Computation of value-at-risk: the fast convolution method, dimension reduction and perturbation theory

by

Petter Wiberg

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Computer Science
University of Toronto

Abstract

Computation of value-at-risk: the fast convolution method, dimension reduction and perturbation theory

Petter Wiberg
Doctor of Philosophy
Graduate Department of Computer Science
University of Toronto
2002

Value-at-risk is a measure of market risk for a portfolio. Market risk is the chance that the portfolio declines in value due to changes in market variables. This thesis is about the computation of value-at-risk for portfolios with derivatives and for models for returns that have a distribution with fat tails.

We introduce a new Fourier algorithm, the fast convolution method, for computing value-at-risk. The fast convolution method is different from other Fourier methods in that it does not require that the characteristic function of the portfolio returns be known explicitly. Our new method can therefore be used with more general return models. In the thesis we present experiments with three return models: the normal model, the asymmetric T model and a model using the non-parametric Parzen density estimator. We also discuss how the fast convolution method can be extended to compute the value-at-risk gradient, present a proof of convergence and illustrate the performance of the method with examples.

We develop and compare two methods for dimension reduction in the computation of value-at-risk. The goal of dimension reduction is to reduce computation time by finding a small model that captures the main dynamics of the original model. We compare the two methods for an example problem and conclude that the method based on mean square error is superior. Finally, we present an optimization example that illustrates that dimension reduction may reduce the time to compute value-at-risk while maintaining good accuracy.

We develop a perturbation theory for value-at-risk with respect to changes in the return model. By considering variational properties, we derive a first-order error bound and find the condition number of value-at-risk. We argue that the sensitivity observed in empirical studies is an inherent limitation of value-at-risk.

Acknowledgements

I would like to thank my two supervisors, Claudio Albanese and Ken Jackson. They have encouraged, supported and challenged me throughout this work. I am thankful for their efforts, but also for what they have not done. Ken and Claudio gave me room to follow my own ideas and intuition and this has been a most valuable lesson.

I would also like to thank my friends at the University of Toronto and the visitors to the Fields Institute during the thematic year on numerical and computational challenges in science and engineering. In particular, I would like to thank Alf Gerisch, Greg Lewis and Des Higham. We have had many interesting discussions and hope to have more in the future. My thesis committee, Jin-Chuan Duan, Wayne Enright, Andrey Feuerverger, and Luis Seco, have raised many questions and I have answered some of them.

Finally, I would like to thank Jessica for her love.

Contents

Chapter 1

Introduction

Companies are exposed to many types of risks, future events that if they occurred would result in the loss of money. The purpose of risk management is to quantify and control these dangers. Value-at-risk (VaR) is a measure of the market risk, the chance of a loss in a company's portfolio caused by unfavorable changes in prices and rates. Several high profile financial disasters[1] [39, 25, 38] serve as reminders of the importance of understanding the sources of risk, some of which were largely due to failure to control exposure to changes in the markets. Minimum risk management standards for financial institutions are set and enforced by national regulators. The Basel Accord [4], the market risk amendment [5, 6] and the recent update [4] contain the international guidelines implemented by the national agencies.[2] Value-at-risk has become the industry standard for quantifying market risk, partly because of its intuitive appeal and, more importantly, because it is endorsed in the Basel Accord.

For a given portfolio, value-at-risk is defined as the maximum loss over a specified holding period and with a given confidence level (see Figure 1). In other words, it is a percentile of the distribution for changes in portfolio value. If $\Delta\Pi$ is the change in portfolio value during the holding period, then value-at-risk is the solution to a non-linear equation:

$$\mathrm{P}\left[\Delta\Pi \leq -\mathrm{VaR}\right] = 1 - \alpha, \tag{1.1}$$

where α is the confidence level. Another interpretation is that in the long term we expect losses exceeding value-at-risk with frequency $1 - \alpha$. For $\alpha = 99\%$, we expect losses exceeding value-at-risk one out of every 100 days. Regulators require value-at-risk to be computed daily with a confidence level of 99% and for a holding period of 10 days. However, since the rules allow for value-at-risk for one day to be scaled to approximate the risk for 10 days, we choose to consider daily holding periods in our examples. There are many review papers about value-at-risk simulation: Stambaugh [54] gives a high-level introduction; for more in-depth, general, algorithmic and mathematical discussions, we have a personal preference for [43, 33, 14].[3]

1.1 A small example

To illustrate this concept, let us consider a small example. Suppose that we have entered into a forward contract where we have agreed to, in six months, pay \$10 Canadian dollars for 7 euros.

[1] A good source of information on financial scandals is Roy Davies's web page at `http://www.ex.ac.uk/ RDavies/arian/scandals/`.

[2] The Basel Accord and related documents are available from the Bank of International Settlements (`www.bis.org`).

[3] The website `www.gloriamundi.org` is an excellent source for information and links to papers on value-at-risk.

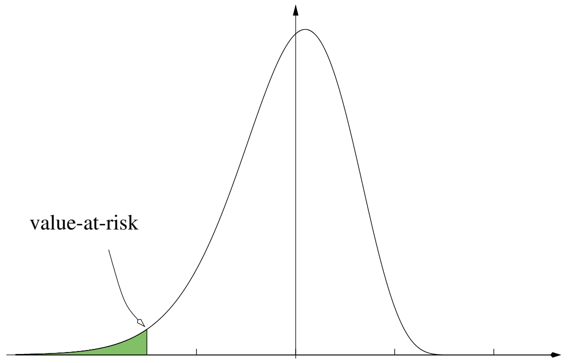


Figure 1.1: The probability that the loss is greater than value-at-risk, the density of shaded region, is equal to $1-\alpha$.

