Parameter estimation of the Vasicek credit risk model

MSc Thesis

Author: Péter Kiss
Actuarial and Financial Mathematics MSc
Faculty of Quantitative Finance

Supervisor: Gergely Mádi-Nagy
Eötvös Loránd University
Department of Operations Research

Contents

Acknowledgement

I would like to express my gratitude to my supervisor, Gergely Mádi-Nagy, for his invaluable advice, patience, and numerous consultations. Without his detailed answers to the questions raised during my research, this thesis would not have been possible.

I am also grateful to my parents and my brother for their unwavering support, patience, and understanding in every situation.

1 Introduction

Managing credit risk is one of the most important tasks for financial institutions to ensure safe operation. The keystone of this is the risk management recommendations issued by the Basel Committee. The Basel II regulation opened the door to the use of IRB (internal rating based) models, which have further developed and become more complex in terms of risk management methods and models. Whether it is Foundation-IRB or Advanced-IRB models, in both cases the probability of default is calculated using their own internal methods. The Vasicek credit risk model [14], which is at the center of my thesis, provides an opportunity for this, as Somanath Chatterjee [4] also demonstrates. The model places great emphasis not only on the probability of default but also on default correlation, and it is capable of integrating the effects of macroeconomic factors as well as the risk of individual borrowers.

The aim of my thesis is to present the parameter estimation methods of the Vasicek model in detail and compare them. Besides the economic assumptions and mathematical theory, I also address the implementation difficulties and apply the presented methods to real historical S&P data.

The 2nd chapter presents the essential concepts and terminologies used in credit risk modeling, thereby providing a solid foundation for the subsequent chapters. Beyond the evolution of Basel regulations, I define and illustrate the most important concepts and demonstrate the assumptions we will use concerning homogeneous risk classes and why default correlation is important.

In Chapter 3, I present the Merton's structural model [12], which is historically considered one of the earliest credit risk models, but its fundamental principle is still used in the most modern models. The basic idea is that a company goes bankrupt when the market value of its assets falls below the level of its liabilities. Besides the assumptions and applications of the model, I also discuss its limitations.

The 4th chapter discusses CreditMetrics [11], a widely used methodology for measuring portfolio credit risk. This chapter covers the model's structure, underlying assumptions, and how it quantifies the risk of credit events in a portfolio context, demonstrating its practical applications and benefits.

The framework for the above three chapters is provided by Jiri Witzany's book "Credit Risk Management" [15], which offers an excellent general insight into the topic of credit risk.

In the 5th chapter, I present the Vasicek model, which gives the title of this thesis. While many introduce it as a single-factor model, I start with a multi-factor approach and demonstrate how it can be reduced to a single-factor model. Besides its basic idea, I explain what economically acceptable assumptions can be additionally used to reduce the number of parameters and thus the degrees of freedom of the model. The restrictions are necessary due to the reliability of parameter estimates, as although portfolios can cover a very large number of companies, the historical data span is never more than 30-40 years. I also discuss its application within the Basel framework.

The 6th chapter focuses on the methods and techniques used to estimate the parameters of the Vasicek model. The method of moments is most commonly used in practice. Gordy [9] used this method to compare the two largest credit risk models, CreditMetrics and CreditRisk+. Another study by Gordy and Heitfield [10] pointed out that this method has a significant downward bias in estimating default correlations when using short-term data, and therefore, they suggested maximum likelihood estimation methods. This recommended method was used by Klaus Düllmann and Harald Scheule [7] on historical data of german companies, supplemented with asymptotic estimators. Later, Paul Demey, Jean-Frédéric Jouanin, Céline Roget, and Thierry Roncalli [5] further developed the method into a tractable multi-factor setup, and they examined homogeneous risk groups by sectors rather than by grades.

In Chapter 7, I present the results of empirical investigations performed on synthetic and historical data. I use the methodology presented by Gordy and Heitfield for generating synthetic data. Using Monte Carlo simulation, I create confidence intervals for the parameter estimates and thereby compare the methods. I also fit the model to the historical default rate data of S&P from 1981 to 2020 under different restrictions. The results obtained are compared with the findings of previous studies that worked with time series up to the early 2000s.

In Chapter 8, I address the challenges encountered during the implementation of the model and the ideas for individual solutions. To my knowledge, previous studies do not cover this part, although I believe that good implementation is essential for the proper use of a financial model. This chapter also touches on topics such as numerical integration, one-dimensional and multi-dimensional optimization. I implemented the model in Python, and the full source code is available in the following GitHub repository:
`https://github.com/PeterKiss18/VasicekPDModel`.

In the final chapter, in addition to drawing conclusions, I also discuss potential future research directions.

2 Fundamental concepts for Credit Risk

Credit is money provided by a creditor to an obligor. Credit risk refers to the possibility that a contracted payment may not be fulfilled. The initial interest in credit risk models stemmed from the need to determine the amount of economic capital required to support a bank's exposures.

2.1 Basel regulations and their evolution

Since the Basel Accord of 1998, minimum capital requirements have been internationally coordinated. Under Basel I, bank assets were allocated into four broad risk categories with risk weightings from 0% to 100%. Corporate loans received a 100% risk weight, while retail mortgages, considered safer, received a 50% weight. The minimum capital was then set in proportion to these weighted assets:

$$\text{Minimum capital requirement} = 8\% \times \sum \text{weighted assets}$$

This approach was criticized for its lack of granularity, failing to capture the cross-sectional distribution of risk. For example, all mortgage loans had the same capital requirement, regardless of borrower risk profiles like loan-to-value or debt-to-income ratios. In response, Basel II introduced a more granular risk weighting approach. Credit risk management techniques under Basel II are classified into:

- **Standardised approach:** A simple categorisation of obligors without considering their actual credit risks, often relying on external credit ratings.

- **Internal Ratings-Based (IRB) approach:** Banks use their internal models to calculate the regulatory capital requirement for credit risk.

These frameworks determine the risk-weighted assets (RWA), which are the denominator in key capitalisation ratios (Total capital, Tier 1, Core Tier 1, Common Equity Tier 1). The IRB approach uses a formula approximating the Vasicek model of portfolio credit risk, detailed in Chapter 5.

Basel III did not change the minimum capital requirement but introduced stricter rules to ensure the quality of capital, including a 4.5% minimum CET1 requirement and usable capital buffers. Although Basel III refined the definition of capital, it retained the Basel II risk-based framework for measuring risk-weighted assets, improving the standardised approach for credit risk and linking it more closely with the IRB approach.

2.2 Probability of Default (PD) and related concepts

In the Basel framework, a key parameter, which is used to calculate the regulatory capital for credit risk, is the probability of default, often indicated with the acronym PD. The **probability of default** (PD) is a financial term describing the likelihood of a default over a particular time horizon. In finance and lending, "default" refers to the failure of a borrower to fulfill their contractual obligations or meet the terms specified in a loan agreement. More specifically, under Basel II, a default event on a debt obligation is said to have occurred if it is unlikely that the obligor will be able to repay its debt to the bank without giving up any pledged collateral and the obligor is more than 90 days past due on a material credit obligation. In order to discuss about PD calculation, it is necessary to precisely define the fundamental concepts related to loss.

When banks estimate the economic capital necessary to support their credit risk activities, they use an analytical framework that relates the required economic capital to the probability density function (PDF) of credit losses, also known as the loss distribution of a credit portfolio. Let us have a portfolio of assets (and liabilities), and let X (a random variable) denote the loss on the portfolio in a fixed time horizon T. Let

$$F_X(x) = \Pr[X \leq x]$$

be the cumulative distribution function of X. Let's define the quantile with given a α probability level, as

$$q_\alpha^X = \inf\{x \mid F_X(x) \geq \alpha\}.$$

These losses will play a role in threshold models, which assume that a default event occurs when the value of assets exceeds the liabilities.

From a statistical perspective, the **Value-at-Risk** (VaR) is nothing more than a quantile of the loss distribution. From an economic standpoint, it expresses the potential maximum absolute loss that can be realized on the probability level α.

The two important components of credit risk are **expected loss** (EL) and **unexpected loss** (UL). The EL represents the amount of credit loss the bank expects to experience over a given time horizon, considered the normal cost of doing business covered by provisioning and pricing policies.

$$\text{EL} = E[X]$$

In contrast, the UL represents the risk of the portfolio, with capital held to offset this risk. Within the IRB methodology, the regulatory capital charge depends only on UL. The unexpected loss (UL) also known as relative Value-at-Risk ($\text{VaR}_\alpha^{\text{rel}}$) is defined as

$$\text{UL} = q_\alpha^X - \text{EL}.$$

These concepts are visually illustrated in Figure 1, which shows an example of a loss density function, depicting the notions of unexpected and expected loss.

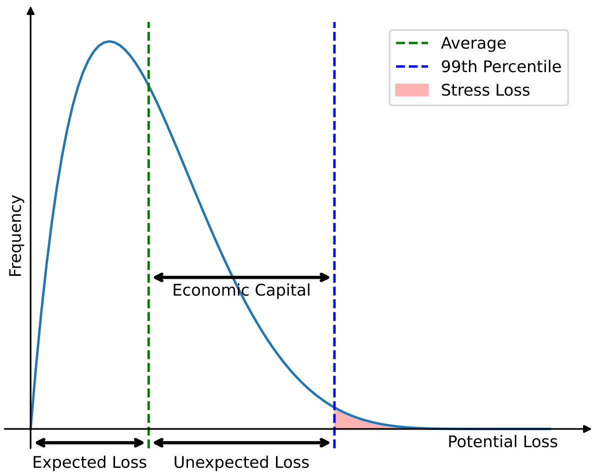


Figure 1: Loss Distribution

The probability of losses beyond the combined amount of expected loss and unexpected loss — meaning the probability that the bank would be unable to fulfill its credit commitments using its profits and capital — is represented by the shaded area on the right-hand side of the curve, which is referred to as stress loss. If capital is set according to the gap between EL and VaR, and EL is covered by provisions or revenues, the likelihood of the bank remaining solvent over a one-year horizon equals the confidence level. Under Basel II, capital is set to maintain a supervisory fixed confidence level, typically 99.9%, meaning an institution is expected to experience losses exceeding its capital once in 1,000 years.

Banks must decide on the time horizon over which they assess credit risk. In the Basel context, a one-year time horizon is used across all asset classes. **Exposure at default (EAD)** is the predicted total amount of credit exposure a bank expects to have at the time a borrower defaults on a loan.
Loss Given Default (LGD) represents the proportion of the exposure that will not be recovered after default. It is expressed as a percentage of EAD.
The expected loss of a portfolio is the product of the proportion of obligors that might default within the time frame, the outstanding exposure at

default (EAD), and the loss given default (LGD). Under the Basel II IRB framework, the probability of default (PD) per rating grade is the average percentage of obligors defaulting over one year. Assuming a uniform LGD value for a given portfolio, EL can be calculated as the sum of individual expected losses in the portfolio:

$$EL = \sum (\mathrm{PD} \cdot \mathrm{EAD} \cdot \mathrm{LGD}) .$$

2.3 Homogeneous risk classes

In the Vasicek model, we assume that all firms have been successfully categorized into a few risk classes that exhibit certain homogeneity properties, as defined below.

