M_t) = 1 - \Phi(-\alpha\theta - \beta M_t) = \Phi(\alpha\theta + \beta M_t)$, thus $\Delta_{i,t}= 1_{U_{i,t}>0}$.

From the distribution $U_{i,t}$ conditioned on (θ, M_t),we get that $\mathbb{E}[U_{i,t}|(\theta, M_t)] = \alpha\theta + \beta M_t$. Using this and the property that a distribution that is conditionally normal is also normally distributed unconditionally, we find the unconditional $U_{i,t}$ to be normally distributed with mean 0 and variance $1 + \alpha^2 + \beta^2$.

This can be seen by calculating:

$$\begin{aligned}
\mathbb{E}\big[U_{i,t}\big] &= \mathbb{E}\big[\mathbb{E}[U_{i,t}\big|(\theta, M_t)]\big] \\
&= \iint \mathbb{E}\big[U_{i,t}\big|(\theta, M_t)\big] f_{\theta,M_t} d(\theta, M_t) = \iint \mathbb{E}\big[U_{i,t}\big|(\theta, M_t)\big] f_\theta f_{M_t} d\theta dM_t \\
&= \iint \mathbb{E}\big[U_{i,t}\big|(\theta, M_t)\big] \frac{1}{\sigma_\theta\sqrt{2\pi}} e^{-(\frac{\theta}{2\sigma_\theta})^2} \frac{1}{\sigma_{M_t}\sqrt{2\pi}} e^{-(\frac{M_t}{2\sigma_{M_t}})^2} d\theta dM_t \\
&= \iint (\alpha\theta + \beta M_t) \frac{1}{\sigma_\theta\sqrt{2\pi}} e^{-\left(\frac{\theta}{2\sigma_\theta}\right)^2} \frac{1}{\sigma_{M_t}\sqrt{2\pi}} e^{-\left(\frac{M_t}{2\sigma_{M_t}}\right)^2} d\theta dM_t \\
&= \iint \alpha\theta \frac{1}{\sigma_\theta\sqrt{2\pi}} e^{-(\frac{\theta}{2\sigma_\theta})^2} \frac{1}{\sigma_{M_t}\sqrt{2\pi}} e^{-(\frac{M_t}{2\sigma_{M_t}})^2} d\theta dM_t \\
&+ \iint \beta M_t \frac{1}{\sigma_\theta\sqrt{2\pi}} e^{-(\frac{\theta}{2\sigma_\theta})^2} \frac{1}{\sigma_{M_t}\sqrt{2\pi}} e^{-(\frac{M_t}{2\sigma_{M_t}})^2} d\theta dM_t \\
&= \int \mathbb{E}[\alpha\theta] \frac{1}{\sigma_{M_t}\sqrt{2\pi}} e^{-\left(\frac{M_t}{2\sigma_{M_t}}\right)^2} dM_t + \int \beta M_t \frac{1}{\sigma_{M_t}\sqrt{2\pi}} e^{-(\frac{M_t}{2\sigma_{M_t}})^2} dM_t = 0.
\end{aligned}$$

And
$var\big(U_{i,t}\big) = \mathbb{E}\big[(U_{i,t} - \mathbb{E}\big[U_{i,t}\big])^2\big] = \mathbb{E}\big[(U_{i,t})^2\big] = \mathbb{E}\big[\mathbb{E}[(U_{i,t})^2|(\theta, M_t)]\big] = \mathbb{E}[var\left(U_{i,t}\middle|(\theta, M_t)\right)] +$
$\mathbb{E}[(\mathbb{E}\big[U_{i,t}\big|(\theta, M_t)\big])^2] = 1 + \mathbb{E}[(\alpha\theta + \beta M_t)^2] = 1 + \mathbb{E}[\alpha^2\theta^2 + 2\alpha\beta\theta M_t + \beta^2 {M_t}^2] = 1 + \alpha^2 + \beta^2.$

In this setup we generate (θ, M, U) given Δ. We are interested in calculating VaRs like the $\text{VaR}_{1-\alpha}^{\text{estimated}}$. The (θ, M) given Δ can be obtained by simply throwing away the U in each sample.

The Gibbs sampler is used to generate from the joint distribution of (θ, M, U) given Δ:

I. Draw from θ given (M, U, Δ)
II. Draw from M given (θ, U, Δ)
III. Draw from U given (θ, M, Δ)

In step one we start with an initial choice of the realization of (M, U), Δ is known. After which we obtain a θ. Now the second step uses the drawn θ and the previous U. The third step uses the drawn (θ, M) in the first and second step to draw an U. After which the first step starts again, repeating these steps sufficiently many times this algorithm will approximate equilibrium and the next draws can be considered to be drawn from the joint distribution of (θ, M, U) given Δ.

Because Δ is a deterministic function of U, Δ becomes superfluous in the first two steps. The two steps become:

I. Draw from θ given (M, U)
II. Draw from M given (θ, U)

The joint distribution (θ, M, U) is multivariate normal:

$$(\theta, M, U) \sim N\left(\begin{pmatrix} \mu_\theta \\ \mu_{M_1} \\ .. \\ \mu_{M_T} \\ \mu_{U_{1,1}} \\ .. \\ \mu_{U_{T,n}} \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{1,1} & .. & \sigma_{1,k} \\ .. & .. & .. \\ \sigma_{k,1} & .. & \sigma_{k,k} \end{pmatrix}\right)$$

$$= N\left(\begin{pmatrix} 0 \\ 0 \\ .. \\ 0 \\ 0 \\ .. \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & .. & \sigma_{1,T+2} & .. & \sigma_{1,k} \\ .. & .. & & & .. \\ & & 1 & & \\ \sigma_{T+2,1} & & 1+\alpha^2+\beta^2 & & \\ .. & & & .. & .. \\ \sigma_{k,1} & .. & & .. & 1+\alpha^2+\beta^2 \end{pmatrix}\right)$$

Where Σ is the k by k covariance matrix (k:=T+nT+1) with on its first T+1 diagonal points,1 and on the remainder diagonal points, $1 + \alpha^2 + \beta^2$. The rest of the covariance matrix is given by:

$$\sigma_{1,2}, .., \sigma_{1,T+1} = Cov(\theta, M_t) = 0 \; for \; all \; t$$

$$\sigma_{1,T+2}, .., \sigma_{1,k} = Cov\left(\theta, U_{i,t}\right) = \alpha \; for \; all \; (i,t)$$

$$\sigma_{2,T+2}, .., \sigma_{2,k}, \sigma_{3,T+2}, .., \sigma_{3,k}, \ldots, \sigma_{T+1,T+2}, .., \sigma_{T+1,k,} = Cov\left(M_s, U_{i,t}\right) = \begin{cases} \beta \; if \; s = t \\ 0 \; if \; s \neq t \end{cases}$$

$$\sigma_{T+2,T+3}, .., \sigma_{T+2,k}, \sigma_{T+3,T+4}, .., \sigma_{T+3,k}, \ldots, \sigma_{T+nT,k} = Cov\left(U_{i,s}, U_{j,t}\right) = \begin{cases} \alpha^2 + \beta^2 & if \; s = t \\ \alpha^2 & if \; s \neq t \end{cases}$$

$$\sigma_{2,3}, .., \sigma_{2,T+1}, \sigma_{3,4}, .., \sigma_{3,T+1}, \ldots, \sigma_{T+1,T+1} = Cov(M_s, M_t) = 0 \; for \; all \; s \neq t$$