It is easy to show that the arbitrage free price is

$$V_0 = -\frac{10}{r_{CDN}} + \frac{7x_{CDN/euro}}{r_{euro}},$$

where $x_{CDN/euro}$ is the current exchange rate for euros to dollars, and r_{CDN} and r_{euro} are the current 6 month interest rates for dollars and euros, respectively. The change in the value over one day is a function of the changes in these three basic risk factors. Suppose that the daily relative changes are normally distributed with zero mean

$$\begin{aligned} R_1 &= \frac{\Delta r_{CDN}}{r_{CDN}} && \sim N(0,\sigma_1^2), \\ R_2 &= \frac{\Delta x_{CDN/euro}}{x_{CDN/euro}} && \sim N(0,\sigma_2^2), \\ R_3 &= \frac{\Delta r_{euro}}{r_{euro}} && \sim N(0,\sigma_3^2), \end{aligned}$$

and that these random variables are independent. Ignoring the effect of time, the value of the contract after one day is

$$\begin{aligned} V_1 &= -\frac{10}{r^{CDN}(1+R_1)} + \frac{7x_{CDN/euro}(1+R_2)}{r_{euro}(1+R_3)} \\ &= V_0 + R_1\frac{10}{r^{CDN}} + R_2\frac{7x_{CDN/euro}}{r_{euro}} - R_3\frac{7x_{CDN/euro}}{r_{euro}} + O(R_1^2+R_2^2+R_3^2). \end{aligned}$$

If the higher-order terms are discarded, it follows that $\Delta V = V_1 - V_0$ is a normal random variable with zero mean and variance

$$\sigma^2(\Delta V) = 100\left(\frac{\sigma_1}{r_{CDN}}\right)^2 + 49\left(\frac{x_{CDN/euro}}{r_{euro}}\right)^2(\sigma_2^2+\sigma_3^2).$$

Therefore, the value-at-risk, the solution to (1.1), is given by

$$\mathrm{VaR} = N^{-1}(\alpha)\sigma(\Delta V)$$

where $N(\cdot)$ is the cumulative distribution function (cdf) for a standard normal random variable. The inverse cdf is easy to either compute using mathematics software libraries or to look up in a table for the normal cdf (see for example [37, 52]).

1.2 Generalizing the method from the example

The method from the example generalizes [43, 33]. Suppose that the vector of relative changes R in the market risk factors is a multivariate normal random variable with the mean μ and the covariance matrix C. Similar to the example, assume that the change in portfolio value can be approximated by an affine function of the relative changes in the risk factors:

$$\Delta\Pi \approx \Xi + \Delta^T R.$$

Since $\Delta\Pi - \Xi$ is a sum of normal random variables it is itself normal. The distribution is determined by its mean and variance,

$$\begin{aligned}\mu(\Delta\Pi - \Xi) &= \mathbb{E}[\Delta\Pi - \Xi] = \Delta^T\mu,\\ \sigma^2(\Delta\Pi - \Xi) &= \mathbb{E}[\Delta^T R R^T \Delta] - \Delta^T\mu\mu^T\Delta = \Delta^T C\Delta.\end{aligned}$$

Hence, the value-at-risk is given by

$$\mathrm{VaR} = N^{-1}(\alpha)\sqrt{\Delta^T C\Delta} - \Xi - \Delta^T\mu.$$

To compute the value-at-risk with this method requires approximately $2n^2$ floating point operations where n is the number of market risk factors.

The linear model with normal relative changes has a closed form solution, but it suffers from two serious problems. First, real world returns have fatter tails than normal distributions. The model will therefore underestimate the likelihood of extreme returns which in consequence may lead to much too conservative estimates of the value-at-risk. Second, for portfolios with derivatives, the change in value is a non-linear function. The local error in the linear approximation will therefore often be unacceptable, a property that is exasperated by dynamic hedging strategies that use the linearization to eliminate risk locally. To compute value-at-risk, for models that take these difficulties into account, is a hard task.

1.3 What is to come

In this thesis we study three problems: fast methods for computing value-at-risk and its gradients, systematic methods for dimension reduction in the computation of value-at-risk, and perturbation theory for value-at-risk.

As illustrated by the example, a value-at-risk model has two main components: a model for changes in risk factors and a pricing model describing the local dynamics of portfolio value. In Chapter 2, we review basic stochastic models for risk factors. The main purpose of this chapter is to introduce the stochastic models that serve as examples in the later chapters. The use of non-parametric Parzen-type estimates for densities appears to be new in this context.

Chapter 3 discusses the second building block, portfolios and pricing of derivatives. In this chapter, we develop a new portfolio-dependent parameter estimation technique for value-at-risk models and we discuss algorithms for this problem.

In Chapter 4 we present a method for computing value-at-risk (see also [1]). A number of methods have been proposed to compute value-at-risk for portfolios that contain derivatives, see [43, 33, 14] for an overview. The algorithms can be divided into three rough categories:

Analytic approximation methods. These methods use ideas from analysis to approximate value-at-risk. Britten-Jones and Schaefer [11] use an approximation to the distribution of a sum of non-central χ^2 random variables. Mina and Ulmer [42] discuss two parametric methods: the Johnson family of distributions and the Cornish-Fisher expansion. El-Jahel et al. [16] propose a method of moments algorithm that can handle fat tailed returns. Feuerverger and Wong [23] explore a saddle point approximation method. They argue that, similar to other methods for normal returns, fat-tails can be incorporated with a mixture model.

Fourier methods. Fourier transform methods are a special class of analytic methods. The method developed in Chapter 4 is in this category. One strategy, studied by several authors, is to use a return model where the characteristic function for $\Delta\Pi$ can be derived explicitly. The distribution for $\Delta\Pi$ can then be computed by inverting the Fourier transform numerically. Mina and Ulmer [42] consider the case where returns are normal. Duffie and Pan [15] explore this idea further, and they extend the model to a fat-tailed mixture model and show how to include credit risk. Glasserman et al. [27] generalize the method to returns with a multivariate Student's T distribution.

Monte Carlo methods. Monte Carlo methods are the most general of the approaches, few assumptions have to be made about the structure of the function for portfolio value or the returns model for the risk factors. Furthermore, since they are easy to implement they are arguably the most commonly used method. The disadvantage of Monte Carlo methods is slow convergence. Therefore, much research in this area is devoted to developing efficient variance reduction techniques to improve the speed of convergence, see for example Glasserman et al. [26, 27].

