ACCURACY AND CONVERGENCE OF THE ASYMPTOTIC SINGLE RISK FACTOR FORMULA IN A LARGE CREDIT PORTFOLIO

by

Madelyn R. Houser

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Science in Mathematical Sciences

Summer 2016

ACCURACY AND CONVERGENCE OF THE ASYMPTOTIC SINGLE RISK FACTOR FORMULA IN A LARGE CREDIT PORTFOLIO

by

Madelyn R. Houser

Approved: ______________________________
Pak-Wing Fok, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: ______________________________
Louis Rossi, Ph.D.
Chair of the Department of Mathematical Sciences

Approved: ______________________________
George H. Watson, Ph.D.
Dean of the College of Arts and Sciences

Approved: ______________________________
Ann L. Ardis, Ph.D.
Senior Vice Provost for Graduate and Professional Education

ACKNOWLEDGMENTS

I would like to thank and acknowledge my thesis advisor, Dr. Fok, without whom this project would have been impossible. His mathematical guidance, his constructive suggestions, and his willingness to take on this project have been invaluable. I would also like to thank my thesis committee, Dr. Edwards and Dr. Fleischhacker, for reviewing and critiquing my work.

This project was initiated at the 2015 Graduate Student Mathematical Modeling Camp (GSMMC) at Rensselaer Polytechnic Institute. I would like to thank Dr. Schwendeman for organizing the conference, as well as my fellow conference collaborators who worked on this problem: Nathan Gold, Hsuan-Wei Lee, Dan Lu, Matthew Moye, Xiaoqi Zhang, and Yingxiang Zhou. A special thanks to Matthew Moye who provided a basis for some of the numerical code used in this project.

I would like to thank Dr. Rossi, Chair of the Mathematics department, and Dr. Bhatnagar, faculty advisor of the Association for Women in Math, for their mentorship, encouragement, and support which have shaped me into the mathematician I am today. A special thanks to the many past and present graduate students with whom I have studied, collaborated, laughed, and, at times, commiserated.

I would be remiss without acknowledging my wonderful and supportive family, particularly my parents. Thank you for encouraging me, praying for me, and loving me unconditionally. Lastly, words fall short in expressing my deep gratitude and thankfulness for Amy Janett. Your friendship, collaboration, and unwavering belief in me have been a blessing and have made the past two years bearable.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

According to international banking standards, all financial institutions must classify the risks associated with the credit portfolios they hold. Mathematical approximations of default distributions are among the most common ways of assessing such risk. Understanding the errors of these approximations is crucial for generating reliable credit portfolio default risk calculations. We consider the error associated with Vasicek's Asymptotic Single Risk Factor model for the cumulative distribution of losses in a portfolio of N companies. We analytically and numerically verify the scaling of the error to be $\mathcal{O}(N^{-1})$, scaling as the reciprocal of the number of companies. Our results provide insight into the error associated with one of the most commonly used credit risk models and serve as a model for future work in examining the errors of more complex, hierarchical structural models.

Chapter 1

INTRODUCTION

A credit portfolio is a group of loans or lines of credit issued by a bank or financial institution generating interest over a fixed time frame. Banks, financial institutions, and investors create credit portfolios to maximize returns while minimizing the risk of losses due to defaults. A default occurs when the assets of a company in the portfolio fall below a threshold level and they are unable to pay back the loan. Because institutional financial risk results from either market risks or credit risks [1], understanding the risks of default associated with credit portfolios remains of crucial importance to financial institutions and portfolio managers [2] [3]. Further, as part of international banking standards, financial institutions often use statistics derived from portfolio loss distributions to set capital reserve requirements to survive periods of economic distress [4].

Indeed, credit rating agencies, such as Standard & Poor's and Moody's [5], exist to independently rate the credit worthiness of companies and loans (i.e. how able or likely a company is to repay a given loan over a fixed time frame and what interest should be charged as a safeguard against default). To distribute and minimize risks, portfolio managers rely on these ratings and other financial models to understand risk concentration within a portfolio; that is, the likelihood of certain groups of companies in a portfolio defaulting together. Companies are known to be linked to each other and the greater economy in a complex system of correlations, meaning companies in a portfolio may be more likely to default in unison as a result of underlying correlation in asset returns. Thus, credit portfolio managers rely on economic and mathematical models, as well as numerical simulations, to gain insight into the default risks inherent in portfolios.

One such family of models is known as structural factor models. These models, based on Wiener processes, allow each company in a portfolio to be correlated through one, two, or more risk factors, such as a global and regional/sector economic risk factor [4], [6]. The simplest model has a single global risk factor correlating all companies within a portfolio [7], but hierarchical multi-factor models have also been explored [8], [9]. In addition, portfolios with high exposure concentration (when a few large loans dominate the portfolio) have been studied using structural factor models [10]. From these models, one can calculate default probabilities for companies in a portfolio, as well as financial indicators such as expected shortfall and value-at-risk, important metrics for assessing the capital needed to cover losses or seasons of extreme financial stress [11].

Numerical simulations of factor models provide an accurate and clear picture of the risks associated with a portfolio, but such simulations can be time consuming and costly [11]. Thus, mathematicians, economists, and financiers create analytic formulas and approximations to key indicators of portfolio health and risk from these models. However errors are always associated with approximations; therefore, calculating and quantifying errors remains a critical part of approximating portfolio risks. In this thesis, we explore the Vasicek Asymptotic Single Risk Factor (ASRF) model [7]. Our main contribution is the exploration and derivation of the truncation error for the ASRF model. We derive the analytic approximation of the loss distribution, the cumulative loss distribution, and the Value-at-Risk in chapter 2. In chapter 3, we explore the error associated with our approximation and derive bounds for the cumulative distribution using results from probability theory. Chapter 4 contains asymptotic analysis of the Vasicek approximation and presents numerical and analytic results of the scaling of the error. We end with a summary of our findings in chapter 5.

Chapter 2

SINGLE RISK FACTOR MODEL

2.1 Derivation of company probability of default for the Asymptotic Single Risk Factor model

Considering a portfolio of N company loans, structural models allow portfolio managers to track the joint evolution of individual companies' assets and default probability when companies are correlated through global, sector, or regional economies [4] [9]. The simplest structure model is given by a single risk factor, i.e. each company's assets are correlated through a global economic risk factor and includes an idiosyncratic risk factor. This is known as the single factor Vasicek model [7] and is given by

$$z_i = \sqrt{\rho}\hat{\epsilon} + \sqrt{1-\rho}\epsilon_i \tag{2.1}$$

$$0 < \rho < 1$$

where z_i is the asset return for company i, $\hat{\epsilon}$ is a global economic factor, and ϵ_i is the company specific idiosyncratic risk factor. Equation (2.1) is derived from an underlying logarithmic Wiener process, so that $\hat{\epsilon} \sim N(0,1)$ and $\epsilon_i \sim N(0,1)$. Further, ρ is the correlation of assets between any two companies and is assumed to be positive.

To calculate the loss distribution of the portfolio, we can calculate the cumulative distribution function (CDF) of the return of the portfolio. From the CDF, we are able to calculate the Value-at-Risk (VaR), which is the inverse of the CDF. With a given confidence level, q, the VaR_q is defined as the q-quantile of the portfolio loss distribution and is given mathematically by [10]

$$\text{VaR}_q = \inf\{x : P(R_N^* \leq x) \geq q\} \tag{2.2}$$

All VaR calculations are associated with a given q confidence level (typically 90, 95, or 99 percent) and will be denoted VaR_q. Common holding periods are 1, 2, 10 days, and 1 month [11] for stock portfolios, but typically much longer for portfolios of bonds. VaR_q is an important financial metric not only for risk management, but also for financial control and reporting [5]. Indeed, the VaR_q forms a major component of the Basel II Accord which regulates how much capital a financial institution must maintain against their institutional and credit risks, and is a possible disclosure method to the U.S. Securities and Exchange Commission for reporting quantitative measures of market risk [11] [12].

Thus, in order for banks and financial institutions to understand the risks associated with portfolios, a full understanding of the loss distribution must be explored. We define the loss of the portfolio with N companies over a given time horizon, T, as

$$R_N^*(T) = \sum_{i=1}^{N} w_i R_i(z_i, T) \tag{2.3}$$

where T is the time horizon, w_i is the weight of the i^{th} loan as a proportion of the entire portfolio (in this thesis, we will always take $w_i = 1/N$), and $R_i(z_i, T)$ is the loss on investment for the i^{th} company in the portfolio. Thus, the loss is simply the fraction of companies to default. Because the $R_i(z_i)$'s are Bernoulli random variables, R_N^* is $\text{Bin}(N, p)$ with p the probability of default.

Hence for a given time horizon, T, we want to find the distribution of

$$R_N^* = \frac{1}{N} \sum_{i=1}^{N} R_i(z_i). \tag{2.4}$$

Conditioned on the global risk, $\hat{\epsilon}$, the R_i's are independent, each with the same distribution function $F_{R_i}(x)$. We can use the central limit theorem to approximate the CDF of $R_N^*|\hat{\epsilon}(x)$, denoted $F_{R_N^*|\hat{\epsilon}}$, by $\Phi\left(\frac{x-m}{\sigma}\right)$, where m is the mean of $F_{R_N^*}(x)$ and σ is the standard deviation of $F_{R_N^*}(x)$. Note that

$$\text{mean}_{R_i} = \alpha_i = \mathbf{E}\left(R_i\right) = p(\hat{\epsilon}) \tag{2.5}$$

and

$$\text{mean} = m = \sum_{i=1}^{N} \alpha_i = p \tag{2.6}$$

where $p(\hat{\epsilon})$ is the probability of default for the i^{th} company. Futher,

$$\begin{aligned} \sigma &= \sqrt{\sigma_1^2 + \cdots + \sigma_N^2} \\ &= \left(\frac{p(1-p)}{N}\right)^{1/2}, \end{aligned} \tag{2.7}$$

we have

$$\Phi\left(\frac{x-m}{\sigma}\right) = \Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right). \tag{2.8}$$

Hence, the conditioned CDF of R_N^*, $F_{R_N^*|\hat{\epsilon}}(x, \hat{\epsilon})$, is approximated by $\Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right)$.

$$F_{R_N^*|\hat{\epsilon}}(x, \hat{\epsilon}) \simeq \Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right) \tag{2.9}$$

But R_N^* is, however, the average of N dependent Bernoulli random variables due to equations (2.1) and (2.11). To find the unconditioned CDF, $F_{R_N^*}$, we start by finding the probability density function (PDF) of R_N^*, denoted $f_{R_N^*}$. We first condition on $\hat{\epsilon}$ to obtain independent Bernoulli random variables, apply the central limit theorem, and then relax the conditioning by using the law of total probability.

We start by calculating the return for each company in the portfolio. Note that by fixing $\hat{\epsilon}$, the asset return of each company becomes independent such that

$$z_i \big| \hat{\epsilon} \sim N(\sqrt{\rho}\hat{\epsilon}, 1-\rho). \tag{2.10}$$

Then the loss distribution for each i company in the portfolio is given by

$$R_i(z_i \big| \hat{\epsilon}) = \begin{cases} 0 & : z_i|\hat{\epsilon} > \theta_i \\ 1 & : z_i|\hat{\epsilon} \leq \theta_i \end{cases} \tag{2.11}$$

where θ_i is the threshold of loss for the i^{th} company in the portfolio (i.e. should the assets for company i fall below θ_i, we would expect a default on the loan). We assume identical distribution for each company resulting in $\theta_i = \theta$ and $p_i(\hat{\epsilon}) = p(\hat{\epsilon})$. This is justified given homogeneous portfolios.

First, notice that the probability of default for each company conditioned on $\hat{\epsilon}$ is

$$p(\hat{\epsilon}) = P(z_i < \theta \big| \hat{\epsilon}) = \int_{-\infty}^{\theta} \frac{1}{\sqrt{2\pi(1-\rho)}} \exp\left\{ -\frac{(t-\sqrt{\rho}\hat{\epsilon})^2}{2(1-\rho)} \right\} dt. \tag{2.12}$$

Making a change of variables in equation (2.12), we obtain

$$p(\hat{\epsilon}) = \frac{1}{\sqrt{1-\rho}} \int_{-\infty}^{\frac{\theta-\sqrt{\rho}\hat{\epsilon}}{\sqrt{1-\rho}}} \frac{1}{\sqrt{2\pi}} \exp\left\{ \frac{-m^2}{2} \right\} \sqrt{1-\rho}\, dm = \Phi\left(\frac{\theta-\sqrt{\rho}\hat{\epsilon}}{\sqrt{1-\rho}} \right), \tag{2.13}$$

where Φ denotes the cumulative distribution function for standard normal distribution $N(0,1)$,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy. \tag{2.14}$$

The portfolio loss distribution conditioned on $\hat{\epsilon}$ is the probability of k out of N companies defaulting and follows a binomial distribution,

$$P(k \text{ defaults}|\hat{\epsilon}) = \binom{N}{k} p_i(\hat{\epsilon})^k (1-p_i(\hat{\epsilon}))^{N-k} \tag{2.15}$$

Because $R_N^*|\hat{\epsilon}$ is a weighted sum of independent Bernoulli random variables, the portfolio loss is binomial distributed, $\text{Bin}(N,p)$. And since the number of Bernoulli random variables (the number of companies in our portfolio) is taken to infinity, we use the central limit theorem to approximate the binomial distribution with a normal distribution to obtain

$$\binom{N}{k} p_i(\hat{\epsilon})^k (1-p_i(\hat{\epsilon}))^{N-k} \simeq \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{\frac{-(k/N-m)^2}{2\sigma_i^2}\right\} \tag{2.16}$$

where m is the mean of $R_N^*|\hat{\epsilon}$ and σ^2 is the variance. Further, because $R_N^*|\hat{\epsilon} = \sum_{i=1}^{N} \frac{R_i(z_i)}{N}$ is $\text{Bin}(N,p)$, we have $m = p(\hat{\epsilon})$ and $\sigma = \sqrt{\frac{p(\hat{\epsilon})(1-p(\hat{\epsilon}))}{N}}$. Thus

$$\binom{N}{k} p(\hat{\epsilon})^k (1-p(\hat{\epsilon}))^{N-k} \simeq \frac{\sqrt{N}}{\sqrt{2\pi p(\hat{\epsilon})(1-p(\hat{\epsilon}))}} \exp\left\{\frac{-N(k/N-p(\hat{\epsilon}))^2}{2p(\hat{\epsilon})(1-p(\hat{\epsilon}))}\right\}. \tag{2.17}$$

Letting $y = \frac{k}{N}$, we have the probability of a fraction, y, of the companies defaulting,

$$\binom{N}{k} p(\hat{\epsilon})^k (1-p(\hat{\epsilon}))^{N-k} \simeq \frac{\sqrt{N}}{\sqrt{2\pi p(\hat{\epsilon})(1-p(\hat{\epsilon}))}} \exp\left\{\frac{-N(y-p(\hat{\epsilon}))^2}{2p(\hat{\epsilon})(1-p(\hat{\epsilon}))}\right\} \tag{2.18}$$

Using the law of total probability, we have the probability density function of R_N^*,

$$\begin{aligned} f_{R_N^*} = P(k \text{ defaults}) &= \int_{-\infty}^{\infty} \binom{N}{k} p_i(\hat{\epsilon})^k (1-p_i(\hat{\epsilon}))^{N-k} \phi(\hat{\epsilon}) d\hat{\epsilon} \\ &\simeq \int_{-\infty}^{\infty} \frac{\sqrt{N}}{\sqrt{2\pi p(\hat{\epsilon})(1-p(\hat{\epsilon}))}} \exp\left\{\frac{-N(y-p(\hat{\epsilon}))^2}{2p(\hat{\epsilon})(1-p(\hat{\epsilon}))}\right\} \phi(\hat{\epsilon}) d\hat{\epsilon} \end{aligned} \tag{2.19}$$

where

$$\phi(\hat{\epsilon}) = \frac{1}{\sqrt{2\pi}} e^{-\hat{\epsilon}^2/2} \tag{2.20}$$

With the unconditioned PDF of R_N^* calculated above, in the subsequent sections we use asymptotic methods to evaluate the integral and derive the CDF of R_N^* [10] [8].