- Every obligor within a given risk class has the same probability of default.
- The correlation between any two aboligors within a c given risk class is constant, i.e.,
$$\rho_{m,n} = \rho_c \quad \forall m, n \in c.$$
- For a given pair of risk classes (c, d), the correlation between two obligors $(m \in c, n \in d)$ depends only on the two classes:
$$\rho_{m,n} = \rho_{c,d} \quad \forall m \in c, n \in d.$$

Risk classes can be constructed based on rating grades, geographical locations, industrial sectors, or a combination of these criteria. Let us introduce the following notations:

- G: the number of risk classes
- $n_{g,t}$: the number of firms in class g at time t
- $d_{g,t}$: the number of firms that defaulted in class g at time t
- $\mu_{g,t} = \frac{d_{g,t}}{n_{g,t}}$ the default rate

We aim to model the default rate variable, but the variance of this default rate for a given risk class heavily depends on the correlation between the defaults of the firms within that class. Intuitively, the higher the correlation, the more likely it is that other firms in the class will default if one defaults at time t. For this reason, we also model the default correlations.

The role of the correlation is illustrated in Figure 2, which shows the distribution of the annual default rate for a risk class consisting of 1,000 firms under different correlations (each firm individually has a 20% probability of default).

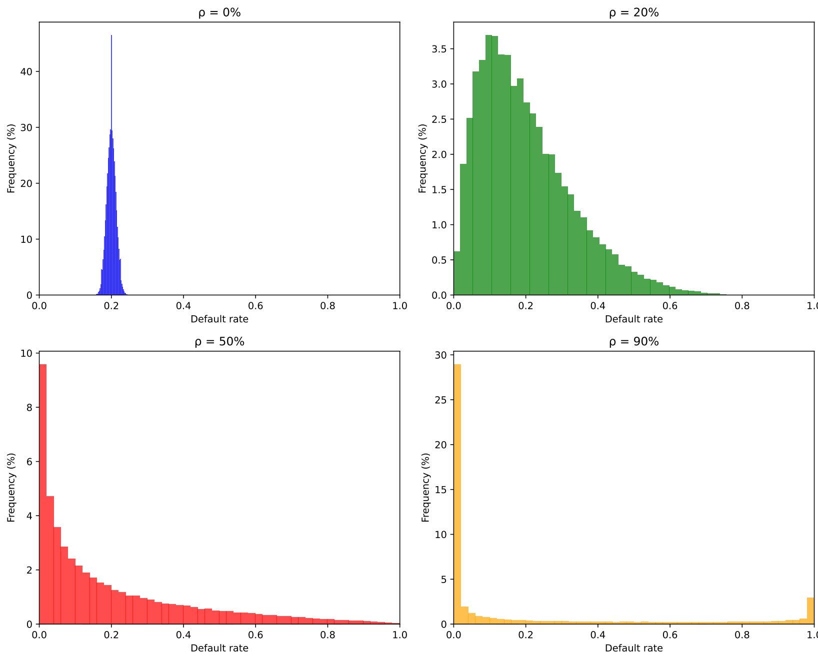


Figure 2: Distribution of annual default rates for different correlation levels with a common default probability of 20% and 1000 obligors in the risk class.

3 Merton's structural model

Merton's structural model is a fundamental framework for assessing credit risk. The model posits that a firm defaults on its debt (D) if the value of its assets (A) falls below the debt, i.e., $A < D$. Based on this criterion, it is categorized among threshold models, which is a common feature with the Vasicek model.

Considering stochastic changes in the firm's equity, we use a stochastic model for $A(t)$, starting at an initial value $A_0 > D$. In a simplified scenario with a single loan and bullet repayment of D at time T, default occurs if $A(T) < D$ (Fig. 3). If $A(T) \geq D$, there is no default, and the remaining shareholders' value is $A(T) - D$.

The model has been formulated not only to theoretically define the probability of default but, in fact, primarily to apply the theory of option valuation to the valuation of debt and equity in the market. The final payoff for creditors at maturity (T) is expressed as

$$D(T) = D - \max(D - A(T), 0).$$

The value of risky debt can theoretically be calculated by subtracting the value of a put option on the firm's assets from the value of risk-free debt.

This put option, with an exercise price D and maturity T, is sold to the shareholders in exchange for the premium paid over the risk-free interest rate. The shareholders' payoff at time T is calculated as

$$E(T) = \max(A(T) - D, 0).$$

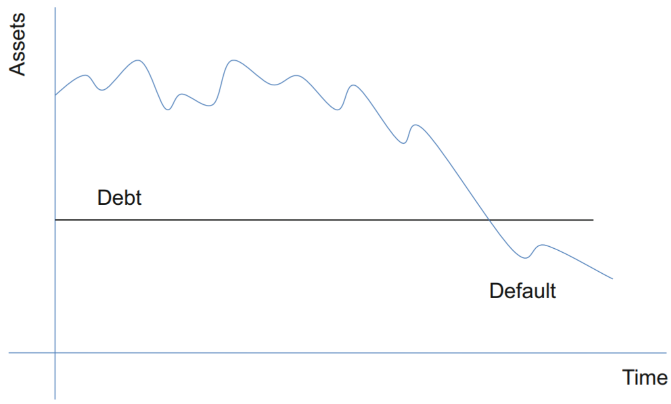


Figure 3: Merton's structural model [1]

Consequently, the equity value $E(0)$ can be theoretically valued akin to the European call option on the assets, with the exercise price D. In cases where the asset value follows a geometric Brownian motion described by the (3.1) stochastic differential equation, the call and put options can be evaluated using the Black-Scholes-Merton formula.

$$dA(t) = \mu A(t)dt + \sigma A(t)dW(t). \tag{3.1}$$

While the model presents a sophisticated theoretical framework, it suffers from several significant limitations. First, the asset market value process is latent, typically unobservable empirically in most situations. However, there exists a functional relationship between equity prices and asset prices as specified by the model, allowing for the transformation of parameters derived from stock data into parameters applicable to latent asset values using stochastic calculus. Additionally, the model inaccurately presupposes that default can only occur at the predetermined time T, whereas in practical scenarios, default could happen at any point up to the maturity date. Consequently, American options would be more suitable than European ones for modeling this reality. Another complication arises from the diverse liquidity of assets (ranging from short-term to long-term and spanning financial to non-financial), necessitating varied discounting approaches during a distress situation.

[1] Source: [15, p. 125, Figure 4.4]

4 CreditMetrics

The CreditMetrics methodology is a risk management approach primarily utilized for the assessment and management of credit risk by financial institutions, including banks and investment firms. Initially published by J.P. Morgan in 1997, the main objective of the methodology is to analyze and manage credit portfolio risk. Originally, the methodology was designed for bonds priced by their market values, but the approach can also be adapted for a loan portfolio, with losses calculated according to accounting provisions that follow a specific classification system.

The model is based on ratings that are assumed to determine the values of individual debt instruments. The model utilizes Monte Carlo simulation and is based on ratings assumed to determine the values of individual debt instruments.

The model's fundamental principles:

- Today's prices of bonds are determined by their current ratings and the term structure of risk-free interest rates, which remains constant during the analysis.
- Future prices of bonds (e.g., in a 1-year horizon) are determined by their future ratings.
- Rating migration probabilities are obtained from historical data.
- Correlations between rating migrations are captured using asset correlations.
- The asset correlations are calculated by mapping the firms into various economic sector indices.
- Lastly, by simulating future ratings and market values for all bonds in the portfolio, the empirical distribution of the portfolio value is obtained, along with measures of expected and unexpected losses at various probability levels.

Explaining the above principle steps in more detail:

Regarding bond valuation, we assume that for each rating s, there is a term structure of zero-coupon interest rate $r_s(t)$ for every maturity t. This allows us to calculate the value of a bond by summing the present values of its cash flows, where we discount with the appropriate points on the zero coupon yield curve for the given rating. The zero coupon rates imply the forward rates, and this allow us to determine the forward price of a bond, conditional on its future rating.

We can simulate rating migration given certain initial ratings, using historical transition probabilities, which are regularly disclosed by major rating agencies in transition matrix.

Another key input factor is the recovery rate (RR) that is the amount of credit recovered through foreclosure or bankruptcy procedures in event of a default, expressed as a percentage of face value. The recovery rate depends on the bond's seniority classes, which define the order in which liabilities are satisfied in the case of bankruptcy (with senior bonds having the highest priority and junior subordinated bonds the lowest). We consider this rate as a random variable, the expected value and standard deviation of which we know from historical data. However, in a simplified approach, it is assumed to be constant, using the average value.

To simulate joint migrations of many bonds from different issuers, we have to take into account their correlations. A crucial aspect of this model is its use of Merton's credit risk option model, which establishes a link between a firm's asset value and its total debt. In CreditMetrics, the rating change of a bond b is modeled by a continuous random variable $r(b)$, which follows the standard normal distribution $N(0,1)$. We can interpret it as an appropriately scaled credit scoring change. Starting with an initial rating, one can establish a series of thresholds for $r(b)$ that, based on the probabilities of rating transitions, are designed to trigger potential rating migrations.

According to the Merton model, the probability of default is determined by the distance of $\ln A(t)$ from $\ln P$. If we look at the logarithm of the asset value instead of the (3.1) equation, it follows a Brownian motion with drift:

$$d(\ln A) = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dW.$$

The larger this distance, the better the credit rating should be. Thus, starting from an initial rating, the rating migration at time 1 is determined by the standardized $N(0,1)$ asset return:

$$r = \frac{1}{\sigma}\left(\ln \frac{A(1)}{A(0)} - \left(\mu - \sigma^2/2\right)\right)$$

If the return is positive, then there is a rating improvement, and if the return is negative, then there is a rating deterioration based on the rating migration thresholds.

Given debtors $i = 1, ..., N$ all we need to know is the matrix Σ of correlations $\rho_{ij} = \rho(r_i, r_j)$ between the standardized asset returns, and the rating migration thresholds. These correlations could be estimated from the equity returns data, if all the companies are liquidly traded on stock markets. Since this is usually not the case, CreditMetrics proposes the use of a single-, or multi-, factor model, breaking down debtors' returns into a combination of systematic factors and independent, idiosyncratic, debtor specific factors.

$$r_i = \sum_{j=1}^{k} w_{i,j} r\left(I_j\right) + w_{i,k+1}\epsilon_i$$

where $r(I_j)$ is the standardized return of the I_j systematic factor (e.g. sector or country index), and ϵ_i is the standardized debtor specific factor. Since the systematic factors can be correlated, it is generally insufficient to require $\sum_{j=1}^{k+1} w_{i,j}^2 = 1$ to standardize r_i. Determining the weights, and thus the correlations, is a crucial step in the model and is often one of the most challenging parts of the CreditMetrics methodology. The Technical Document suggests an expert approach: estimating the weights of the systematic factors and the complementary idiosyncratic factor. When more systematic sector or country factors are involved, the "participation" of the debtor in these indices must be specified and combined appropriately.

