UNIVERSITY OF TURKU
Turku School of Economics

The Vasicek Credit Risk Model and CDS Default Probabilities in Portfolio Loss Distribution Analysis

An empirical study from CDS market of the United States

Bachelor's Thesis in
Accounting and Finance

Author:
Tuomas Kerkkonen

Supervisor:
D.Sc. Antti Miihkinen

8.4.2024
Turku

Bachelor's thesis

Subject: Accounting and Finance
Author(s): Tuomas Kerkkonen
Title: The Vasicek Credit Risk Model and CDS Default Probabilities in Portfolio Loss Distribution Analysis
Supervisor(s): D.Sc. Antti Miihkinen
Number of pages: 33
Date: 26.3.2024

This thesis studies the changes in a loan or a bond portfolio by quantifying the portfolio cumulative distribution function by using the Vasicek credit risk model and Credit Default Swaps and their Probabilities of Default (PD) as a proxy. The Vasicek credit risk model is a well-known risk model introduced by Oldřich Alfons Vašíček. It can be used to model the portfolio cumulative distribution while holding certain assumptions. The model has solidified its place in credit risk modelling and in regulatory frameworks. The Credit Default Swaps are credit derivatives used in transferring risk.

The empirical study is conducted using the Vasicek model's cumulative distribution function equation and data from the Credit Default Swap markets of the United States. Due to the lack of sufficient bond market default data, the CDS contracts and their default probabilities are used as a proxy for the portfolio modelling. Number of CDS contracts in this study is 272 with the terms of five years. In addition to the Probability of Default, correlation coefficient between any default event and the loss threshold value are used to create different scenarios affecting the portfolio. To further emphasize the possible losses in the portfolio, the study presents the inverses of the loss distributions. These values represent the probabilities of losses exceeding the threshold value of the portfolio.

The results of this study illustrate significant differences between the different parameters used in the loss distribution. As expected, increase in the default probability in the proxy data increases the probability of the portfolio losses exceeding the threshold value. An increase in the loss threshold value means tightening requirement for the desirable losses. The results show that this causes an increase in the inverse loss distributions. The study notices that the correlation coefficient provides inconsistent results. In addition to being an extremely difficult parameter to estimate accurately, its effects on the inverse distribution do not indicate similar behaviour as the other parameters do. In some scenarios, larger correlations cause inverse probabilities to lower while in some scenarios the direction is the opposite.

Key words: Vasicek credit risk model, Credit Default Swap, Probability of Default, credit risk modelling, Cumulative Distribution Function

Kanditutkielma

Oppiaine: Laskentatoimi ja rahoitus
Tekijä: Tuomas Kerkkonen
Otsikko: The Vasicek Credit Risk Model and CDS Default Probabilities in Portfolio Loss Distribution Analysis
Ohjaaja: KTT Antti Miihkinen
Sivumäärä: 33
Päivämäärä: 26.3.2024

Tutkielma tarkastelee laina- tai joukkovelkakirjaportfolion muutoksia kvantifioimalla portfolion kumulatiivista jakaumafunktiota käyttäen Vasicekin luottoriskimallia sekä luottoriskinvaihtosopimuksien maksukyvyttömyyden todennäköisyyksiä proxynä. Vasicekin luottoriskimalli on laajalti tunnettu riskienhallintamittari, jonka esitteli Oldřich Alfons Vašíček. Mallia voidaan käyttää portfolion kumulatiivisen jakaumafunktion mallintamisessa tiettyjen olettamusten ollessa voimassa. Vasicekin luottoriskimalli on kiinteyttänyt asemansa luottoriskimallinnuksessa sekä sääntelykehikoissa. Luottoriskinvaihtosopimukset ovat luottojohdannaisia, joita käytetään riskin siirtämisessä.

Empiirinen tutkimus on suoritettu käyttämällä Vasicekin luottoriskimallia sekä Yhdysvaltojen luottoriskinvaihtosopimusmarkkinoilta kerätyllä datalla. Joukkovelkakirjamarkkinoiden puutteellisen maksukyvyttömyysdatan takia tutkielmassa käytetään luottoriskivaihtosopimusten maksukyvyttömyyden todennäköisyyksiä proxynä portfolion mallinnuksessa. Luottoriskinvaihtosopimusten määrä on 272, ja ne ovat kestoajaltaan viisi vuotta. Maksukyvyttömyyden todennäköisyyden lisäksi mallinnuksessa käytetään maksukyvyttömyyksien välistä korrelaatiota sekä suotavan tappion raja-arvoa luomaan eri skenaarioita, jotka vaikuttavat portfolioon. Havainnollistaakseen portfolion mahdollisia tappioita, tutkielma esittelee kumulatiivisten jakaumien käänteisjakaumat. Nämä arvot edustavat todennäköisyyttä, jolla portfolion tappiot ylittävät suotavan tappion määrän.

Tutkielman tulokset kertovat merkittävistä eroista parametrien tuottamien arvojen välillä. Odotetusti maksukyvyttömyyden todennäköisyyden nousu aiheuttaa portfolion tappioiden todennäköisyyden kasvua. Tappion raja-arvon nousu tarkoittaa käytännössä vaatimusten tiukentumista portfolion suotavien tappioiden suhteen. Tulokset osoittavat, että tämä aiheuttaa tappioiden todennäköisyyden kasvua. Tutkimuksessa huomataan, että maksukyvyttömyyksien välinen korrelaatiokerroin tuottaa epätasaisia tuloksia. Ollessaan äärimmäisen vaikea tarkasti arvioitava parametri, sen vaikutuksen portfolion tappion todennäköisyyteen eivät osoita samanlaista käyttäytymistä kuin muut parametrit. Joissakin skenaarioissa suuremmat korrelaatiot aiheuttavat pienempiä todennäköisyyksiä, kun taas toisissa skenaarioissa suunta on vastakkainen.

Avainsanat: Vasicekin luottoriskimalli, luottoriskinvaihtosopimus, maksukyvyttömyyden todennäköisyys, luottoriskimallinnus, kumulatiivinen jakaumafunktio

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

1.1 Background

Risk is often affiliated with hazards, uncertainty, or otherwise unwanted consequences. In finance, risk is usually discussed as the downside of something, usually the losses an investment might face. It is the uncertainty of the future returns of an investment. If risk wasn't present in the world of finance, most of the industry wouldn't even exist (McNeil et al. 2015), and this thesis along with numerous other studies wouldn't be needed. One can decompose risk into different factors, such as credit, operational, liquidity, and market risk. One can also reduce risks by aggregation, which equals being well diversified (Hull 2018). Risk is closely associated with randomness, which means that the outcomes of some entity are not specifically known. Even though financial markets are continuously analysed, it is not possible to perfectly predict fluctuations. This randomness can be observed through interest rate movements, stock prices and commodity prices, to name a few.

According to the fundamentals of financial theory, investors want to optimize their risk taking to the expected returns at this predetermined level of risk. Basically, this means that the investor, an institution, or an individual, wants to reach the maximum expected returns at the risk level they are willing to take. Therefore, risk can be considered as the price to be taken to reach higher returns. For financial institutions, risk management is one of the key cornerstones for the success of their businesses. According to past studies and general opinion, investing in a well-functioning risk management function will help to improve firm performance (Christoffersen 2011).

Credit risk is a risk which is omnipresent for any financial institution. It is the counterparty risk in crediting business, which means the risk that the debtor will not be able to fulfil their obligations (McNeil et al. 2015). Being a centrepiece in financial risk management, academic studies of credit risk management and its different measures, instruments and factors are numerous. Credit risk modelling has its roots in 1970s in the form of option-pricing techniques and the studies of corporate liabilities. The goal of credit risk modelling is to find a link between some statistical model describing the default events and an economic-pricing model. Credit risk modelling aims to create distributions of these default events, and thereafter deduce prices from these models (Black & Scholes 1973; Lando 2009; Merton 1973).