2.2 Laplace's Method

Laplace's method is an asymptotic approach to approximating the value of Laplace-type integrals. It states that if some function $\tau(t) \in \mathcal{C}^4$ has a minimum at an interior point $t^* \in (a, b)$ then for a large parameter $N \gg 1$,

$$\int_a^b \alpha(t)\exp\{-N\tau(t)\}\,dt \;\sim\; \alpha(t^*)\exp\{-N\tau(t^*)\}\sqrt{\frac{2\pi}{\tau''(t^*)k}} + \mathcal{O}(k^{-3/2}) \tag{2.21}$$

where $\tau''(t^*) \geq 0$, $\tau'(t^*) = 0$, and $\alpha(t) \in \mathcal{C}^2$ [13] [14]. Making a change of variables in equation (2.19), we obtain a Laplace-type integral that can be evaluated used Laplace's method. Letting

$$\tau(\hat{\epsilon}) = \frac{(y - p(\hat{\epsilon}))^2}{2p(\hat{\epsilon})(1 - p(\hat{\epsilon}))} \tag{2.22}$$

and

$$\alpha(\hat{\epsilon}) = \frac{\sqrt{N}}{\sqrt{2\pi p(\hat{\epsilon})(1 - p(\hat{\epsilon}))}}\phi(\hat{\epsilon}) = \frac{\sqrt{N}}{\sqrt{2\pi p(\hat{\epsilon})(1 - p(\hat{\epsilon}))}}\frac{1}{\sqrt{2\pi}}\exp\left\{\frac{-\hat{\epsilon}^2}{2}\right\}, \tag{2.23}$$

then the default probability of the portfolio given by (2.19) simplifies to

$$f_{R_N^*} = \int_{-\infty}^{\infty} \alpha(\hat{\epsilon})\exp\{-N\tau(\hat{\epsilon})\}\,d\hat{\epsilon} \tag{2.24}$$

and we apply Laplace's method.

First note that for all $\hat{\epsilon}$, $\tau(\hat{\epsilon}) \geq 0$ so the minimum of $\tau(\hat{\epsilon})$ occurs when $y = p(\hat{\epsilon})$ (thus the minimum of $\tau(\hat{\epsilon})$ is zero). Let $\hat{\epsilon} = \hat{\epsilon}^*$ be the location of the minimum of $\tau(\hat{\epsilon})$. Then,

$$\hat{\epsilon}^* = p^{-1}(y) = \frac{\sqrt{1-\rho}\,\Phi^{-1}(y) - \theta}{-\sqrt{\rho}}. \tag{2.25}$$

Since we have a minimum, $\hat{\epsilon}^*$, by the Laplace approximation, the integral from (2.24), in terms of the large parameter N, is given by

$$f_{R_N^*} \sim \alpha(\hat{\epsilon}^*)\exp\{-N\tau(\hat{\epsilon}^*)\}\sqrt{\frac{2\pi}{N\tau''(\hat{\epsilon}^*)}} \tag{2.26}$$

We then compute $\tau''(\hat{\epsilon})$ and evaluate at $\hat{\epsilon}^*$ for which the only non-zero term becomes

$$\tau''(\hat{\epsilon}^*) = \frac{p'(\hat{\epsilon}^*)^2}{p(\hat{\epsilon}^*)(1 - p(\hat{\epsilon}^*))}. \tag{2.27}$$

Note that

$$p'(\hat{\epsilon}) = \frac{-1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\theta - \sqrt{\rho}\hat{\epsilon}}{\sqrt{1-\rho}}\right)^2} \left(\frac{\sqrt{\rho}}{\sqrt{1-\rho}}\right). \tag{2.28}$$

Since $y = \Phi\left(\frac{\theta - \sqrt{\rho}\hat{\epsilon}^*}{\sqrt{1-\rho}}\right) = p(\hat{\epsilon}^*)$, then $\Phi^{-1}(y) = \frac{\theta - \sqrt{\rho}\hat{\epsilon}^*}{\sqrt{1-\rho}}$ and we have

$$p'(\hat{\epsilon}^*) = -\frac{1}{\sqrt{2\pi}} \frac{\sqrt{\rho}}{\sqrt{1-\rho}} e^{-\frac{1}{2}\left[\Phi^{-1}(y)\right]^2} \tag{2.29}$$

Then evaluating the Laplace approximation from equation (2.24), we have

$$f_{R_N^*} \sim \sqrt{1-\rho} \cdot e^{-\frac{1}{2}\left(\frac{\sqrt{1-\rho}\Phi^{-1}(y) - \theta}{\sqrt{\rho}}\right)^2} \cdot \frac{1}{\sqrt{\rho} e^{-\frac{1}{2}[\Phi^{-1}(y)]^2}} \tag{2.30}$$

$$\implies f_{R_N^*} \sim \frac{\sqrt{1-\rho}}{\sqrt{\rho}} \exp\left\{\frac{\left[\Phi^{-1}(y)\right]^2 - \left[\frac{\sqrt{1-\rho}\Phi^{-1}(y) - \theta}{\sqrt{\rho}}\right]^2}{2}\right\}. \tag{2.31}$$

Therefore, the density for the loss is given by

$$f_{R_N^*}(y) \sim \sqrt{\frac{1-\rho}{\rho}} \exp\left\{\frac{1}{2}\left[\Phi^{-1}(y)\right]^2 - \frac{\left[\sqrt{1-\rho}\Phi^{-1}(y) - \theta\right]^2}{2\rho}\right\} \tag{2.32}$$

as $N \to \infty$.

2.3 Value-at-Risk

Now that we have found the probability distribution of the loss of the portfolio, we are able compute the CDF and, from that, the VaR_q. Making a change of variables in equation (2.32) such that $u = \sqrt{1-\rho}\Phi^{-1}(y) - \theta$ and integrating the PDF, we obtain $P\{R_N^* < x\} = F_{R_N^*}(x)$:

$$
\begin{aligned}
F_{R_N^*}(x) &= \int_0^x \sqrt{\frac{1-\rho}{\rho}} \exp\left\{ \frac{1}{2}\left[\Phi^{-1}(y)\right]^2 - \frac{\left[\sqrt{1-\rho}\Phi^{-1}(y) - \theta\right]^2}{2\rho} \right\} dy \\
&= \int_{-\infty}^{\sqrt{1-\rho}\Phi^{-1}(x)-\theta} \sqrt{\frac{1-\rho}{\rho}} \exp\left\{ \frac{1}{2}\left(\frac{u+\theta}{\sqrt{1-\rho}}\right)^2 - \frac{u^2}{2\rho} \right\} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(\frac{u+\theta}{\sqrt{1-\rho}}\right)^2 \right\} \frac{du}{\sqrt{1-\rho}} \\
&= \int_{-\infty}^{\sqrt{1-\rho}\Phi^{-1}(x)-\theta} \frac{1}{\sqrt{2\pi\rho}} \exp\left\{ \frac{-u^2}{2\rho} \right\} du.
\end{aligned}
\tag{2.33}
$$

Making a second change of variables and simplifying terms yields

$$
\begin{aligned}
F_{R_N^*}(x) &= \int_{-\infty}^{\frac{\sqrt{1-\rho}\Phi^{-1}(x)-\theta}{\sqrt{\rho}}} \frac{1}{\sqrt{2\pi}} e^{-v^2/2} dv \\
&= \Phi\left(\frac{\sqrt{1-\rho}\Phi^{-1}(x) - \theta}{\sqrt{\rho}} \right).
\end{aligned}
\tag{2.34}
$$

To find VaR_q, we compute the inverse of the CDF. Thus

$$
\begin{aligned}
F_{R_N^*}(x) = \Phi\left(\frac{\sqrt{1-\rho}\Phi^{-1}(x) - \theta}{\sqrt{\rho}} \right) &= q \\
\Longrightarrow \frac{\sqrt{1-\rho}\Phi^{-1}(x) - \theta}{\sqrt{\rho}} &= \Phi^{-1}(q) \\
\Longrightarrow \Phi^{-1}(x) &= \frac{\sqrt{\rho}\Phi^{-1}(q) + \theta}{\sqrt{1-\rho}}
\end{aligned}
$$

Hence

$$
F_{R_N^*}^{-1}(q) = \text{VaR}_q = \Phi\left(\frac{\theta + \sqrt{\rho}\phi^{-1}(q)}{\sqrt{1-\rho}} \right) \tag{2.35}
$$

In summary, the PDF, CDF, and VaR_q of the portfolio loss respectively are given by

$$
f_{R_N^*}(y) = \sqrt{\frac{1-\rho}{\rho}} \exp\left\{ \frac{1}{2}\left[\Phi^{-1}(y)\right]^2 - \frac{\left[\theta - \sqrt{1-\rho}\Phi^{-1}(y)\right]^2}{2\rho} \right\}, \tag{2.36}
$$

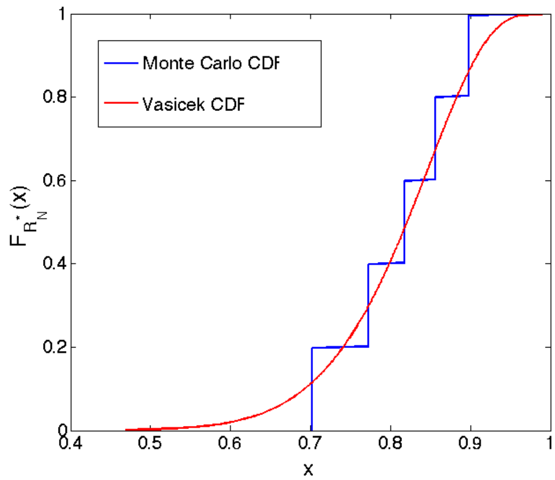
$$F_{R_N^*}(x) = \Phi\left(\frac{\sqrt{1-\rho}\Phi^{-1}(x) - \theta}{\sqrt{\rho}}\right), \tag{2.37}$$

$$\mathrm{VaR}_q = \Phi\left(\frac{\theta + \sqrt{\rho}\phi^{-1}(q)}{\sqrt{1-\rho}}\right). \tag{2.38}$$

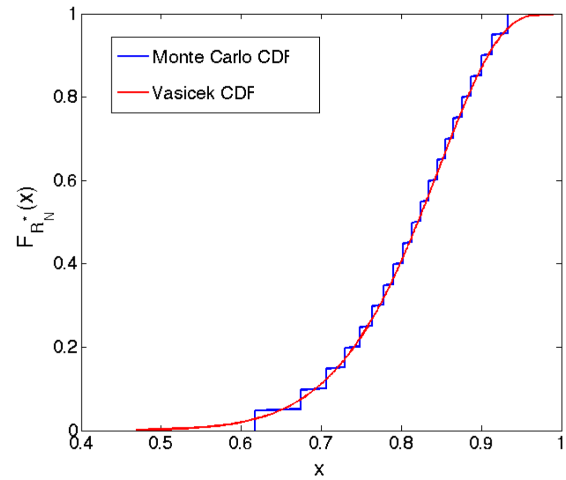
Thus Vasicek's ASRF model provides us with an analytic approximation of market risks associated with a credit portfolio. Using this approach, portfolio managers can use these analytic formulas as approximations for numerical simulations which are generally much more computationally intensive. It should be noted that above model is an implicit copula model based on the multivariate Gaussian distribution of asset value processes and that such applications of copulas to credit portfolio modeling have become widely used in risk management [15].

2.4 Monte Carlo Simulations

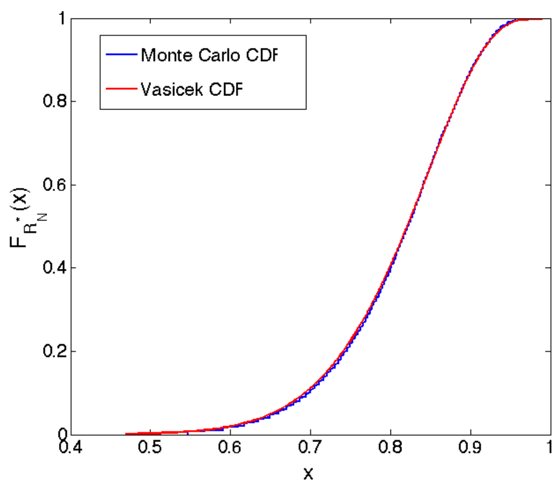
With the PDF, CDF, and VaR_q calculated above (2.36) - (2.38), we run Monte Carlo simulations and compare the approximated and simulated distributions. The plots below show the CDF and VaR_q of the Monte Carlo simulations and the Vasicek formulas derived above. Note that the Monte Carlo simulated CDF and VaR_q are step functions due to the discreteness of R_N^* in (2.4). As the number of companies, N, increases, the Monte Carlo simulated CDF and VaR_q become better approximations of the analytic Vasicek formulas.



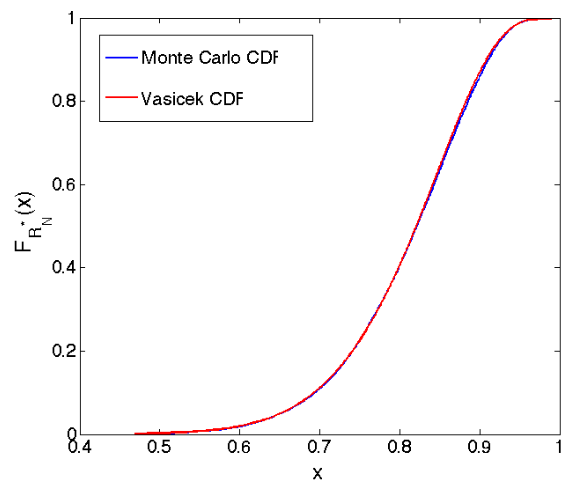
(a) 5 companies



(b) 20 companies

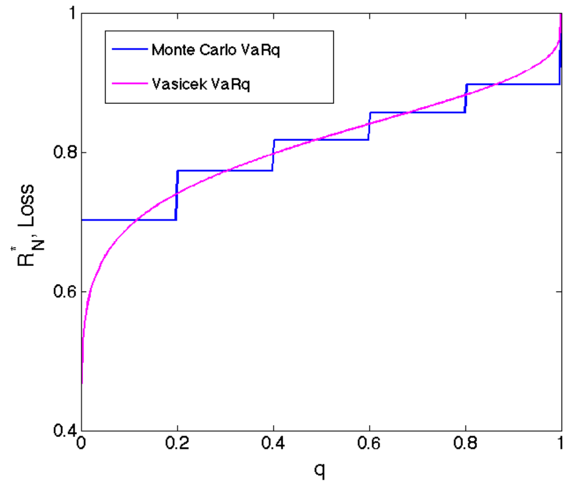


(c) 100 companies

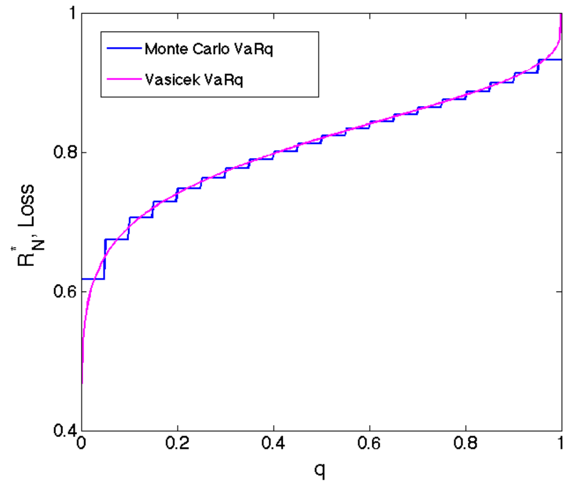


(d) 1,000 companies

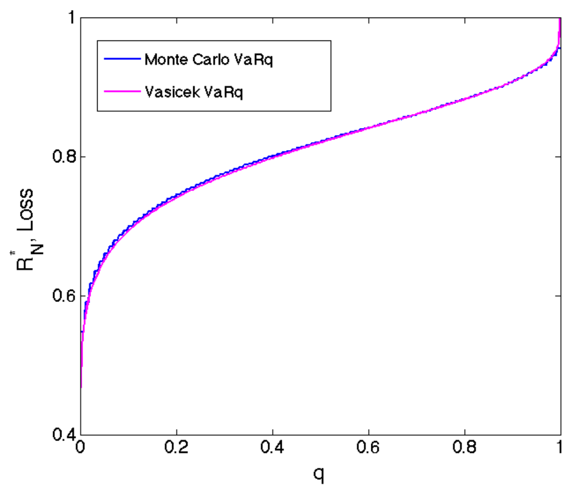
Figure 2.1: Plots of the Vasicek approximated CDF against the Monte Carlo simulated CDF. Plots generated using $\rho = 0.9058$, $\theta =$ -0.8730, and 50,000 Monte Carlo realizations.