Given the asset return correlation matrix Σ, scenarios can be generated by sampling a vector of standardized normal variables u, and multiplying it by the Cholesky matrix A such that $r = Au$, where A is a lower triangular matrix with the property that $AA^\top = \Sigma$. We then determine the rating migrations based on the thresholds and the simulated portfolio value $V(r)$ at the end of the period. By repeating this procedure, we obtain a large number of sampled values $V_1, V_2, \ldots, V_M$, and an empirical distribution of the portfolio market value. In the case that there is only one systematic factor, or a few of them, it is computationally more efficient to sample first of all the systematic factors, and then the independent idiosyncratic factors generating r_i for all the exposures.

The empirical distribution is used to estimate the α-quantile, representing the Value at Risk (VaR) at a specified confidence level. The confidence intervals of these estimations must also be considered. Typically, the precision of a plain Monte Carlo simulation is of the order $\frac{1}{\sqrt{M}}$, where M is the number of simulations. Therefore, at least 10,000 simulations are required to achieve satisfactory precision.

Similarly, in the Vasicek model a standard normal distribution variable will determine the defaults, and there this variable will be also divided into systematic and idiosyncratic components.

5 Vasicek model

The credit risk model proposed by Vasicek in 1987 [14] serves as the foundation for Basel's regulatory capital requirements and enjoys widespread use in the financial industry.

Let T_i denote the time when the i-th obligor defaults. It is customary to assume that everyone will defaults once, but the timing is uncertain, so $T_i < \infty$ is regarded as a random variable. Let Q_i be the cumulative probability distribution of T_i. Let $Y_i = \Phi^{-1}\left(Q_i\left(T_i\right)\right)$ denote the variable obtained after the quantile-to-quantile transformation from T_i to a standardized normal distribution, where Φ is the cumulative distribution function of the standard normal distribution.

Using the previous notations, the i-th obligor defaults within 1 year if and only if

$$Y_i \leq \Phi^{-1}(PD),$$

where $\mathrm{PD} = Q_i(1)$ represents the 1-year probability of default.

Suppose the latent variable Y_i, which represents the normalized return of an obligor, can be written in the following form:

$$Y_i = \omega_i \cdot \mathbf{Z} + \xi_i \cdot \epsilon_i \tag{5.1}$$

where $\mathbf{Z}$ is a K-vector of systematic risk factors. These factors capture unforeseen shifts in economy-wide such as interest rates, GDP growth rates, commodity prices and stock market indices, which affect asset returns across different sectors. It is assumed that $\mathbf{Z}$ is a mean-zero normal random vector with a Ω variance matrix. The sensitivity of obligor i to $\mathbf{Z}$ is measured by a ω_i vector of factor loadings. The term ϵ_i represents obligor-specific risk, where each ϵ_i is assumed to follow a standard normal distribution, being independent across obligors and also independent of $\mathbf{Z}$. For simplicity, the Ω covariance matrix is considered to have ones on its main diagonal (implying each Z_k has a standard normal marginal distribution), and the weights ω_i and ξ_i are scaled so that Y_i has a mean of zero and a variance of one.

As mentioned, the obligor defaults if Y_i is below $\Phi^{-1}(\mathrm{PD})$ threshold, and if we consider Y_i as representing the normalized return of the obligor, we can recognize the basic idea of the threshold model seen in Chapter 3. From now on, I will denote this default threshold by γ.

To enable model calibration using historical data typically available from rating agencies, obligors are grouped into G homogeneous "buckets" indexed by g. In the ensuing applications, these buckets correspond to an ordered set of rating grades. However, a bucketing system can theoretically be defined along multiple dimensions. For instance, a bucket might consist of obligors of a given rating in a specific industry and country. Assume that within a bucket, all obligors have the same default threshold γ_g, meaning the PD of obligors in grade g is:

$$\bar{p}_g = \Phi(\gamma_g).$$

The vector of factor loadings is assumed to be constant across all obligors in a grade, thus we can rewrite the equation for Y_i as one-factor Gaussian copula model:

$$Y_i = w_g \cdot X_g + \sqrt{1 - w_g^2} \cdot \epsilon_i, \tag{5.2}$$

where

$$X_g = \frac{\sum_{k=1}^{K} Z_k \omega_{g,k}}{\sqrt{\omega_g' \Omega \omega_g}} \tag{5.3}$$

is a univariate bucket-specific common risk factor. So Y_i is driven by a X_g common systematic (macroeconomic) factor, and an ϵ_i idiosyncratic independent factor, where both X_g and ϵ_i follow independent standard Gaussian distributions.

The G-vector $\mathbf{X} = (X_1, \ldots, X_G)$ has a multivariate normal distribution. Let σ_{gh} denote the covariance between X_g and X_h. Intuitively, we expect $\sigma_{gh} > 0$. The factor loading on X_g for obligors in bucket g is:

$$w_g = \sqrt{\omega_g' \Omega \omega_g}. \tag{5.4}$$

which is bounded between zero and one. Note that due to the scaling convention, which sets the variance of Y to 1, we can eliminate the ξ_i variable from equation (5.1). Thus, in equation (5.2) we can express the weights of both factors as a function of w_g. In the following, I will refer to X_g as risk factor, elements of Z as structural risk factors, w_g as factor loading, and ω_g as structural factor loadings.

Default rate correlations

Default rate correlations among obligors are driven by correlations in the risk factors. It can be demonstrated that the default correlation between an obligor in bucket g and another in bucket h is:

$$\rho_{gh} = \frac{F(\gamma_g, \gamma_h; w_g w_h \sigma_{gh}) - \bar{p}_g \bar{p}_h}{\sqrt{\bar{p}_g(1 - \bar{p}_g)}\sqrt{\bar{p}_h(1 - \bar{p}_h)}}, \tag{5.5}$$

where $F(z_1, z_2; c_{12})$ denotes the joint CDF for a mean-zero bivariate normal random vector with unit variances and covariance c_{12}. In the special case where both obligors are in the same bucket, the within-bucket default correlation is:

$$\rho_g = \frac{F(\gamma_g, \gamma_g; w_g^2) - \bar{p}_g^2}{\bar{p}_g(1 - \bar{p}_g)}. \tag{5.6}$$

Possible restrictions

Given sufficient data, it is possible to estimate all $G(G+1)/2$ default correlations defined by equations (4) and (5). However, with limited data, many parameters may be unidentified or poorly identified, necessitating the imposition of ex-ante restrictions on factor loadings and risk factor correlations. Equations (5.5) and (5.6) define $\frac{G(G-1)}{2}$ default correlations, which we could theoretically estimate if we had sufficient data. However, in reality, data is always scarce, making the parameters unidentified or poorly identified. To avoid this, we impose ex-ante restrictions on the factor loadings and risk factor correlations.

One Risk Factor restriction : $$X_1 = X_2 = \ldots = X_G. \tag{R1}$$

R1 is equivalent to requiring that $\sigma_{gh} = 1$ for all (g, h) bucket pairs. A sufficient condition for R1 is that there is exactly one structural risk factor (i.e., $K = 1$). This condition is not always met, for example, if there are obligors working in different sectors or countries, they are likely influenced by different structural factors, but if the portfolio is relatively homogeneous, then this constraint can be a reasonable approximation.

The strongest restriction one can impose on the factor loadings is to assume that they are constant across all obligors:

Same Factor Loading Restriction: $w_g = w_h \quad \forall g, h$ bucket pairs. (R2)

The combination of R1 and R2 restrictions imply that the structural factor loadings are constant across buckets.

Using for loss calculation

Let's assume we have $j = 1, 2, \ldots, J$ obligors with the same probability of default PD in a homogeneous bucket:

$$Y_j = w \cdot X + \sqrt{1 - w^2}\epsilon_j.$$

Using Monte Carlo simulation, we generate data by first generating the value of the systematic factor, $x \sim N(0, 1)$, and then generating values $\epsilon_j \sim N(0, 1)$

for each borrower: If J is sufficiently large, by the law of large numbers, the conditional probability of default is:

$$\Pr\left[wX + \sqrt{1-w^2}\epsilon_j \leq \Phi^{-1}(\text{PD}) \mid X = x\right] = \Pr\left[T_j \leq 1 \mid X = x\right] = \text{PD}_1(x);$$

i.e., the rate of default PD_1, conditional on the systematic factor $X = x$.

Ordering the probability by X_j and utilizing the fact that X_j follows a standard normal distribution, the conditional probability can be expressed as:

$$\text{PD}_1(x) = \Pr\left[wX + \sqrt{1-w^2}\epsilon_j \leq \Phi^{-1}(\text{PD}) \mid X = x\right] =$$

$$= \Pr\left[\epsilon_j \leq \tfrac{\Phi^{-1}(\text{PD}) - wx}{\sqrt{1-w^2}}\right] = \Phi\left(\tfrac{\Phi^{-1}(\text{PD}) - w \cdot x}{\sqrt{1-w^2}}\right)$$

Since the conditional portfolio rate of default, $\text{PD}_1(m)$, depends monotonically only on the single factor, $m \sim N(0,1)$, we can use the quantiles of m to determine the unexpected default rate at any desired probability level, α (e.g., 99%), by setting $m = \Phi^{-1}(1-\alpha) = -\Phi^{-1}(\alpha)$. Thus, the unexpected default rate at the probability level α can be expressed as:

$$\text{UDR}_\alpha(\text{PD}) = \Phi\left(\frac{\Phi^{-1}(\text{PD}) + w \cdot \Phi^{-1}(\alpha)}{\sqrt{1-w^2}}\right). \tag{5.7}$$

The same argument may be applied to a portfolio of exposures with different probabilities of default, driven by a rating scale, assuming that for each rating grade the number of exposures is large. The total losses of a portfolio (L) consisting of J clients can be calculated by summing the individual losses, which is expressed as:

$$L = \sum_{j=1}^{J} \text{EAD}_j \cdot \text{LGD}_j \cdot \text{D}_j(X_j, \epsilon_j)$$

where EAD_j is the exposure at default, LGD_j is the loss given default (defined in Section 2.2) and D_j is a dummy variable that indicates if the clients defaults. The losses conditional to a macroeconomic scenario are given by:

$$E[L|x] = \sum_{j=1}^{J} \text{EAD}_j \cdot \text{LGD}_j \cdot \Phi\left(\frac{\Phi^{-1}(\text{PD}) - w \cdot x}{\sqrt{1-w^2}}\right).$$

As we can see, the individual-specific risk completely disappears. However, in the general scenario of multi-factor and non-granular portfolios, closed-form expressions are unavailable, necessitating the use of Monte Carlo methods or approximate formulas.

Basel II/III Capital Formula

Even though the first Consultative Papers [2] proposed measures to capture portfolio granularity by penalizing less diversified portfolios, the final version of the Basel capital calculation formula is portfolio invariant. It is based on the following equation:

$$\begin{aligned} \mathrm{RWA} &= \mathrm{EAD} \cdot W, \\ W &= K \cdot 12.5, \\ K &= (\mathrm{UDR}_{99.9\%}(\mathrm{PD}) - \mathrm{PD}) \cdot \mathrm{LGD} \cdot \mathrm{MA}\,. \end{aligned} \tag{5.8}$$

The variables in the equation are defined as follows:

- RWA (Risk-Weighted Assets): The total value of assets, weighted according to credit risk.
- W: Risk Weight applied to the exposure.
- K: Capital requirement ratio.
- $\mathrm{UDR}_{99.9\%}(\mathrm{PD})$: Unexpected default rate at the 99.9% confidence level.
- MA (Maturity Adjustment): Adjusts the risk weight based on the maturity of the exposure.