Where $Cov\left(\theta, U_{i,t}\right)$, $Cov\left(M_s, U_{i,t}\right)$ and $Cov\left(U_{i,s}, U_{j,t}\right)$ are found by seeing $U_{i,t}$ as an independent standard normal variable $\varepsilon_{i,t}$ relocated at $\alpha\theta + \beta M_t$:

$$Cov\left(\theta, U_{i,t}\right) = Cov\left(\theta, \varepsilon_{i,t} + \alpha\theta + \beta M_t\right) = Cov\left(\theta, \varepsilon_{i,t}\right) + Cov(\theta, \alpha\theta) + Cov(\theta, \beta M_t) = \alpha$$

$$Cov\left(M_s, U_{i,t}\right) = Cov\left(M_s, \varepsilon_{i,t} + \alpha\theta + \beta M_t\right) = Cov\left(M_s, \varepsilon_{i,t}\right) + Cov(M_s, \alpha\theta) + Cov(M_s, \beta M_t)$$

$$= \begin{cases} \beta \; if \; s = t \\ 0 \; if \; s \neq t \end{cases}$$

$$Cov\left(U_{i,s}, U_{j,t}\right) = Cov\left(\varepsilon_{i,s} + \alpha\theta + \beta M_s, \varepsilon_{j,t} + \alpha\theta + \beta M_t\right) = Cov(\alpha\theta, \alpha\theta) + Cov(\beta M_s, \beta M_t) =$$

$$= \begin{cases} \alpha^2 + \beta^2 & if \; s = t \\ \alpha^2 & if \; s \neq t \end{cases}$$

The conditional distributions $(\theta|M, U)$ and $(M|\theta, U)$ are normal as well and fully defined given the joint distribution.

The following theorem page 1014 [26] helps us to derive the distribution of the conditional distribution.

Let $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}$ be multivariate normally distributed, $\boldsymbol{x}_1$ a subset of and $\boldsymbol{x}_2$ the remaining variables out of $\boldsymbol{x}$. Partition $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ likewise so that $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$. Then the conditional distribution of $\boldsymbol{x}_1|\boldsymbol{x}_2$ is normal as well: $\boldsymbol{x}_1|\boldsymbol{x}_2 \sim N(\boldsymbol{\mu}_{\boldsymbol{x}_1|\boldsymbol{x}_2}, \boldsymbol{\Sigma}_{\boldsymbol{x}_1|\boldsymbol{x}_2})$ with $\boldsymbol{\mu}_{\boldsymbol{x}_1|\boldsymbol{x}_2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{\boldsymbol{22}}^{\boldsymbol{-1}}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)$ and $\boldsymbol{\Sigma}_{\boldsymbol{x}_1|\boldsymbol{x}_2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{\boldsymbol{22}}^{\boldsymbol{-1}}\boldsymbol{\Sigma}_{21}$. $\boldsymbol{\Sigma}_{\boldsymbol{x}_1|\boldsymbol{x}_2}$ is also known as the Schur complement of the block $\boldsymbol{\Sigma}_{22}$ of the matrix $\boldsymbol{\Sigma}$.

Using this result one can find the mean vector and covariance matrix of $(\theta|M, U)$.

The mean of $(\theta|M, U) = \mu_\theta + (\sigma_{1,2} \quad \cdot\cdot \quad \sigma_{1,k})$

$$\begin{pmatrix} \sigma_{2,2} & \cdot\cdot & \sigma_{2,k} \\ \cdot\cdot & \cdot\cdot & \cdot\cdot \\ \sigma_{k,2} & \cdot\cdot & \sigma_{k,k} \end{pmatrix}^{-1} \left(\begin{pmatrix} M_1 \\ \cdot\cdot \\ M_T \\ U_{1,1} \\ \cdot\cdot \\ U_{T,n} \end{pmatrix} - \begin{pmatrix} \mu_{M_1} \\ \cdot\cdot \\ \mu_{M_T} \\ \mu_{U_{1,1}} \\ \cdot\cdot \\ \mu_{U_{T,n}} \end{pmatrix} \right) = \mu_\theta + (\sigma_{1,2} \quad \cdot\cdot \quad \sigma_{1,k})$$

$$\frac{1}{\begin{vmatrix} \sigma_{2,2} & \cdot\cdot & \sigma_{2,k} \\ \cdot\cdot & \cdot\cdot & \cdot\cdot \\ \sigma_{k,2} & \cdot\cdot & \sigma_{k,k} \end{vmatrix}} adj \begin{pmatrix} \sigma_{2,2} & \cdot\cdot & \sigma_{2,k} \\ \cdot\cdot & \cdot\cdot & \cdot\cdot \\ \sigma_{k,2} & \cdot\cdot & \sigma_{k,k} \end{pmatrix} \begin{pmatrix} M_1 - \mu_{M_1} \\ \cdot\cdot \\ M_T - \mu_{M_T} \\ U_{1,1} - \mu_{U_{1,1}} \\ \cdot\cdot \\ U_{T,n} - \mu_{U_{T,n}} \end{pmatrix}$$

Which is cumbersome and not possible to simplify analytically further, the mean and covariance of the conditional distribution will be calculated in Matlab.

Equivalently for step 2 $(M|\theta, U)$ is calculated in the same fashion.

For step III we need to draw from U given (θ, M, Δ). We can represent U as $U_{i,t} = \alpha\theta + \beta M_t + e_{i,t}$, where each $e_{i,t} \sim N(0,1)$ independent of (θ, M). Since Δ is available we hold certain information about U. At the moment $\Delta_{i,t} = 1$, by definition $U_{i,t} > 0$. So that $\alpha\theta + \beta M_t + e_{i,t} > 0$ or $e_{i,t} > -(\alpha\theta + \beta M_t)$. Idem for $\Delta_{i,t} = 0$.

We get:

$$\begin{cases} e_{i,t} > -\alpha\theta - \beta M_t, & if\ \Delta_{i,t} = 1 \\ e_{i,t} \leq -\alpha\theta - \beta M_t, & if\ \Delta_{i,t} = 0 \end{cases}$$

There is no further information on $e = (e_{i,t})$ given (θ, M, Δ). Thus $e_{i,t}$ is sampled from the standard normal distribution restricted to $(-\alpha\theta - \beta M_t, \infty)$ if $\Delta_{i,t} = 1$ and restricted to $(-\infty, -\alpha\theta - \beta M_t)$ if $\Delta_{i,t} = 0$, for $e_{i,t}$ independently across (i, t).

Such a distribution is called a truncated distribution and the probability density function of X restricted to (a,b) is given by:

$$f(x|a < X < b) = \frac{g(x)}{G(b) - G(a)}$$

Where g(.) is the unrestricted probability density function and G(.) is the unrestricted cumulative distribution.