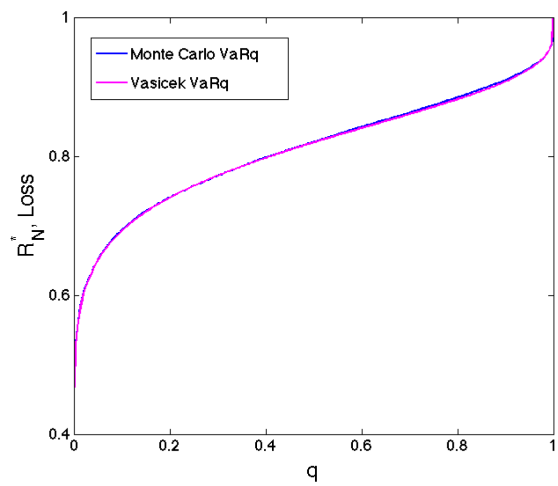
(a) 5 companies



(b) 20 companies



(c) 100 companies



(d) 1,000 companies

Figure 2.2: Plots of the Vasicek approximation of the VaR_q against the Monte Carlo simulated VaR_q. Plots generated using ρ = 0.9058, θ = -0.8730, and 50,000 Monte Carlo realizations.

In the next section, we examine the error associated with our approximation of the CDF as the number of companies, N, in the portfolio increases. That is, as N goes to infinity, what is the scaling of the error between the Vasicek approximation and the Monte Carlo simulated CDF? We explore the scaling and find a bound for the error using analytic and numerical techniques.

Chapter 3

THE BERRY-ESSEEN BOUND

Recall from the derivation of the approximation for the portfolio loss, we used the central limit theorem to approximate a binomial distribution with a normal distribution. The central limit theorem is a fundamental result in probability theory and determines when the sum of many random variables has a probability distribution that is approximately normal [16]. In this chapter, we discuss the central limit theorem, derive an error bound associated with approximating a binomial distribution with a normal distribution, and examine the scaling of the error as it relates to Vasicek's analytic approximation of portfolio default risk.

3.1 Central Limit Theorem

The central limit theorem is a well-studied branch of probability theory and states:

Theorem 3.1.1. *Let X_1, $X_2, \cdots$ be a sequence of independent and identically distributed random variables, each having mean μ and variance $\sigma^2 < \infty$. Then the distribution of*

$$Z_n(x) = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal distribution as $n \to \infty$. [16]

DeMoivre put forth the first version of the central limit theorem, proving it for the case of Bernouilli random variables with $p = \frac{1}{2}$. Laplace generalized the theorem to binomial distributions, and then to general distributions; however, a rigorous proof of the generalized central limit theorem was not obtained until the 20th century by Alexander Liapounoff [16].

Once the central limit theorem was proven, many mathematicians have dedicated their energies to studying the convergence rate of Z_N. In other words, what is the error in approximating the PDF of Z_N with the normal distribution? Andrew Berry and Carl-Gustav Esseen independently calculated the bound for the error in 1942 [17], [18]. Since then, Esseen and others continued to refine the error estimation to understand how well the normal distribution approximates other distributions when n is large.

3.2 The Berry-Esseen Bound

As stated, Esseen calculated a rate for how quickly the binomial CDF converges to the normal distribution and provided a bound on the error of the approximation. Let $F(x)$ be the distribution function for each $\{X_k\}$ defined in Theorem 3.1.1 and let $F_N(x)$ be the CDF for Z_N also defined in Theorem 3.1.1. The Berry-Esseen Theorem [19] states that the error of the approximation of the binomial CDF is bounded such that

$$|F_N(x) - \Phi(x)| \leq \frac{1.88 \max(\frac{\mu_3}{\mu_2})}{\sigma(\hat{\epsilon})} \tag{3.1}$$

where $\Phi(x)$ is the Gaussian distribution function as defined in (2.14), μ_3 is the third absolute moment of $F(x)$, μ_2 is the second absolute moment of $F(x)$, and $\sigma(\hat{\epsilon})$ is the standard deviation of $F(x)$.

We will show that this error bound should scale like $\mathcal{O}(N^{-1/2})$. Conditioned on global risk $\hat{\epsilon}$, the total portfolio loss over time horizon T for a portfolio of N

companies is $R_N^*(T) = \frac{1}{N}\sum_{i=1}^{N} R_i(z_i)$, $i = 1, \cdots, N$, with R_i independent Bernoulli random variables. If $\theta_i = \theta$ and $p_i = p$, the R_i's are identically distributed and

$$\text{second moment} = \mu_2 = \mathbf{E}\left(\frac{R_i^2}{N^2}\right) = \frac{p(\hat{\epsilon})}{N^2} \tag{3.2}$$

$$\text{third moment} = \mu_3 = \mathbf{E}\left(\frac{R_i^3}{N^3}\right) = \frac{p(\hat{\epsilon})}{N^3} \tag{3.3}$$

where $p(\hat{\epsilon})$ is defined as in (2.13) as

$$p(\hat{\epsilon}) = \Phi\left(\frac{\theta - \sqrt{\rho}\hat{\epsilon}}{\sqrt{1-\rho}}\right) = 1/2\left[1 + \operatorname{erf}\left(\frac{\theta - \sqrt{\rho}\hat{\epsilon}}{\sqrt{2(1-\rho)}}\right)\right]. \tag{3.4}$$

Thus, as $\hat{\epsilon} \to \infty$, we have $p(\hat{\epsilon}) \to 0$, i.e. as the health of the economy strengthens, company default becomes less likely. Similarly, as $\hat{\epsilon} \to -\infty$, we have $p(\hat{\epsilon}) \to 1$.

Further

$$\sigma_i^2 = \operatorname{Var}(R_i/N) = \frac{p(\hat{\epsilon})}{N^2} - \left(\frac{p(\hat{\epsilon})}{N}\right)^2 = \frac{p(\hat{\epsilon})(1-p(\hat{\epsilon}))}{N^2}. \tag{3.5}$$

Therefore

$$\begin{aligned} \sigma &= \sqrt{\sigma_1^2 + \cdots + \sigma_N^2} \\ &= \left(\frac{p(1-p)}{N}\right)^{1/2} \end{aligned} \tag{3.6}$$

So, from the Berry-Esseen bound we have

$$\begin{aligned} \left|F_{R_N^*|\hat{\epsilon}}(x,\hat{\epsilon}) - \Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right)\right| &\le \frac{1.88\max(\mu_3/\mu_2)}{\sigma} \\ &= \frac{1.88/N}{\sqrt{p(1-p)}}\sqrt{N} \\ &= \frac{1.88}{\sqrt{Np(1-p)}} \end{aligned} \tag{3.7}$$

where $F_{R_N^*|\hat{\epsilon}}(x,\hat{\epsilon})$ is the CDF of R_N^* conditioned on $\hat{\epsilon}$, $\Phi(\cdot)$ is the Gaussian cumulative normal distribution defined as in (2.14), and $\Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right)$ is the conditioned CDF from the central limit theorem given in (2.8).

Thus the error associated with the central limit theorem should be bounded by a function scaling like $\mathcal{O}(N^{-1/2})$. Given the Berry-Esseen bound, the CDF of R_N^* conditioned on $\hat{\epsilon}$ is bounded by

$$\Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right) - \frac{1.88}{\sqrt{Np(1-p)}} \leq F_{R_N^*|\hat{\epsilon}}(x,\hat{\epsilon}) \leq \Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right) + \frac{1.88}{\sqrt{Np(1-p)}}. \tag{3.8}$$

To find the bounds for the unconditioned cumulative density, we have

$$\int_{-\infty}^{\infty}\left(\Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right) - \frac{1.88}{\sqrt{Np(1-p)}}\right)\phi(\hat{\epsilon})d\hat{\epsilon} \leq \int_{-\infty}^{\infty} F_{R_N^*|\hat{\epsilon}}(x,\hat{\epsilon})\phi(\hat{\epsilon})d\hat{\epsilon} \tag{3.9}$$

and

$$\int_{-\infty}^{\infty} F_{R_N^*|\hat{\epsilon}}(x,\hat{\epsilon})\phi(\hat{\epsilon})d\hat{\epsilon} \leq \int_{-\infty}^{\infty}\left(\Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right) + \frac{1.88}{\sqrt{Np(1-p)}}\right)\phi(\hat{\epsilon})d\hat{\epsilon}. \tag{3.10}$$

Thus the error of the unconditioned CDF is bounded such that

$$\left|\int_{-\infty}^{\infty} F_{R_N^*|\hat{\epsilon}}(x;\hat{\epsilon})\phi(\hat{\epsilon})d\hat{\epsilon} - \int_{-\infty}^{\infty}\Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right)\phi(\hat{\epsilon})d\hat{\epsilon}\right| \leq \frac{1.88}{\sqrt{N}}\int_{-\infty}^{\infty}\frac{\phi(\hat{\epsilon})d\hat{\epsilon}}{\sqrt{p(1-p)}}. \tag{3.11}$$

We now check the convergence of the integral on the right-hand side of (3.11).

Considering the integrand of the right-hand side of (3.11) we have

$$\frac{1.88e^{-\hat{\epsilon}^2/2}}{\sqrt{2\pi Np(\hat{\epsilon})(1-p(\hat{\epsilon}))}}. \tag{3.12}$$

Again, since $p(\hat{\epsilon}) \to 0$ as $\hat{\epsilon} \to \infty$, we have $\sqrt{2\pi Np(1-p)} \to 0$ and $e^{-\hat{\epsilon}^2/2} \to 0$. Thus we must consider the behavior of both the numerator and denominator to see whether the integral will converge. Note that the denominator should scale like $\sqrt{p(\hat{\epsilon})}$ as $p \to 0$. Recall from (3.4) that $p(\hat{\epsilon})$ is a constant plus the erf function. Recall the asymptotic formula for the error function for large arguments,

$$\operatorname{erf}(x) \sim 1 - \frac{1}{x\sqrt{\pi}e^{x^2}} + \cdots, \quad x \to +\infty. \tag{3.13}$$

The erf function is odd, hence -erf(x) = erf$(-x)$, and we have

$$\operatorname{erf}(-x) = -\operatorname{erf}(x) \sim -1 + \frac{1}{x\sqrt{\pi}e^{x^2}} + \cdots, \quad x \to -\infty \tag{3.14}$$

for large negative arguments. Therefore, we have $x \mapsto \frac{-\sqrt{\rho}\hat{\epsilon}}{\sqrt{2(1-\rho)}}$ and as $\hat{\epsilon} \to \infty$ we use (3.14) to obtain

$$\begin{aligned} p(\hat{\epsilon}) &= 1/2\left[1 + \operatorname{erf}\left(\frac{-\sqrt{\rho}\hat{\epsilon}}{\sqrt{2(1-\rho)}}\right)\right] \\ &\sim 1/2\left[1 + \left(-1 + \frac{1}{\frac{\sqrt{\pi\rho}\hat{\epsilon}}{\sqrt{2(1-\rho)}}\exp\left\{\left(\frac{\sqrt{\rho}\hat{\epsilon}}{\sqrt{2(1-\rho)}}\right)^2\right\}} + \cdots\right)\right] \\ &\sim \frac{1}{\frac{2\sqrt{\pi\rho}\hat{\epsilon}}{\sqrt{2(1-\rho)}}\exp\left\{\frac{\rho\hat{\epsilon}^2}{2(1-\rho)}\right\}} \end{aligned} \tag{3.15}$$

But the denominator of (3.12) scales like $\sqrt{p}$. Therefore the denominator scales like

$$\sqrt{p(\hat{\epsilon})} \sim \left(e^{\frac{-\rho\hat{\epsilon}^2}{2(1-\rho)}}\right)^{1/2} = e^{\frac{-\rho\hat{\epsilon}^2}{4(1-\rho)}} \tag{3.16}$$

while the numerator scales like $e^{-\hat{\epsilon}^2/2}$.

Similarly, for large negative arguments, we have that as $\hat{\epsilon} \to -\infty$, $x \to \infty$. Using (3.13) and the fact that erf is odd,

$$\begin{aligned} p(\hat{\epsilon}) &= 1/2\left[1 + \operatorname{erf}\left(\frac{-\sqrt{\rho}\hat{\epsilon}}{\sqrt{2(1-\rho)}}\right)\right] \\ &\sim 1/2\left[1 + \left(1 - \frac{1}{\frac{\sqrt{\pi\rho}\hat{\epsilon}}{\sqrt{2(1-\rho)}}\exp\left\{\left(\frac{\sqrt{\rho}\hat{\epsilon}}{\sqrt{2(1-\rho))}}\right)^2\right\}} + \cdots\right)\right] \\ &\sim \frac{1}{\frac{2\sqrt{\pi\rho}\hat{\epsilon}}{\sqrt{2(1-\rho)}}\exp\left\{\frac{\rho\hat{\epsilon}^2}{2(1-\rho)}\right\}} \end{aligned} \tag{3.17}$$

Therefore, as $\hat{\epsilon} \to -\infty$ we have $p - 1 \to 0$ and the denominator scaling like

$$\sqrt{p(\hat{\epsilon})} \sim \left(e^{\frac{-\rho\hat{\epsilon}^2}{2(1-\rho)}}\right)^{1/2} = e^{\frac{-\rho\hat{\epsilon}^2}{4(1-\rho)}} \tag{3.18}$$

while the numerator scales like $e^{-\hat{\epsilon}^2/2}$.

As $\hat{\epsilon} \to \infty$ and $\hat{\epsilon} \to -\infty$, the integral on the right-hand side of (3.11) scales like

$$\exp\left\{-\left(\frac{1}{2} - \frac{\rho}{4(1-\rho)}\right)\hat{\epsilon}^2\right\}$$

Therefore, in order for the integral to converge, we require

$$\frac{1}{2} - \frac{\rho}{4(1-\rho)} > 0 \tag{3.19}$$

Solving for ρ, we find that $\frac{2}{3} > \rho$ in order for (3.19) to hold. Thus, for $\rho \in [0, 2/3)$, the integral on the right-hand side of (3.11) will converge. The main result of this section is that the error bound of the unconditioned CDF is bounded by a function that scales like $\mathcal{O}(N^{-1/2})$. In other words,

$$\left|\int_{-\infty}^{\infty} F_{R_N^*|\hat{\epsilon}}(x;\hat{\epsilon})\phi(\hat{\epsilon})d\hat{\epsilon} - \int_{-\infty}^{\infty} \Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right)\phi(\hat{\epsilon})d\hat{\epsilon}\right| \leq C/\sqrt{N} \tag{3.20}$$

where

$$C = 1.88 \int_{-\infty}^{\infty} \frac{\phi(\hat{\epsilon}) d\hat{\epsilon}}{\sqrt{p(\hat{\epsilon})(1 - p(\hat{\epsilon}))}} < \infty \tag{3.21}$$

for $\rho \in (0, 2/3)$.