Originally, the Risk Weights were determined from the ratings using a regulatory table, and this was replaced by the (5.8) formula. The number 12.5 in the formula refers to the reciprocal of 8%, which was a typical leverage ratio under previous regulatory frameworks.

Banks applying the Internal Ratings-Based (IRB) approach must estimate the Probability of Default (PD) parameter and, in the advanced approach, the Loss Given Default (LGD) and Exposure at Default (EAD) parameters using their internal models that satisfy a number of qualitative requirements. The correlation parameter in the (5.7) formula is provided by regulation and varies according to different segments. For corporate, sovereign, and bank exposures, the correlation is set as a weighted average between 0.12 and 0.24, depending on the PD:

$$w = 0.12 \cdot \left(\frac{1 - e^{-50 \cdot \mathrm{PD}}}{1 - e^{-50}} \right) + 0.24 \cdot \left(\frac{e^{-50 \cdot \mathrm{PD}} - e^{-50}}{1 - e^{-50}} \right). \tag{5.9}$$

The correlation is slightly reduced for Small and Medium-sized Enterprises (SME) exposures, reflecting their lower size and higher diversification. Similarly, for consumer loans, the correlation is a weighted average between 0.03 and 0.16, while for mortgages it is fixed at 0.15, and for revolving loans (e.g., credit cards), it is 0.04. The maturity adjustment (MA) is applied only to corporate, sovereign, and bank exposures to differentiate risk between shorter and longer maturities:

$$\mathrm{MA} = \frac{1 + (\mathrm{M} - 2.5) \cdot b}{1 - 1.5 \cdot b}, \tag{5.10}$$

$$b = (0.1182 - 0.05478 \cdot \ln(\mathrm{PD}))^2. \tag{5.11}$$

For retail receivables, the adjustment is not used (MA = 1). Regulators have set a high confidence level ($\alpha = 99.9\%$), which is considered over-conservative compared to analogous regulations in the insurance industry, such as Solvency II, which uses a confidence level of 99.5%. This high confidence level might compensate for imperfect diversification not reflected in the model. The correlation coefficients are adjusted to reflect different levels of diversification, with lower coefficients for SMEs and revolving credits, where higher diversification is expected.

Advantages

1. **Risk Sensitivity**: The Basel II formula offers improved risk sensitivity compared to the Basel I RWA calculation.

2. **Simplicity**: Despite its improvements, the formula remains relatively simple, allowing banks to estimate key parameters without sophisticated simulation or analytical portfolio modeling.

3. **Encourages Higher Standards**: The intention behind the IRB approach is to motivate banks to adopt higher credit risk management standards, with IRB capital requirements typically being slightly below the Standardized Approach requirements for the same portfolio.

Disadvantages

1. **Simplifying Assumptions**: The Vasicek model, which underpins the Basel formula, makes significant simplifying assumptions, such as assuming a perfectly diversified portfolio and ignoring the effects of low diversification due to limited exposures.

2. **Unexpected Recovery Risk**: The model underestimates unexpected recovery risk, which can be significant. Empirical studies [1] show that recovery rates are negatively correlated with the probability of default, leading to potential underestimation of economic capital during downturns.

3. **Pro-Cyclicality**: Basel II capital requirements are pro-cyclical, being relatively low during economic booms and high during recessions. This can exacerbate economic cycles, encouraging more lending during booms and restricting it during downturns.

Overall, while the Basel II/III capital formulas represent a significant advancement in risk sensitivity and regulatory standards, they also introduce new challenges and complexities to their predecessors that need ongoing refinement and management.

6 Parameter estimation of Vasicek model

6.1 Moment based estimation

The concept of method of moments parameter estimation is that the first and second moments of the conditional $p(x)$ probability of default match the time series of default rates for $T \to \infty$. The unconditional PD is estimated by the average $\overline{p}$ of the time series of default rates.

$$E[p(x)] = \overline{p} \tag{6.1}$$

Let $\hat{d}_t$ be the number of defaults during year t, and let $\hat{n}_t$ denote the number of obligors at the start of year t. Let $\hat{p}_t$ denote the observed default frequency $\hat{d}_t/\hat{n}_t$. Assuming that $\hat{n}_t$ is independent of the realization of x_t, the conditional probability's variance V can be derived as follows.

$$V[\hat{p}] = E[V[\hat{p} \mid p(x), \hat{n}]] + V[E[\hat{p} \mid p(x), \hat{n}]] \tag{6.2}$$

In equation (6.2), we used the law of total variance for $V[\hat{p}]$. The variable $\hat{d}_t$ follows a $\mathrm{Binom}(\hat{n}_t, p(x_t))$ binomial distribution because the obligors defaults are assumed to be independent conditional on x. Hence, we obtain the expected value of the conditional variance of $\hat{p}$ using equation (6.3a).

$$
\begin{aligned}
E\left[\mathrm{V}\left[\hat{p} \mid p(x), \hat{n}\right)\right] &= E\left[\mathrm{V}\left[\hat{d} \mid p(x)\right] / \hat{n}^2\right] & (6.3a)\\
&= E\left[p(x)\left(1 - p(x)\right) / \hat{n}\right] & (6.3b)\\
&= E\left[1/\hat{n}\right]\left(E\left[p(x)\right] - \left(\mathrm{V}\left[p(x)\right] + E\left[p(x)\right]^2\right)\right) & (6.3c)\\
&= E\left[1/\hat{n}\right]\left(\bar{p}\left(1 - \bar{p}\right) - \mathrm{V}\left[p(x)\right]\right) & (6.3d)
\end{aligned}
$$

The expression in (6.3a) is obtained by substituting the variance of the binomial distribution into (6.3b), where the mutual independence of x and $\hat{n}$ and the equality $E[p(x)^2] = -V[p(x)] + E[p(x)]^2$ are exploited to derive the expression in (6.3c). Finally, by substitution of (6.1), we obtain the expression in (6.3d).

$$E[\hat{p} \mid p(x), \hat{n}] = p(x) \tag{6.4}$$

The equation (6.4) holds true by definition, and it implies equation (6.5).

$$V[E[\hat{p} \mid p(x), \hat{n}]] = V[p(x)] \tag{6.5}$$

Replacing equations (6.5) and (6.3d) into equation (6.2) yields equation (6.6).

$$V[\hat{p}] = E\left[1/\hat{n}\right]\left(\bar{p}\left(1 - \bar{p}\right) - \mathrm{V}\left[p(x)\right]\right) + V[p(x)] \tag{6.6}$$

By rearrangement, equation (6.6) leads to the expression for $V[p(x)]$.

$$\mathrm{V}\left[p(x)\right] = \frac{\mathrm{V}\left[\hat{p}\right] - E\left[1/\hat{n}\right]\bar{p}\left(1-\bar{p}\right)}{1 - E\left[1/\hat{n}\right]} \tag{6.7}$$

On the right-hand side of equation (6.7), for historical data, each term is known, thus we calculate the unconditional variance using this formula.

Now that we know how to calculate the unconditional variance from historical data, we want to express this quantity as a function of the sought parameters p and w. Let y_1 and y_2 the 2 latent variable for 2 obligors in the same grade. Assume that the obligors have the same w factor loading in our one systematic risk factor model.

$$\begin{aligned} y_1 &= w \cdot x + \sqrt{1-w^2} \cdot \epsilon_1 \\ y_2 &= w \cdot x + \sqrt{1-w^2} \cdot \epsilon_2 \end{aligned} \tag{6.8}$$

The default events of the two obligors are independent conditional on x, thus the probability that both y_1 and y_2 are below the given cut-off value C, i.e., both obligors default, can be calculated as in equation (6.9).

$$\begin{aligned} Pr(y_1 < C \ \& \ y_2 < C \mid) &= Pr(y_1 < C \mid x) \cdot Pr(y_2 < C \mid x) \\ &= \Phi\left(\frac{C - xw}{\sqrt{1-w^2}}\right)^2 = p(x)^2 \end{aligned} \tag{6.9}$$

Then we express the unconditional variance as the difference between the second moment, in order to substitute $p(x)^2$ based on equation (6.9).

$$\begin{aligned} \mathrm{V}\left[p(x)\right] &= E[p(x)^2] - E[p(x)]^2 \\ &= E[Pr(y_1 < C \ \& \ y_2 < C \mid x)] - E[p(x)]^2 \end{aligned} \tag{6.10}$$

Using the fact that y_1 and y_2 are standard normally distributed with correlation w^2, the unconditional expectation equals the value of a two-dimensional normal cumulative distribution function:

$$E[Pr(y_1 < C \ \& \ y_2 < C \mid x)] = \mathrm{BIVNOR}(C, C, w^2), \tag{6.11}$$

where $\mathrm{BIVNOR}(z_1, z_2, w^2)$ is the bivariate normal cdf for $Z \equiv [z_1 \ z_2]'$ such that

$$E[Z] = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad V[Z] = \begin{bmatrix} 1 & w^2 \\ w^2 & 1 \end{bmatrix}.$$

Substituting equations (6.11) and (6.1) into equation (6.10), we obtain:

$$\mathrm{V}[p(x)] = \mathrm{BIVNOR}(C, C, w^2) - \overline{p}^2. \tag{6.12}$$

It is evident that in the variance calculation, the square of the w factor loading appears always, so the equations would hold true for both $+w$ and $-w$ cases. However, from an economic perspective, we always choose the positive w since intuitively and empirically, there appears to be a positive correlation between the x common systematic factor and the obligor's default-inducing variable y.

The cut-off value C determines the unconditional probability in a grade, so in order to match the first moment, the equation $\overline{p} = \Phi(C)$ (where Φ is the standard normal cumulative distribution function) must be satisfied.

In summary, first, we set the C threshold value based on the average historical default rate, $C = \Phi^{-1}(\overline{p})$. For fitting the parameter w factor loading, first, we calculate the unconditional variance from historical data based on equation (6.7) (from realized variance, the value of $E[1/\hat{n}]$, and the value of $\overline{p}$), then adjust the value of w to satisfy equation (6.12).

In Chapter 7, I compare this MM parameter estimation on synthetic simulated data with the maximum likelihood estimation (discussed in the following Chapter 6.2), and also apply them to historical S&P data.

6.2 Maximum likelihood

Our goal is to determine those w factor loadings and the γ default threshold parameters from historical data that maximize the value of the likelihood function. Historical data includes the number of obligors for G rating grades at the beginning of each year and how many of them defaulted by the end of the year over T years. The model assumes that w_g and γ_g are constant over time, and that the values of the risk factors X are independent over time. In the following, I will use the conventional notation:

- d_g: the number of defaults in grade g
- n_g: the number of obligors in grade g
- $\mathbf{d}$: G-long vector whose elements are n_g
- $\mathbf{n}$: G-long vector whose elements are n_g
- w_g: factor loading for grade g
- γ_g: default threshold for grade g.