1.2 Motivation and research

After the Financial Crisis in 2007-2009, risk management has taken a significant step in importance and many argued for this function to take a bigger role in financial institutions (Miller 2018). One of the most important applications of the Vasicek credit risk model is its role in regulatory framework. The model is included in the Basel and Solvency regulatory frameworks, and its overall role in the field of credit risk management is undisputable (Hull 2018; Vasicek 2002).

The purpose of this thesis is to examine the portfolio cumulative distribution function and inverse distribution risk behaviour by using the Vasicek credit risk model. The main research question is: *How do different scenarios affect cumulative portfolio loss distribution functions and inverse distributions?* The answer to this question is sought by utilizing Credit Default Swap data and creating a scenario analysis by changing the Vasicek equation parameters, which are the Probability of Default (PD), correlation coefficient and the loss threshold value. To be precise, the study presents annual values for the loss distributions and inverse distributions.

The results of this study provide insight into the loss probability behaviour, which can be useful in the context of portfolio management and credit risk management in general. The Vasicek credit risk model analysis and the implementation of Credit Default Swaps give information about the portfolio loss behaviour in a simplified but useful way. While being simplistic, this approach can mimic theoretical situations that portfolio managers could face. The results indicate clear trends for each parameter change. While the increase in probability of default and the loss threshold value indicates clear growth in riskiness, correlation coefficient provides more ambiguous results.

1.3 Structure

This thesis continues with the second chapter, which discusses the theoretical background needed to understand the models and methods used in this study. The second chapter includes the basics of risk management, in-depth analysis of credit risk management, Value at Risk, and the introduction of the Vasicek credit risk model. The Probability of Default (PD) and the tail risk are introduced to broaden the view and create depth. The third chapter is for the introduction of the data and how the actual research is conducted. The chapter further specifies how the modelling is done. The goal is to get the reader to understand the methods of this study, and to familiarize themselves with the Credit Default Swap proxy data. Mathematical background used in the empirical segment is also specified. The fourth chapter introduces the results of this study and discusses them. Finally, the fifth and last chapter concludes this study. The final chapter summarizes the topic and discusses if there is a need or possibility for further research.

2 Theoretical background

2.1 Basic concepts of risks and risk management

Businesses in general, regardless of their industry, will face certain risks related to their lines of business and operations. For financial institutions, this means that they will face risks related closely to their own business models, which are called financial risks. These are the types of institutions whose business models by and large consist of lending, crediting, and a broad variety of different investing activities. Financial risk management is a vast field with diverse and changing components (Andersen et al. 2013). Therefore, it can be considered an important factor in the global economy. In the context of financial risks, it can be perceived to include market risk, credit risk, and operational risk (Jorion 2007). This thesis further examines credit risk using the Vasicek credit risk model and the portfolio cumulative loss distribution.

Risk management is a beneficial component to possess for any company. Understanding the risks your company might face can possibly prevent significant losses from realizing. Utilizing risk management accordingly will help an organization to enhance its effectiveness in its operations (Hopkin 2018). When speaking of losses, risk management is usually associated with potential losses, not profits. Since risk management is an essential part of companies' functionality, development of various financial risk evaluation models has been created and put into further use. Some of those models are further explained later in this thesis.

In addition to the benefits of using financial risk models, different regulatory frameworks require financial institutions to implement these models in their risk management processes. When reaching a certain size and significance to the whole financial markets, large financial institutions are sometimes considered too big to fail by the regulators and other stakeholders (Goddard et al. 2009). This can lead to guarantees for these institutions granted by governments and central banks. In the event of a crisis, this could be viewed as justified and even necessary. However, the limited level of downside or risk of losses might allow financial institutions to benefit from risk taking and therefore contribute to the possibility of a new crisis. Solution to this hazard is often sought through regulation (Poutanen 2017).

2.2 Credit risk

Credit risk is the risk of a financial loss showing up from the failure of a counterparty to fulfil its contractual obligations (McNeil et al. 2015). It traditionally signifies the risk that a lender may not receive owed principal and interest, leading to an interruption of cash flows. Credit risk is also often referred to as default risk, performance risk or counterparty risk (Brown & Moles 2014). For entities that are engaged in lending business, for example banks, credit risk is continuously present and requires constant evaluation. As McNeil et al. (2015) pointed out, credit risk management is also present for insurance companies, which are exposed to significant credit risk through their investment portfolios and counterparty risk in their reinsurance treaties. For further context, credit risk management can be interpreted as a principal factor in the Finnish financial sector due to the heavy presence of bank financing, current pension insurance system and other insurance companies. An important characteristic of credit risk management is its time frame compared to other risk management subareas. Since credit risk is related to bonds, loans or similar types of instruments, the period is expressed in their maturities or terms which are often years. In contrast, market risk is more often glanced in shorter time periods, for example daily (Coleman 2012).

Managing credit risk can be strongly related to other basic methods and theories known in finance. Beginning from portfolio diversification, the investors can restrict the risks they face when making an investment with the cost of lower expected returns. Ideally, an investor's portfolio will try to diversify away all idiosyncratic risk, which is the risk which is specific to some asset or group of assets and therefore diversifiable, leaving us only with systematic, non-diversifiable risk (Wilson 1998). The systematic risk or the market risk is the risk carried in the whole market or a market segment and cannot be reduced by diversification.

Collateralization is a method used by the creditor against the risk of default of the party taking credit (Brown & Moles 2014). Use of collateral basically means that the party taking the credit will provide a security against default. This means that the lender has a claim on some predetermined asset of the borrower. This asset can be for example real estate or different securities, which hold some financial value in the case of borrower defaulting. In addition, using collateral helps reduce the problems which arise in the form of moral hazards. It helps in aligning the interests of both parties, lenders and borrowers, for example in a situation where the borrower might be careless or negligent regarding their obligations (Jiménez & Saurina 2004). Being closely related to moral hazards, the problem of asymmetric information between the borrower and lender can be cut down with collateralization.

In credit risk management, credit derivatives are contracts traded in financial markets, which are used to transfer risk from the party wanting to lessen its risk to the party willing to carry the risk (Brown & Moles 2014). Examples of different credit derivatives include Credit Default Swaps (CDS), Collateralized Debt Obligations (CDO) and Credit Spread Options or Forwards. The contracts are traded *over-the-counter*, and the payoffs are dependent on the credit quality of a firm. This specified firm is usually not a party to the contract. The instruments underlined in the credit derivatives are often corporate bonds, bank loans or smaller loans, for example credit card receivables. Credit derivatives can be interpreted as more conveniently repackaged instruments with which risks can be traded (Duffee & Zhou 2001). Thus, they create significant flexibility in credit risk management and give alternatives for overall risk management.

Like many other subjects in the field of finance, credit risk can be quantified. Presented below is the widely used formula for expected credit loss calculation (Brown & Moles 2014)

$$Credit\, risk = Exposure \times Probability\, of\, Default \times (1 - Recovery\, rate).$$

This formula can also be presented in the form specified by Volarević & Varović (2018)

$$ECL = EAD \times LGD \times PD,$$

where ECL stands for expected credit loss, EAD or $Exposure$ equals exposure at default, which is the total sum invested and exposed to the credit risk. LGD or $(1 - Recovery\, rate)$ is the $Loss\, Given\, Default$, and it implicates the sum to be lost in case of a default. PD stands for the $Probability\, of\, Default$, which is the likelihood of a default. The PD will be further exhibited later in this thesis.

2.3 Value at Risk (VaR)

The need of uniform measure of risk lead to the development of the Value at Risk (VaR), which means the expected worst loss over some time period at some given confidence level (Jorion 1996). It is one of the most used risk measurement tools in financial institutions, if not the most used one. In addition, Value at Risk has an essential role in the Basel regulatory framework and notable influence in Solvency II (McNeil et al. 2015). The model was developed as a risk metric at J.P Morgan in the late 1980s to early 1990s and has solidified its position as the standard measure for risk measurement among finance professionals (Hull 2018). The usefulness of VaR lies in its relative simplicity, since it gives one metric which summarizes the risk of an investment portfolio (Miller 2018).