3.3 Numerical results

From the above discussion, we note that the error is bound by a function that scales like $\mathcal{O}(N^{-1/2})$. Thus, we likely expect the error of the CDF to scale like $\mathcal{O}(N^{-1/2})$. However, numerical simulations indicate that the scaling of the error of the CDF may be quite different. For default thresholds close to zero, a numerical investigation of the error finds the scaling to be $\mathcal{O}(N^{-1})$ which is smaller than expected. The plots that follow demonstrate the scaling error between the Monte Carlo simulated CDF and the analytic CDF derived in (2.37).

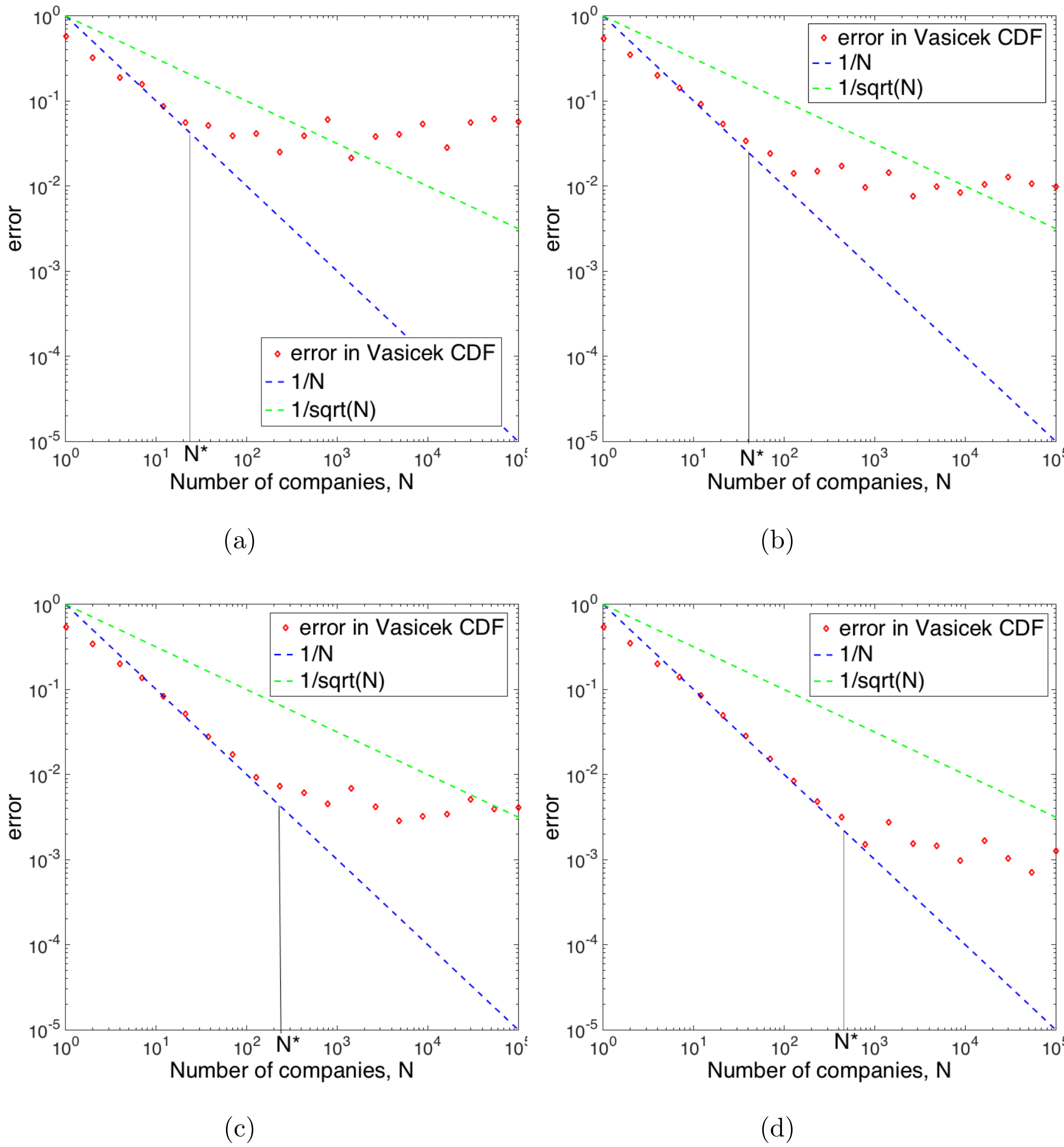


Figure 3.1: Plots of the error between the Monte Carlo simulated CDF and the Vasicek approximation of the CDF in (2.37) as the number of companies in the portfolio is increased. Figure 3.1a corresponds to 500 Monte Carlo realizations, $\rho = 0.35$, and $\theta = -0.1$. Figures 3.1b, 3.1c, and 3.1d correspond to 5K, 50K, and 500K respectively with the same ρ and θ as in 3.1a.

The flattening out of the error for large number of companies is a result of the error associated with Monte Carlo simulations and can be shown to be $\mathcal{O}\left(\frac{1}{\text{number of trials}^{1/2}}\right)$.

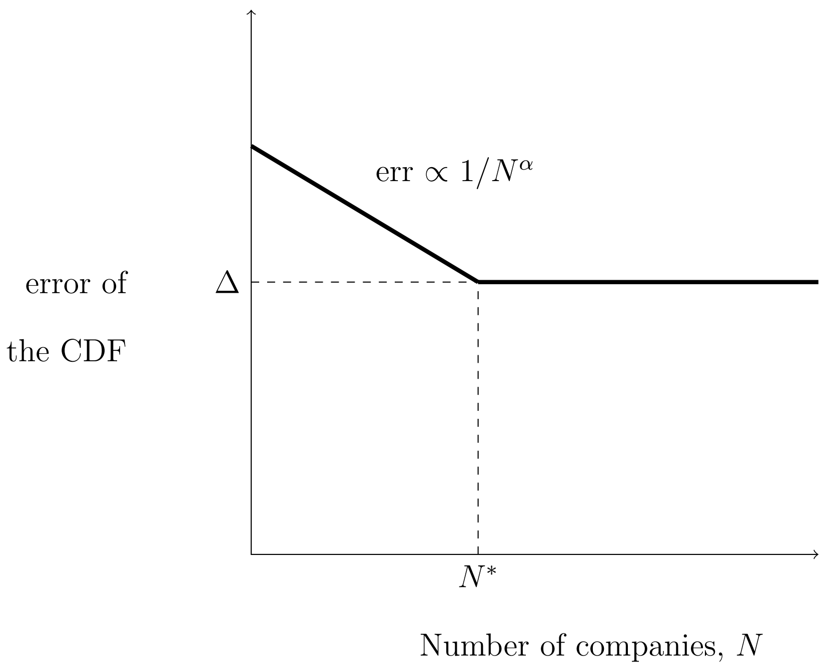


Figure 3.2: Monte Carlo simulation error

Note that

$$\Delta = \text{Monte Carlo error} \propto \frac{1}{(\text{number of trials})^{1/2}} \tag{3.22}$$

Therefore

$$\frac{1}{N^{*\alpha}} = \frac{1}{(\text{number of trials})^{1/2}} \tag{3.23}$$

$$N^{*} = (\text{number of trials})^{1/2\alpha} \tag{3.24}$$

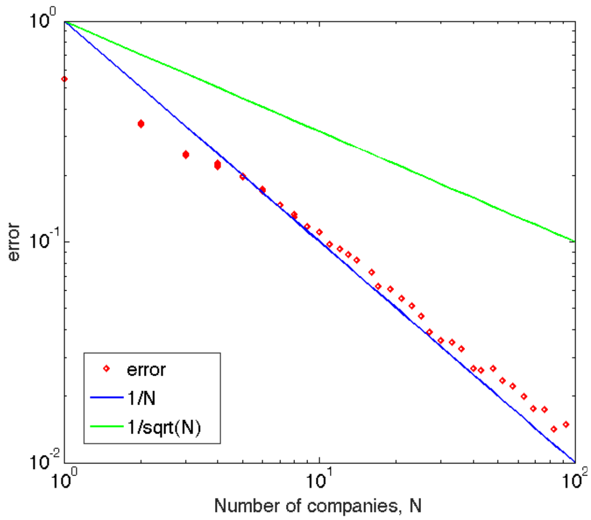
Thus, the $1/N^{\alpha}$ scaling is valid providing $N \ll N^* = (\text{number of trials})^{1/2\alpha}$. So for the plots in Figure 3.1, where $\alpha = 1$, we have

Monte Carlo trails	approximate N^*
500	22
5,000	70
50,000	225
500,000	700

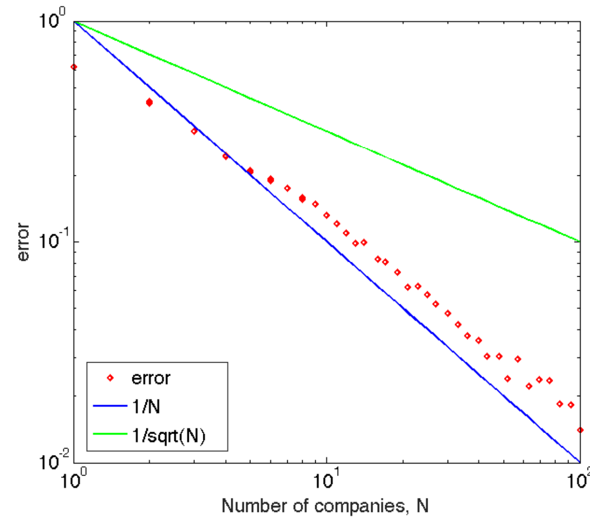
Table 3.1: N^* values for given Monte Carlo realizations.

Plots in Figure 3.1 demonstrate a leveling off at roughly these figures; therefore, the transition point of where the CDF error flattens out is due to the finite sample size and moves to the right as the number of Monte Carlo realizations are increased. Hence, as the number of Monte Carlo realizations increases, this error becomes small and the total error of the CDF converges to the true scaling as seen in the plots below, shown to be $\mathcal{O}(N^{-1})$.

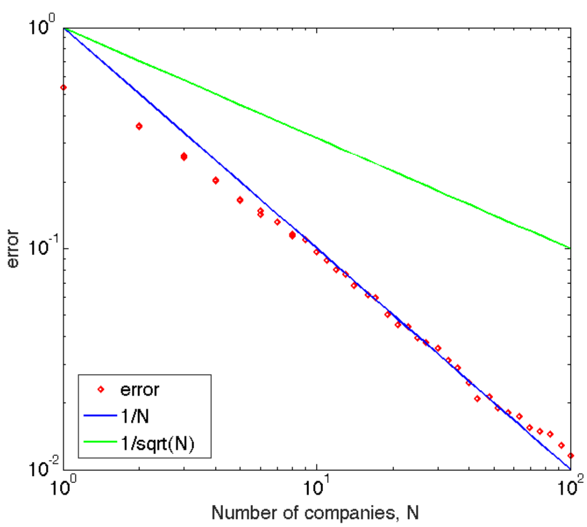
Because of the numerical error associated with the Monte Carlo simulation, for the rest of this thesis we consider the error of the CDF over 10^0 to 10^2 companies using 100,000 Monte Carlo realizations. From the above discussion, this should exhibit the $\mathcal{O}(N^{-1})$ scaling. Next, we verify the scaling of the error for values of $\rho \in [0.3, 0.6]$.



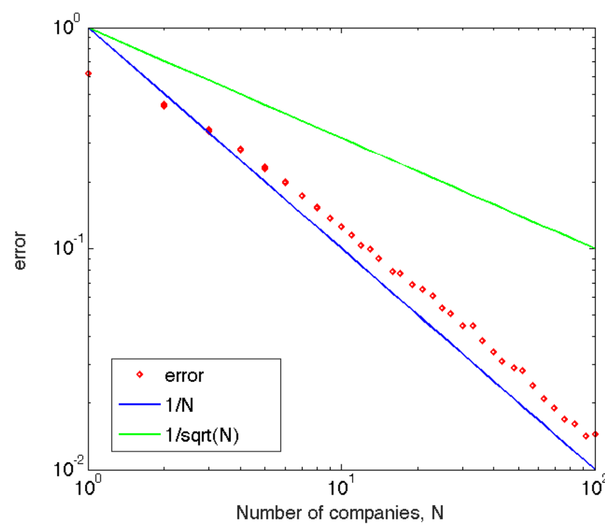
(a) ρ = 0.3, θ = -0.1



(b) ρ = 0.3, θ = -0.3

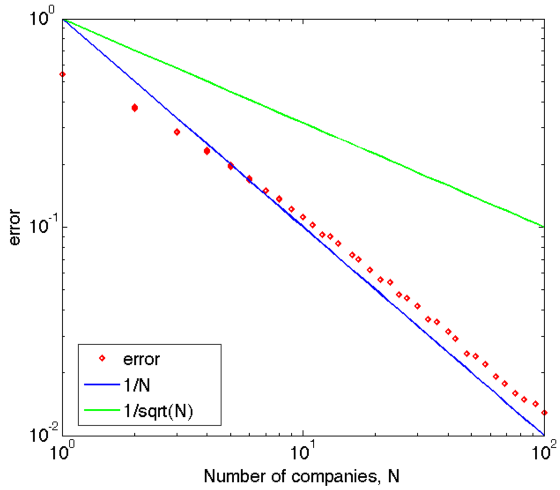


(c) ρ = 0.4, θ = -0.1

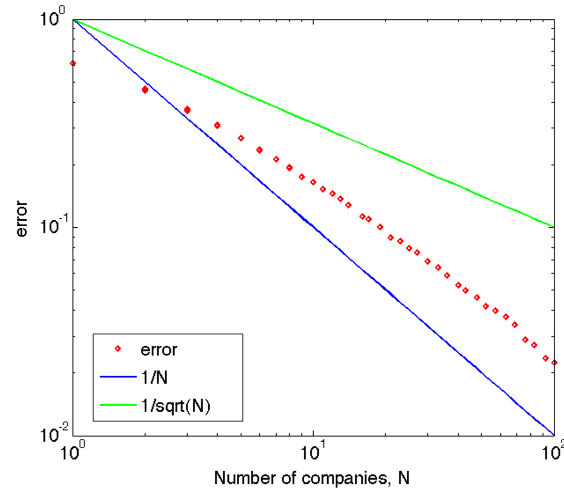


(d) ρ = 0.4, θ = -0.3

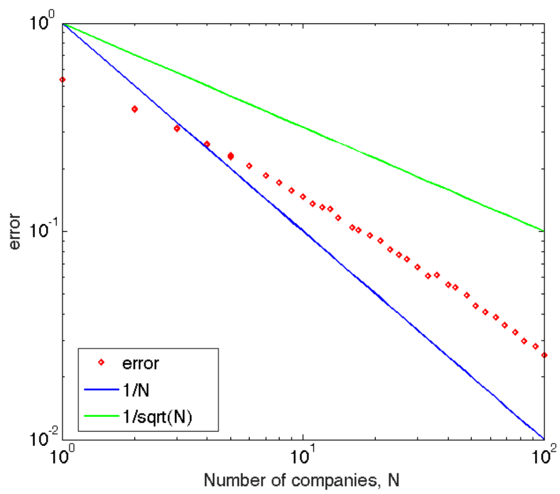
Figure 3.3: Scaling of the error between the Monte Carlo simulated CDF and the analytic Vasicek CDF derived from 100,000 Monte Carlo realizations.