To distinguish our method from other Fourier methods, we call it the *fast convolution method.* The fast convolution method is different from existing methods (see [42, 15, 27]) in that it does not require explicit knowledge about the characteristic function. It therefore enjoys greater flexibility in the type of return models that can be handled. In Chapter 4, we prove that the method converges and show how the method can be extended to compute the gradient of value-at-risk. Furthermore, we present three computational examples that illustrate the performance of the method. In the last example, we consider an optimization problem, to hedge a non-linear portfolio by minimizing the value-at-risk.

In Chapter 5, we develop and compare two methods for dimension reduction (see also [2]). The goal of dimension reduction is to extract a small model from the original high-dimensional model such that the small model has the same main properties as the large model. Hence, the computation time can be reduced by replacing the high-dimensional model by the smaller reduced model. This problem has been examined by other authors, e.g., Hull [33] gives an example for an interest rate model; Kreinin et al. [36] develop a method for linear portfolios with normal returns; and Reimers and Zerbs [49] discuss a method based on principal component analysis for asset blocks. The methods we present are more appropriate for non-linear portfolios and non-normal return models. After comparing the performance of our two new methods for an example problem, we conclude that the method based on mean square error is superior. The chapter ends with an optimization example that shows that dimension reduction may reduce computation time while retaining acceptable accuracy.

In Chapter 6, we derive a new result about the sensitivity of value-at-risk to changes in the return model. It has been observed that value-at-risk is sensitive to seemingly small changes in the return model, either in going from one class of models to another or to perturbations of the parameters of a fixed model, see for example Beder [7] or Jorion [34]. By considering variational properties of the value-at-risk equation (1.1), we derive a first order bound for the error and find the condition number of value-at-risk. This study leads to the conclusion that the sensitivity observed in empirical studies is indeed an inherent limitation of value-at-risk; for small confidence levels, large relative errors are to be expected since they are intrinsic to the problem formulation.

Chapter 2

Risk factor models

In financial markets, risk is caused by uncertainty about the value of an investment in the future. The value of a portfolio is a function of a set of *risk factors.* Risk factor is the generic term for a financial variable, for example, the price of a stock, an interest rate or a foreign exchange rate. Market risk is the risk that the value of a portfolio declines as a consequence of changes in the risk factors. Therefore, to model market risk we need to understand how risk factors evolve over time.

Our starting point is that the changes in risk factors are random. This is consistent with the *efficient market hypothesis*—prices reflect all available information at all times. Therefore, it is not possible to device a strategy that consistently gives higher returns than the market. In its weak form, the hypothesis suggests that technical analysis, using past stock prices, is doomed to fail. To what extent markets are efficient is a thoroughly studied question; see for example the excellent reviews [20, 21], an econometric study in [12], and a discussion from an investor's perspective in [40]. Although the hypothesis is the subject of controversy, most evidence suggests that the opportunities to consistently beat the market are limited and that luck plays a large part.

Although historical data is of limited use to predict changes in risk factors, it can be used to guide the choice of stochastic process used to model the dynamics of risk factors and to estimate the parameters of such models. For statistical analysis, the price for risk factors must be observable and a sufficient amount of data must be available for processing. We assume that the risk factors are basic securities. That is, their value is not a function of other risk factors, i.e., they are not derivatives. In our examples, we use stocks, although our approach applies to a wide range of financial instruments. In this chapter, most material is a review of important results used in the thesis. We will emphasize procedure and highlight the new extensions, such as the use of non-parametric Parzen distributions.

2.1 Discrete random walk models

In this section, we review the standard discrete random walk model for risk factors. The relative return on a risk factor, from time 0 to t, is defined as

$$R_t = \frac{S_t - S_0}{S_0}$$

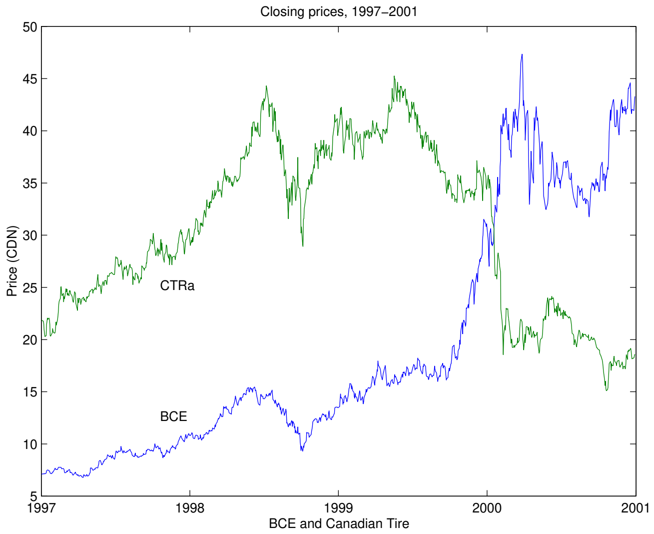


Figure 2.1: Daily closing prices for BCE and Canadian Tire from January 1997 to December 2000.

where S_t is the price for the risk factor.[1] Because the return is dimensionless, the quantity does not have a unit, return models are preferred over models for prices. We consider a model in which the returns, sampled at equally spaced points in time, form a sequence $\{R_i\}_{i=1}^{\infty}$ of independent and identically distributed random variables. This means that stock price are discrete time Markov chains with an infinite state space [52]. Choosing different distributions gives different models in this family.

By examining time series of stock prices, we get an indication about how the distribution should be chosen for a realistic model. Figure 2.1 shows the daily closing prices over 4 years for two Canadian stocks traded on the Toronto Stock Exchange (TSX), Bell Canada Enterprises (BCE) and Canadian Tire (CTRa). The scatter plot in Figure 2.2 shows that the daily returns form a cloud of samples around the origin in what resembles a multivariate unimodal distribution. The time series can be divided into pieces with the same time span as the returns in the model $\{R_i\}_{i=1}^{\infty}$. For each piece, the relative return can be computed as

$$r_i = \frac{s_i - s_{i-1}}{s_{i-1}}, \qquad i = 1, \ldots, d,$$

where s_{i-1} and s_i are the prices at the beginning and end of the time interval. Since the returns $\{R_i\}_{i=1}^{\infty}$ in the model are independent and identically distributed, the computed returns r_i are viewed, rightly or wrongly, as independent samples from the same distribution. After settling

[1] An alternative definition is to use the continuously compounded log-return, $\tilde{R}_t = \log S_t - \log S_0$, which has the advantage that returns can be aggregated over time by addition. Of course the two are closely related. The difference

$$\tilde{R}_t - R_t = \frac{1}{2}\left(\frac{\Delta S_t}{S_0}\right)^2 + O\left(\left[\frac{\Delta S_t}{S_0}\right]^3\right)$$

is small for small relative changes. Either type of return can be approximated by the other. By choosing daily returns we avoid this approximation in the later sections which simplifies the presentation. In our examples, we use daily relative returns.