The main reason to introduce VaR in this thesis is its connection and usage in the Vasicek credit risk model, which will be comprehensively explained later. The results from the Vasicek model can be effectively used in the calculation of Value at Risk (Vasicek 2002). In simple terms, Value at Risk is a measure of how much a portfolio of investments might lose with some level of probability and period. For example, a portfolio with 95% VaR of $1 million means that it has a 5 % probability of decreasing in value by over $1 million over a one-day period, given there is no trading during that period. VaR can be expressed mathematically as follows:

$$VaR_\alpha = VaR_\alpha(L) = inf\{l \in \mathbb{R}: P(L > l) \leq 1 - \alpha\} = inf\{l \in \mathbb{R}: F_L(l) \geq \alpha\}.$$

The equation states that the Value at Risk at confidence level $\alpha \in (0,1)$ is the smallest loss *l* such that the probability of the actual loss exceeding *l* is less than or equal to the risk level $1 - \alpha$. Equivalently, it's the smallest loss *l* such that the cumulative probability up to l is greater than $\alpha \in (0,1)$. L stands for the loss at given confidence level. *Inf* or *infimum* is a term which means the greatest lower boundary of a set, in this context the smallest value for *l*. $P(L > l)$ is the probability that the loss *L* exceeds certain level *l*.

2.4 Credit Default Swaps

Credit default swaps (CDS) are a form of credit derivatives which are used in credit risk management, but also do have a somewhat controversial reputation. Their usage has become more popular over time as protection against defaults, and they are often regarded as the most popular credit derivative instruments (Hull 2017; Weistroffer 2009). The basic functionality of a credit default swap is transferring the credit exposure of fixed income products, like bonds. A simple way to understand their principle is to think of an insurance contract against the default of some company, which is called the *reference entity* (Hull 2017). As with insurance contracts, credit default swaps require periodic premium payments, which in the context of credit default swaps is the spread, which the bondholder loses if the company of which the bonds are doesn’t default. On the contrary, if the company does default, the credit default swap allows the bondholder to exchange these bonds for the principal amount of them. This is called *physical delivery,* which is the first form of settlement. *Cash settlement* or *auction* may also be used as settlement. It is important to notify that the CDS seller is a third party, and not the bond issuer.

Their controversial reputation originates from the subprime mortgage crisis in 2007–2009, where their role alongside mortgage-backed securities (MBS) and collateralized debt obligations (CDO) has been widely recognized as one of the key reasons for the crisis. However, while they do have controversy surrounding them, their role as a complementary piece of the financial markets should not be overlooked (Weistroffer 2009). The credit default swaps can be used for multiple purposes, which include hedging, speculation, or arbitrage. Being a multi-trillion dollar, global market, strengthens their essential role in the financial markets (Augustin et al. 2016). According to (Blanco et al. 2005), single-name credit default swaps account for approximately half of the entire credit derivatives market. The single-name credit default swaps will also be later used in this study as a proxy in determining default probabilities.

The value of the CDS can be calculated from the estimation of the default of the reference entity. To be exact, this is the risk-neutral probability and does not equal the real-world probability. The value of the CDS also depends on the contract details, and in this case the amount to be paid back could be for example principal amount minus the bond’s current value (Stulz 2010). The probabilities of default used to value Credit Default Swaps should be risk-neutral probabilities and not real-world ones. As an alternative, the probabilities can be implied from CDS quotes (Hull 2017).

2.5 Probability of Default (PD)

The probability of default is the likelihood of the borrower or some asset defaulting. In practice, this means the probability that the borrower is unable to pay their loan repayments to the lender. In this situation, the asset’s value goes to zero and it will not yield any return to its holder (CFI Team 2024). Under Basel regulatory framework the probability of default is an important parameter, relating to the 2008 Financial crisis. (Volarević & Varović 2018).

The parameter plays a significant role in risk management and asset pricing. The probability of default is often used by ratings agencies, like Moody’s and Standard & Poor’s, as a metric to illustrate riskiness of some asset or a company to the investors. In addition, the PD is also linked to corporate bond credit spread and loan rate measurements. The PD is also a core risk parameter used in stress testing and comprehensive capital analysis and review (CCAR) (Li et al. 2023). It is present in Credit Default Swaps to imply the default probability or the credit risk of some company or other entity. However, their complete unbiasedness as a PD proxy should be treated with caution because they might not provide sufficient analysis of the impact of risk premium, counterparty risk or market frictions (Jarrow 2012).

Four basic factors have been identified affecting to the corporate probability of default: (i) size of the company; (ii) measures of financial structure; (iii) measures of performance; (iv) measures of current liquidity (Ohlson 1980). These measures can be interpreted and connected relatively easily to how they might affect the financial success of a company.

Similarly to other financial metrics, the probability of default is affected by macroeconomic factors. In previous studies it was concluded that Gross Domestic Product (GDP) and the Repo rate indeed affected to the probability of default of a portfolio consisting of retail loans (Antonsson 2018).

2.6 Tail risk

The tail risk refers to the risk of fluctuations in the ends of the probability distributions of returns. These can and usually are outside of the normal range of outcomes. The probability of the tail risk realizing is small, but often the losses in the tail are the most catastrophic ones (Brown & Moles 2014). Such events might occur during times of financial crises, natural disasters, or major geopolitical events such as war. Although it has a small probability of occurring, it is important for financial institutions and portfolio managers to recognize and counter this risk with their best efforts.

In credit risk management, portfolio managers are often interested in specifically the tail risk. Often the question in risk management is, what is the probability of loss, and not what the probability of profits is. Thus, the tail risk, or the worst-case scenario, is what is under observation. The tail risk is under observation in the common metrics of risk management like the conditional Value at Risk (cVaR) or Expected Shortfall (ES) (Nakagawa & Ito 2021).

2.7 The Vasicek credit risk model

2.7.1 Introduction to the model

The Vasicek credit risk model is a financial risk model originating from 1987 and again published in 2002. Introduced by Oldrich Alfons Vasicek, it is a central model used in quantifying credit risk in portfolios. At the core of its function, the model stresses that the amount of capital needed to support a portfolio of different debt securities is dependent on the probability distribution of the portfolio loss (Vasicek 2002). The probability distribution of portfolio losses, alongside the capital required supporting the portfolio, has numerous other applications. It is especially useful in calculating other risk measures, for example Value at Risk (VaR), portfolio optimization and structuring and pricing debt portfolio derivatives (Schönbucher 2003). In addition, it is used in regulatory reporting, particularly in its application in Basel frameworks. The model has also its uses in other areas of modern finance, such as in machine learning, which can be applied in calibrating the outputs of the Vasicek model (García-Céspedes & Moreno 2022).

Having a portfolio which is partly financed by equity capital and partly by borrowed funds, the credit quality of the notes of the lender depends on the probability that the portfolio loss will *exceed* the amount of equity capital. Simply put, in this situation the lender will not be able to cover its portfolio losses with equity capital. As Vasicek points out, to keep or reach the credit quality of the notes at a predetermined level, the lender needs to limit the probability of default (PD) at a level needed to qualify for this credit rating (For example Baa on rating agency scale). In other words, the equity capital in the portfolio must be equal to the percentile of the distribution of the portfolio losses that match the sought-after probability (Vasicek 2002).

The Vasicek credit risk model has certain assumptions, the first of these being that all loans included in the portfolio have the same maturities. The second assumption is that the correlations between all the loans, to be precise between any two loans, are the same. The third assumption is that the probability of default is the same for all the loans. Finally, it is assumed that the portfolio consisting of n loans are in equal amounts of some currency (Vasicek 2002).

The assumptions create restrictions in the model's interpretation. In the real world, these assumptions will not be fully satisfied. To begin with, the default correlations are extremely difficult or even impossible to estimate accurately. This is due to the fact that there are numerous different factors affecting the default events, which are characteristic to their industries, regions, credit qualities and other factors (Nagpal & Bahar 2001). Additionally, usually all the portfolio's assets will not have the same maturities, and the portfolio will not consist of equal amounts of investments to each asset.