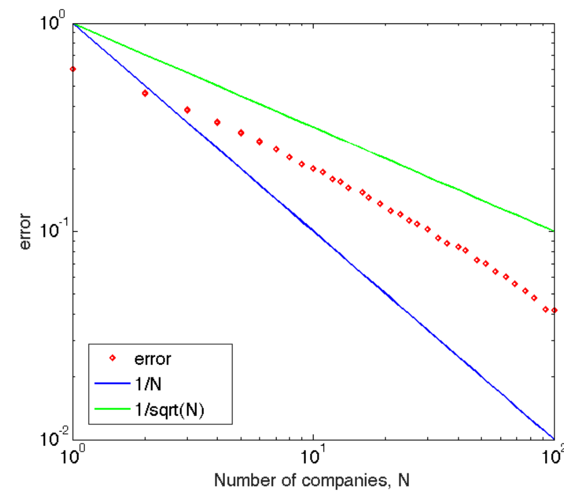
(a) ρ = 0.5, θ = -0.1



(b) ρ = 0.5, θ = -0.3



(c) ρ = 0.6, θ = -0.1



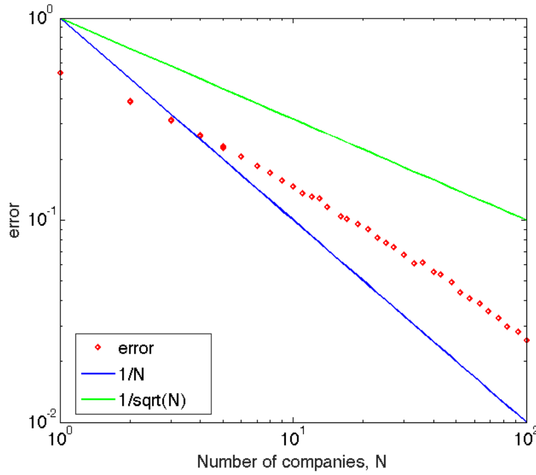
(d) ρ = 0.6, θ = -0.3

Figure 3.4: Scaling of the error between the Monte Carlo simulated CDF and the analytic Vasicek CDF derived from 100,000 Monte Carlo realizations.

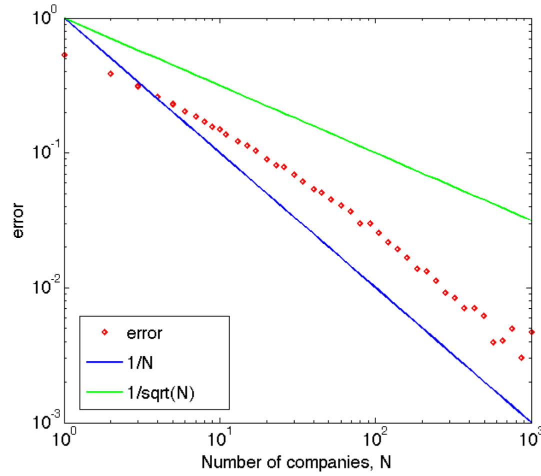
Note that for higher values of ρ, the scaling initially seems to shift from away from $\mathcal{O}(N^{-1})$ when N is moderate. However, when we consider larger portfolios, we see the error scaling converging to $\mathcal{O}(N^{-1})$. From the central limit theorem, we know that the binomial distribution, $\text{Bin}(N, p)$, is well approximated with a normal distribution for values of p not too close to 0 or 1.

For values of p near 0 or 1, the central limit theorem still holds but N must be sufficiently large for the normal distribution to be a good approximation. Note that as ρ increases, $p \to 0$. Recall that in (3.6), for p close to 0 or 1, N must be taken large in order to make $\frac{1.88}{\sqrt{Np(1-p)}}$ small. Thus larger values of N are needed to see the scaling converge to $\mathcal{O}(N^{-1})$ as shown in Figures 3.3 and 3.4: as N increases, the error becomes parallel to the $1/N$ diagonal.

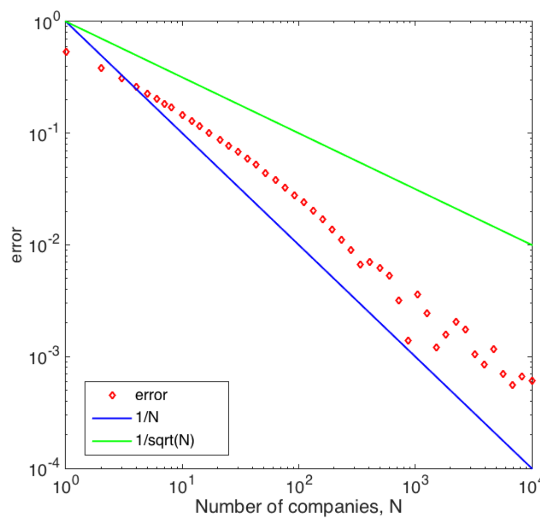
We verify this numerically, by plotting the error when $\rho = 0.6$ as the scale in N is increased. As we increase the number of companies, N, we also increase the number of Monte Carlo trials to account for the simulation error discussed above. We see that as the scaling increases, the error becomes parallel to the $1/N$ diagonal.



(a)



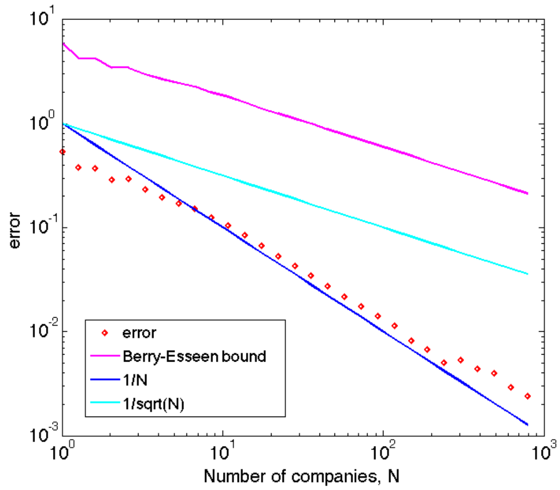
(b)



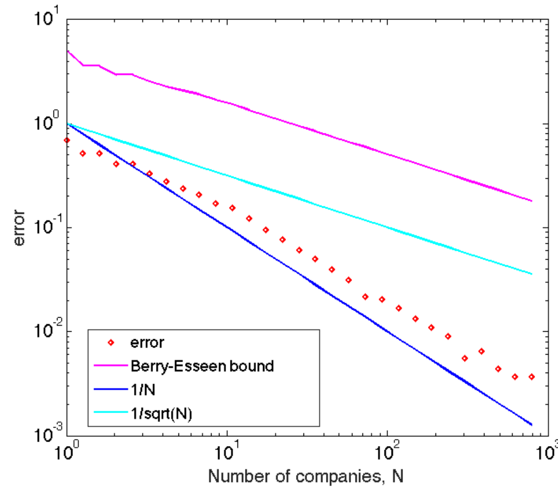
(c)

Figure 3.5: Scaling of the error between the Monte Carlo simulated CDF and the analytic Vasicek CDF as the number of companies increases when ρ = 0.6 and θ = -0.1. Plots 3.5a and 3.5b use 100,000 Monte Carlo realizations, while 3.5c uses 1,000,000 realizations.

Plotting the Berry-Esseen bound given by the right-hand side of equation (3.11), we observe that while the bound always scales like $\mathcal{O}(N^{-1/2})$, the error of the CDF scales like $\mathcal{O}(N^{-1})$.



(a) ρ = 0.5, θ = -0.1



(b) ρ = 0.3, θ = -0.5

Figure 3.6: Plots depict the Berry-Esseen bound and error of the CDF for given ρ and θ generated by 100,000 Monte Carlo realizations.

3.4 Rao-Chattacharya Bound

The CDF of the portfolio loss, however, has discontinuities due to the discreteness of the portfolio. To account for these discontinuous jumps in the CDF and improve upon the Berry-Esseen bound, a correction term is added to the bound. Rao and Bhattacharya [20] derive the following bound for the discontinuous CDF

$$\left| F_{R_N^*|\hat{\epsilon}}(x;\hat{\epsilon}) - \Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right) \right| \leq \frac{3\sigma^2 + |\mu_3|}{6\sqrt{2\pi N}\sigma^3} \tag{3.25}$$

where μ_3 is the third moment of $R_N^*|\hat{\epsilon}$ and σ^2 is the variance (derived in chapter 4).

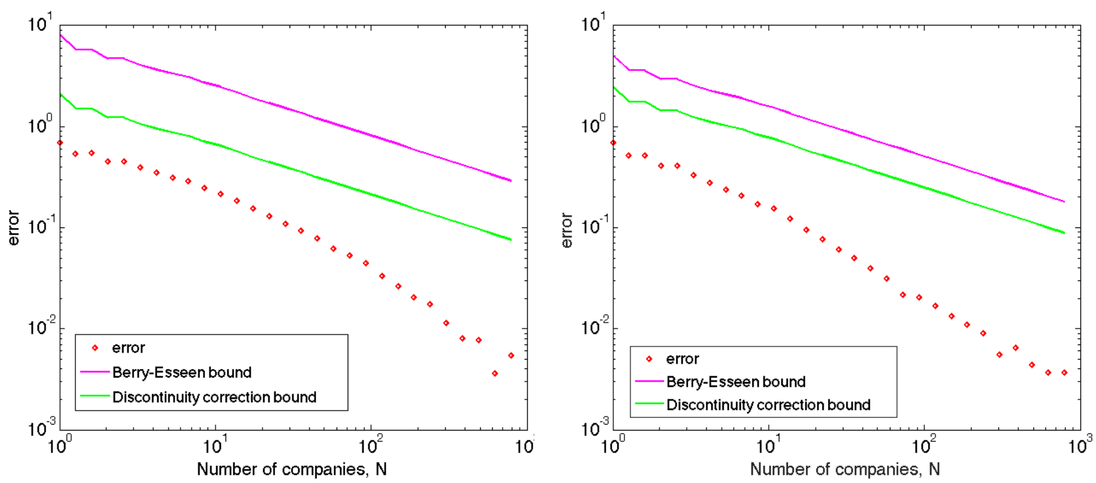
Again, this is the bound of the error of the CDF conditioned on $\hat{\epsilon}$. To find the unconditioned bound, we multiply by $\phi(\hat{\epsilon})$ and integrate over $\hat{\epsilon}$. Therefore,

$$\left| \int_{-\infty}^{\infty} F_{R_N^*|\hat{\epsilon}}(x;\hat{\epsilon})\phi(\hat{\epsilon})d\hat{\epsilon} - \int_{-\infty}^{\infty} \Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right)\phi(\hat{\epsilon})d\hat{\epsilon}\right| \leq \frac{1}{\sqrt{N}}\int_{-\infty}^{\infty} \frac{3\sigma^2+|\mu_3|}{6\sqrt{2N\pi}\sigma^3}\phi(\hat{\epsilon})d\hat{\epsilon} \tag{3.26}$$

In the context of the problem of the large portfolio, we have

$$\left| \int_{-\infty}^{\infty} F_{R_N^*|\hat{\epsilon}}(x;\hat{\epsilon})\phi(\hat{\epsilon})d\hat{\epsilon} - \int_{-\infty}^{\infty} \Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right)\phi(\hat{\epsilon})d\hat{\epsilon}\right| \leq \frac{1}{\sqrt{N}}\int_{-\infty}^{\infty} \frac{3p(1-p)+|p-3p^2+2p^3|}{6\sqrt{2N\pi}(p(1-p))^{3/2}}\phi(\hat{\epsilon})d\hat{\epsilon} \tag{3.27}$$

Numerically investigating this bound, we find that it is a much tighter bound for the error of the CDF, though it still exhibits the incorrect scaling. In the next section, we asymptotically analyze the scaling of the error.



(a) ρ = 0.5, θ = -0.5 (b) ρ = 0.3, θ = -0.5

Figure 3.7: Plots show the error of the CDF, the continuous Berry-Esseen bound, and the corrected discontinuity bound for given ρ and θ values, and 100,000 Monte Carlo realizations.

Chapter 4

ASYMPTOTIC ANALYSIS

4.1 The approximation of a Binomial distribution with the Normal distribution

Recall from our model derivation that the portfolio loss, $R_N^*(T)$, is a weighted sum of Bernoulli random variables for a fixed global risk, $\hat{\epsilon}$. When the weights are equal, the portfolio loss is binomially distributed, $\mathrm{Bin}(N, p)$. Because the number of Bernoulli random variables (the number of companies in our portfolio) is taken to infinity, we use the central limit theorem to approximate the binomial distribution with a normal distribution. To derive this approximation, we employ characteristic functions.

First, let $R_i \sim \mathrm{Bernoulli}(p)$. Then

$$R_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } (1-p) \end{cases} \tag{4.1}$$

Thus $\mathbf{E}(R_i) = p$ and $\mathrm{Var}(R_i) = \sigma^2 = \mathbf{E}(R_i^2) - \mathbf{E}(R_i)^2 = p(1-p)$. Now consider the total loss of a portfolio consisting of N companies,

$$R_N^*|\hat{\epsilon} = \sum_{i=1}^{N} \frac{R_i}{N} \tag{4.2}$$

Let

$$Y_i = \frac{R_i - p}{\sigma} \tag{4.3}$$

with

$$Y_i = \begin{cases} \frac{1-p}{\sigma} & \text{with probability } p \\ \frac{-p}{\sigma} & \text{with probability } (1-p) \end{cases} \tag{4.4}$$

Define

$$Z_N = \sum_{i=1}^{N} \frac{Y_i}{\sqrt{N}}. \tag{4.5}$$

Then from Theorem 3.1.1,

$$Z_N = \frac{\sum_{i=1}^{N} X_i - Np}{\sqrt{Np(1-p)}} = \frac{\sum_{i=1}^{N} X_i}{\sqrt{Np(1-p)}} - \frac{Np}{\sqrt{Np(1-p)}} \tag{4.6}$$

tends to the standard normal distribution as $n \to \infty$. Notice that

$$\frac{Z_N}{\sqrt{N}} = \frac{\sum_{i=1}^{N} X_i}{N\sqrt{p(1-p)}} - \frac{p}{\sqrt{p(1-p)}} \tag{4.7}$$

$$\begin{aligned} \implies \frac{Z_N\sqrt{p(1-p)}}{\sqrt{N}} &= \frac{\sum_{i=1}^{N} X_i}{N} - p \\ &= R_N^* - p. \end{aligned} \tag{4.8}$$

Hence

$$\frac{Z_N\sqrt{p(1-p)}}{\sqrt{N}} + p = R_N^*. \tag{4.9}$$

Therefore

$$F_{R_N^*|\hat{\epsilon}}(x) = F_{Z_N}\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right). \tag{4.10}$$

Moreover, note that

$$\mathbf{E}(Z_N) = \frac{1}{\sqrt{N}}\mathbf{E}\left(\sum_{i=1}^{N} Y_i\right) = \frac{1}{\sqrt{N}} N\mathbf{E}(Y_i) = 0, \tag{4.11}$$

and

$$\text{Var}(Z_n) = \mathbf{E}(Z_N) - \mathbf{E}(Z_N)^2 = 1. \tag{4.12}$$

We first seek to find the distribution function, F_{Z_N}, using an Edgeworth expansion [17]. To approximate using the central limit theorem, we have $F_{Z_N} \to N(0,1) + \gamma(N)$, where γ is the error term. Using properties of characteristic functions, note that

$$\begin{aligned}\varphi_{Z_N}(t) &= \varphi_{Y_1/\sqrt{N}+\cdots+Y_n/\sqrt{N}}(t) \\ &= \varphi_{Y_1}(t/\sqrt{N})\varphi_{Y_2}(t/\sqrt{N})\ldots\varphi_{Y_N}(t/\sqrt{N}) \\ &= \left[\varphi_{Y_1}(t/\sqrt{N})\right]^N\end{aligned} \tag{4.13}$$

since the Y_i's are independently and identically distributed. However,

$$\begin{aligned}\varphi_{Y_1}(t/\sqrt{N}) = \mathbf{E}(e^{itY_i}) &= pe^{it\frac{(1-p)}{\sigma}} + (1-p)e^{\frac{-ipt}{\sigma}} \\ &= pe^{\frac{it}{\sigma}}e^{\frac{-itp}{\sigma}} + e^{\frac{-itp}{\sigma}} - pe^{\frac{-itp}{\sigma}} \\ &= e^{\frac{-itp}{\sigma}}\left[pe^{\frac{it}{\sigma}} + 1 - p\right].\end{aligned} \tag{4.14}$$