In the multivariate case, S_t is a vector of prices and the division (or logarithm) is taken componentwise.

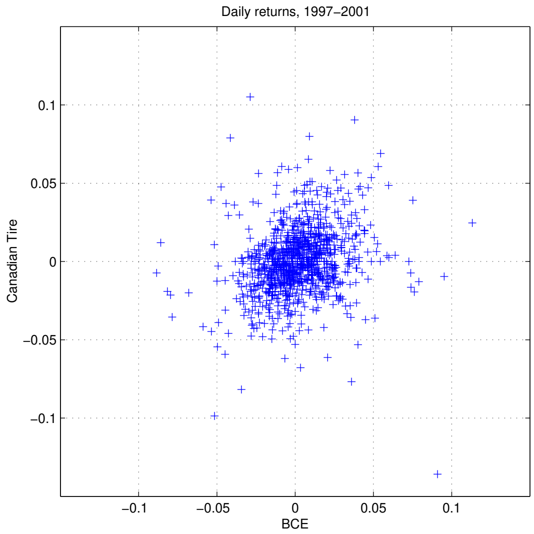


Figure 2.2: Scatter plot of relative returns for BCE and Canadian Tire.

on a family of distributions for the random walk, the parameters of this distribution can be estimated from the time series of returns $\{r_i\}_{i=1}^{d}$.

Many generalizations of the random walk model have been proposed to correct shortcomings revealed in empirical studies (the monograph [12] contains an excellent empirical review). This simple model assumes that the returns $\{R_i\}_{i=1}^{\infty}$ are independent and identically distributed. Firstly, for time periods spanning more than a few years, the returns are not identically distributed. Nevertheless, for shorter time periods it is a reasonable assumption. For this reason, it is standard practice to use only recent data or to use a weighting scheme when estimating the parameters in the model. Secondly, there is empirical evidence that suggests that some risk factors have significant dependence structure over time. Unlike the first one, this problem can not be solved without introducing more complex models, such as ARCH or GARCH models [18, 9, 44, 33]. If we want to use the model for simulating a sequence of returns this property might become important. However, in our application, simulating value-at-risk, we are interested in a one-step simulation and it is arguably less significant for the final result. Despite its shortcomings, the random walk model is a good model that accurately describes many of the real world behaviors of financial markets. In Chapter 6, we return to study perturbations to the return model.

2.2 Three instances of the random walk model

Recall that, in the random walk model, returns are modeled as a sequence $\{R_i\}_{i=1}^{\infty}$ of independent and identically distributed random variables. In this section, we discuss three different instances of this model, three different alternatives for the distribution of the random variables. The first distribution, the normal random walk, is a standard model which has been studied extensively. We believe that the other two distributions, the asymmetric Student's T distribution and the non-parametric density estimator due to Parzen [46], are new in this context. The methods will be generalized to the multivariate case in the next section.

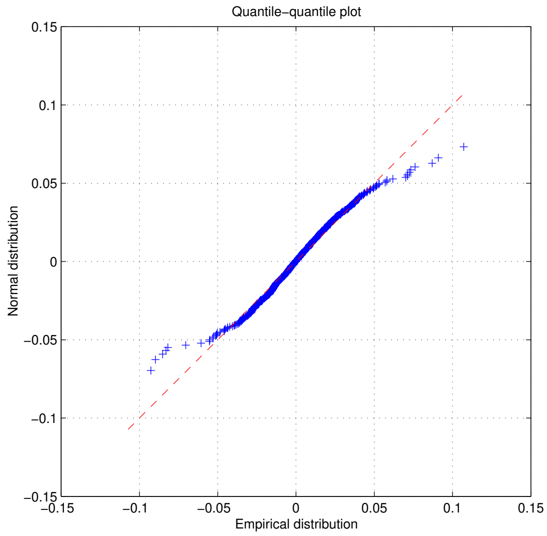


Figure 2.3: Quantile-quantile plot for the normal random walk with parameters estimated from 4 years of daily returns for BCE.

2.2.1 The normal model

In the normal random walk model,[2] the random variables in the sequence of returns are normal,

$$R_i \sim N(\mu, \sigma^2), \qquad i = 1, 2, \ldots.$$

The probability density function (pdf) for all R_i is

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Given a time series of returns $\{r_i\}_{i=1}^{d}$, μ and σ^2 can be estimated with standard statistical methods [37]: the mean can be estimated by

$$\widehat{\mu} = \frac{1}{d} \sum_{i=1}^{d} r_i,$$

and the variance by

$$\widehat{\sigma}^2 = \frac{1}{d-1} \sum_{i=1}^{d} (r_i - \widehat{\mu})^2.$$

Some authors advocate using estimators that give more weight to recent returns than to old ones (see for example [43, 33]).

To illustrate the performance, we estimate the parameters $\widehat{\mu}$ and $\widehat{\sigma}^2$ for daily returns for BCE. Figure 2.3 shows the quantile-quantile plot[3] for the fitted normal distribution. It is clear that the normal model is a good approximation for small returns, but for both the negative

[2] In the case of log-returns, this model is often called the log-normal model, stock prices are log-normally distributed. We caution the reader that some authors use the terminology "normal model" for a model where stock price changes are normal.

[3] A quantile-quantile plot is a method for comparing two distributions. Given a set of observations, we use it

and positive tails the distribution does not fit the data. Fat-tails are typical for stock returns, and to estimate value-at-risk, where we need to compute tail quantiles, the normal model is less suitable. The next two subsections explore different approaches to construct random walk models with more realistic tails.