Conditional on X_g, the default events of obligors in grade g are independent and can be considered individually as Bernoulli distributed with the following probabilities:

$$p_g(X_g) = \Phi\left(\frac{\gamma_g - w_g X_g}{\sqrt{1 - w_g^2}}\right). \tag{6.13}$$

Therefore, d_g the number of defaults in grade g conditional on X_g follows a binomial distribution

$$L(d_g \mid X_g) = \binom{n_g}{d_g} p_g(X_g)^{d_g} (1 - p_g(X_g))^{n_g - d_g}. \tag{6.14}$$

The joint likelihood of conditional $\mathbf{d}$ defaults on $\mathbf{X}$ is the product of the $\mathbf{G}$ conditional likelihoods specified in (6.14), because defaults are conditionally independent across grades

$$L(\mathbf{d} \mid \mathbf{X}) = \prod_{g=1}^{G} \binom{n_g}{d_g} p_g(x_g)^{d_g} (1 - p_g(x_g))^{n_g - d_g}.$$

The unconditional likelihood for $\mathbf{d}$ is thus expressed as:

$$L(\mathbf{d}) = \int_{\mathbb{R}^G} \prod_{g=1}^{G} \binom{n_g}{d_g} p_g(x_g)^{d_g} (1 - p_g(x_g))^{n_g - d_g} dF(\mathbf{x}). \tag{6.15}$$

Here, $F(\mathbf{x})$ denotes the multivariate normal cumulative distribution function of $\mathbf{X}$. The parameters in Equation (6.15) are $\mathbf{w} = (w_1, \ldots, w_G)$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_G)$, and Σ, which is the variance matrix of $\mathbf{X}$ that includes $\frac{G(G-1)}{2}$ free covariance parameters.

Note that this (6.15) likelihood corresponds only to a single time point, and the values of the $\mathbf{X}$ risk factors are assumed to be independent over time. Therefore, the overall likelihood function over the entire time period is the product of the likelihood functions at individual time points, as we will see later in the (8.1) formula.

Theoretically, we could maximize the product of (6.15) over T observations concerning all $2G + \frac{G(G-1)}{2}$ free parameters simultaneously, providing unrestricted full information maximum likelihood estimates of the parameters. However, this approach is computationally feasible only for a small G.

MLE1

Unless the common factor covariance parameters are specifically of interest, a limited information approach that does not require estimating the elements of Σ is preferable. By integrating X_g out of equation (6.14), we derive the marginal likelihood:

$$L(d_g) = \int_{\mathbb{R}} \binom{n_g}{d_g} p_g(x)^{d_g}(1 - p_g(x))^{n_g - d_g} d\Phi(x). \tag{6.16}$$

This function depends only on the parameters w_g and γ_g, allowing us to estimate w and γ by maximizing the marginal likelihood for each grade independently. This method provides our least restrictive maximum likelihood estimator (**MLE1**), which imposes no constraints on the parameters of the default model described in Chapter 5. This estimator does not use information about potential correlations in default rates across grades, making it asymptotically inefficient (except in the unrealistic special case where $\sigma_{gh} = 0$ for all $g \neq h$).

MLE2

Assumption (R1) implies that the effect of $\mathbf{X}$ on all obligors can be represented by a single standard normal scalar variable X. Under this restriction, we can rewrite (6.15) as:

$$L(\mathbf{d}) = \int_{\mathbb{R}} \prod_{g=1}^{G} \binom{n_g}{d_g} p_g(x)^{d_g}(1 - p_g(x))^{n_g - d_g} d\Phi(x) \tag{6.17}$$

Maximizing this (6.17) likelihood over $\mathbf{w}$ and $\boldsymbol{\gamma}$ that becomes our second maximum likelihood estimator (**MLE2**).

MLE3

Similarly, (R1) and (R2) can be imposed by replacing the vector $\mathbf{w}$ in equation (6.17) with a single loading w and maximizing the resulting likelihood concerning $\boldsymbol{\gamma}$ and the scalar w. This method will be referred to as our third maximum likelihood estimator (**MLE3**).

If both (R1) and (R2) hold, then all the maximum likelihood estimators described are consistent, and the (MLE3) estimator achieves the lowest possible asymptotic variance, making it the most efficient among consistent estimators. It is crucial to note, however, that in finite samples, some or all of these maximum likelihood estimators may be biased.

In the next chapter, we use Monte Carlo simulations to investigate the small sample properties of these estimators and compare them with the method of moments estimator, which was described in 6.1.

7 Empirical results

In this chapter, we examine the properties of the parameter estimates previously introduced. However, since all data are scarce in reality, we analyze the asymptotic properties on synthetically generated data. If we had data over a sufficiently long period, the choice between models would be a tradeoff, as the more constrained MLE methods are more precise (if their assumptions hold true), but the less constrained models are more robust to specification errors. The use of generated data helps not only in studying asymptotic properties but also in examining the bias of parameter fittings, as unlike with historical data, we know the theoretical values of the parameters here.

In the analysis, I use the 4 methods described earlier:

- MM: momentum estimation for each grade
- MLE1: maximum likelihood estimation for each grade
- MLE2: maximum likelihood estimation with (R1) assumption
- MLE3: maximum likelihood estimation with (R1) and (R2) assumptions

7.1 Synthetic dataset

With the satisfaction of (R1) and (R2), I generate the synthetic data using Monte Carlo simulations, where the input parameters are the T length of the time period , the w common factor loading, the PDs of the grades and the number of obligors (which for simplicity, I assumed to be constant over time).

I use the input parameters and method proposed by Gordy [10] for constructing the data, creating three grades where Grade A corresponds to S&P grades A and BBB, Grade B corresponds to the S&P BB grade, and Grade C corresponds to the S&P B grade. The real parameters for the three grades can be read from Table 1.

Grade	Factor loading	PD	Gamma	Number of obligors
A	0.45	0.0015	-2.9677	400
B	0.45	0.0100	-2.3263	250
C	0.45	0.0500	-1.6449	100

Table 1: Generated Rating Grades

In the equation (7.1), the variables X_t and $\epsilon_{i,t}$ are generated from a standard normal distribution. If the resulting value of $Y_{i,t}$ is less than the threshold value corresponding to the grade of the i-th obligor, then a default event occurs.

$$Y_{i,t} = w \cdot X_t + \sqrt{1 - w^2} \cdot \epsilon_{i,t}, \tag{7.1}$$

where $Y_{i,t}$ is the latent variable of the i-th obligor at time t, X_t is the value of the common risk factor at time t, and $\epsilon_{i,t}$ is the idiosyncratic risk value associated with the i-th obligor at time t.

To investigate the bias that varies with time horizon size, I generate data for 20, 40, 80, and 160 years, although in reality, we never have historical default rates for 80 and 160 years. For each T, we conduct 500 simulations, that is, we generate 500 datasets, and for each dataset, we perform parameter fitting with the four methods described above. By taking the 5% and 95% quantiles of the 500 fitted estimates, we obtain the confidence interval for the parameter estimation with the given method. The results are shown in Figure 4, where the median of the estimates is marked in blue, and the true parameter is indicated by a green dashed line.

As expected, increasing T results in the estimated factor loadings' means converging to 0.45 for all grades and estimators. The Table 8 shows the mean, standard deviation, percentiles, and Root Mean Squared Errors (RMSEs) calculated from the true factor loading for T=20 years. We observe that introducing more restrictions leads to more accurate estimates. The evolution of these descriptive statistics with increasing time horizons can be seen in the Table 6. Interestingly, when bias is present, it is always negative, suggesting that risk might be underestimated and potentially exposing us to greater danger than originally planned. Not surprisingly, both variance and RMSE decrease with increasing T, indicating that the properties of all four estimators improve as T increases.

For small T, the accuracy of the MM method lags significantly behind the others, and MLE1 also underperforms compared to the more restrictive estimators, and this trend becomes more evident in the percentiles. It is worth noting that, in practice, such data sizes are typically available.

The descriptive statistics for the default threshold parameter estimates are presented in Tables 7 and 9. Here, we observe that even for small T, all four estimators accurately estimate the default threshold value and, consequently, the probability of default (PD).

Overall, it can be said that the MLE3 estimator performed the best.

In reality, if we do not use the parameter restrictions, we may obtain significant bias. However, if we do use the restrictions and the assumptions are not true, the estimates will be weaker due to specification error.

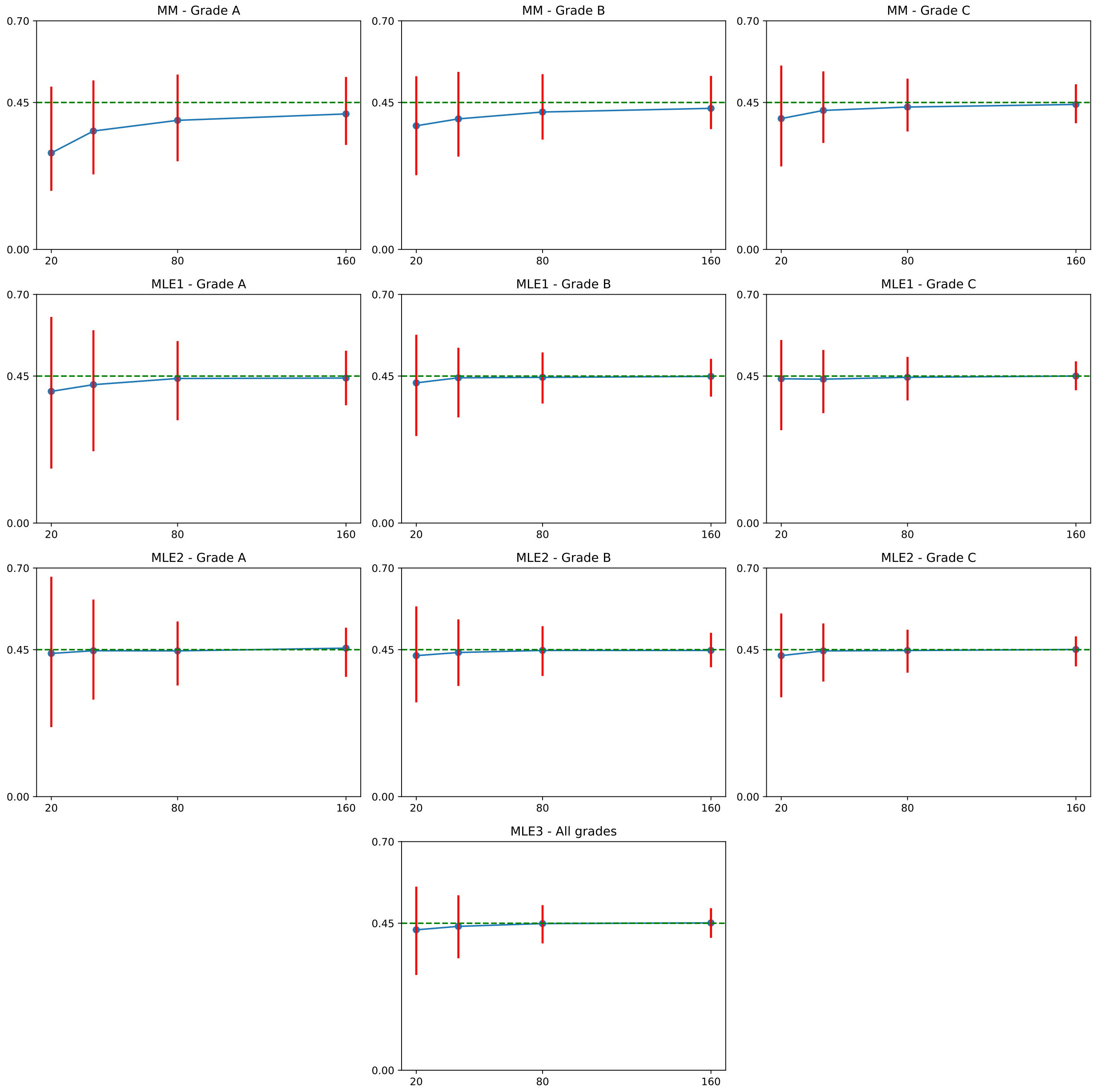


Figure 4: Estimated factor loadings by sample size

7.2 S&P historical dataset

In this chapter, I employ the previously introduced maximum likelihood and method of moments estimation techniques on historical rating data obtained from S&P. The historical default rates between 1981 and 2020 can be seen in Table 10 of the appendix. However, data on the temporal evolution of the number of examined companies were not available, so I assumed the number of obligors to be constant over time, corresponding to the 2020 data, which can be read from Table 2.