Thus

$$\left[\varphi_{Y_1}\left(t/\sqrt{N}\right)\right]^N = \left[e^{-itp/\sqrt{N}\sigma}\left(pe^{it/\sqrt{N}\sigma} + 1 - p\right)\right]^N \tag{4.15}$$

$$\left[\varphi_{Y_1}\left(t/\sqrt{N}\right)\right]^N = e^{-itp\sqrt{N}/\sigma}\left[pe^{it/\sqrt{N}\sigma} + 1 - p\right]^N \tag{4.16}$$

$$\Longrightarrow \ln\varphi_{Z_N}(t) = \frac{-it\sqrt{N}p}{\sigma} + N\ln\left[1 - p(1 - e^{it/\sigma\sqrt{N}})\right]. \tag{4.17}$$

Expanding the right-hand side in Taylor series for $N \gg 1$ and for fixed t,

$$\begin{aligned}\ln\varphi_{Z_N}(t) &= \frac{-it\sqrt{N}p}{\sigma} - Np(1 - e^{it/\sigma\sqrt{N}}) - \frac{Np^2}{2}(1 - e^{it/\sigma\sqrt{N}})^2 - \frac{Np^3}{3}(1 - e^{it/\sigma\sqrt{N}})^3 + \mathcal{O}(t^4) \\ &= \frac{(-p+p^2)t^2}{2\sigma^2} - \frac{p - 3p^2 + 2p^3}{6\sqrt{N}\sigma^3}it^3 + \mathcal{O}\left(\frac{t^4}{N}\right)\end{aligned} \tag{4.18}$$

Recall that $\sigma^2 = p(1-p)$. Thus

$$\ln \varphi_{Z_N}(t) = \frac{-t^2}{2} - \frac{(p - 3p^2 + 2p^3)}{6\sqrt{N}\sigma^3} it^3 + \mathcal{O}\left(\frac{t^4}{N}\right) \tag{4.19}$$

Letting $\frac{(p-3p^2+2p^3)}{6\sqrt{N}\sigma^3} = \alpha$, and expanding once more using the series expansion of e^x, we have

$$\begin{aligned}
\varphi_{Z_N}(t) &= \exp\left\{\frac{-t^2}{2} - \frac{(p - 3p^2 + 2p^3)}{6\sqrt{N}\sigma^3} it^3 + \mathcal{O}(t^4/N)\right\} \\
&= e^{-t^2/2} e^{-\alpha i t^3} e^{\mathcal{O}(t^4/N)} \\
&= e^{-t^2/2}\left(1 - \alpha i t^3 + \mathcal{O}(t^4/N)\right) \\
&= e^{-t^2/2} - \alpha i t^3 e^{-t^2/2} + \mathcal{O}(t^4/N e^{t^2/2})
\end{aligned} \tag{4.20}$$

Taking the inverse Fourier transform, substituting in α, and integrating over x, we have

$$F_{Z_N}(x) = \Phi(x) + \frac{(p - 3p^2 + 2p^3)}{6\sqrt{2\pi N}(p(1-p))^{3/2}} \left[1 - x^2\right] e^{-x^2/2} \tag{4.21}$$

Recalling that $F_{R_N^*|\hat{\epsilon}}(x) = F_{Z_n}\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right)$ yields

$$F_{R_N^*|\hat{\epsilon}}(x) = \Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right) + \frac{(p - 3p^2 + 2p^3)}{6\sqrt{2\pi N}(p(1-p))^{3/2}} \left[1 - \left(\frac{(x-p)^2 N}{p(1-p)}\right)\right] e^{-\frac{(x-p)^2 N}{2p(1-p)}} \tag{4.22}$$

Thus a Binomial distribution function can be approximated with a standard normal distribution, $\Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right)$, plus a correction term that scales like $N^{-1/2}$. As shown, when $N \to \infty$, the Binomial distribution converges to a normal distribution as the central limit theorem mandates. Because (4.20) is only valid for small t and the inverse Fourier transform requires an integral over all $t \in \mathbb{R}$, obtaining (4.21) from (4.20) is non-rigorous. When the characteristic function is defined only on a small interval, the CDF is not uniquely determined [17]. Thus, (4.22) is not the only CDF with (4.20) as the characteristic function. In the next section, we discuss one of the candidates for the CDF with the given characteristic function.

4.2 A discontinuity correction

Again, equation (4.22) is only valid provided the characteristic function of R_i in (4.1) is not periodic and defined for all $t \in \mathbb{R}$. Because the portfolio loss is a lattice random variable, i.e. the CDF is a discontinuous function with uniform discrete jumps, there is an additional correction term of $\mathcal{O}(N^{-1/2})$ that accounts for the discrete jumps in the distribution function. Esseen proves the following theorem [19]

Theorem 4.2.1. *Let X_1, $X_2, \cdots, X_n$ be a sequence of independent random variables with the same distribution function $F(x)$, the mean value zero, the $\sigma^2 \neq 0$, and the third moment α_3. Suppose further that X_is are discrete random variables so that $F(x)$ is a CDF with discontinuities separated by a distance d. Then*

$$F_n(x) = \Phi(x) + \frac{\alpha_3}{6\sigma^3\sqrt{2\pi n}}(1-x^2)e^{-x^2/2} + \frac{d}{\sigma\sqrt{2\pi n}}\psi_n(x)e^{-x^2/2} + \mathcal{O}(n^{-1/2}). \quad (4.23)$$

We have shown the derivation of the first two terms and will now explore the third term $\frac{d}{\sigma\sqrt{2\pi n}}\psi_n(x)e^{-x^2/2}$. Esseen [17] defines the following:

$$\psi_n(x) = Q\left(\frac{(x-\xi_n)\sigma\sqrt{n}}{d}\right) \quad (4.24)$$

$$Q(x) = [x] - x + \frac{1}{2} \quad (4.25)$$

$$\xi_n = \frac{d}{\sigma\sqrt{n}}\left\{\frac{nx_0}{d} - \left[\frac{nx_0}{d}\right]\right\} \quad (4.26)$$

where $[\cdot]$ is the floor function and ξ_n is the "least non-negative discontinuity point." of $F_n(x)$ [17]. In the context of the problem above, recall that

$$Z_N = \frac{R_1 + \cdots + R_N}{\sigma\sqrt{N}} - \frac{\sqrt{N}p}{\sigma} \quad (4.27)$$

with $R_i \sim \text{Bernoulli}(p)$ random variables with mean zero.

Since each R_i can be zero or one, Z_n has support on

$$\left\{\frac{-\sqrt{N}p}{\sigma}, \frac{1}{\sigma\sqrt{N}} - \frac{\sqrt{N}p}{\sigma}, \frac{2}{\sigma\sqrt{N}} - \frac{\sqrt{N}p}{\sigma}, \cdots, \frac{N}{\sqrt{N}\sigma} - \frac{\sqrt{N}p}{\sigma}\right\}.$$

Simplified, we obtain that Z_N has support on

$$\left\{\frac{-\sqrt{N}p}{\sigma} + \frac{k}{\sigma\sqrt{N}}\right\}, \quad \text{for } k \in \{0, 1, \cdots, N\}.$$

Thus, ξ_N occurs for the ceiling of k, denoted $\lceil k \rceil$, i.e. when k satisfies

$$0 = \frac{-\sqrt{N}p}{\sigma} + \frac{k}{\sigma\sqrt{N}}. \tag{4.28}$$

Hence $k = Np$ and since the ceiling of k is given by $[Np] + 1$, we have

$$\xi_N = \frac{-\sqrt{N}p}{\sigma} + \frac{[Np] + 1}{\sigma\sqrt{N}} = \frac{-Np + [Np] + 1}{\sigma\sqrt{N}}. \tag{4.29}$$

Since the CDF of R_i has discontinuitites at $-p$ and $(1-p)$, we have that $d = 1$, the discontinuity correction term becomes

$$\frac{1}{\sqrt{2\pi Np(1-p)}} Q\left(\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}} - \frac{-Np + [Np] + 1}{\sqrt{p(1-p)}\sqrt{N}}\right)\sqrt{p(1-p)}\sqrt{N}\right) e^{-\frac{(x-p)^2 N}{2p(1-p)}}.$$

Moreover,

$$\begin{aligned} Q\left(\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}} - \frac{-Np + [Np] + 1}{\sqrt{p(1-p)}\sqrt{N}}\right)\sqrt{p(1-p)}\sqrt{N}\right) &= Q\left((x-p)N - (1 + [Np] - Np)\right) \\ &= Q\left(xN - 1 - [Np]\right). \end{aligned} \tag{4.30}$$

So (4.23) becomes

$$F_{Z_N}(x) = \Phi(x) + \frac{p - 3p^2 + 2p^3}{6(p(1-p))^{3/2}\sqrt{2\pi N}}(1 - x^2)e^{-x^2/2} + \frac{Q\left(xN - 1 - [Np]\right)}{\sigma\sqrt{2\pi N}} e^{-x^2/2} + \mathcal{O}(N^{-1/2}). \tag{4.31}$$

since $\sigma = [p(1-p)]^{1/2}$ and $\alpha_3 = p - 3p^2 + 2p^3$ when $R_i \sim \text{Bernoulli}(p)$. Again, since $F_{R_N^*|\hat{\epsilon}}(x) = F_{Z_n}\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right)$, we have that the CDF of $F_{R_N^*}$ conditioned on $\hat{\epsilon}$ is given by

$$\begin{aligned} F_{R_N^*|\hat{\epsilon}}(x) = \Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right) &+ \frac{(p-3p^2+2p^3)}{6\sqrt{N}(p(1-p))^{3/2}}\left(1 - \frac{(x-p)^2 N}{p(1-p)}\right) e^{-\frac{(x-p)^2 N}{2p(1-p)}} \\ &+ \frac{1}{\sqrt{2\pi N p(1-p)}} Q\left(xN - 1 - [Np]\right) e^{-\frac{(x-p)^2 N}{2p(1-p)}} \end{aligned} \tag{4.32}$$

To find the unconditioned CDF, we multiply by $\phi(\hat{\epsilon})$ and integrate over all values of $\hat{\epsilon}$.

$$\begin{aligned} \int_{-\infty}^{\infty} F_{R_N^*|\hat{\epsilon}}(x)\phi(\hat{\epsilon})d\hat{\epsilon} = &\int_{-\infty}^{\infty} \Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right)\phi(\hat{\epsilon})d\hat{\epsilon} \\ &+ \int_{-\infty}^{\infty} \frac{(p-3p^2+2p^3)}{6\sqrt{N}(p(1-p))^{3/2}}\left(1 - \frac{(x-p)^2 N}{p(1-p)}\right) e^{-\frac{(x-p)^2 N}{2p(1-p)}}\phi(\hat{\epsilon})d\hat{\epsilon} \\ &+ \int_{-\infty}^{\infty} \frac{Q\left(xN - 1 - [Np]\right) e^{-\frac{(x-p)^2 N}{2p(1-p)}}}{\sqrt{2\pi N p(1-p)}}\phi(\hat{\epsilon})d\hat{\epsilon}. \end{aligned} \tag{4.33}$$

It can be shown that the Vasicek approximation to the CDF is asymptotic to

$$\int_{-\infty}^{\infty} \Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right)\phi(\hat{\epsilon})d\hat{\epsilon}.$$

To show this, let

$$F_N(x) = \int_{-\infty}^{\infty} \Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right)\phi(\hat{\epsilon})d\hat{\epsilon}. \tag{4.34}$$

Taking the derivative with respect to x yields

$$F_N'(x) = \int_{-\infty}^{\infty} \phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right)\frac{\sqrt{N}}{\sqrt{p(1-p)}}\phi(\hat{\epsilon})d\hat{\epsilon} \tag{4.35}$$

$$\implies F_N'(x) \sim \int_{-\infty}^{\infty} e^{\frac{-N(x-p)^2}{2p(1-p)}} \frac{\sqrt{N}}{\sqrt{2\pi p(1-p)}}\phi(\hat{\epsilon})d\hat{\epsilon} \tag{4.36}$$

Notice that this integral is the same as in (2.19). Following the procedure in Section 2.2 using Laplace's method and (2.21),

$$\begin{aligned} F'_N(x) &\sim \sqrt{N}\left\{\alpha(\hat{\epsilon}^*)\exp\{-N\tau(\hat{\epsilon}^*)\}\sqrt{\frac{2\pi}{N\tau''(\hat{\epsilon}^*)}} + \mathcal{O}(N^{-3/2})\right\} \\ &\sim \alpha(\hat{\epsilon}^*)\exp\{-N\tau(\hat{\epsilon}^*)\}\sqrt{\frac{2\pi}{\tau''(\hat{\epsilon}^*)}} + \mathcal{O}(N^{-1}) \end{aligned} \tag{4.37}$$

where $\alpha(\hat{\epsilon}) = \frac{\phi(\hat{\epsilon})}{\sqrt{2\pi p(\hat{\epsilon})(1-p(\hat{\epsilon}))}}$ and $\tau(\hat{\epsilon})$, $\tau''(\hat{\epsilon})$ are defined as in (2.22) and (2.27), respectively. The $\mathcal{O}(N^{-1})$ are given by

$$\frac{1}{N}\sqrt{\frac{2\pi}{\tau''(\hat{\epsilon}^*)}}\left\{\frac{\alpha''(\hat{\epsilon}^*)}{2\tau''(\hat{\epsilon}^*)} + \frac{\alpha(\hat{\epsilon}^*)d\tau^{(4)}/d\hat{\epsilon}^4(\hat{\epsilon}^*)}{8[\tau''(\hat{\epsilon}^*)]^2} + \frac{\alpha'(\hat{\epsilon}^*)\tau'''(\hat{\epsilon}^*)}{2[\tau''(\hat{\epsilon}^*)]^2} - \frac{5[\tau'''(\hat{\epsilon}^*)]^2\alpha(\hat{\epsilon}^*)}{24[\tau''(\hat{\epsilon}^*)]^3}\right\} \tag{4.38}$$

Integrating over x, we obtain the Vasicek approximated CDF found in (2.34) plus an error term of $\mathcal{O}(N^{-1})$, as $N \to \infty$. Hence,

$$\underbrace{\int_{-\infty}^{\infty}\Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right)\phi(\hat{\epsilon})d\hat{\epsilon}}_{\text{First term on the right-hand side of (4.32)}} = \underbrace{\Phi\left(\frac{\sqrt{1-\rho}\Phi^{-1}(x)-\theta}{\sqrt{\rho}}\right)}_{\text{Vasicek's CDF, equation (2.37)}} + E_N(x) \tag{4.39}$$

where the $E_N(x)$ error term of $\mathcal{O}(N^{-1})$ is given by

$$\begin{aligned} E_N(x) &= \int_0^x \frac{1}{N}\sqrt{\frac{2\pi}{\tau''(\hat{\epsilon}^*)}}\left\{\frac{\alpha''(\hat{\epsilon}^*)}{2\tau''(\hat{\epsilon}^*)} + \frac{\alpha(\hat{\epsilon}^*)d\tau^{(4)}/d\hat{\epsilon}^4(\hat{\epsilon}^*)}{8[\tau''(\hat{\epsilon}^*)]^2} + \frac{\alpha'(\hat{\epsilon}^*)\tau'''(\hat{\epsilon}^*)}{2[\tau''(\hat{\epsilon}^*)]^2} - \frac{5[\tau'''(\hat{\epsilon}^*)]^2\alpha(\hat{\epsilon}^*)}{24[\tau''(\hat{\epsilon}^*)]^3}\right\}dx' \\ &= \frac{1}{N}\sqrt{\frac{2\pi}{\tau''(\hat{\epsilon}^*)}}\left\{\frac{\alpha''(\hat{\epsilon}^*)}{2\tau''(\hat{\epsilon}^*)} + \frac{\alpha(\hat{\epsilon}^*)d\tau^{(4)}/d\hat{\epsilon}^4(\hat{\epsilon}^*)}{8[\tau''(\hat{\epsilon}^*)]^2} + \frac{\alpha'(\hat{\epsilon}^*)\tau'''(\hat{\epsilon}^*)}{2[\tau''(\hat{\epsilon}^*)]^2} - \frac{5[\tau'''(\hat{\epsilon}^*)]^2\alpha(\hat{\epsilon}^*)}{24[\tau''(\hat{\epsilon}^*)]^3}\right\} \end{aligned} \tag{4.40}$$

Plotting the CDF obtained from Monte Carlo simulations against the right-hand side of (4.33) shows that the asymptotic approximation of the CDF converges as the number of companies in the portfolio gets large.