2.2.2 The asymmetric Student's T model

Student's T distribution has fat-tails. The density for a T distributed random variable is

$$p_T(x;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

the mean is $\mu = 0$, and the variance is

$$\sigma^2 = \frac{\nu}{\nu-2}.$$

The degrees of freedom ν controls the fatness of the tails; as $\nu \to \infty$ the distribution converges to the normal distribution.

An alternative to the normal model is therefore to define a random walk with T distributed random variables. Since the fatness of the tails can be different for negative and positive returns, we generalize this idea and let each random variable in the sequence $\{R_i\}_{i=1}^{\infty}$ be distributed as

$$A = m + \sigma\sqrt{(1-\rho)\left(\frac{\nu_+-2}{\nu_+}\right)}B|T_+| + \sigma\sqrt{\rho\left(\frac{\nu_--2}{\nu_-}\right)}(B-1)|T_-|. \tag{2.1}$$

The random variables T_+ and T_- are T distributed with degrees of freedom ν_+ and ν_-, respectively The random variable B is a Bernoulli random variable; B takes the values 0 or 1 with probability 0.5. The random variables T_-, T_+ and B are independent. We say that A is an asymmetric Student's T distributed random variable. The density, figuratively a density made up of a Student's T pdf cut in half, is

$$p(x) = \begin{cases} \dfrac{p_T\left(\sqrt{\frac{\nu_-}{\nu_--2}}\frac{(x-m)}{\sigma\sqrt{\rho}};\nu_-\right)\sqrt{\frac{\nu_-}{\nu_--2}}}{\sigma\sqrt{\rho}} & \text{if } x \le m, \\ \dfrac{p_T\left(\sqrt{\frac{\nu_+}{\nu_+-2}}\frac{(x-m)}{\sigma\sqrt{1-\rho}};\nu_+\right)\sqrt{\frac{\nu_+}{\nu_+-2}}}{\sigma\sqrt{1-\rho}} & \text{if } x > m. \end{cases}$$

Since the two regions make up half of the density each, m is the median of the distribution, and, with a little algebra, it is easy to derive moment properties relative to the median.

Proposition 2.1. *Suppose that $\nu_- > 4$ and $\nu_+ > 4$. Then, an asymmetric Student's T distributed random variable, defined by* (2.1)*, satisfies the following moment properties:*

(i). The expectation is

$$\eta^{(1)} = \mathbb{E}[A-m] = \sigma\left[\frac{\Gamma\left(\frac{\nu_++1}{2}\right)\sqrt{(1-\rho)(\nu_+-2)}}{\Gamma\left(\frac{\nu_+}{2}\right)\sqrt{\pi}(\nu_+-1)} - \frac{\Gamma\left(\frac{\nu_-+1}{2}\right)\sqrt{\rho(\nu_--2)}}{\Gamma\left(\frac{\nu_-}{2}\right)\sqrt{\pi}(\nu_--1)}\right].$$

to compare the empirical distribution and a distribution fitted to this data. Sorting the observations gives the empirical cumulative distribution functions (cdf). Each observation, which corresponds to a quantile, and the corresponding quantile for the fitted distribution are marked in the plot. If the two distributions are the same, the points fall on the diagonal reference line.

(ii). The second moment is

$$\eta^{(2)} = \mathbb{E}[(A-m)^2] = \sigma^2.$$

(iii). The second conditional moments are, for negative values,

$$\eta_-^{(2)} = \mathbb{E}[(A-m)^2 \mid A \le m] = 2\sigma^2\rho$$

and, for positive values,

$$\eta_+^{(2)} = \mathbb{E}[(A-m)^2 \mid A > m] = 2\sigma^2(1-\rho).$$

(iv). The fourth conditional moments are, for negative values,

$$\eta_-^{(4)} = \mathbb{E}[(A-m)^4 \mid A \le m] = 2\sigma^4\rho^2\left[3 + \frac{6}{\nu_- - 4}\right]$$

and, for positive values,

$$\eta_+^{(4)} = \mathbb{E}[(A-m)^4 \mid A > m] = 2\sigma^4(1-\rho)^2\left[3 + \frac{6}{\nu_+ - 4}\right].$$

Once the moment properties are known, estimating the parameters in the model is straight forward. The first step is to compute the median $\widehat{m}$ of the observed returns $\{r_i\}_{i=1}^d$, by sorting the samples and taking $\widehat{m}$ to be the order k value, if $d = 2k+1$ is odd, or the average of the order k and $k+1$ values, if $d = 2k$ is even. Then, find the sample estimate for the second moment

$$\widehat{\sigma}^2 = \frac{1}{d-1}\sum_{i=1}^{d}(r_i - \widehat{m})^2.$$

We then estimate the contribution to the second moment $\widehat{\rho}$ from the negative and the positive halves. Let $d = d_- + d_+$ where d_- and d_+ are the number of observations less than and greater than $\widehat{m}$. Then,

$$\widehat{\rho} = \frac{1}{2\widehat{\sigma}^2 d_-}\sum_{r_i \le \widehat{m}}(r_i - \widehat{m})^2.$$

Finally, using the sample estimates for the fourth moments,

$$\widehat{\eta}_-^{(4)} = \frac{2}{d_-}\sum_{r_i \le \widehat{m}}(r_i - \widehat{m})^4 \qquad \text{and} \qquad \widehat{\eta}_+^{(4)} = \frac{2}{d_+}\sum_{r_i > \widehat{m}}(r_i - \widehat{m})^4,$$

we can solve for estimates of the degrees of freedom ν_+ and ν_-,

$$\widehat{\nu}_- = \frac{6}{\frac{\widehat{\eta}_-^{(4)}}{2\widehat{\sigma}^4\widehat{\rho}^2} - 3} + 4 \qquad \text{and} \qquad \widehat{\nu}_+ = \frac{6}{\frac{\widehat{\eta}_+^{(4)}}{2\widehat{\sigma}^4(1-\widehat{\rho})^2} - 3} + 4.$$

The advantage of the asymmetric T model over the normal model is that, as illustrated by the quantile-quantile plot in Figure 2.4, the tails of the empirical distribution can be reproduced more accurately. However, this improvement comes at a price since the pdf has a discontinuity at the center. The jump is counter intuitive and the implementation of this model is more difficult, but in comparison to the advantage of increased accuracy these are minor concerns.