2.7.2 Mathematical discourse: *The limiting distribution of portfolio losses*

The theoretical background of the model begins by observing asset value. Having a maturity of T, if the borrower's assets fall below the contractual value of B, the loan will default. A_i being the asset value of the i -th borrower, the process is

$$dA_i = \mu_i A_i dt + \sigma_i A_i dx_i.$$

This can be represented as, at some maturity T

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

Therefore, the probability of default on i -th loan is

$$p_i = P[A_i(T) < B_i] = P[X_i < c_i] = N(c_i),$$

where the parameter c_i is

$$c_i = \frac{logB_i - logA_i - \mu_i T + \frac{1}{2}\sigma_i^2 T}{\sigma_i \sqrt{T}}.$$

The probability of portfolio loss can be evaluated as the expectation over a common factor Y of the conditional probability given Y. Basically, the following equation calculates the conditional probability of a single loan given the realization of this common factor Y. The loss on one specified loan is given by

$$p(Y) = P[L = 1|Y] = N\left(\frac{N^{-1}(p) - Y\sqrt{\rho}}{\sqrt{1-\rho}}\right),$$

where $p(Y)$ is the probability of default (PD) under some given scenario. Here, N^{-1} is the inverse of cumulative distribution function of a standard normal distribution, ρ is the asset default event correlation and p is the probability of default (PD) averaging on the conditional probabilities over some scenarios (Vasicek 2002).

As continuation, Vasicek introduces a loss distribution formula for larger portfolios. Simply put, this formula will output the probability that the portfolio loss will not exceed some predetermined amount x. Later in this thesis, this value will be called the threshold value. The formula for the representation of loss distribution in a large portfolio when given certain parameters and conditions is

$$F(x;p,\rho) = P[L \leq x] = N\left(\frac{\sqrt{1-\rho}\,N^{-1}(x) - N^{-1}(p)}{\sqrt{\rho}}\right),$$

where $F(x;p,\rho) = P[L \leq x]$ calculates that the probability of loss L does not exceed a certain value x as a function of x, p and ρ. Here, N^{-1} is the inverse of cumulative distribution function of a standard normal distribution, ρ is the asset default event correlation and p is the probability of default (PD). The distribution is continuous and is concentrated between interval $0 \leq x \leq 1$ and forms two-parameter family with $0 < p, \rho < 1$, which are the default probability and correlation, respectively. This formula will be later used in this thesis when determining the portfolio losses under certain conditions (Vasicek 2002).

The inverse of the distribution function specified before allows to assess the probability of the threshold-exceeding losses of the portfolio at given parameters. The inverse is given by

$$L_\alpha = F(\alpha; 1-p, 1-\rho) = 1 - P[L \leq x].$$

Here, the function represents the α-percentile value of L, which stands for loss (Vasicek 2002).

3 Research data and design

3.1 CDS Data

The proxy data used in this thesis is gathered from LSEG Refinitiv Workspace. The data consists of Credit Default Swaps (CDS) from the market region of the United States. The reason for using Credit Default Swaps as a proxy in this thesis is that the default data available from bonds or loans was not sufficient to make a satisfactory study. The terms of the Credit Default Swaps in this data are 5 years. The data includes the CDS' issuer names, tiers, terms, probabilities of default, bid prices and ask prices. The primary focus in this study is on the Probabilities of Default (PD) through the instrument's term, which were available in the data stream.

From the data, the most extreme values of Probability of Default have been sorted away for the sake of consistency and relevance. However, as Table 1 shows, the maximum values of the PD data are still high in the context of corporate loans or bonds. The most extreme values in Probabilities of Default imply that the reference entity is close to defaulting. In the data, it is worth noting that larger, more prominent public corporations possess much lower default probabilities in comparison to other, smaller companies.

Table 1: Descriptive statistics of the Credit Default Swap PD data

The table presents the descriptive statistics of the entire CDS data Probabilities of Default gathered from Refinitiv Workspace.

	Mean	Min	Median	Max	Kurtosis	Skewness	St.Dev	N
CDS	7,77%	1,01%	5,07%	37,84%	3,9702	2,0098	0,0043	272

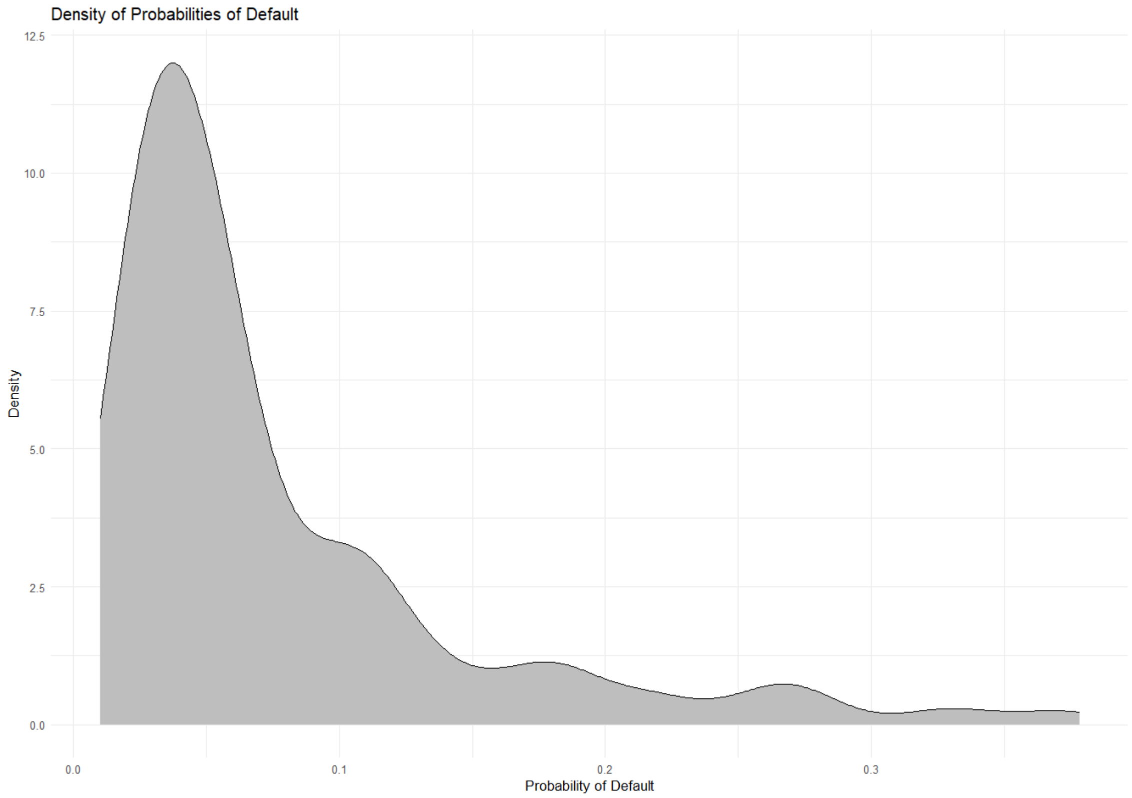

Figure 1: Distribution of default probabilities

The figure visualizes the distribution of default probabilities within the data. The density is represented on the y- axis, and the Probability of Default values are on the x- axis. As the distribution shows, most default probabilities lie in the 0.0-0.1 range, which causes the skewed distribution.

3.1.1 Probability of Default annualization

To calculate the one-year cumulative loss distribution using CDS proxy data and the Vasicek model, it is mandatory to estimate the one-year PD. The five-year Probabilities of Default have been converted to annual PD's by utilizing the Hazard Rate by utilizing formulas specified by Hull (2017). The annual PD's can be calculated from the term of the CDS and the Probability of Default

$$Hazard\ Rate\ \lambda = -\frac{\ln(1 - PD_t)}{t},$$

The annualized Probability of Default can be calculated when the Hazard Rate λ is known, and is

$$PD_t = 1 - e^{-\lambda t}.$$

As shown below, the Probability of Survival is complementary to the Probability of Default. The survival probability is

$$S_t = e^{-\lambda t}.$$

In the data, the annualized Probability of Default and Probability of Survival values and their complementary nature are verified by calculating their sum, which should equal as 1,

$$PD_t + S_t = 1.$$

Using the formulas presented above and eventually ensuring that the results are indeed calculated properly, it is possible to proceed to use these results in scenario analysis.