Rating	**AAA**	**AA**	**A**	**BBB**	**BB**	**B**	**CCC/C**
# of rating	8	322	1,432	1,855	1,289	2,078	238

Table 2: Number of ratings as of 1/1/2020

The quantities required for the MM estimation are calculated in Table 3. The data contains information on 7 ratings from AAA to CCC/C, but in the top 2 grades, there were either no defaults or too few defaults to reliably estimate the PD. Therefore, I will only model the ratings from A to CCC/C. Certainly, the AAA and AA grades are important in practice, but they often require additional methodologies that are beyond the scope of this thesis.

	$\overline{p}$	$E[1/\hat{n}]$	$\sqrt{V[\hat{p}]}/\overline{p}$	$\sqrt{V[p]}/\overline{p}$
AAA	0.0000	0.0092	0.0000	-
AA	0.0001	0.0030	0.0006	0.3175
A	0.0005	0.0017	0.0010	0.6048
BBB	0.0019	0.0026	0.0025	0.5760
BB	0.0086	0.0038	0.0098	0.9325
B	0.0419	0.0041	0.0320	0.7017
CCC/C	0.2492	0.0360	0.1165	0.3377

Table 3: Empirical Default Frequency and Volatility

The estimated factor loading parameters are shown in Table 4, while the estimated factor loading parameters are displayed in Table 5. We find that there is no monotonic relationship between the default correlation and the grades, and thus the default thresholds that determine them.

Some previous studies did not directly estimate the w factor loading but assumed that it can be expressed as:

$$w_g = f(\lambda(\gamma_g))$$

where f is a continuous monotonic function that maps real numbers to the interval $[-1, 1]$, and λ maps the default thresholds to the real number line. For example, the following functions are often assumed:

$$\begin{aligned} f(\lambda) &= \frac{2}{\pi} \arctan(\lambda), \\ \lambda(\gamma) &= \beta_0 + \beta_1 \gamma + \beta_2 \gamma^2. \end{aligned} \tag{7.2}$$

In this case, the beta coefficients would be estimated. For MLE2, the g function would be linear (with the constraint $\beta_2 = 0$), while for MLE3, it would be constant (with the constraints $\beta_1 = 0$ and $\beta_2 = 0$). In my dataset, I have not used such index functions, as the data did not strongly support their applicability. Future studies might explore this avenue further.

	A	**BBB**	**BB**	**B**	**CCC/C**
MM	0.3208	0.3053	0.3443	0.3280	0.3519
MLE1	0.5379	0.5072	0.4023	0.3454	0.3333
MLE2	0.2580	0.3081	0.2866	0.3296	0.2340
MLE3	0.3004	0.3004	0.3004	0.3004	0.3004

Table 4: S&P estimated w factor loadings

The estimated default thresholds according to the four methods are relatively close to each other, which is consistent with the results observed in the Monte Carlo simulations. Compared to previous studies [10] fitted on earlier datasets, I obtained similar parameters for the A, BBB, and BB rating grades. However, while the B grade has a slightly lower, the CCC/C grade has significantly higher default probabilities, which can also be observed in historical data.

	A	**BBB**	**BB**	**B**	**CCC/C**
MM	-3.2741	-2.8865	-2.3842	-1.7289	-0.6770
MLE1	-3.1696	-2.8031	-2.3656	-1.7281	-0.6574
MLE2	-3.2573	-2.8874	-2.3834	-1.7303	-0.6764
MLE3	-3.2549	-2.8902	-2.3833	-1.7328	-0.6738

Table 5: S&P estimated γ default thresholds

In the following, I will present the methodologies related to numerical implementation that emerged during the model implementation.

8 Numerical methods

In this chapter, I discuss the problems encountered during the implementation of parameter estimations and elaborate on my own solutions in detail. The literature often does not address this part, although I believe it is at least as important to obtain numerically stable solutions as to be familiar with the fundamental assumptions of the model. Therefore, I dedicate an entire chapter to it as a supplement, to ensure that my results are fully reproducible.

8.1 Integration in the Likelihood function

As previously mentioned, the essence of maximum likelihood parameter estimation is to maximize the likelihood value in equation (8.1) with respect to the parameters $\boldsymbol{\gamma}$ and $\mathbf{w}$:

$$L(\mathbf{d}) = \prod_{t=1}^{T} \int_{\mathbb{R}} \prod_{g=1}^{G} \binom{n_{g,t}}{d_{g,t}} p_g(x)^{d_{g,t}} (1 - p_g(x))^{n_{g,t} - d_{g,t}} d\Phi(x), \tag{8.1}$$

where $d_{g,t}$ is the number of defaults in grade g at time t, and $n_{g,t}$ is the number of obligors in grade g at time t.

Even at a single point in time, the likelihood value is relatively small for larger grade buckets, and this is further compounded by multiplying the values calculated for all time points. Therefore, we typically work with the log-likelihood value for numerical manageability:

$$LL(\mathbf{d}) = \sum_{t=1}^{T} \log \left(\int_{\mathbb{R}} \prod_{g=1}^{G} \binom{n_{g,t}}{d_{g,t}} p_g(x)^{d_{g,t}} (1 - p_g(x))^{n_{g,t} - d_{g,t}} d\Phi(x) \right). \tag{8.2}$$

There is no analytical formula for calculating the integral, so we must use numerical integration tools. One approach is to utilize the relationship in equation (8.3), where ϕ is the density function of the standard normal random variable, and calculate the value in equation (8.4).

$$\int_{-\infty}^{\infty} f(x)\, d\Phi(x) = \int_{-\infty}^{\infty} f(x)\phi(x)\, dx \tag{8.3}$$

$$LL(\mathbf{d}) = \sum_{t=1}^{T} \log \left(\int_{\mathbb{R}} \prod_{g=1}^{G} \binom{n_{g,t}}{d_{g,t}} p_g(x)^{d_{g,t}} (1 - p_g(x))^{n_{g,t} - d_{g,t}} \cdot \phi(x) dx \right) \tag{8.4}$$

The problem with this is that we still integrate over the entire real line while the integrand function takes non-negligible values only in a small neighborhood around 0. The numerical integration functions built into popular Python packages calculate the integral value based on an equidistant division by default, so for such a function concentrated around 0, the numerical absolute error is of the same order of magnitude as the integral value itself.

Inverse CDF method

Partly for this reason, the so-called Inverse CDF method [6] is used in practice to numerically calculate the integral, the essence of which is the transformation in equation (8.5).

$$\int_{-\infty}^{\infty} f(x)\phi(x)\,dx = \int_{0}^{1} f\left(\Phi^{-1}(y)\right)\,dy \tag{8.5}$$

Thus, the log-likelihood function value is calculated based on the below equation:

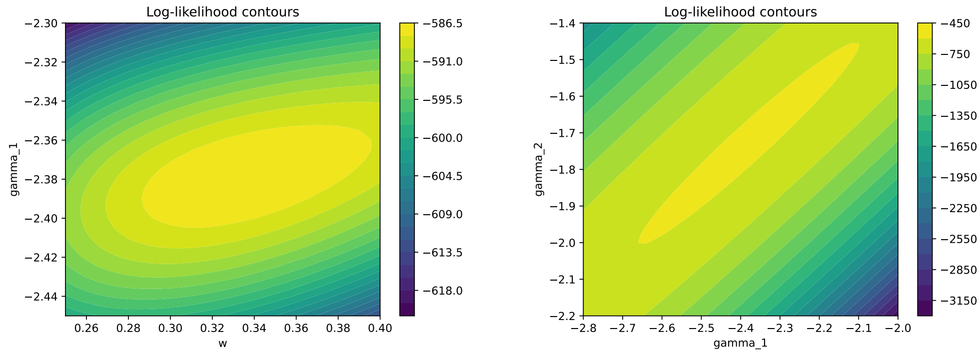
$$LL(\mathbf{d}) = \sum_{t=1}^{T} \log\left(\int_{0}^{1} \prod_{g=1}^{G} \binom{n_{g,t}}{d_{g,t}} p_g(\Phi^{-1}(y))^{d_{g,t}}(1 - p_g(\Phi^{-1}(y)))^{n_{g,t}-d_{g,t}}dy\right).$$

8.2 The unimodality of the Log-Likelihood function

Given a dataset, $\mathbf{d}$ is fixed, so this log-likelihood value can be considered a function of $\boldsymbol{\gamma}$ and $\mathbf{w}$. In the case of MLE1 method, we examine the grade classes separately, so this is a 2-parameter optimization (now w and γ are scalars) for each bucket. In the case of MLE2 estimator, if we have G grade buckets, then we need to optimize with respect to $2 \cdot G$ parameters. In the case of MLE3, we assume a common factor loading and different threshold values, so here we need to maximize the log-likelihood value with respect to $G + 1$ parameters.

Before starting the optimization, it is worth examining the shape of the multidimensional log-likelihood function to see if it is unimodal, meaning whether the **local maximum equals the global maximum**. It is worth examining the contour lines in the cross-sectional plots of the multidimensional log-likelihood function because, for non-negative functions, concavity is a special case of log-concavity, which is a part of quasi-concavity, and all of these are unimodal (if they have a local maximum on their domain).

If we have 2 grades and consider the log-likelihood function with respect to 2 thresholds and a common factor loading, then the cross-sectional contour lines of the function are shown in Figure 5. The contour lines appear convex, suggesting that the function is unimodal, although it should be noted that unimodality is not generally proven. It should be noted that the success of some optimizers is fundamentally dependent on convexity.