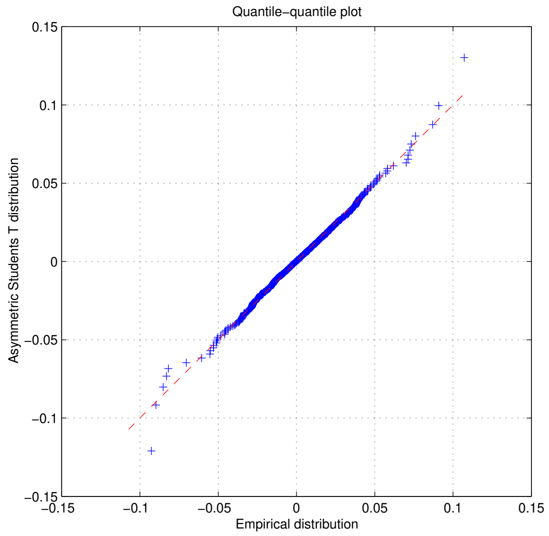


Figure 2.4: BCE quantile-quantile plot for the random walk model with the asymmetric T model.

2.2.3 The Parzen model

A non-parametric density estimator is an alternative to using a parametric method, such as either of the first two examples. Let $\{r_i\}_{i=1}^d$ be samples from a distribution with an unknown pdf, $p(x)$. In [46] Parzen develops and analyzes a family of estimates of the form

$$\widehat{p}_d(x) = \frac{1}{dh}\sum_{i=1}^{d} K\left(\frac{x - r_i}{h}\right), \tag{2.2}$$

initially suggested by Rosenblatt in [51]. In our examples, we use the weighting function [58]

$$K(x) = \frac{15}{16}(1 - x^2)^2 \qquad \text{for } |x| \leq 1.$$

Parzen shows that, if $p(x)$ is sufficiently smooth, $\widehat{p}_d(x)$ is asymptotically unbiased and, for an optimal sequence of h-values, the mean square error converges to zero as[4]

$$\mathbb{E}[(\widehat{p}_d(x) - p(x))^2] = O(d^{-\frac{4}{5}}).$$

We refer to a random walk using the Parzen estimate (2.2) for the pdf as the Parzen model.

Similar to the asymmetric T model, the Parzen model can recreate the fat tails more accurately than the normal model, and it also seems to have a slight advantage over the asymmetric

[4]Parzen presents a theory for density estimates of the form (2.2) with general weighting functions $K(x)$. Let $h_d \to 0$ as the number of samples $d \to \infty$. He shows that density estimates of the form (2.2) converges (pointwise in a mean square sense) to a continuous pdf as $d \to \infty$, More precisely, given a sequence of smoothing parameters $\{h_d\}_{d=1}^{\infty}$ with $\lim_{d\to\infty} h_d = 0$ and $\lim_{d\to\infty} dh_d = \infty$. Then,

$$\mathbb{E}[\widehat{p}_d(x) - p(x)]^2 \to 0 \text{ as } d \to \infty.$$

The sequence of smoothing parameters giving optimal rate of convergence depends both on the point x and the pdf $p(x)$, as well as the weighting function $K(x)$. See Parzen [46] for conditions for, examples of and details about general weighting functions.

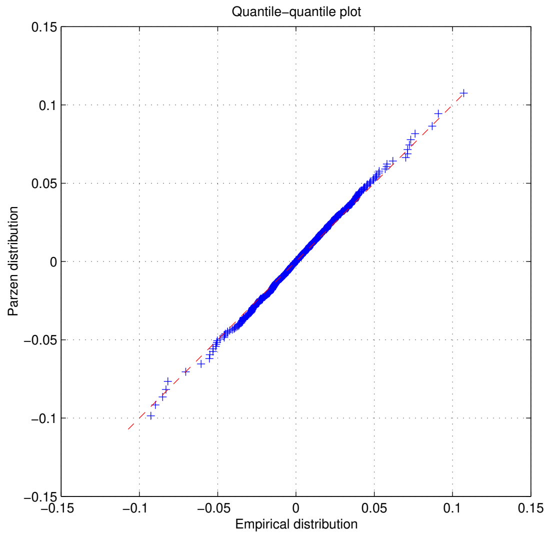


Figure 2.5: BCE quantile-quantile plot for the random walk model with the Parzen density estimate.

T model, as illustrated by the quantile-quantile plot in Figure 2.5. The advantage of using a non-parametric model is that it does not rely on specific assumptions about the shape of the density. There are three disadvantages to the Parzen model. Firstly, the optimal smoothing parameter h is unknown. While experimenting with different stocks, we have found that taking h equal to the standard deviation works well.[5] Secondly, for our choice of weighting function, the density estimate has compact support. However, the support covers the region of interest for value-at-risk calculations so it should have a minor influence on the result. Thirdly, evaluating (2.2), or the corresponding cumulative distribution function (cdf), for different values of x is expensive for large samples. In our implementation, we avoid summing over all sample points by using cubic splines to approximate the cdf and the pdf.

2.3 Multivariate models

So far we have only considered models for the return on a single risk factor. In general, portfolios depend on many risk factors. Therefore we must extend the one-dimensional random walk models, presented in the previous section, to the multivariate case.

In the multivariate random walk, $\{R_i\}_{i=0}^{\infty}$ is a sequence of N-vectors of random variables. The random vectors are independent and identically distributed. The difficulty in constructing a realistic multivariate model is that returns on the risk factors are typically dependent, as exemplified by Figure 2.6. To approximate the dependence structure, without introducing an overly complex model, we restrict our attention to multivariate models where the random vectors $\{R_i\}_{i=1}^{\infty}$ satisfy

$$R_i = A^{-1}X_i + b. \tag{2.3}$$

Moreover, we assume that the random vector X has independent components and the pdf is a

[5]This choice of h may work well in our examples, but it is not a satisfactory solution in general since a fixed smoothing parameter does not give convergence as the number of samples $d \to \infty$. The estimate converges for a sequence of smoothing parameters that decrease to zero as the number of samples increases (see [46] for details).