Table 2: Descriptive statistics of the annualized Credit Default Swap PD data

The table presents the descriptive statistics of the annualized CDS data Probabilities of Default.

	Mean	Min	Median	Max	Kurtosis	Skewness	St.Dev	N
CDS	1,72%	0,20%	1,04%	9,07%	4,0362	2,0512	0,001	272

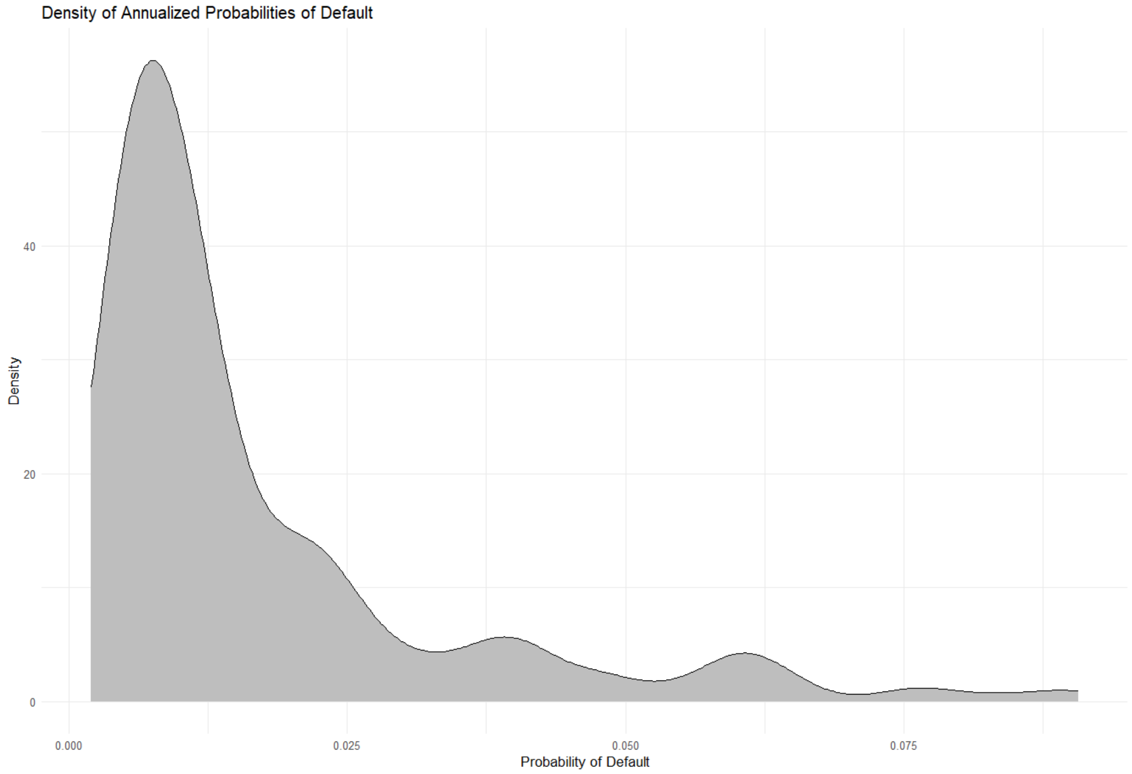

Figure 2: Distribution of annualized default probabilities

The figure visualizes the distribution of annualized default probabilities within the data. The density is represented on the y- axis, and the Probability of Default values are on the x- axis. As the distribution shows, most default probabilities lie in the 0.000-0.025 range, which causes the skewed distribution.

An important factor to notice is the positive skewness of the Probabilities of Default. This is visualized as a long right tail of the distributions. In this study, the annualized default probabilities have been sorted into quantiles for easier evaluation of its effects in the loss distribution. As the Tables and Figures above present, most of the PD data centres in the lower values, but in the skinny right tail of the distribution function the largest PD values are relatively high. This can also be notified as a large jump in Figure 3 below, which represents the data quantiles.

Table 3: Percentage quantiles for the Probability of Default data

The table presents the percentage quantiles which are used in this study for determining the portfolio loss distribution. Percentiles are acquired by using Excel Data Analysis tool.

	Min	25th	Median	75th	Max
Percent	0,20%	0,62%	1,04%	2,16%	9,07%

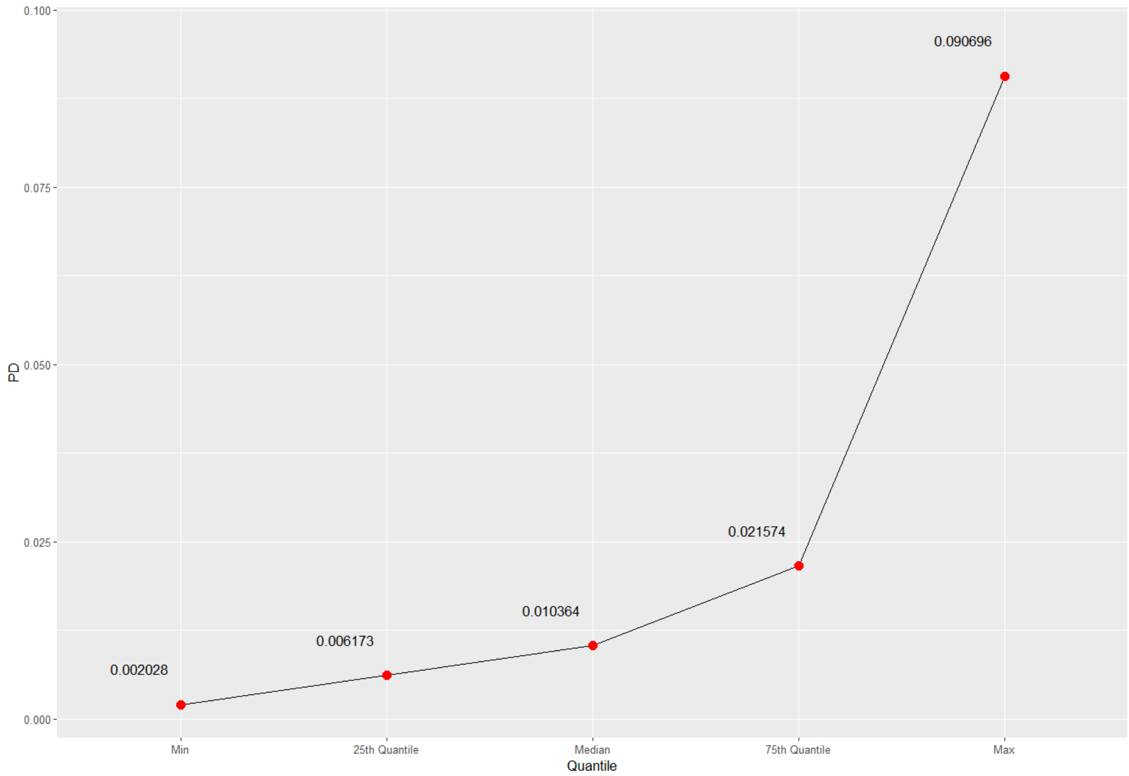

Figure 3: Probability of Default quantile values

The figure visualizes the quantile values in the data, which are marked as red dots. The vertical axis represents the PD value, and the horizontal axis represents the quantiles.