(a) w and gamma1 (b) gamma1 and gamma2

Figure 5: Plane sections of the log-likelihood function

8.3 Variable change

To enhance the numerical stability of the optimization, I use a variable change introducing **a** and **b** as variable vectors instead of $\boldsymbol{\gamma}$ and **w**, as shown in equation (8.6).

$$\begin{aligned} a_g &= -\frac{w_g}{\sqrt{1-w_g^2}} \\ b_g &= \frac{\gamma_g}{\sqrt{1-w_g^2}} \end{aligned} \tag{8.6}$$

This simplifies the $p_g(x)$ function from the form in equation (6.13) to the form in equation (8.7).

$$p_g\left(x\right) = \Phi\left(a_g \cdot x + b_g\right). \tag{8.7}$$

Then, we can consider the likelihood function as a function of **a** and **b**. After maximizing the log-likelihood with respect to **a** and **b**, we can recover the estimated **w** and $\boldsymbol{\gamma}$ parameters based on the equations in (8.8).

$$\begin{aligned} w_g &= -\frac{a_g}{\sqrt{a_g^2+1}} \\ \gamma_g &= b_g \cdot \sqrt{1-w_g^2} \end{aligned} \tag{8.8}$$

8.4 Finding the maximum in multiple dimensions

Since calculating the log-likelihood function is computationally intensive and the gradient can only be calculated numerically, likely magnifying the numerical error, I do not recommend a gradient-based optimizer to find the optimal parameters. Several multidimensional optimizers were tested, leading to the selection of the Nelder-Mead method, specifically its SciPy implementation based on the paper by Fuchang Gao and Lixing Han [8], and now I will briefly introduce this method.

Nelder-Mead method

The Nelder-Mead simplex algorithm is a popular direct search method used for solving unconstrained optimization problems. It is especially useful when the gradient of the objective function is unavailable or unreliable. This algorithm, introduced by John Nelder and Roger Mead [13], is designed to minimize a scalar-valued function $f : \mathbb{R}^n \to \mathbb{R}$.

The algorithm uses a simplex, which is a geometric figure consisting of $n+1$ vertices in n-dimensional space, to iteratively approximate the minimum of the objective function.

Each iteration of the algorithm involves evaluating the function at the vertices of the simplex and updating the simplex based on these evaluations. The algorithm operates using four primary operations: reflection, expansion, contraction, and shrinkage. These operations adjust the position of the simplex to explore the search space and find the minimum of the f objective function.

Algorithm Description:

- *Initialization:* The algorithm begins with an initial simplex defined by $n + 1$ vertices, $\{x_1, x_2, \ldots, x_{n+1}\}$. These vertices are ordered such that $f(x_1) \leq f(x_2) \leq \cdots \leq f(x_{n+1})$. The centroid x_c of the best n vertices is computed as:

$$x_c = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

- *Reflection:* The reflection point x_r is calculated as:

$$x_r = x_c + \alpha(x_c - x_{n+1}),$$

  where $\alpha > 0$ is the reflection coefficient. If $f(x_1) \leq f(x_r) < f(x_n)$, then x_r replaces x_{n+1} and the algorithm proceeds to the next iteration.

- *Expansion:* If $f(x_r) < f(x_1)$, an expansion is performed to explore further in the reflected direction. The expansion point x_e is computed as:

  $$x_e = x_c + \beta(x_r - x_c),$$

  where $\beta > 1$ is the expansion coefficient. If $f(x_e) < f(x_r)$, then x_e replaces x_{n+1}; otherwise, x_r replaces x_{n+1}.

- *Contraction:* If $f(x_n) \leq f(x_r) < f(x_{n+1})$, an outside contraction is performed:

  $$x_{oc} = x_c + \gamma(x_r - x_c),$$

  where $0 < \gamma < 1$ is the contraction coefficient. If $f(x_{oc}) \leq f(x_r)$, x_{oc} replaces x_{n+1}. If $f(x_r) \geq f(x_{n+1})$, an inside contraction is performed:

  $$x_{ic} = x_c - \gamma(x_r - x_c).$$

  If $f(x_{ic}) < f(x_{n+1})$, x_{ic} replaces x_{n+1}; otherwise, a shrinkage step is executed.

- *Shrinkage:* If none of the above steps produce a better point, the simplex is contracted towards the best vertex:

  $$x_i = x_1 + \delta(x_i - x_1), \quad i = 2, \ldots, n+1,$$

  where $0 < \delta < 1$ is the shrinkage coefficient. This step reduces the size of the simplex and ensures convergence.

In the standard implementation, the parameters for these operations are typically chosen as $\{\alpha, \beta, \gamma, \delta\} = \{1, 2, 0.5, 0.5\}$.

Stopping Criteria: The algorithm terminates when the standard deviation of the function values at the vertices of the simplex falls below a predefined tolerance.

Convergence Properties: The convergence of the Nelder-Mead algorithm is well-studied in low dimensions but less understood in higher dimensions. For strictly convex functions in two dimensions, the algorithm converges to the global minimum under certain conditions. However, there are known examples where the algorithm fails to converge to a stationary point, particularly for higher-dimensional or poorly scaled problems.

Adaptive Variants: To address the limitations of the standard Nelder-Mead algorithm in high dimensions, adaptive variants have been proposed. These variants adjust the algorithm's parameters based on the problem's dimensionality and the current iteration's performance. For instance, adaptive schemes may modify the reflection, expansion, contraction, and shrinkage coefficients to improve convergence rates and robustness.

For the MLE2 and MLE3 maximum likelihood estimators, I used an adaptive variant of the Nelder-Mead method [8].

8.5 One-dimensional search

The advantage of the method of moments is that it relies on far fewer assumptions than the maximum likelihood methods. Additionally, thanks to analytical formulas, it is much more manageable both numerically and in terms of computational demand. I only used an one-dimensional search to find the optimal factor loading. The objective function is given by the square of the difference between the two sides of equation (6.12), where the left side is calculated from the empirical data, and the right side is a function of the factor loading.

I used the Brent's algorithm [3] to find the minimum point of the function, which combines the bisection method with the secant method. Furthermore it uses inverse parabolic interpolation when possible to speed up convergence of golden section method.

The objective function is convex, so the local minimum found using Brent's algorithm is also the global minimum.

9 Conclusion

This thesis has explored the Vasicek credit risk model, focusing on its parameter estimation methods and their applications to synthetic and historical data. The results underscored that MLE, particularly when incorporating more restrictions, offered superior accuracy and reduced bias, aligning with previous studies.

While the method of moments is numerically well manageable, the somewhat more sophisticated maximum likelihood method posed numerical challenges, and it was instructive to tackle these.

Beyond understanding the Vasicek framework, i.e., the Gaussian copula's mathematical background and the business context, it was fascinating to see what numerical methods the model implementation requires.

In my further work/research, I will likely encounter this methodology again, and the following questions can still be explored:

- it is interesting to examine whether non-global optimization can always be used, i.e., whether the log-likelihood function is always unimodal
- investigating other methods, such as Gordy and Heitfield [10], who modeled default correlation as a function of the default threshold.

References

[1] Altman, E., Resti, A., & Sironi, A. (2004). Default recovery rates in credit risk modelling: a review of the literature and empirical evidence. Economic Notes, 33(2), 183-208.

[2] Basle Committee on Banking Supervision. (1999). A New Capital Adequacy Framework: Consultative Paper. The Committee.

[3] Brent, R. P. (1971). An algorithm with guaranteed convergence for finding a zero of a function. The computer journal, 14(4), 422-425.

[4] Chatterjee, S. (2015). Modelling credit risk. Centre for Central Banking Studies, Bank of England.

[5] Demey, P., Jouanin, J. F., Roget, C., & Roncalli, T. (2007). Maximum likelihood estimate of default correlations. Available at SSRN 1032590.

[6] Devroye, L. (1986). Sample-Based Non-Uniform Random Variate Generation. Institute of Electrical and Electronics Engineers (IEEE).

[7] Düllmann, K., & Scheule, H. (2003). Determinants of the asset correlations of German corporations and implications for regulatory capital. Deutsches Bundesbank.

[8] Gao, F., & Han, L. (2012). Implementing the Nelder-Mead simplex algorithm with adaptive parameters. Computational Optimization and Applications, 51(1), 259-277.

[9] Gordy, M. B. (2000). A comparative anatomy of credit risk models. Journal of Banking & Finance, 24(1-2), 119-149.

[10] Gordy, M., & Heitfield, E. (2002). Estimating default correlations from short panels of credit rating performance data. Working Paper, Federal Reserve Board.

[11] Gupton, G. (1997). Credit metrics technical document.

[12] Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. The Journal of Finance, 29(2), 449-470.

[13] Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. The Computer Journal, 7(4), 308-313.

[14] Vasicek, O. A. (1984). Credit valuation.

[15] Witzany, J. (2017). Credit Risk Management. Springer.

Appendix

	MM			MLE1			MLE2			MLE3
	A	B	C	A	B	C	A	B	C	All
T=20										
Mean	0.3275	0.3817	0.4050	0.4062	0.4284	0.4322	0.4390	0.4319	0.4320	0.4293
Std. Dev.	0.0949	0.0950	0.0934	0.1446	0.0972	0.0798	0.1396	0.0872	0.0767	0.0790
RMSE	0.1549	0.1169	0.1036	0.1510	0.0994	0.0817	0.1398	0.0890	0.0787	0.0816
T=40										
Mean	0.3586	0.4057	0.4273	0.4166	0.4361	0.4398	0.4473	0.4409	0.4437	0.4381
Std. Dev.	0.0947	0.0776	0.0680	0.1183	0.0663	0.0576	0.0923	0.0621	0.0552	0.0604
RMSE	0.1315	0.0893	0.0717	0.1228	0.0677	0.0584	0.0922	0.0628	0.0555	0.0615
T=80										
Mean	0.3987	0.4272	0.4390	0.4392	0.4459	0.4458	0.4440	0.4464	0.4468	0.4479
Std. Dev.	0.0780	0.0619	0.0493	0.0769	0.0475	0.0403	0.0599	0.0460	0.0396	0.0365
RMSE	0.0933	0.0659	0.0505	0.0776	0.0476	0.0405	0.0601	0.0461	0.0397	0.0366
T=160										
Mean	0.4163	0.4371	0.4451	0.4449	0.4481	0.4501	0.4521	0.4488	0.4479	0.4519
Std. Dev.	0.0649	0.0488	0.0366	0.0525	0.0358	0.0278	0.0472	0.0319	0.0297	0.0278
RMSE	0.0731	0.0504	0.0369	0.0526	0.0358	0.0278	0.0470	0.0318	0.0296	0.0279