The cumulative distribution of the truncated distribution can be easily found:

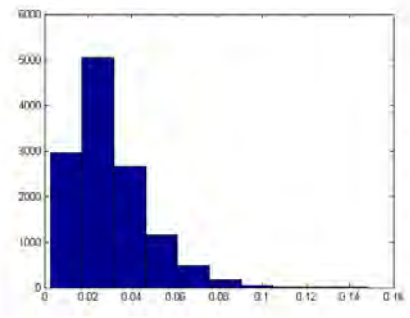
$$F(x|a < X < b) = \int_{t=a}^{x} \frac{g(t)}{G(b) - G(a)} dt = \frac{G(x) - G(a)}{G(b) - G(a)}$$

The distribution function of the truncated normal distribution restricted to $[\mu, \infty)$ is $\frac{\Phi(x) - \Phi(\mu)}{1 - \Phi(\mu)}$ for $x \geq \mu$. Thus its quantile function is $u \mapsto \Phi^{-1}(\Phi(\mu) + u(1 - \Phi(\mu)))$. A sample from the left-truncated normal can now be generated by drawing from an uniform(0,1), u, and filling it in the quantile function.

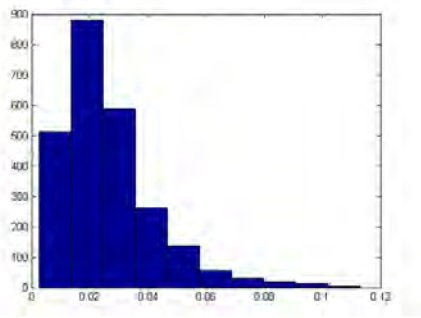
Similarly one obtains a right truncated distribution function at $(-\infty, \mu]$, which is $\frac{\Phi(x)}{\Phi(\mu)}$. We can find a sample form the right truncated normal by drawing from an uniform u, and filling it in the quantile function $u \mapsto \Phi^{-1}(u\Phi(\mu))$ for $x \leq \mu$.

*6.2.1 BURN IN SIZE GIBBS SAMPLER**

The size of the burn in for the implementation of the Gibbs sampler needs to be determined. First we vary the number of obligors and years of data to look at the form of the posterior distribution.

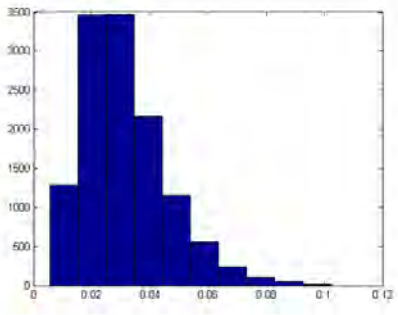


n=100,T=5

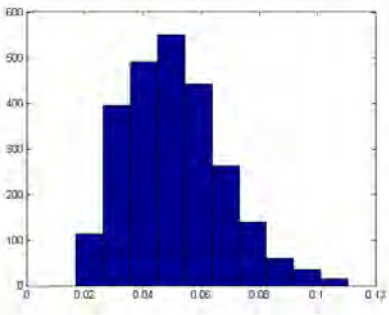


n=200,T=5

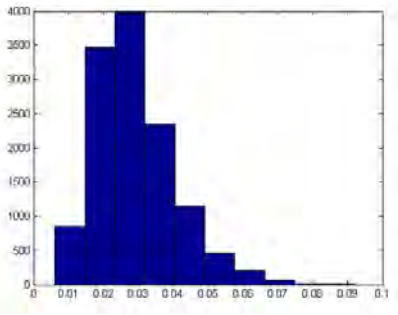
* This section is based on [27].



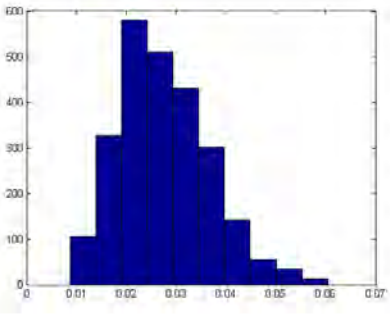
n=100,T=7



n=200,T=7



n=100,T=10



n=200,T=10

As we can see there is clearly one peak in the resulting distribution of the posterior PD. Markov chains are said to be poor mixing when a chain stays in small parameter regions for a long time, which can arise due to multiple peaks of the convergent distribution. This is clearly not the case here.

Determining the convergence of the Markov Chain

In the above Gibbs sampler runs, we have done ten thousands Markov Chain updates and kept the last 2500 values as the sample of the posterior PD distribution.

In this section we determine the size of the burn-in for the Gibbs sampler in a more appropriate way, first looking at possible autocorrelations to avoid dependency. We perform a QQ-plot (quantiles against quantiles plot) along with the two sample Kolmogorov Smirnov test and obtain two independent samples by using 'batch sampling'. This means that we take a subset, $y^{(1)}, y^{(2)}, \ldots$ out of the chain of values $x^{(1)}, x^{(2)}, \ldots$ by only retaining those values that satisfy $y^{(i)} = x^{(ki)}$ for some fixed k. The higher the lag k, the lower the dependency and the (auto)correlation is. However a small k is preferred so that waste is minimized. To determine an appropriate k we look at the autocorrelation with lags 2,5, 10 and 25 of one sample path of the Gibbs sampler (T=10 and n=100), going through the Gibbs sampler 13,000 times. After throwing away the first 1000 values, welook at the 1000th to 13000th value of the chain with different lags.

1 lag	2 lags	5 lags	10 lags	25 lags
0.0590	0.0556	0.0554	0.0438	-0.0065

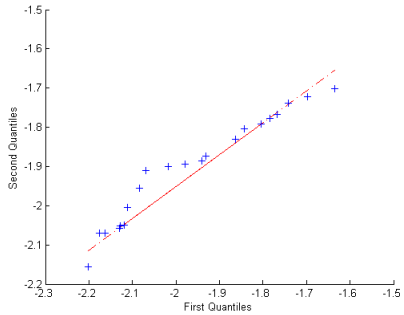
Table 7 Autocorrelations

We choose $k = 25$, since the correlation is minimal. This will be used for all simulations in the next section and will get us independent samples for the Kolmogorov Smirnov testing procedure.

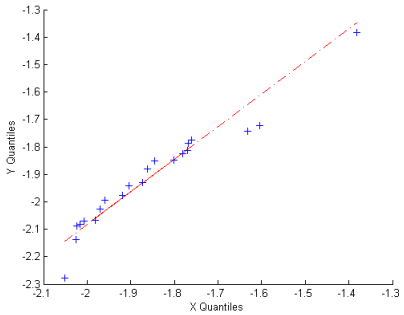
We continue with the simulation run and extend the iteration through the Gibbs sampler to above 20,000 times. The k-s test is runned on the Markov Chain along with a QQplot on the two samples (again after throwing away the first 1000 values) on three different sets.