3.1.2 Determining the correlation coefficient

The default correlation evaluation is an important measure in credit risk and risk management, but it is never an easy task since the default correlations cannot be measured directly (Zhou 1997). However, historic studies indicate that default events are indeed correlated (Nagpal & Bahar 2001). The problematic nature of incorporating flawless correlation measure stems from multiple different reasons, which include lack of sufficient data, region, industry, and time horizon. Therefore, in this thesis, the scenario analysis and its correlation coefficients are set from 0.1 to 0.5. Simply put, the analysis examines how different levels of correlation affect the loss distribution within the portfolio.

3.1.3 Determining the threshold value x for the portfolio loss distribution

The parameter x represents the proportion of the maximum potential loss we are considering for the portfolio. It allows us to evaluate the portfolio losses at different levels. For example, when choosing $x = 0.05$, we are assessing the probability of experiencing portfolio losses that are 5% *or less* of the maximum potential loss. Commonly in financial risk management, for example in Value at Risk calculations, the x threshold is set to 95%, 99% and 99.9%, which correspond as x values 0.05, 0.01 and 0.001% (Luenberger 2014). Thus, these values are used in this thesis as thresholds.

3.2 Research methods

This study aims to model the changes in the cumulative portfolio loss distribution by scenario analysis, or stress testing. The analysis is conducted by changing the previously presented variables Probability of Default p, Correlation coefficient ρ and threshold loss value x. The variables are inserted into R Studio and Excel Data Table, and the end results are converted into a data frame for easier evaluation. Ultimately, the objective is to illustrate the effects that the variable changes have in the portfolio loss distribution. The formulas used in this study are the following, as already presented by Vasicek and in chapter 2.7.3

$$F(x; p, \rho) = P[L \leq x] = N\left(\frac{\sqrt{1-\rho}\, N^{-1}(x) - N^{-1}(p)}{\sqrt{\rho}}\right),$$

which calculates the cumulative loss distribution of a portfolio as a function of the variables presented previously (Vasicek 2002). Having calculated the cumulative loss distributions in different scenarios, this thesis covers the portfolio inverse loss distributions, which is the inverse of the distribution measured above. The second function used is the inverse distribution, which outputs the α-percentile value of L, is

$$L_\alpha = F(\alpha; 1 - p, 1 - \rho) = 1 - P[L \leq x].$$

By assessing the portfolio inverse loss distribution, it is easier to understand the probability of losing *more* than the predetermined x, and therefore it is presented in this thesis. The results are also visualized to enhance interpretation.

4 Research results

4.1 Scenario analysis

This chapter will go through the test results obtained from the scenario analysis or stress testing. Prior to analysing the results, it should be noted that the assumptions made in this study are relatively simplistic and might not fully mirror the events occurring in the real world. However, the results do indicate the effects each parameter has in the loss distributions and inverse loss distributions, and therefore provide interesting insight to the model's behaviour.

The results are presented as a tables and figures below and are sorted according to the parameters. Calculations were conducted with both R Studio and Excel. Visual representations were conducted with R Studio using the *ggplot2* package.

Table 4: Portfolio cumulative loss distribution values

The table reports the cumulative loss distribution values obtained from R Studio and Excel Data Table sensitivity analysis, in which the portfolio loss distributions were calculated by changing the variables. The results are sorted according to PD Quantile, correlation ρ and loss threshold value x.

		x = 0,05				
		ρ = 0,1	ρ = 0,2	ρ = 0,3	ρ = 0,4	ρ = 0,5
Min	p = 0,202821%	99,998359%	99,914416%	99,687295%	99,428478%	99,222328%
25th Quantile	p = 0,617274%	99,854844%	98,942192%	98,009215%	97,391092%	97,086543%
Median	p = 1,036359%	99,133254%	97,009102%	95,638874%	94,975858%	94,803638%
75th Quantile	p = 2,157426%	92,791222%	89,106745%	88,091768%	88,158806%	88,782980%
Max	p = 9,069582%	23,940049%	38,161510%	47,110857%	53,928724%	59,685616%

		x = 0,01				
		ρ = 0,1	ρ = 0,2	ρ = 0,3	ρ = 0,4	ρ = 0,5
Min	p = 0,202821%	98,250706%	96,190083%	95,478685%	95,492434%	95,887123%
25th Quantile	p = 0,617274%	82,467531%	82,695426%	84,486514%	86,585399%	88,727699%
Median	p = 1,036359%	63,120370%	69,817002%	74,832716%	79,041461%	82,757128%
75th Quantile	p = 2,157426%	27,958798%	44,798114%	55,511175%	63,619489%	70,318154%
Max	p = 9,069582%	0,295517%	4,803322%	13,275049%	23,085976%	33,131814%

		x = 0,001				
		ρ = 0,1	ρ = 0,2	ρ = 0,3	ρ = 0,4	ρ = 0,5
Min	p = 0,202821%	42,734739%	59,693224%	70,066027%	77,608464%	83,493460%
25th Quantile	p = 0,617274%	8,718063%	27,908184%	43,951597%	56,806144%	67,302766%
Median	p = 1,036359%	2,519666%	15,657742%	30,937669%	44,919208%	57,170895%
75th Quantile	p = 2,157426%	0,201554%	4,860502%	15,191243%	27,851545%	40,892493%
Max	p = 9,069582%	0,000023%	0,070646%	1,129389%	4,730340%	11,503846%

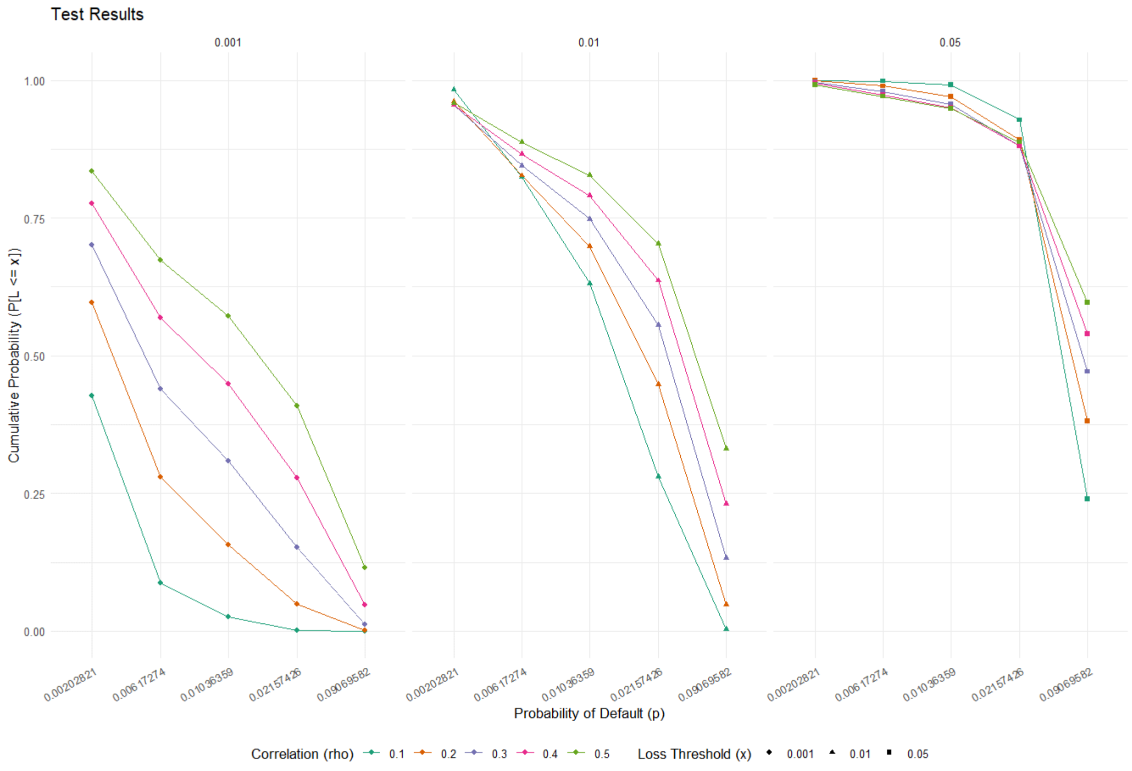

Figure 4: Visual presentation of the portfolio cumulative loss distribution values

The figure visualizes the results. The graphs are sorted by the loss threshold x. Inside the graphs, different correlations are provided with different colours. The y -axis represents the cumulative probability distribution, and the x -axis represents the different Probabilities of Default.