Table 6: Distribution of estimated factor loadings by sample size

	MM			MLE1			MLE2			MLE3		
	A	B	C	A	B	C	A	B	C	A	B	C
T=20												
Mean	-3.0180	-2.3562	-1.6658	-2.9698	-2.3308	-1.6451	-3.0003	-2.3434	-1.6596	-3.0015	-2.3486	-1.6631
Std. Dev.	0.1906	0.1501	0.1315	0.1600	0.1424	0.1255	0.2516	0.1364	0.1224	0.1854	0.1436	0.1323
RMSE	3.4732	2.8102	2.1198	3.4235	2.7844	2.0989	3.4594	2.7967	2.1132	3.4564	2.8023	2.1172
T=40												
Mean	-2.9915	-2.3430	-1.6525	-2.9844	-2.3372	-1.6502	-2.9809	-2.3351	-1.6529	-2.9958	-2.3340	-1.6508
Std. Dev.	0.1185	0.0991	0.0883	0.1225	0.0997	0.0859	0.1199	0.0944	0.0835	0.1376	0.1039	0.0972
RMSE	3.4436	2.7947	2.1044	3.4365	2.7890	2.1020	3.4330	2.7867	2.1045	3.4485	2.7860	2.1031
T=80												
Mean	-2.9789	-2.3301	-1.6445	-2.9729	-2.3279	-1.6445	-2.9679	-2.3249	-1.6419	-2.9711	-2.3298	-1.6467
Std. Dev.	0.0904	0.0711	0.0599	0.0922	0.0698	0.0614	0.0886	0.0703	0.0612	0.0871	0.0704	0.0639
RMSE	3.4301	2.7810	2.0954	3.4241	2.7787	2.0954	3.4191	2.7757	2.0928	3.4223	2.7807	2.0976
T=160												
Mean	-2.9755	-2.3292	-1.6483	-2.9629	-2.3242	-1.6413	-2.9698	-2.3248	-1.6468	-2.9635	-2.3243	-1.6417
Std. Dev.	0.0624	0.0479	0.0429	0.0656	0.0529	0.0464	0.0638	0.0550	0.0443	0.0648	0.0527	0.0459
RMSE	3.4260	2.7796	2.0987	3.4136	2.7747	2.0922	3.4204	2.7754	2.0973	3.4141	2.7748	2.0922

Table 7: Distribution of estimated default thresholds by sample size.

	MM			MLE1			MLE2			MLE3
	Grade A	Grade B	Grade C	Grade A	Grade B	Grade C	Grade A	Grade B	Grade C	All grade
Mean	0.3275	0.3817	0.4050	0.4062	0.4284	0.4322	0.4390	0.4319	0.4320	0.4293
Std. Dev.	0.0949	0.0950	0.0934	0.1446	0.0972	0.0798	0.1396	0.0872	0.0767	0.0790
RMSE	0.1549	0.1169	0.1036	0.1510	0.0994	0.0817	0.1398	0.0890	0.0787	0.0816
Percentile										
2.5	0.1423	0.1954	0.2294	0.1461	0.2344	0.2774	0.1684	0.2717	0.2871	0.2698
5.0	0.1794	0.2273	0.2542	0.1667	0.2664	0.2842	0.2127	0.2886	0.3042	0.2915
50.0 (Med.)	0.2955	0.3786	0.4007	0.4032	0.4292	0.4417	0.4384	0.4317	0.4317	0.4299
95.0	0.4985	0.5303	0.5631	0.6312	0.5765	0.5605	0.6734	0.5825	0.5608	0.5622
97.5	0.5416	0.5734	0.5958	0.6690	0.6057	0.5725	0.7145	0.6013	0.5813	0.5788

Table 8: Distribution of estimated factor loadings for $T = 20$.

	MM			MLE1			MLE2			MLE3		
	A	B	C	A	B	C	A	B	C	A	B	C
Mean	-3.0180	-2.3562	-1.6658	-2.9698	-2.3308	-1.6451	-3.0003	-2.3434	-1.6596	-3.0015	-2.3486	-1.6631
Std. Dev.	0.1906	0.1501	0.1315	0.1600	0.1424	0.1255	0.2516	0.1364	0.1224	0.1854	0.1436	0.1323
RMSE	3.4732	2.8102	2.1198	3.4235	2.7844	2.0989	3.4594	2.7967	2.1132	3.4564	2.8023	2.1172
Percentile												
2.5	-3.3706	-2.6788	-1.9189	-3.2658	-2.6470	-1.9035	-3.3681	-2.6268	-1.9229	-3.4212	-2.6241	-1.9447
5.0	-3.3706	-2.6045	-1.8882	-3.2259	-2.5727	-1.8609	-3.2944	-2.5657	-1.8685	-3.3498	-2.5902	-1.8758
50.0 (Med.)	-2.9944	-2.3495	-1.6621	-2.9670	-2.3261	-1.6406	-2.9872	-2.3316	-1.6468	-2.9856	-2.3458	-1.6643
95.0	-2.7337	-2.1107	-1.4572	-2.7015	-2.1050	-1.4437	-2.7098	-2.1263	-1.4802	-2.7335	-2.1298	-1.4574
97.5	-2.6521	-2.0790	-1.4081	-2.6457	-2.0730	-1.4160	-2.6660	-2.0854	-1.4587	-2.6920	-2.0885	-1.4125

Table 9: Distribution of default thresholds for $T = 20$.

Year	AAA	AA	A	BBB	BB	B	CCC/C
1981	0.00	0.00	0.00	0.00	0.00	2.33	0.00
1982	0.00	0.00	0.21	0.35	4.24	3.18	21.43
1983	0.00	0.00	0.00	0.34	1.16	4.70	6.67
1984	0.00	0.00	0.00	0.68	1.14	3.49	25.00
1985	0.00	0.00	0.00	0.00	1.50	6.53	15.38
1986	0.00	0.00	0.18	0.34	1.33	8.45	23.08
1987	0.00	0.00	0.00	0.00	0.38	3.13	12.28
1988	0.00	0.00	0.00	0.00	1.05	3.68	20.37
1989	0.00	0.00	0.18	0.61	0.73	3.40	33.33
1990	0.00	0.00	0.00	0.58	3.57	8.56	31.25
1991	0.00	0.00	0.00	0.55	1.69	13.84	33.87
1992	0.00	0.00	0.00	0.00	0.00	6.99	30.19
1993	0.00	0.00	0.00	0.00	0.70	2.62	13.33
1994	0.00	0.00	0.14	0.00	0.28	3.09	16.67
1995	0.00	0.00	0.00	0.17	0.99	4.59	28.00
1996	0.00	0.00	0.00	0.00	0.45	2.91	8.00
1997	0.00	0.00	0.00	0.25	0.19	3.52	12.00
1998	0.00	0.00	0.00	0.41	0.82	4.64	42.86
1999	0.00	0.17	0.18	0.20	0.95	7.31	33.82
2000	0.00	0.00	0.27	0.37	1.16	7.71	35.96
2001	0.00	0.00	0.27	0.34	2.98	11.56	45.45
2002	0.00	0.00	0.00	1.02	2.90	8.20	44.44
2003	0.00	0.00	0.00	0.23	0.59	4.07	32.93
2004	0.00	0.00	0.08	0.00	0.44	1.45	16.30
2005	0.00	0.00	0.00	0.07	0.31	1.75	9.09
2006	0.00	0.00	0.00	0.00	0.30	0.82	13.33
2007	0.00	0.00	0.00	0.00	0.20	0.25	15.24
2008	0.00	0.38	0.39	0.49	0.81	4.11	27.27
2009	0.00	0.00	0.22	0.55	0.75	11.03	49.46
2010	0.00	0.00	0.00	0.00	0.58	0.87	22.83
2011	0.00	0.00	0.00	0.07	0.00	1.68	16.42
2012	0.00	0.00	0.00	0.00	0.30	1.58	27.52
2013	0.00	0.00	0.00	0.00	0.10	1.65	24.67
2014	0.00	0.00	0.00	0.00	0.00	0.78	17.51
2015	0.00	0.00	0.00	0.00	0.16	2.42	26.67
2016	0.00	0.00	0.00	0.06	0.47	3.76	33.17
2017	0.00	0.00	0.00	0.00	0.08	1.00	26.56
2018	0.00	0.00	0.00	0.00	0.00	0.99	27.18
2019	0.00	0.00	0.00	0.11	0.00	1.49	29.76
2020	0.00	0.00	0.00	0.00	0.93	3.52	47.48

Table 10: Historical Default Rates by Credit Rating

Source: Standard & Poor's Global Ratings, Annual Default Study

Magyar nyelvű összefoglaló

A szakdolgozatom célja a Vasicek hitelkockázati modell bemutatása és paraméterbecsléseinek vizsgálata. A modell lényege a csődvalószínűségek és ezek korrelációjának modellezése egy egyfaktoros Gauss-kopula segítségével. A módszer annyira elterjedt, hogy a Bázel II óta a tőkeszámítás is ezen alapszik.

A dolgozatom elején a legfontosabb hitelkockázati fogalmak definiálása mellett kitérek a tőkeszámítás alapjaira is a Bázeli keretrendszerben. Ezután Merton strukturált modelljét mutatom be, amely a csődök modellezésének alapját adja. Feltételezi, hogy akkor csődöl be egy cég, ha eszközeinek értéke a kötelezettségeinek mértéke alá csökken. A következő fejezetben a CreditMetrics módszertanáról írok, amely mára a portfólió hitelkockázat mérésének egyik standardjává vált. Alapfeltevése, hogy az egyes adósságinstrumentumok értékét a kockázati besorolásuk határozza meg.

A Vasicek modellt először többfaktoros modellként vezetem be, majd visszavezetem egyfaktorosra. A modell alapfeltevésein túl kifejtem azt is, hogy milyen további korlátozásokat lehetne bevezetni a szabadságfokok csökkentése érdekében, hogy pontosabb paraméterbecsléseket kapjunk. A gyakorlatban egyszerűsége és könnyű használata miatt a momentum módszer az elterjedt eljárás a paraméterek becslésére. Gordy és Heitfield [10] rámutatott, hogy rövid időtávú adatok esetén ez jelentős negatív torzítással rendelkezik, ezért maximum likelihood becslési módszereket mutattak be. Ezen három maximum likelihood esztimátor egyre több alapfeltevést használ, és ezáltal egyre kevesebb paraméterrel rendelkezik.

Az említett módszereket implementáltam Pythonban, és generált adatokon megvizsgáltam a becslések tulajdonságait. Az előző tanulmányokkal egybehangzó eredményeket kaptam, vagyis ha igazak az alapfeltevések, akkor a leginkább korlátozott modell adja a legpontosabb becslést. Azonban a gyakorlatban a feltételezések megkérdőjelezhetők, és ezáltal modell specifikációs hiba torzíthatja a becslést. Az S&P által közzétett 1981 és 2020 közötti csődráták adathalmazára is alkalmaztam a paraméterbecsléseket, és a csődvalószínűségeket meghatározó küszöbértékekre hasonló értékeket kaptam, mint a korábbi tanulmányok, amelyek még 2002-ig tartó adatokkal dolgoztak.

A dolgozatomban külön fejezetet szenteltem az implementálás során előjött numerikus módszereknek, ahol a numerikus integráláson túl érintem az egy- illetve többdimenziós optimalizálás témakörét is.

További kutatási irány lehet a log-likelihood függvény logkonkavitásának vizsgálata, hogy megbizonyosodjunk arról, hogy az optimalizálással kapott lokális optimum egyben globális optimum is.