- 1^{th}, 26^{th} ,…,501^{th} values against the sample of 501^{th} , 1026^{th},…,1001^{th} values (A)
- 5001^{th} , 5026^{th} ,…,5501^{th} values against the 5501^{th} , 5526^{th},…, 6001^{th} values (B)
- 20001^{th} , 20026^{th} ,…,20501^{th} values against the 20501^{th} , 20526^{th},…, 21001^{th} values (C)

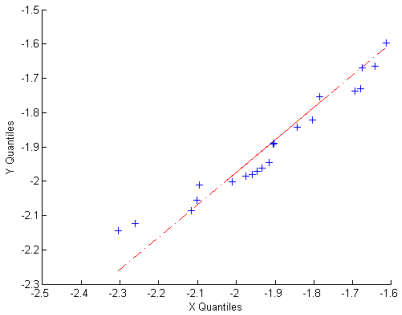
$$H_0: The\ two\ samples\ are\ from\ the\ same\ distribution$$
$$H_1: The\ two\ samples\ are\ not\ from\ the\ same\ distribution$$



(A)
K-S test: $H_0\ not\ rejected, p = 0.1545$



(B)
K-S test: $H_0\ not\ rejected, p = 0.3038$



(C)
K-S test: $H_0\ not\ rejected, p = 0.7974$

This will enable us get some insight in the required burn-in sample size before convergence. As a result we see that the k-s test is never rejected and there is no clear evidence from this test the samples are not from the same distribution. However the QQplot of (A) seems to show some abnormality. So to be on the safe side, we will let our burn-in size be followed by (B), a sample size of 5000.

7 RESULTS

In this chapter two questions will be answered. The first question concerns the impact of parameter uncertainty in terms of the VaR and the second question zooms in the performance of the alternative estimator incorporating parameter uncertainty. Prior to the results the choices made in this simulation study are fully specified.

Here parameter uncertainty concerns solely the uncertainty around the PD and its impact on the VaR measure. More specifically in the regulatory capital framework such that $\alpha = 0.1\%$ and VaR is calculated in terms of a fraction of the portfolio. That is for eventual exposures incorporating recovery in money units, the numbers need to be multiplied with LGD and EAD. It is assumed that the true PD of the portfolios is 0.03 and data is generated following the model of Vasicek specified in the chapter about the ASRF model.

As parameter uncertainty depends on the size of available datasets and is especially present in small portfolio sizes, several portfolios with varying sizes are considered. We vary the portfolios along the cross sections or number of obligors ($n = 50, 100 or\ 200$) in combination of different time series lengths or number of years of data (T = 5, 10 to 20). So that for each combination of n and T, nine different sizes are considered. The choice to set T for a minimum of 5 years is based on the Basel guidelines requiring estimates of the risk components on at least 5 years of data for sovereign, corporate and bank exposures.

Five hundred different datasets (five hundred realizations of the model) are generated consistently for each portfolio size. That is 500 datasets are generated in total for all nine portfolio sizes by using the same common random numbers (same seed) for each portfolio size. I.e. for the default realizations of the portfolio with 100 obligors, the same defaults applies to the first 100 obligors in the 200 obligors portfolio. So that the total number of individual defaults equals 500 times 200 (=maximum n) times 20 (=maximum of T).

The same common random numbers reduces sampling noise of the simulation for comparison, this can be seen by the variance of the difference between the estimates ϑ_1 and ϑ_2 is:

$$var(\vartheta_1 - \vartheta_2) = var(\vartheta_1) + var(\vartheta_2) - 2correlation(\vartheta_1, \vartheta_2)$$

The variance of the difference is minimized by maximizing the correlation. Using the same common random numbers maximizes the correlation and performing the calculations such also ensures that the results are reproducible.

To answer the two questions, the theory given the common factor is used for the simulation. It is assumed that the common factor is estimated perfectly from data and further more is generated independently from year to year. This assumption of independency is not of influence in this setting where the common factor is known. But might be in the setting where the common factor is unknown and when in a sequence of high (generated) valued common factors, i.e. consecutive booming economy states, in combination of small cross sections and low PD causes no observed defaults. To compare datasets with different time series We have noticed the influence of particular realizations of the common factor. In a sample of 20 standard normal draws, there will always be some realizations far positive. In case this high values are obtained

for years after year 5, the comparison of a 5 year dataset with a 20 year dataset is distorted. This poses an additional difficulty of estimating the VaR of a portfolio with zero defaults. I choose to generate 5 standard normal variables, none extreme, and let the longer time series consist of these 5 standard normal variables. That is the common factor M_t for year t is unique for the first five years: $M_1, M_2, .., M_5$. While the following common factors $M_6, M_7, .., M_{20}$ are equal to the sequence $M_1, M_2, .., M_5, M_1, M_2, .., M_5$, ... This ensures the same conditions for the comparison of portfolios. Thus for T=20, it consists of four times the block of common factors for T=5.

The advantage of the framework given the common factor is the freedom in the choice of the a priori distribution for the PD as opposed to the framework without knowing the common factor where this factor is assumed to be $uniform(0,1)$. Strictly speaking the only correct choice for the a priori distribution with no data other than the defaults and common factors at hand is this uniform prior. However next to the data on the defaults and the common factors there is additional information available such as expert opinions and certain information on the characteristics of the obligors in the portfolios. Because of this data an uniform prior is on the pessimistic sideand causes priors to put too much mass on the higher PDs. However connecting this data to the prior or in the Bayesian updating mechanism in an universal compatible method is not straightforward. We assume that incorporating this additional information translates in a dispersed belief around the true PD. This is shaped in the Beta distribution with parameters 3 and 97 while the mean of this distribution is chosen to be 0.03.

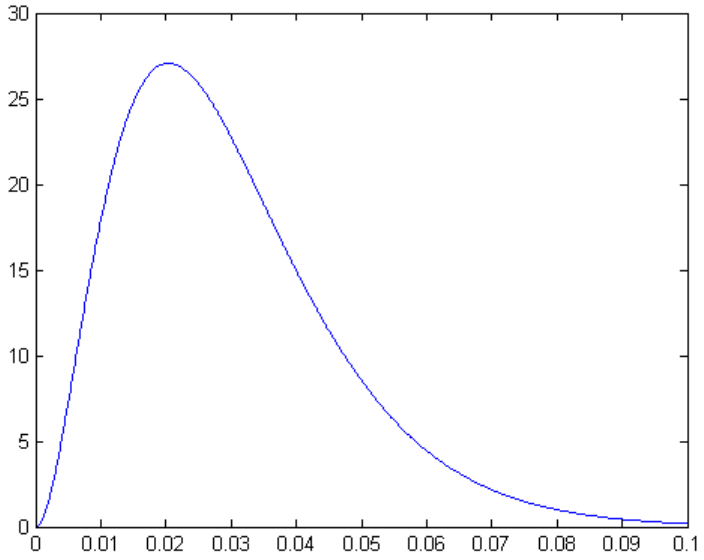


Figure 4 The a priori belief dispersed around the true PD

The dispersion is great around the mean, the probability that the true PD is lower than 0.02 is 32% while probability that the true PD is larger than 0.04 is 24% so the belief is 56% that the PD is within 0.01 deviance of the true PD.