Table 5: Portfolio distribution inverse values

The table reports the cumulative loss distribution values obtained from R Studio and Excel Data Table sensitivity analysis, in which the portfolio loss distributions were calculated by changing the variables. The results are sorted according to PD Quantile, correlation ρ and loss threshold value x. The inverse loss distribution captures the probability that the portfolio value will exceed the given loss threshold value x.

	x = 0,05					
		ρ = 0,1	ρ = 0,2	ρ = 0,3	ρ = 0,4	ρ = 0,5
Min	p = 0,202821%	0,001641%	0,085584%	0,312705%	0,571522%	0,777672%
25th Quantile	p = 0,617274%	0,145156%	1,057808%	1,990785%	2,608908%	2,913457%
Median	p = 1,036359%	0,866746%	2,990898%	4,361126%	5,024142%	5,196362%
75th Quantile	p = 2,157426%	7,208778%	10,893255%	11,908232%	11,841194%	11,217020%
Max	p = 9,069582%	76,059951%	61,838490%	52,889143%	46,071276%	40,314384%
	x = 0,01					
		ρ = 0,1	ρ = 0,2	ρ = 0,3	ρ = 0,4	ρ = 0,5
Min	p = 0,202821%	1,749294%	3,809917%	4,521315%	4,507566%	4,112877%
25th Quantile	p = 0,617274%	17,532469%	17,304574%	15,513486%	13,414601%	11,272301%
Median	p = 1,036359%	36,879630%	30,182998%	25,167284%	20,958539%	17,242872%
75th Quantile	p = 2,157426%	72,041202%	55,201886%	44,488825%	36,380511%	29,681846%
Max	p = 9,069582%	99,704483%	95,196678%	86,724951%	76,914024%	66,868186%
	x = 0,001					
		ρ = 0,1	ρ = 0,2	ρ = 0,3	ρ = 0,4	ρ = 0,5
Min	p = 0,202821%	57,265261%	40,306776%	29,933973%	22,391536%	16,506540%
25th Quantile	p = 0,617274%	91,281937%	72,091816%	56,048403%	43,193856%	32,697234%
Median	p = 1,036359%	97,480334%	84,342258%	69,062331%	55,080792%	42,829105%
75th Quantile	p = 2,157426%	99,798446%	95,139498%	84,808757%	72,148455%	59,107507%
Max	p = 9,069582%	99,999977%	99,929354%	98,870611%	95,269660%	88,496154%

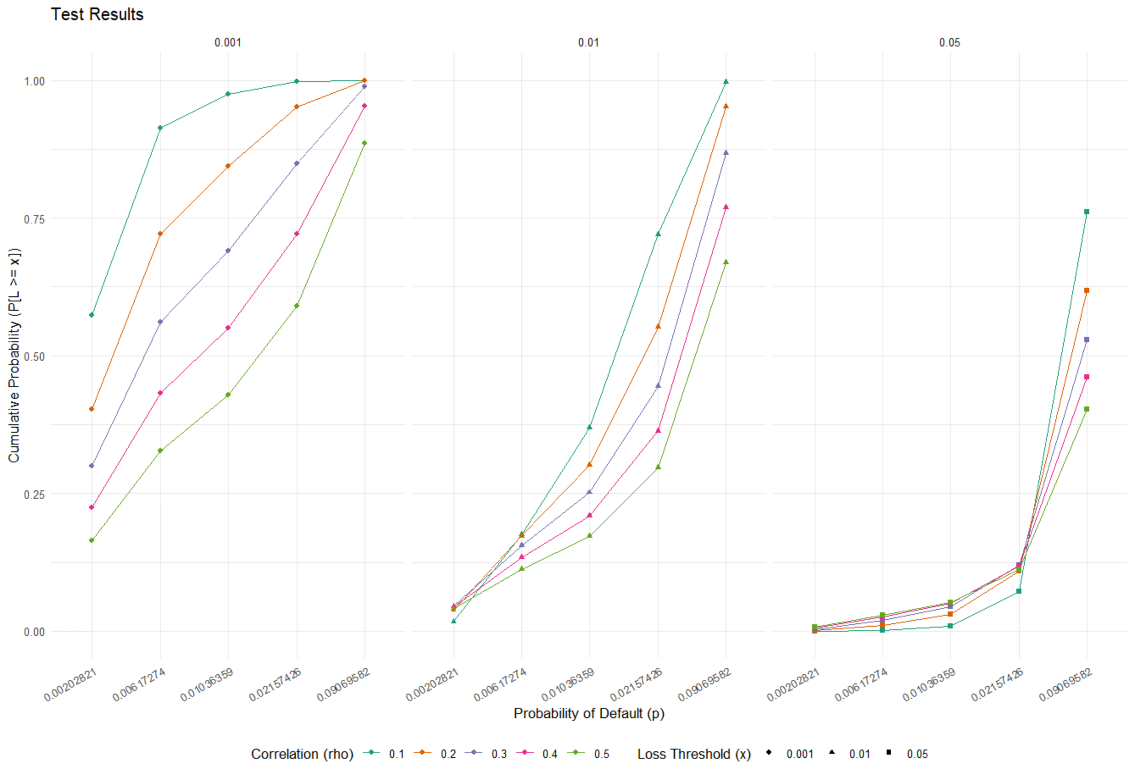

Figure 5: Visual presentation of the portfolio inverse distribution values

The Figure visualizes the portfolio inverse loss distribution. The graphs are sorted by the loss threshold x. Inside the graphs, different correlations are provided with different colours. The y- axis represents the cumulative probability distribution, and the x- axis represents the different Probabilities of Default.

4.2 Cumulative loss distribution and inverse loss distribution assessment

The inverse distributions presented in Table 5 and Figure 5 draw the same conclusions as the cumulative loss distributions did in Table 4 and Figure 4. The reason for presenting this information is that the inverse values help draw conclusions of the possible portfolio losses. Simply put, this helps in understanding the *probability of threshold-exceeding losses*.

In both methods, the characteristics and behaviour of the model behaviour are clearly visible. The effects each parameter causes differ in both magnitude and shape. However, the parameter changes generally do create visible trends, according to the figures presented above, and thus interpretations in each scenario can be made.

The distances between the extreme values in the charts are noticeably big. To make logical conclusions of these results, it should be noted that their significance lies rather in the step-by-step, relative movements than in their extreme values. From the perspective of loan portfolio management, it would be more feasible to assess the differences between adjacent quantiles of the default probabilities rather than the minimum and maximum values.

4.3 Model sensitivity to the parameters

4.3.1 Loss threshold value x

As Figure 4 presents, when the threshold loss value decreases, the cumulative probability $P[L < x]$ decreases simultaneously. At the same time, the inverse distribution values increase. When the threshold value does decrease, this can be interpreted as tightening or more strict demands considering the losses we don't want to face. Therefore, it is logical that the probability that our losses will *not* exceed the x will decrease, as the x itself decreases. As shown in the figure above, the model can be considered sensitive to the changes in x, although it should be notified that the x values do not change proportionally. Since the test results have significant differences between x values, it can be considered as a decisive factor for portfolio management decisions.

4.3.2 Probability of default p

Sensitivity to the changes in Probability of Default is relatively difficult to express since the graphs are of different shapes for each x. The main trend is that the distribution function value decreases as the CDS Probability of Default p increases. Thus, the inverse distribution values increase. This phenomenon relates to the fact that when individual PD's increase, the overall risk of loss in the portfolio increases. This is important for portfolio managers who will evaluate their overall portfolio riskiness and individual loan or bond risks. In most cases, the jump from 75th quantile to the maximum value in the loss distribution is significantly larger than other jumps between quantiles. The difference is explained by the biggest difference in PD values, which is from the 75th quantile to the maximum value.