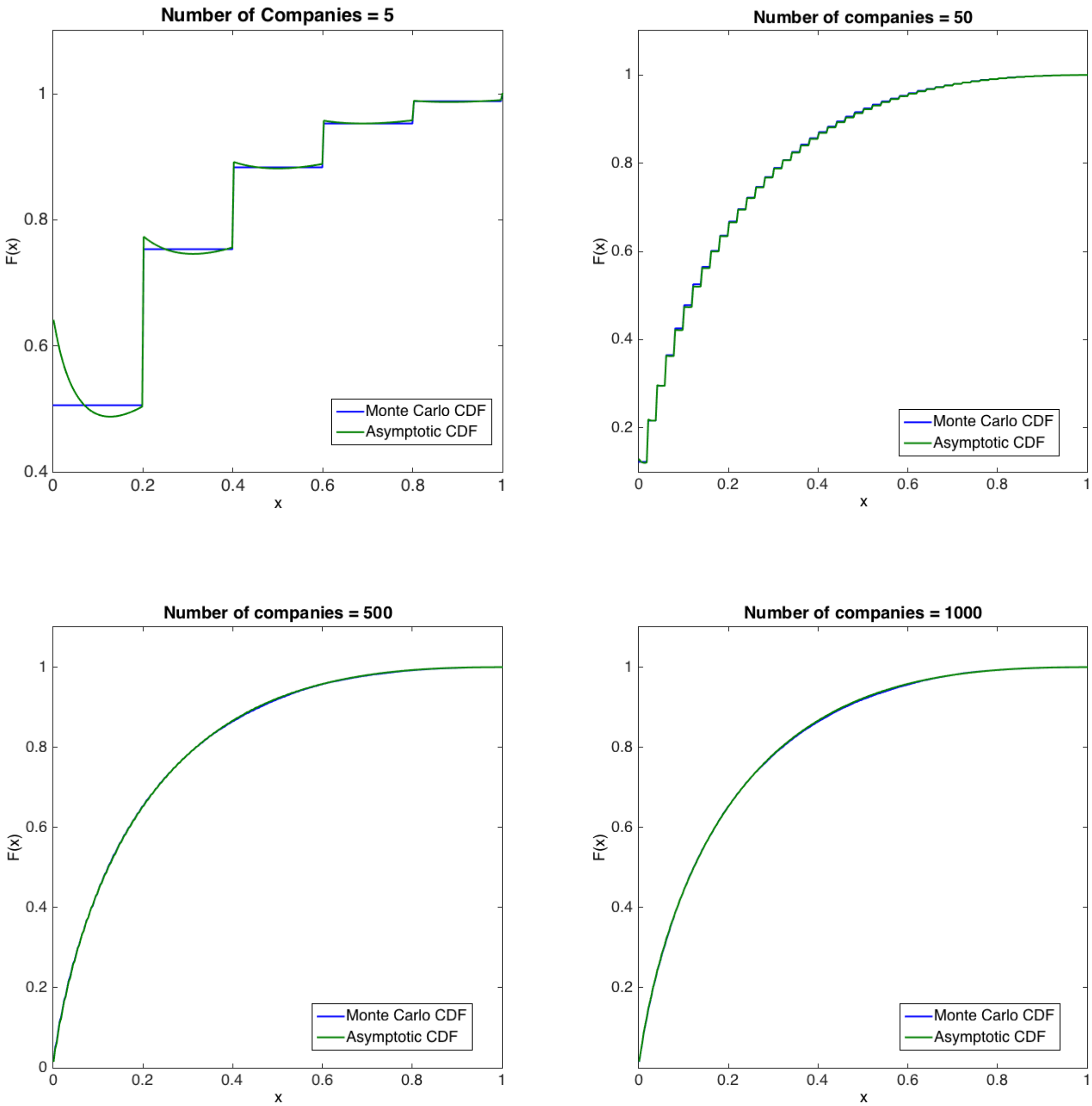


Figure 4.1: A comparison of the CDF generated by Monte Carlo simulations versus the CDF of the asymptotic approximation for different sized portfolios. The green lines plot the right-hand side of equation (4.33) while the blue lines plot the Monte Carlo simulated left-hand side of equation (4.33). Plots were generated using ρ = 0.4, θ = -0.9, and 50,000 Monte Carlo realizations.

4.3 Scaling of the asymptotic error

Recall that $F_{R_N^*}$ in (4.32) is the CDF of R_N^* conditioned on $\hat{\epsilon}$. In order to find the unconditioned form, we must multiply by $\phi(\hat{\epsilon})$ and integrate over all values of $\hat{\epsilon}$. Thus to find the scaling of the error, we consider

$$I(N) = \int_{-\infty}^{\infty} \frac{(p - 3p^2 + 2p^3)}{6\sqrt{N}(p(1-p))^{3/2}} \left(1 - \frac{(x-p)^2 N}{p(1-p)}\right) e^{-\frac{(x-p)^2 N}{2p(1-p)}} \phi(\hat{\epsilon}) d\hat{\epsilon} + \int_{-\infty}^{\infty} \frac{Q\left(xN - 1 - [Np]\right) e^{-\frac{(x-p)^2 N}{2p(1-p)}}}{\sqrt{2\pi N p(1-p)}} \phi(\hat{\epsilon}) d\hat{\epsilon}. \tag{4.41}$$

Let

$$I_1(N) = \int_{-\infty}^{\infty} \frac{(p - 3p^2 + 2p^3)}{6\sqrt{N}(p(1-p))^{3/2}} e^{-\frac{(x-p)^2 N}{2p(1-p)}} \phi(\hat{\epsilon}) d\hat{\epsilon} \tag{4.42}$$

$$I_2(N) = \int_{-\infty}^{\infty} \frac{(p - 3p^2 + 2p^3)}{6\sqrt{N}(p(1-p))^{3/2}} \frac{(x-p)^2 N}{p(1-p)} e^{-\frac{(x-p)^2 N}{2p(1-p)}} \phi(\hat{\epsilon}) d\hat{\epsilon} \tag{4.43}$$

and

$$I_3(N) = \int_{-\infty}^{\infty} \frac{Q\left(xN - 1 - [Np]\right)}{\sqrt{2\pi N p(1-p)}} e^{-\frac{(x-p)^2 N}{2p(1-p)}} \phi(\hat{\epsilon}) d\hat{\epsilon} \tag{4.44}$$

For $N \to \infty$, we can use Laplace's method to approximate the integrals.

Considering $I_1(N)$ first, we transform this integral into a Laplace-type so as to use Laplace's method. First, let $\tau(\hat{\epsilon}) = \frac{(x - p(\hat{\epsilon}))^2}{2p(\hat{\epsilon})(1-p(\hat{\epsilon}))}$. Then $\tau(\hat{\epsilon})$ has a minimum at $\hat{\epsilon}^*$ where $\hat{\epsilon}^*$ is defined as in (2.25). Thus, $\tau(\hat{\epsilon}^*) = 0$.

Hence we have

$$I_1(N) = \int_{-\infty}^{\infty} \frac{(p(\hat{\epsilon}) - 3p(\hat{\epsilon})^2 + 2p(\hat{\epsilon})^3)}{6\sqrt{N}(p(\hat{\epsilon})(1-p(\hat{\epsilon})))^{3/2}} e^{-N\tau(\hat{\epsilon})} \phi(\hat{\epsilon}) d\hat{\epsilon} \tag{4.45}$$

Letting

$$\alpha_1(\hat{\epsilon}) = \frac{(p(\hat{\epsilon}) - 3p(\hat{\epsilon})^2 + 2p(\hat{\epsilon})^3)}{6\sqrt{N}(p(\hat{\epsilon})(1-p(\hat{\epsilon})))^{3/2}} \phi(\hat{\epsilon}) \tag{4.46}$$

Then we have the Laplace-type integral,

$$I_1(N) = \int_{-\infty}^{\infty} \alpha_1(\hat{\epsilon}) e^{-N\tau(\hat{\epsilon})} d\hat{\epsilon} \tag{4.47}$$

Note that $\tau(\hat{\epsilon})$ is defined as in (2.22), and using that $\tau(\hat{\epsilon}^*) = 0$ along with (2.25) and (2.27)-(2.29) we have by Laplace's method,

$$\begin{aligned} I_1(N) &\sim \alpha_1(\hat{\epsilon}^*) \exp\{-N\tau(\hat{\epsilon}^*)\} \sqrt{\frac{2\pi}{N\tau''(\hat{\epsilon}^*)}} \\ &= \frac{(p(\hat{\epsilon}^*) - 3p(\hat{\epsilon}^*)^2 + 2p(\hat{\epsilon}^*)^3)}{6\sqrt{N}(p(\hat{\epsilon}^*)(1-p(\hat{\epsilon}^*)))^{3/2}} \phi(\hat{\epsilon}^*) \sqrt{\frac{2\pi}{N\tau''(\hat{\epsilon}^*)}} \\ &= \frac{p(\hat{\epsilon}^*) - 3p(\hat{\epsilon}^*)^2 + 2p(\hat{\epsilon}^*)^3}{6Np(\hat{\epsilon}^*)(1-p(\hat{\epsilon}^*))} \frac{e^{-\frac{1}{2}\hat{\epsilon}^{*2}}}{p'(\hat{\epsilon}^*)} \end{aligned} \tag{4.48}$$

which is $\mathcal{O}(N^{-1})$.

Similarly, for the second integral in $I_2(N)$, we define $\tau(\hat{\epsilon}^*)$ as above and let

$$\begin{aligned} \alpha_2(\hat{\epsilon}) &= \frac{p(\hat{\epsilon}) - 3p(\hat{\epsilon})^2 + 2p(\hat{\epsilon})^3}{6(p(\hat{\epsilon})(1-p(\hat{\epsilon})))^{3/2}} \tau(\hat{\epsilon})\phi(\hat{\epsilon}) \\ &= \frac{p(\hat{\epsilon}) - 3p(\hat{\epsilon})^2 + 2p(\hat{\epsilon})^3}{6(p(\hat{\epsilon})(1-p(\hat{\epsilon})))^{3/2}} \tau(\hat{\epsilon})\phi(\hat{\epsilon}) \end{aligned} \tag{4.49}$$

Then we have the integral

$$I_2(N) = \sqrt{N} \int_{-\infty}^{\infty} \alpha_2(\hat{\epsilon}) e^{-N\tau(\hat{\epsilon})} d\hat{\epsilon} \tag{4.50}$$

Since $\alpha(\hat{\epsilon}^*)$ vanishes, we must expand the integral using higher-order Laplace terms. Using [13], the higher order approximation of $I_2(N)$ using Laplace's method are given by

$$\begin{aligned} I_2(N) \sim & \sqrt{N} \sqrt{\frac{2\pi}{N\tau''(\hat{\epsilon}^*)}} e^{-N\tau(\hat{\epsilon}^*)} \left\{ \alpha_2(\hat{\epsilon}^*) + \frac{1}{N} \left[\frac{-\alpha_2''(\hat{\epsilon}^*)}{2\tau''(\hat{\epsilon}^*)} \right.\right. \\ & \left.\left. + \frac{\alpha_2(\hat{\epsilon}^*)(d^4\tau/d\hat{\epsilon}^4)(\hat{\epsilon}^*)}{8[\tau''(\hat{\epsilon}^*)]^2} + \frac{\alpha_2'(\hat{\epsilon}^*)\tau'''(\hat{\epsilon}^*)}{2[\tau''(\hat{\epsilon}^*)]^2} - \frac{5[\tau'''(\hat{\epsilon}^*)]^2\alpha_2(\hat{\epsilon}^*)}{24[\tau''(\hat{\epsilon}^*)]^3} \right] \right\} \end{aligned} \tag{4.51}$$

as $N \to \infty$.

In our case, the $\alpha_2(\hat{\epsilon}^*)$ terms vanish and the term with the first derivative in α_2 will not contribute because it is multiplied against a Gaussian and integrated through a symmetric interval during Laplace's method. Hence using the fact that $\alpha_2(\hat{\epsilon}^*) = 0$ and the definitions in (2.25) and (2.27)-(2.29), we have

$$\begin{aligned} I_2(N) &\sim \sqrt{\frac{2\pi}{\tau''(\hat{\epsilon}^*)}} \frac{1}{N} \frac{-\alpha_2''(\hat{\epsilon}^*)}{2\tau''(\hat{\epsilon}^*)} \\ &\sim \frac{\sqrt{2\pi}\phi(\hat{\epsilon}^*)}{N} \frac{-\alpha_2''(\hat{\epsilon}^*)}{2\left(\frac{p'(\hat{\epsilon}^*)^2}{p(\hat{\epsilon}^*)(1-p(\hat{\epsilon}^*))}\right)^{3/2} \phi(\hat{\epsilon}^*)} \\ &\sim \frac{\sqrt{2\pi}\phi(\hat{\epsilon}^*)}{Np'(\hat{\epsilon}^*)} \frac{-\alpha_2''(\hat{\epsilon}^*)[p(\hat{\epsilon}^*)(1-p(\hat{\epsilon}^*))]^{3/2}}{2\phi(\hat{\epsilon}^*)p'(\hat{\epsilon}^*)^2} \end{aligned} \tag{4.52}$$

which is $\mathcal{O}(N^{-1})$.