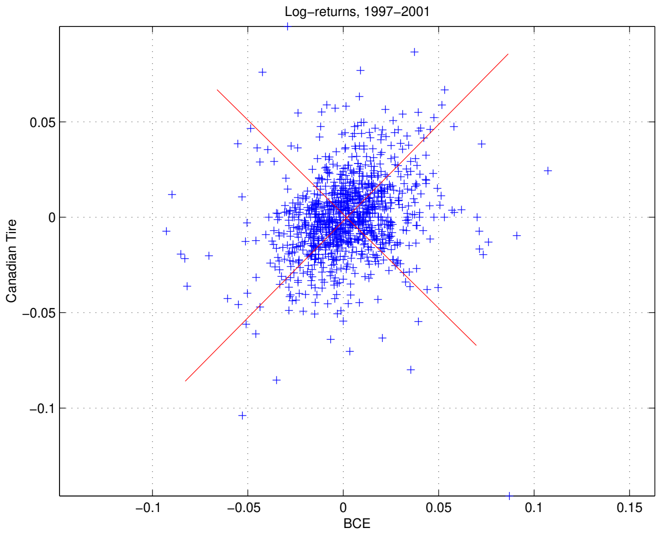


Figure 2.6: Principal components superimposed on the scatter plot for the returns on BCE and CTRa.

product of one-dimensional density functions:

$$p(x) = p_1(x_1) \cdots p_1(x_n).$$

We postpone the discussion about how to choose the linear transformation A to Section 3.4, after discussing pricing of derivatives.

2.4 A brief comment on continuous time models

To find a stochastic process to model stock prices in continuous time is a more difficult problem. Returns are often modeled by stochastic differential equations (SDE). *Brownian motion* is the natural continuous-time generalization of a random walk with normal increments. In this model, the return process is a constant coefficient SDE,

$$dR = \mu dt + \sigma dW,$$

where dW is a Wiener process. Like the normal model for stock prices, geometrical Brownian motion underestimates the likelihood of large returns: it does not have fat tails.

Many different types of continuous-time models have been proposed and studied in the literature, in particular for pricing derivatives. If the returns are a stationary Markov process then the sequence $\{r_i\}_{i=1}^d$ of historical returns can be used to find an estimate for the transition density—the time dependent probability density $p(r, t)$ representing the density for the return r at time t. Figure 2.7 shows the Parzen estimate for the transition density for the stock BCE. A good model must have a transition density that is close to this estimate.

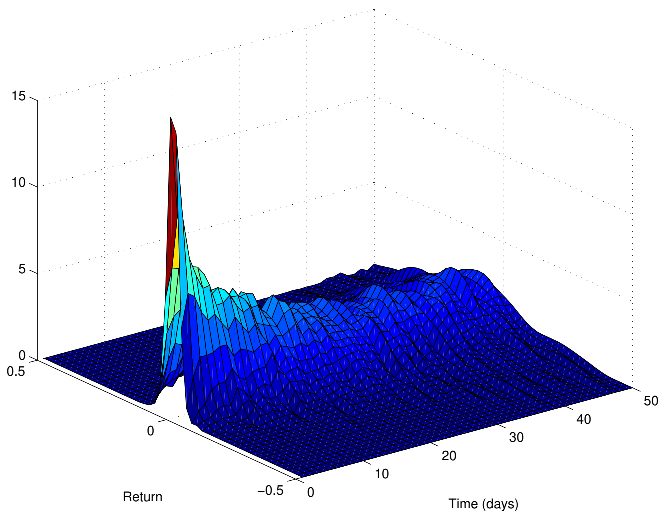


Figure 2.7: Parzen estimate for BCE to the transition density.

Chapter 3

Portfolios and derivative securities

In this chapter, we discuss portfolios, give a brief review of pricing of derivative securities, and introduce a new method for portfolio dependent parameter estimation. The purpose of this chapter is first to connect the idea of a portfolio and the changes in its value to the risk factor models considered in the previous section. The second purpose is to give a detailed presentation of the portfolio dependent estimation procedure that we have developed and to discuss the related computational issues.

3.1 Portfolios

A portfolio is a set of investments. It is represented as a k-vector θ where θ_i is the quantity of the ith security. If we consider a market without derivatives, then the number of securities k and the number of risk factors n are the same, $k = n$. In a market with derivatives, the number of securities is greater than the number of risk factors, $k > n$. In practice, θ_i is an integer, but, in modeling portfolios, it is convenient to let θ_i be a real number. Furthermore, in a market where short selling is allowed, θ_i can be either positive or negative.[1]

Let V_t^i be the price of the ith security at time t. If the security is a direct investment in the risk factor, then V_t^i satisfies the identity

$$V_t^i = S_t^i = S_0^i(1 + R_t^i).$$

If on the other hand it is a derivative security, then V_t^i is a function of the risk factors. Assuming that the portfolio remains unchanged, the dollar value of a portfolio θ at time t is the sum

$$\Pi_t = \sum_{i=1}^{k} \theta_i V_t^i. \tag{3.1}$$

The change in the value from time 0 to time t is

$$\Delta\Pi_t = \Pi_t - \Pi_0 = \sum_{i=1}^{k} \theta_i (V_t^i - V_0^i).$$

[1] An investor is said to be *long* the asset if she has purchased the asset, and, vice versa, she is said to be *short*, if she has borrowed the asset.

3.2 Arbitrage-free pricing of derivatives

How to determine the price for a derivative security is a thoroughly studied problem in finance. Arbitrage-free pricing provides an answer to this question. It is based on the assumption that in an ideal market there can be no arbitrage opportunities. An arbitrage opportunity is a trading strategy that, for no initial cost and without any risk of losing money, yields a positive expected pay-off in the future. The justification for this assumption is that such opportunities are very attractive to all investors and therefore must vanish rapidly as all investors strive to exploit them. In this section, we review the fundamental theorem of finance, in its continuous time version; and we discuss how this theory applies to the models we introduce for value-at-risk.