Next to the advantage of freedom in choice of prior distribution, the framework without the common factor requires significantly more time to compute. Running the simulation under one particular prior requires given the common factor more than 24 hours on an Intel Core 2 Duo processor E8500 @ 3.16 GHZ in combination of 3.49 GB of memory to run. While running the same framework without the common factor scheme requires more than thirty times as much time. The framework not given the common factor will be used for the real portfolio data as only one realization in combination with a fixed number of obligors and years of data needs to be calculated. Furthermore the knowledge of the common factor or on the economic state is valuable as capitalizing this information causes more accurate estimation of the PD. The assumption knowing the common factor will be assessed as well as the particular choice of the prior in the next sections.

7.1 Assessing the impact of probability of default uncertainty

We give some insight in the reduction in uncertainty around the PD using the above framework and looking at the first dataset (first common random number) that is used. We show a graph with a fixed T and increasing n and vice versa.

The belief of the PD after updating is as follows for increasing T starting at the smallest portfolio n=50 and T=5.

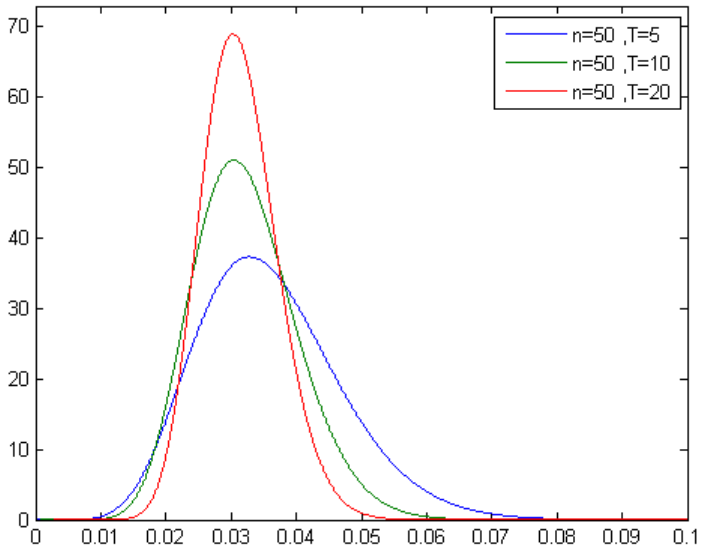


Figure 5 Probability of Default belief updated with increasing years

In the below table we summarized the distribution by looking at the probability within 0.01 interval of the true PD of 0.03 and outside these border (results are rounded to one decimal).

Probability of default	<0.02	0.02-0.04	>0.04
Posterior under T=5, n=50	20.7%	71.1%	8.2%
Posterior under T=10, n=50	23.8%	74.5%	1.6%
Posterior under T=20, n=50	8.3%	91.1%	0.6%

Table 8 Probability of Default belief updated with increasing years

Similar we look at the PD uncertainty from the smallest portfolio n=50 and T=5 and increasing the number of obligors.

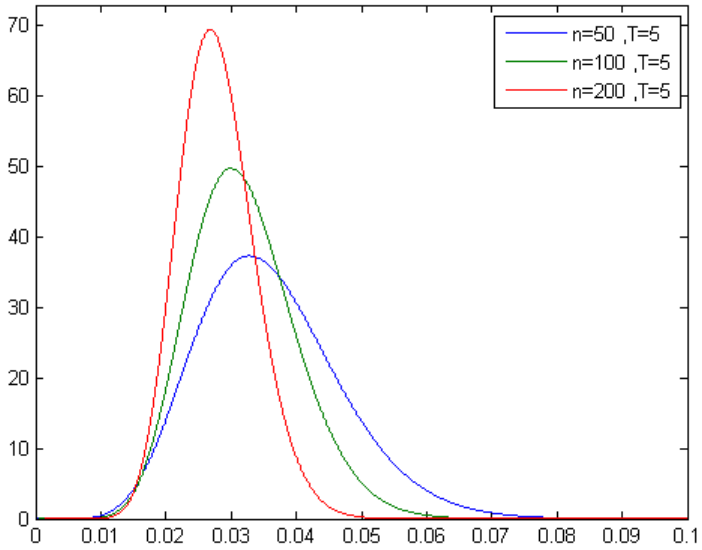


Figure 6 Probability of Default belief updated with increasing obligors

Probability of default	<0.02	0.02-0.04	>0.04
Posterior under T=5, n=50	20.7%	71.1%	8.2%
Posterior under T=5, n=100	7.9%	85.7%	6.4%
Posterior under T=5, n=200	2.9%	95.2%	2.0%

Table 9 Probability of Default belief updated with increasing obligors

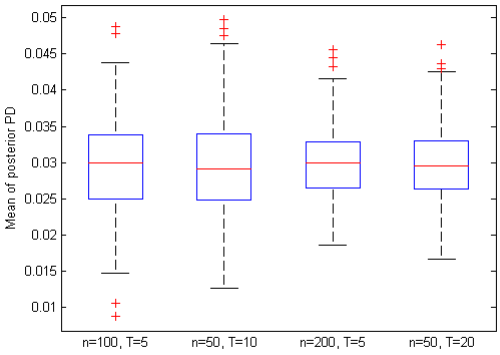
Looking at the mass figures in the interval 0.02-0.04, for increasing the number of obligors more mass is concentrated around this area which can be interpreted as less uncertainty around the true PD.

A question is whether this particular seed caused more reduction in uncertainty in increasing the cross sections. This is particular interesting as Tarashev found the opposite result, that uncertainty decreases more with increasing years of data. Although it must be kept in mind that Tarashevs framework is specified differently.

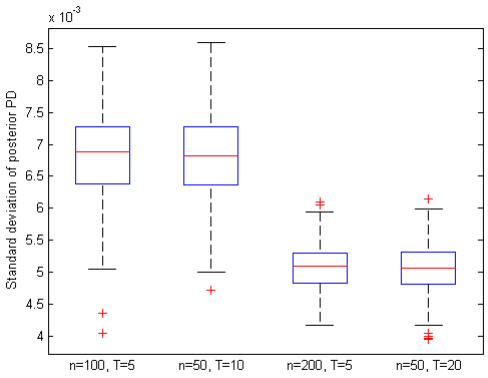
7.1.1 THE EFFECT OF THE DIMENSIONS OF THE DATASET

To tackle the previous concern we perform an experiment using 500 different realizations of the data. The posterior distribution for all 500 realizations is obtained and the smallest portfolio is increased in size two times and 4 times in either the cross sections or in the time series. The comparison concerns the portfolios with n=100 and T=5 versus n=50 and T=10 and the portfolios with n=200 and T=5 versus n=50 and T=20. Each of the 500 realizations for each portfolio consists of the posterior distribution of the PD. To compare the difference in the portfolio the mean and the standard deviation of the posterior is obtained. We get two paired samples (due to the same seed) for each comparison containing 'sample means' and 'sample standard deviations'.

Below the plots of the two means and standard deviations are shown, where the boxplots of the first two and the last two must be compared.