4.3.3 Correlation coefficient ρ

The behaviour to changes in correlation differs from the other parameters. When implementing the smallest x value, the smaller the correlation, the smaller the distribution function value. However, when increasing x, bigger correlation values might result in bigger distribution function values. This is especially true, when $x = 0.05$. Thus, according to the results, the model’s dependence on the correlation value is not linear. However, in credit risk management non- linearity is often present, and correlations can be dynamic. Overall, the correlations between loan or bond defaults are related to many different factors and overall market conditions. For example, during turbulent times it would be logical to assume that default numbers will increase in the close proximity of each other, thus increasing the correlation between them.

5 Conclusion

This thesis examined the usage of Credit Default Swaps and their Probabilities of Default as a proxy in the Vasicek credit risk model. The proxy data is of Credit Default Swaps in the market region of the United States, where the CDS quantities were sufficient and diverse. The study also included other parameters in the cumulative distribution function, correlation coefficient ρ and loss threshold value x. The main objective of this study is to model and mimic different scenarios, which a credit portfolio might face in real-world situations. Aiming for as comprehensive insight as possible into different situations, this thesis might prove useful for credit risk management purposes.

The theoretical segment of this thesis begins with the basics of risk management, after which the attention centres towards credit risk management theory. The study aims to introduce credit risk management and its sections, including Credit Default Swaps (CDS), Probability of Default, and tail risk. Popular risk management measures and tools are presented in the theory section, such as the Value at Risk (VaR) and the Vasicek credit risk model, which is also implemented in the empirical section of the study.

The empirical study focuses on the Credit Default Swap default probabilities and using them as a proxy in estimating the cumulative loss distribution function obtained by using the equation specified by Vasicek. Due to the lack of sufficient default data of bonds and loans, it was determined by the author that the CDS data would be more relevant to use. The results provided valuable insight on how the parameter changes would affect the cumulative loss distributions in the portfolio, as well as the inverse distributions of the portfolio.

This topic is interesting for further study. The Vasicek model is widely used in credit risk management and is also included in regulatory frameworks. The model has already been studied further, for example in the context of machine learning (García-Céspedes & Moreno 2022) and in multi-period approaches (García-Céspedes & Moreno 2017). Additionally, this thesis could be taken further with implementation of, for example, expected losses, unexpected losses, and economic capital. At the dawn of the new era of Artificial Intelligence and Machine Learning, this topic has the potential for further implementation.

References

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., & Diebold, F. X. (2013). Financial risk measurement for financial risk management. In *Handbook of the Economics of Finance*, Vol. 2, 1127-1220.

Antonsson, H. (2018) *Macroeconomic factors in Probability of Default: A study applied to a Swedish credit portfolio*. Master's thesis. KTH, School of Industrial Engineering and Management (ITM), Stockholm.

Augustin, P., Subrahmanyam, M. G., Tang, D. Y., & Wang, S. Q. (2016) Credit Default Swaps: Past, Present, and Future. *Annual Review of Financial Economics*, Vol. 8 (1), 175–196.

Black, F. & Scholes, M. (1973) The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, Vol. 81 (3), 637–654.

Blanco, R., Brennan, S., & Marsh, I. W. (2005) An Empirical Analysis of the Dynamic Relation between Investment-Grade Bonds and Credit Default Swaps. *The Journal of Finance*, Vol. 60 (5), 2255–2281.

Brown, K., & Moles, P. (2014). Credit risk management. K. Brown & P. Moles, *Credit Risk Management*, 16.

CFI Team (2024). *Probability of Default.* Corporate Finance Institute <https://corporatefinanceinstitute.com/resources/career-map/sell-side/capital-markets/probability-of-default/>, retrieved 7.3.2024.

Christoffersen, P. F. (2011) *Elements of Financial Risk Management*. Academic press.

Coleman, T. S. (2012). *Quantitative Risk Management,+ Website: A Practical Guide to Financial Risk* (Vol. 669). John Wiley & Sons.

Duffee, G. R. & Zhou, C. (2001) Credit derivatives in banking: Useful tools for managing risk? *Journal of Monetary Economics*, Vol. 48 (1), 25–54.

García-Céspedes, R. & Moreno, M. (2017) An approximate multi-period Vasicek credit risk model. *Journal of Banking & Finance*, Vol. 81, 105–113.

García-Céspedes, R. & Moreno, M. (2022) The generalized Vasicek credit risk model: A Machine Learning approach. *Finance Research Letters*, Vol. 47, 102669.

Goddard, J., Molyneux, P. & Wilson, J. O. S. (2009) The financial crisis in Europe: Evolution, policy responses and lessons for the future. *Journal of Financial Regulation and Compliance*, Vol. 17 (4), 362–380.

Hopkin, P. (2018) *Fundamentals of Risk Management: Understanding, Evaluating and Implementing Effective Risk Management*. Kogan Page Publishers.

Hull, J. C. (2017) *Options, Futures, and Other Derivatives*. 9th ed., Pearson Education.

Hull, J. C. (2018) *Risk Management and Financial Institutions*. 5th ed., John Wiley & Sons.

Jarrow, R. A. (2012). Problems with using CDS to infer default probabilities. *Journal of Fixed Income*, Vol. 21 (4), 6.

Jiménez, G. & Saurina, J. (2004) Collateral, type of lender and relationship banking as determinants of credit risk. *Journal of Banking & Finance*, Vol. 28 (9), 2191–2212.

Jorion, P. (1996) Risk2: Measuring the Risk in Value at Risk. *Financial Analysts Journal*, Vol. 52 (6), 47–56.

Jorion, P. (2007). *Financial risk manager handbook.* Vol. 406, John Wiley & Sons.

Lando, D. (2009). Credit risk modeling. In *Handbook of Financial Time Series* (pp. 787-798). Springer Berlin Heidelberg.

Li, S., Tian, S., Yu, Y., Zhu, X., & Lian, H. (2023). Corporate Probability of Default: A Single-Index Hazard Model Approach. *Journal of Business & Economic Statistics*, Vol. 41 (4), 1288-1299.

Luenberger, D. G. (2014). *Investment science*. 2nd ed., Oxford University Press.

McNeil, A. J. et al. (2015) *Quantitative risk management: concepts, techniques and tools*. Revised ed., Princeton University Press.

Merton, R. C. (1973) Theory of Rational Option Pricing. *The Bell Journal of Economics and Management Science*, Vol. 4 (1), 141–183.

Miller, M. B. (2018) *Quantitative Financial Risk Management*. John Wiley & Sons.

Nagpal, K., & Bahar, R. (2001). Measuring default correlation. *Risk*, Vol. 14 (3), 129-132.

Nakagawa, K. & Ito, K. (2021) Taming Tail Risk: Regularized Multiple β Worst-Case CVaR Portfolio. *Symmetry*, Vol. 13 (6), 922.

Ohlson, J. A. (1980) Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, Vol. 18 (1), 109–131.

Poutanen, O. (2017) *Structural approach for credit risk modeling: An empirical analysis on European corporate bond prices*. Master's Thesis. Turku School of Economics, Turku.

Schönbucher, P. J. (2003) *Credit Derivatives Pricing Models: Models, Pricing and Implementation*. John Wiley & Sons.

Stulz, R. M. (2010) Credit Default Swaps and the Credit Crisis. *Journal of Economic Perspectives*, Vol. 24 (1), 73–92.

Vasicek, O. (2002). The distribution of loan portfolio value. *Risk*, Vol. 15 (12), 160–162.

Volarević, H. & Varović, M. (2018) Internal model for IFRS 9 -Expected credit losses calculation. *Ekonomski Pregled*, Vol. 69 (3), 269–297.

Weistroffer, C. (2009) Credit default swaps: Heading towards a more stable system. *Deutsche Bank Research.*

Wilson, T. (1997). Portfolio credit risk I. *Risk*, Vol. 10 (9), 111-117.

Zhou, C. (1997) Default Correlation: An Analytical Result. *Finance and Economics Discussion Series*.