Lastly, we consider $I_3(N)$ from (4.44). Defining $\tau(\hat{\epsilon})$ as above and

$$\alpha_3(\hat{\epsilon}) = \frac{Q\left(Nx - 1 - [Np(\hat{\epsilon})]\right)}{\sqrt{2\pi Np(\hat{\epsilon})(1-p(\hat{\epsilon}))}} \phi(\hat{\epsilon}) \tag{4.53}$$

we have

$$I_3(N) = \int_{-\infty}^{\infty} \alpha_3(\hat{\epsilon}) e^{-N\tau(\hat{\epsilon})} d\hat{\epsilon} \tag{4.54}$$

However, Laplace's method requires $\alpha_3(\hat{\epsilon})$ to be sufficiently smooth and $Q\left(Nx - 1 - [Np(\hat{\epsilon})]\right)$ has discontinuities. Thus we first bound $I_3(N)$ by noting

$$\begin{aligned} |I_3(N)| &= \left| \int_{-\infty}^{\infty} \frac{Q\left(Nx - 1 - [Np(\hat{\epsilon})]\right)}{\sqrt{2\pi Np(\hat{\epsilon})(1-p(\hat{\epsilon}))}} e^{-N\tau(\hat{\epsilon})} \phi(\hat{\epsilon}) d\hat{\epsilon} \right| \\ &\leq \int_{-\infty}^{\infty} \left| \frac{Q\left(Nx - 1 - [Np(\hat{\epsilon})]\right)}{\sqrt{2\pi Np(\hat{\epsilon})(1-p(\hat{\epsilon}))}} e^{-N\tau(\hat{\epsilon})} \phi(\hat{\epsilon}) \right| d\hat{\epsilon} \end{aligned} \tag{4.55}$$

By Cauchy-Schwarz,

$$\begin{aligned}
|I_3(N)| &\leq \int_{-\infty}^{\infty} \left| \frac{Q\left(Nx - 1 - [Np(\hat{\epsilon})]\right)}{\sqrt{2\pi N p(\hat{\epsilon})(1-p(\hat{\epsilon}))}} e^{-N\tau(\hat{\epsilon})} \phi(\hat{\epsilon}) \right| d\hat{\epsilon} \\
&\leq \int_{-\infty}^{\infty} \left| Q\left(Nx - 1 - [Np(\hat{\epsilon})]\right) \right| \left| \frac{\phi(\hat{\epsilon})}{\sqrt{2\pi N p(\hat{\epsilon})(1-p(\hat{\epsilon}))}} e^{-N\tau(\hat{\epsilon})} \right| d\hat{\epsilon} \\
&\leq \int_{-\infty}^{\infty} \max\left(\left| Q\left(Nx - 1 - [Np(\hat{\epsilon})]\right) \right| \right) \left| \frac{\phi(\hat{\epsilon})}{\sqrt{2\pi N p(\hat{\epsilon})(1-p(\hat{\epsilon}))}} e^{-N\tau(\hat{\epsilon})} \right| d\hat{\epsilon} \\
&\leq \frac{1}{2} \int_{-\infty}^{\infty} \frac{\phi(\hat{\epsilon})}{\sqrt{2\pi N p(\hat{\epsilon})(1-p(\hat{\epsilon}))}} e^{-N\tau(\hat{\epsilon})} d\hat{\epsilon}
\end{aligned} \tag{4.56}$$

since $|Q\left(Nx - 1 - [Np(\hat{\epsilon})]\right)| \leq \ 1/2$.

We now use Laplace's method by defining $\alpha_3(\hat{\epsilon})$ such that

$$\alpha_3(\hat{\epsilon}) = \frac{\phi(\hat{\epsilon})}{\sqrt{2\pi N p(\hat{\epsilon})(1-p(\hat{\epsilon}))}}. \tag{4.57}$$

Using the definitions given in (2.25) and (2.27)-(2.29) yields

$$\begin{aligned}
I_3(N) &\sim \frac{1}{2} \frac{\phi(\hat{\epsilon}^*)}{\sqrt{2\pi N p(\hat{\epsilon}^*)(1-p(\hat{\epsilon}^*))}} \sqrt{\frac{2\pi}{N\tau''(\hat{\epsilon}^*)}} \exp\{-N\tau(\hat{\epsilon}^*)\} \\
&= \frac{\phi(\hat{\epsilon}^*)}{2Np'(\hat{\epsilon}^*)}
\end{aligned} \tag{4.58}$$

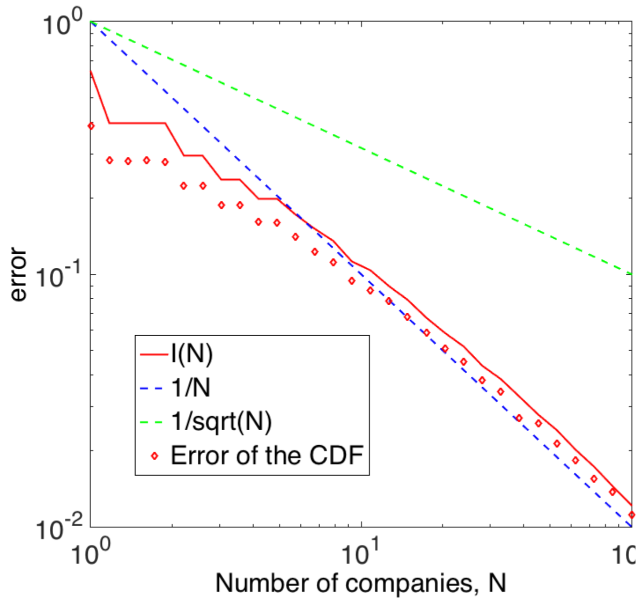
which is $\mathcal{O}(N^{-1})$. Therefore, the integral in (4.41) is asymptotic to

$$\frac{\sqrt{2\pi}\phi(\hat{\epsilon}^*)}{Np'(\hat{\epsilon}^*)} \left\{ \frac{p(\hat{\epsilon}^*) - 3p(\hat{\epsilon}^*)^2 + 2p(\hat{\epsilon}^*)^3}{6p(\hat{\epsilon}^*)(1-p(\hat{\epsilon}^*))} - \frac{\alpha_2''(\hat{\epsilon}^*)[p(\hat{\epsilon}^*)(1-p(\hat{\epsilon}^*))]^{3/2}}{2\phi(\hat{\epsilon}^*)p'(\hat{\epsilon}^*)^2} + \frac{1}{2\sqrt{2\pi}} \right\} \tag{4.59}$$

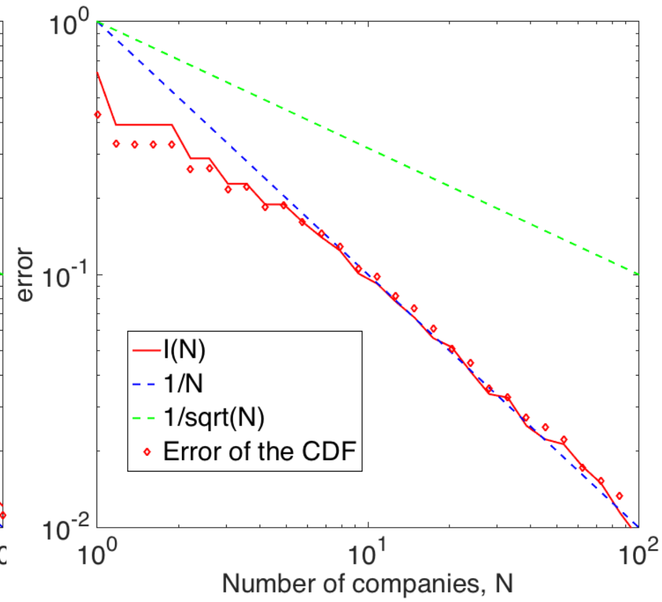
This confirms the $\mathcal{O}(N^{-1})$ scaling of the error between the Monte Carlo simulated CDF and the Vasicek approximated CDF seen numerically in section 3.3. Moreover, we numerically verify the scaling for

$$\left| \int_{-\infty}^{\infty} F_{R_N^*|\hat{\epsilon}}(x)\phi(\hat{\epsilon})d\hat{\epsilon} - \int_{-\infty}^{\infty} \Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}} \right) \phi(\hat{\epsilon})d\hat{\epsilon} \right| \tag{4.60}$$

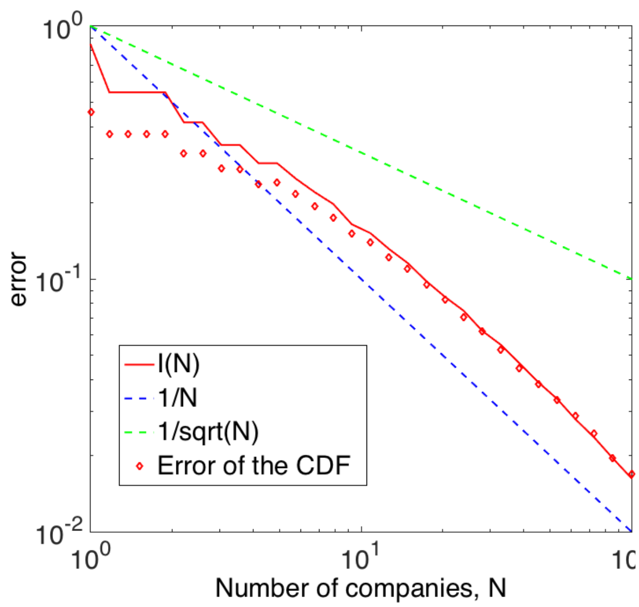
represented by light blue lines in the plots to follow compared with $I(N)$ in (4.41) represented by the red lines. Notice that as N increases, the two lines converge as expected, i.e. the asymptotic approximation improves for larger N. Moreover, the scaling is consistent with $\mathcal{O}(N^{-1})$.



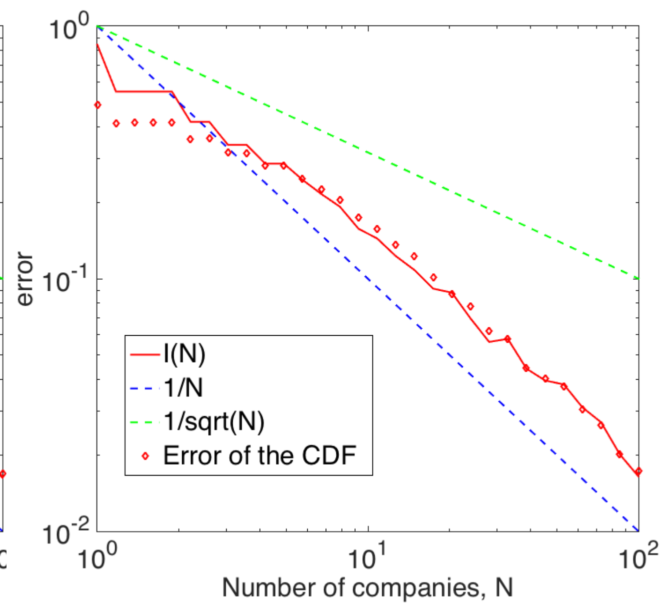
(a) θ = -0.3, ρ = 0.5



(b) θ = -0.5, ρ = 0.4



(c) θ = -0.7, ρ = 0.4



(d) θ = -0.9, ρ = 0.3

Figure 4.2: The error of the Monte Carlo simulated CDF against the asymptotic approximated CDF. The red line represents the asymptotic terms given in equation (4.41). Plots were generated using 100,000 trials Monte Carlo trials.

Chapter 5

CONCLUSIONS

Error quantification of approximated distributions remains an important study in mathematics, particularly in mathematical finance where such approximations are used to guard against financial losses during times of economic stress. In this thesis, we build upon the work of Esseen and others who seek to quantify the error of the approximation of a sum of N independent Bernoulli random variables with a normal distribution. In this case, the error incurred in the CDF was $\mathcal{O}(N^{-1/2})$. Our main contribution is studying this error when the random variables summed are not independent. We find the scaling of error both analytically and numerically to be, surprisingly, $\mathcal{O}(N^{-1})$.

We started by deriving the Vasicek CDF for cumulative portfolio loss, as well as the associated approximation of VaR_q. In chapter 3, we analytically and numerically explored the bound of the error between our analytic approximation and the Monte Carlo simulated CDF. Lastly, we quantified the scaling of the error due to the central limit theorem using asymptotic analysis.

We found that the total error of (4.60) is comprised of two terms - the error associated with Laplace's method in estimating the Vasicek approximation and the error derived in chapter 4 associated with the approximation of the CDF using the central limit. Recall that (4.40) gives the integral approximation using Laplace's method and the associated error which scales like $\mathcal{O}(N^{-1})$. From chapter 4, we find that the error associated with the central limit theorem approximation is given by (4.59). Therefore, the total error of the Vasicek approximation is $\mathcal{O}(N^{-1})$ and is given by

$$\begin{aligned}\int_{-\infty}^{\infty} F_{R_N^*|\hat{\epsilon}}(x)\phi(\hat{\epsilon})d\hat{\epsilon} &= \int_{-\infty}^{\infty} \Phi\left(\frac{(x-p)\sqrt{N}}{\sqrt{p(1-p)}}\right)\phi(\hat{\epsilon})d\hat{\epsilon} + \underbrace{(4.59)}_{\text{Central Limit theorem error}} \\ &= \underbrace{\Phi\left(\frac{\sqrt{1-\rho}\Phi^{-1}(x)-\theta}{\sqrt{\rho}}\right)}_{\text{Vasicek approximation}} + \underbrace{(4.40)}_{\text{Laplace approximation error}} + \underbrace{(4.59)}_{\text{Central Limit theorem error}}\end{aligned} \tag{5.1}$$

With an understanding of the error scaling for the Vasicek model, future work includes exploring the scaling of the error for other methods of calculating VaR_q, such as through historical simulations or the Delta-Normal approach [11]. Further work also remains to extend this analysis to the hierarchical multi-factor model, a structural model allowing for correlation among companies on a global and sector/regional scale.

BIBLIOGRAPHY

[1] E Brau. International monetary fund financial risk in the fund and the level of precautionary balances background paper. *Prepared by the Finance Department*, pages 429–452, 2004.

[2] Arnaud De Servigny and Olivier Renault. *Measuring and managing credit risk*. McGraw Hill Professional, 2004.

[3] Michel Denault. Coherent allocation of risk capital. *Journal of risk*, 4:1–34, 2001.

[4] William Morokoff, Liming Yang, and Saiyid Islam. Approximating correlation matrices. 2012.

[5] Michel Crouhy, Dan Galai, and Robert Mark. A comparative analysis of current credit risk models. *Journal of Banking & Finance*, 24(1):59–117, 2000.

[6] Michael Pykhtin. Portfolio credit risk multi-factor adjustment. *RISK-LONDON-RISK MAGAZINE LIMITED-*, 17(3):85–90, 2004.

[7] Oldrich A Vasicek. *Probability of loss on loan portfolio*. Wiley Online Library, 1987.

[8] Pak-Wing Fok, Xiuling Yan, and Guangming Yao. Analysis of credit portfolio risk using hierarchical multifactor models. *Journal of Credit Risk*, 10(4):45–70, 2014.

[9] Philipp J Schönbucher. Factor models: Portfolio credit risks when defaults are correlated. *The Journal of Risk Finance*, 3(1):45–56, 2001.

[10] Xinzheng Huang, Cornelis W Oosterlee, and JAM Van Der Weide. Higher order saddlepoint approximations in the vasicek portfolio credit loss model. Technical report, Delft University of Technology, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft Institute of Applied Mathematics, 2006.

[11] Thomas J Linsmeier and Neil D Pearson. Value at risk. *Financial Analysts Journal*, 56(2):47–67, 2000.

[12] Banking Supervision. Basel committee on banking supervision, 2006.

[13] Steven A Orszag and CM Bender. Advanced mathematical methods for scientists and engineers. *Mac Graw Hill*, 1978.

[14] Mark J Ablowitz and Athanassios S Fokas. *Complex variables: introduction and applications*. Cambridge University Press, 2003.

[15] Christian Bluhm, Ludger Overbeck, and Christoph Wagner. *Introduction to credit risk modeling*. Crc Press, 2016.

[16] Ross Sheldon et al. *A first course in probability*. Pearson Education India, 2002.

[17] Carl-Gustav Esseen. Fourier analysis of distribution functions. a mathematical study of the laplace-gaussian law. *Acta Mathematica*, 77(1):1–125, 1945.

[18] Andrew C Berry. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the American Aathematical Society*, 49(1):122–136, 1941.

[19] Carl-Gustaf Esseen. *On the Liapounoff limit of error in the theory of probability*. Almqvist & Wiksell, 1942.

[20] Rabindra N Bhattacharya and Ramaswamy Ranga Rao. *Normal approximation and asymptotic expansions*, volume 64. SIAM, 1986.