Boxplot of the means of the PD



Boxplot of the standard deviations of the PD

From the boxplots one can observe that the dispersion of the posterior PD in terms of mean and standard deviation look similar, we strengthen this observation with the Wilcoxon signed-rank test which is a non parametric paired two sample test.

The hypothesis for the means is:

$$H_0: The\ means\ sample\ are\ identical\ distributed$$
$$H_1: One\ of\ the\ means\ sample\ is\ stochastically\ larger$$

And for the standard deviations is:

$$H_0: the\ standard\ deviations\ samples\ are\ identical\ distributed$$
$$H_1: One\ of\ the\ standard\ deviations\ sample\ is\ stochastically\ larger$$

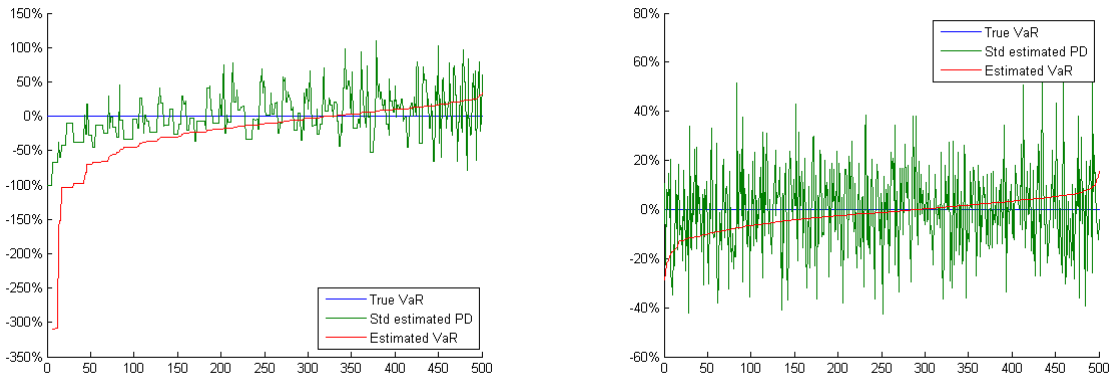
With a p-values of 0.9790 for the means and 0.9914 for the standard deviations for n=100 and T=5 versus n=50 and T=10, the null hypothesis is not rejected at confidence level of 5%.

High p-values for n=200 and T=5 versus n=50 and T=20 of 0.7476 for the means and 0.7954 for the standard deviations means that the null hypothesis is not rejected as well in the second case at 5% confidence level.

We conclude that the influence of increasing the cross sections or time series has the same influence in the framework where the common factor is known.

7.1.2 *The gap between the true VaR and estimated VaR*

In the below graphs uncertainty due to sampling noise is exposed. The red line displays the relative error in percentage of the computed estimated VaR for 500 realizations, i.e. -100% on the y-axis means that of the estimated VaR an additional 100% of this estimated VaR must be added to get to the true VaR level. The graph is sorted from the lowest estimated VaR till the highest. As one can see the estimated method tends to underestimate the true VaR, more than half of the calculated estimated VaR is lower than the true VaR and on top the lower calculated estimated VaRs go relatively 'deeper' down. This follows from section 3.2.1. The two discrepancies from the true VaR is shown graphically for the smallest portfolio, where one expects the largest uncertainty impact and the largest portfolio where one expects the smallest uncertainty impact. This reduction is large as the smallest portfolio has a range around -300% to 40% while a 4 times bigger portfolio and a 4 times longer time series gets us a range around -20% to +20%.



n=50, T=5 n=200, T=20

Figure 7 Discrepancy between true and estimated VaR

The smallest portfolio had 5 out of the 500 datasets no observations (these are runs 139, 184, 235, 382 and 484), so that the estimated estimated PD was 0 in those times. This caused the red line in the graph to start the drawing from the sixth observation while this did not occur in the rest of the portfolios. Observe that the largest portfolio's red line does start from the y-axis.

Additionally we have plotted the standard deviation of the observed unconditional default rates (green line), which shows no clear tendency of the standard deviation to move with the discrepancy between the true and estimated VaR.

The average underestimation of the estimated VaR relative to the true VaR is around the 22% to 2% of its computed estimated VaR for T=5,10 and 20 as can be seen in the table. Due to large number of realizations, these numbers can be considered pretty accurate for our purpose. The percentages represent how much additional capital in the regulatory framework on average is needed due to parameter uncertainty.

	n=50	n=100	n=200
T=5	-21.7%	-10.0%	-4.4%
T=10	-13.2%	-5.9%	-2.4%
T=20	-9.4%	-4.3%	-1.8%

Table 10 Average error of $\mathbf{VaR}_{1-\alpha}^{\mathbf{estimated}}$

The underestimation issue gets smaller as data increases.

The absolute error disregards the sign of the error and is thus absolutely larger. The absolute error of the estimated VaR ranges from 30%-5%.

	n=50	n=100	n=200
T=5	30.3%	16.8%	10.1%
T=10	18.9%	10.7%	6.4%
T=20	13.2%	8.0%	4.8%

Table 11 Absolute error of $\mathbf{VaR}_{1-\alpha}^{\mathbf{estimated}}$

Before we stated that the influence of the dimensions of the dataset is not of influence in the posterior distribution of the PD. A first look shows a difference in both average and absolute error to the true VaR when one increases the number of years of the smallest portfolio (vertical direction) has a smaller error than increasing the number of obligors (horizontal direction). The reason is due to the influence of specification of the conversion from conditional to unconditional defaults using the common factor. This conversion is more precise when the number of obligors gets larger. To see this observe how the observed default rate is determined:

$$\delta_t = \frac{1}{n}\sum_{i=1}^{n}\Delta_{i,t} \text{ or observed default rate} = \frac{\#defaults\ in\ portfolio}{n}$$

As n gets larger the discrete observed default rate gets more and more continuous with smaller intervals in consecutive rates. The mapping from this observed conditional default rate to the unconditional default rate thus gets more and more accurate and hence the additional error in

this mapping gets small so that the observed average and absolute error towards the true VaR gets smaller. The same argumentation applies to δ_t^*. Increasing the time series doesn't help this mapping to be more refined.

Tarashev found that increasing the number of years decreased the uncertainty more than increasing the number of obligors. The essential difference is that the common factor is assumed known in our setting while this not in the Tarashev's paper. The reason why uncertainty reduction was larger in the number of years is because the influence of a couple high valued common factors, resulting in a higher probability of no defaults, is smaller with more years of data.

Furthermore the information on the common factor is useful to convert from conditional defaults to unconditional defaults leading to more certain estimates of the unconditional PD.

7.2 Assessing the performance of the different VaR estimators

In this section the performance of $\text{VaR}_{1-\alpha}^{\text{estimated}}$ is set against $\text{VaR}_{1-\alpha}^{\text{pu}}$. Below is the plot for the smallest portfolio. As can be seen the noise in the parameter uncertainty incorporating measure is lower and closer to the true VaR, the estimated VaR is far worse.

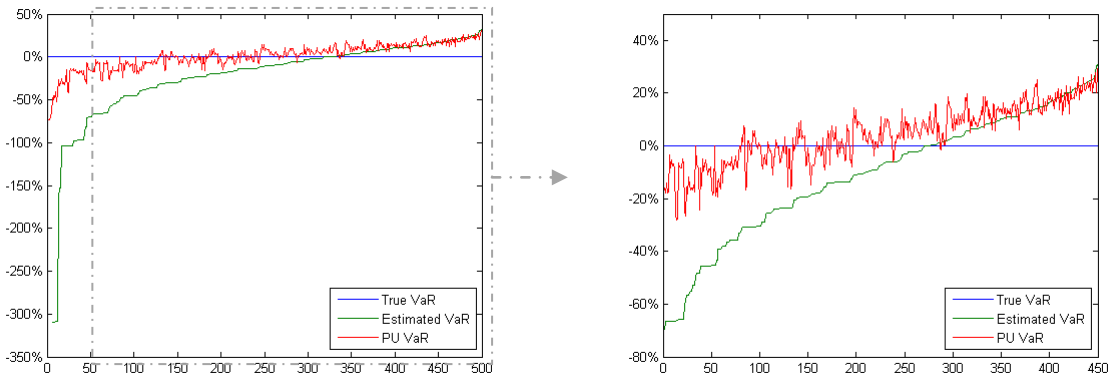


Figure 8 Impact uncertainty on the smallest portfolio

For $\text{VaR}_{1-\alpha}^{\text{pu}}$, the obvious underestimation bias present in the estimated VaR is gone. Observe that the plot is sorted from the lowest $\text{VaR}_{1-\alpha}^{\text{estimated}}$ and that $\text{VaR}_{1-\alpha}^{\text{pu}}$ is relatively higher than $\text{VaR}_{1-\alpha}^{\text{estimated}}$ when $\text{VaR}_{1-\alpha}^{\text{estimated}}$is relatively more lower to the true VaR. On the other hand the correction upwards by $\text{VaR}_{1-\alpha}^{\text{pu}}$ is not present in the less frequent case that $\text{VaR}_{1-\alpha}^{\text{estimated}}$overestimates the true VaR. This is a characteristic that is desirable and also makes clear one cannot just increase the $\text{VaR}_{1-\alpha}^{\text{estimated}}$ with a fixed percentage. Furthermore the absolute errors are smaller, so that this measure is more accurate in estimating the true VaR. Between apprentices in the absolute table is the percentage error reduction.

	n=50	n=100	n=200
T=5	1.0%	0.2%	-0.1%
T=10	0.3%	-0.1%	-0.0%
T=20	-0.2%	-0.2%	-0.0%

Table 12 Average Error of PU VaR

	n=50	n=100	n=200
T=5	11.9%	9.6%	7.6%
T=10	9.7%	7.2%	5.2%
T=20	7.8%	6.0%	4.4%

Table 13 Absolute error of PU VaR

7.2.1 ADDITIONAL CAPITAL REQUIREMENT FOR PARAMETER UNCERTAINTY

Here the additional capital requirement in the regulatory framework above the estimated VaR is calculated to get to the level of $\text{VaR}_{1-\alpha}^{\text{pu}}$. In the smallest portfolio, 5 times no defaults are observed so that the five times the estimated VaR is zero. These 5 values are taken out in the calculations below.

	n=50	n=100	n=200
T=5	21.5%	11.1%	8.0%
T=10	23.3%	7.7%	5.5%
T=20	16.0%	6.9%	4.5%

Table 14 Average additional capital on top of $\text{VaR}_{1-\alpha}^{\text{estimated}}$

Overall the additional capital requirement ranges for the small portfolio around 24% to less than 5% for the largest portfolio.

7.2.2 THE INFLUENCE OF THE PRIOR

We will look at the influence of the prior on $\text{VaR}_{1-\alpha}^{\text{pu}}$. It is known that the prior effect is smaller as more data is available, the influence is assessed comparing the uniform prior against the beta prior for all 9 portfolio sizes.

In the following figures We plot the $\text{VaR}_{1-\alpha}^{\text{pu}}$ resulting from the uniform prior and the Beta prior for two portfolios. The prior choice will have the largest influence on the smallest portfolio with 50 obligors and 5 years of data and the smallest influence for the largest portfolio with 200 obligors and 20 years of data.

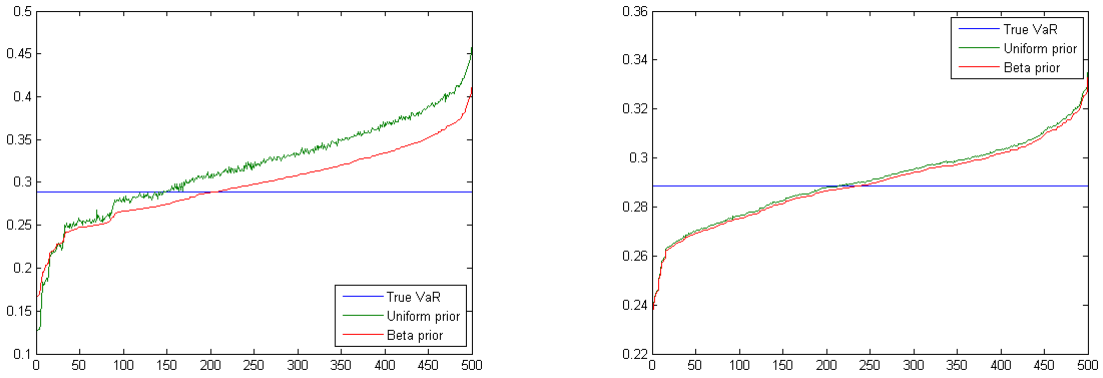


Figure 9 Influence of Uniform over the Beta prior for portfolio n=50, T=5(left) and n=200, T=20 (right)

As expected the green line representing the uniform prior with mean 0.5 is overall higher than the Beta prior with the lower mean. In terms of $VaR_{1-\alpha}^{pu}$ the difference is by average 6.5% in the smallest dataset but quickly decreases to small levels.

	n=50	n=100	n=200
T=5	6.5%	3.4%	1.8%
T=10	3.4%	1.8%	0.9%
T=20	1.7%	0.9%	0.5%

Table 15 Influence Uniform over Beta prior in % of the Beta prior

In case of the largest portfolio the difference is negligible with n=200 and T=20 only 0.5% extra VaR.

We conclude that the influence decreases quickly with the size of the dataset. In relation to the size of the uncertainty, the effect is substantial especially for the smaller portfolios. From the plot the overestimation of the uniform over the beta prior looks stable so that optionally one could subtract a percentage additional VaR from the uniform calculated VaR.

7.3 APPLICATION IN PRACTICE: REAL BANK PORTFOLIOS <CLASSIFIED>*

* This section is classified and is concealed in the published version.

8 CONCLUSION

As credit risk is important for banks to hold capital to protect themselves against unexpected losses, the real capital requirement can differ substantially without incorporating the parameter uncertainty that is present in small datasets. In this thesis the impact in the regulatory framework using the ASRF model is investigated, which is the model for the internal ratings based approach of Basel II. The Value at Risk (VaR) is calculated according to the ASRF model in a simulation study. In the model the PD determines the fraction of the portfolio to be set aside, where the correlation parameter has a relatively smaller impact and is in the Basel accords a function of the PD.

In this thesis two methodologies are shown. The first conditions on the knowledge of a common factor that represents a factor that influences all obligors in a specific portfolio in a homogeneou way, such a common factor can be regarded as a systematic economic indicator. The second assumes this latent variable to be unknown.

Both methodologies adopt a Bayesian approach in which the parameter PD embodies the uncertainty by modeling it as a probability distribution. A prior must be chosen after which it is updated to a posterior distribution by conditioning on the data. The more data is present the less the uncertainty there is, so that we choose to investigate for different amounts of data. Since the methodology not capitalizing on the common factor takes significant more time and some information is present about the common factor, we investigate the impact capitalizing on this information. To investigate the impact of the uncertainty, an 'estimated VaR' $\mathrm{VaR}_{1-\alpha}^{\mathrm{estimated}}$is compared to an alternative VaR formula incorporating parameter uncertainty$\mathrm{VaR}_{1-\alpha}^{\mathrm{pu}}$. It is shown that the first VaR tends to underestimate the VaR by -21.7% to -1.8% on average, which uses the mean of the historical default rates as an estimate of the PD. The effect of this assumption on the common factor must be investigated properly as it is important in both methodologies.

Assuming the validity of the framework, $\mathrm{VaR}_{1-\alpha}^{\mathrm{pu}}$is shown to be superior to $\mathrm{VaR}_{1-\alpha}^{\mathrm{estimated}}$. When the estimated VaR underestimates the true VaR, $\mathrm{VaR}_{1-\alpha}^{\mathrm{pu}}$ is higher and thus closer to the true VaR. In the cases the estimated VaR is higher than the true VaR, $\mathrm{VaR}_{1-\alpha}^{\mathrm{pu}}$ is not much different than $\mathrm{VaR}_{1-\alpha}^{\mathrm{estimated}}$. This observed property is desirable and it will depend on the data how much the additional capital requirement is. On average the additional capital requirement over $\mathrm{VaR}_{1-\alpha}^{\mathrm{estimated}}$ to the parameter uncertainty incorporating measure $\mathrm{VaR}_{1-\alpha}^{\mathrm{pu}}$ ranges from +24% for the smaller portfolios to + 4.5% for the largest portfolio.

In the above results the true VaR is known as a simulation was performed to generate data. For two portfolios using real default data, the methodology conditioned on the common factor and without the common factor are performed to calculate the difference between $\mathrm{VaR}_{1-\alpha}^{\mathrm{estimated}}$ and $\mathrm{VaR}_{1-\alpha}^{\mathrm{pu}}$.

The results have to be taken as a possible warning of a lack of capital in small portfolios. The assumption of a homogeneous portfolio might not be the case for small portfolios in practice and further research on the impact of estimating the common factor is needed.

Finally the VaR is in practice not estimated with data only on the mean of the historical default rates (and the common factor). Instead mathematical models are used in combination with expert judgments. Additionally conservatism is included to account for uncertainties and includes parameter uncertainty, so that the calculated additional capital must be set off against the conservatism. Further research can be conducted by looking at these mitigating effects and possibly in combination with the mathematical models used to estimate the PD in practice.

9 REFERENCES

[1] Bank for International Settlements, BIS website, http://www.bis.org

[2] The Dutch Central Bank (2010), De Solvabiliteitsratio van het Nederlandse Bankwezen, DNB Statistisch Bulletin, september 2010.

[3] Charles Goodhart (2011), The Basel Committee on Banking Supervision: History of the Early Years first edition, Cambridge University Press.

[4] G.G. Kaufman (1992), Banking Structures in Major Countries, Loyola Unversity of Chicago, Chicago.

[5] J.D. Becker (1976), International Insolvency: The Case of Herstatt, American Bar Association Journal page 1290-1295, oktober 1976, volume 62.

[6] Basel Committee on Banking Supervision (1988), Basel I: International Convergence of Capital Measurement and Capital Standards, July 1988.

[7] Basel Committee on Banking Supervision (2006), Basel II: International Convergence of Capital Measurement and Capital Standards, June 2006.

[8] Basel Committee on Banking Supervision (2005), An Explanatory Note on the Basel II IRB Risk Weight Functions, July 2005.

[9] Basel Committee on Banking Supervision (2011), Progress report on Basel III implementation, October 2009.

[10] Basel Committee on Banking Supervision (2010), Group of Governors and Heads of Supervision announces higher global minimum capital standards, Press release September 2010.

[11] A.J. McNeil, R. Frey and P. Embrechts, Quantitative Risk Management: concepts, techniques and tools, Princeton University Press, 2005.

[12] O. Vasicek (2002), Loan Portfolio Value, Risk December 2002.

[13] S. Ross, A First Course in Probability, Pearson International Edition, 2006.

[14] N. A. Tarashev (2010), Measuring portfolio credit risk correctly: why parameter uncertainty matters, Journal of Banking & Finance(34) 2010 2065-2076.

[15] Standard & Poor's (1999), Ratings performance 1998: Stability and Transition.

[16] E. Altman and V. Kishore (1996), Almost everything you wanted to know about recoveries on defaulted bonds, Financial Analysts Journal 52, 57-64.

[17] G. Löffler, The Effects of Estimation error on Measures of Portfolio Credit Risk, Journal of Banking & Finance 27 (2003) 1427-1453, University of Frankfurt.

[18] N.A. Tarashev and H. Zhu (2008), Specification and Calibration Errors in Measures of Portfolio Credit Risk: The Case of the ASRF Model, International Journal of Central Banking June 2008.

[19] M.B. Gordy (2003), A Risk-Factor Foundation for Ratings-Based Bank Capital Rules, Journal of Financial Intermediation Volume 12 Issue 3 pages 199-232, July 2003.

[20] A. Korteweg and N. Polson (2010), Corporate Credit Spreads under Parameter Uncertainty, Working paper, July 2010.

[21] Heitfield (2009), Parameter uncertainty and the credit risk of collateralized debt obligations, Division of Research and Statistics Federal Reserve Board Washington, January 2009.

[22] A. van der Vaart et al. (2010), Dictaat Algemene Statistiek Chapter 4.5, Vrije Universiteit Amsterdam.

[23] W. Bolstad, Introduction to Bayesian Statistics, John Wiley & Sons, 2007.
[24] J.H. Albert and S. Chib, Bayesian Analysis of Binary and Polychotomous Response Data, Journal of the American Statistical Association, Vol. 88, No. 422, pages 669-679, June 1993.
[25] G. Casella and E.I. George (1992), Explaining the Gibbs Sampler, The American Statistician, Vol. 46, No. 3 pages 167-174, August 1992.
[26] W.H. Greene (2008), *Econometric Analysis Sixth edition*, New York University.
[27] C. P. Robert and G. Casella, Introducing Monte Carlo Methods with R, Springer, 2010.
[28] Centraal Bureau voor de Statistiek (CBS), website: http://www.cbs.nl