3.2.1 The fundamental theorem of finance

The fundamental theorem was first introduced for discrete time models for securities markets by Harrison and Kreps [30], and later it was extended to a market with continuous trading by Harrison and Pliska [31]. We give a brief review of the result for continuous time and refer to [13, 35, 17, 45] for the details.[2] It should be noted that the assumptions can be relaxed to prove more general forms of this theorem, see for example the introductory article [48] and the references therein.

Consider a market with asset prices given by the SDE

$$dS_t^i = \mu_t^i S_t^i dt + S_t^i \sum_{j=1}^{m} \sigma_t^{ij} dW_t^j, \qquad \text{for } i = 1, \dots, n \tag{3.2}$$

where dW is an n-dimensional standard Brownian motion with respect to a probability measure P. In addition to the risky assets, there is a risk free asset which can be thought of as a money market account that pays an instantaneous interest on the investment. The value of the risk free asset is given by a differential equation

$$dR_t = r_t R_t dt \quad \text{with} \quad R_0 = 1. \tag{3.3}$$

In general, the expected returns μ_t^i, the volatilities σ_t^{ij} and the interest rate r_t are stochastic processes.

We define a portfolio to be a process $\theta_t = (\theta_t^0, \theta_t^1, \dots, \theta_t^n)$. The first component θ_t^0 is the quantity invested in the risk free asset at time t, and the remaining components θ_t^i are the quantities invested in the risky assets S_t^i with $i = 1, \dots, n$. The value of a portfolio is

$$V_t = \theta_t^0 R_t + \sum_{i=1}^{n} \theta_t^i S_t^i.$$

It is said to be *self-financing* if

$$0 = R_t d\theta_t^0 + \sum_{i=1}^{n} S_t^i d\theta_t^i.$$

This condition simply means that a change in a position must be offset by selling or buying other securities. Or equivalently, in terms of the differential for the value,

$$dV_t = \theta_t^0 dR_t + \sum_{i=1}^{n} \theta_t^i dS_t^i,$$

[2]Below, we have suppressed several technical assumptions about the growth and measurability of the stochastic processes involved.

meaning that any change in the value of the portfolio is an effect of changes in the prices alone, i.e, there are no deposits or withdrawals from the portfolio. As noted above, an arbitrage opportunity is defined as a self-financing portfolio[3] $\theta(t)$, with no initial cost $V_0 = 0$, that has a non-negative pay-off, $V_T \geq 0$ almost surely, and the probability of a positive pay-off is non-zero, $P[V_T > 0] > 0$.

Define the discounted stock price $\overline{S}_t^i = S_t^i / R_t$ and note that

$$d\overline{S}_t^i = \left(\mu_t^i - r_t\right) \overline{S}_t^i dt + \overline{S}_t^i \sum_{j=1}^{m} \sigma_t^{ij} dW_t^j.$$

follows from (3.2) and (3.3). Equipped with these definitions, we now state the fundamental theorem of finance. If there is a process q_t, the market price of risk, that satisfies

$$\begin{bmatrix} \sigma_t^{11} & \cdots & \sigma_t^{1m} \\ \vdots & \ddots & \vdots \\ \sigma_t^{n1} & \cdots & \sigma_t^{nm} \end{bmatrix} \begin{bmatrix} q_t^1 \\ \vdots \\ q_t^m \end{bmatrix} = \begin{bmatrix} \mu_t^1 - r_t \\ \vdots \\ \mu_t^n - r_t \end{bmatrix}, \tag{3.4}$$

then the market has no arbitrage. In addition, it can be shown that the process

$$d\widetilde{W}_t^i = q_t^i dt + dW_t^i$$

is an n-dimensional standard Brownian motion with respect to a measure[4] $Q \sim P$. The measure Q is called the *risk neutral measure*. A consequence is that the discounted stock prices $\overline{S}_t^i$ is a martingale under Q. In this situation, the discounted value $\overline{V}_t = V_t / R_t$ of a self-financing portfolio satisfies

$$d\overline{V}_t = \sum_{i=1}^{n} \theta_t^i \overline{S}_t^i \sum_{j=1}^{m} \sigma_t^{ij} d\widetilde{W}_t^i.$$

Therefore, it too is a martingale under the risk neutral measure Q.

Consider a European derivative that pays $f(S_T)$ at the time of maturity T. Suppose that there is a self-financing portfolio θ_t that *replicates* the derivative, they have the same value at maturity,

$$V_T = f(S_T).$$

Now, since both the portfolio and the derivative gives the holder the same pay-off at maturity, the value of the derivative must be the same as the value of the portfolio. Using the martingale property it follows that their value today is

$$V_0 = \overline{V}_0 = \mathbb{E}^Q \left[\overline{V}_T\right] = \mathbb{E}^Q \left[\frac{f(S_T)}{R_t}\right]. \tag{3.5}$$

Thereby, the fundamental theorem of finance provides a method for pricing derivatives in complete markets.

[3]For technical reasons, the set of portfolios must be restricted to the set of *admissible* portfolios— portfolios for which V_t is bounded from below except possibly on a set of measure zero.

[4]If the linear system (3.4) is nonsingular, the market price of risk is uniquely defined and the risk neutral measure Q is unique. For a singular system, there may not exist a solution or there may be an infinite number of solutions. If there is no solution, then the market contains arbitrage opportunities. If there is an infinite number of solutions, then Q is not uniquely defined and the market is said to be incomplete.

3.2.2 The Black-Scholes model

The Black-Scholes model for pricing options predates the fundamental theorem [8]. It corresponds to the case where all parameters are in (3.2) and (3.3) constant. The Black-Scholes model has been tremendously successful since its introduction, but, at the same time, most researchers agree that it does not capture several important properties of prices observed in the markets. Much of current research into pricing models tries to remedy these problems by including features such as stochastic volatility and jump processes. Such features lead to more complex pricing algorithms and we refer to [33, 60, 59] for details.

In our examples, we use portfolios with European call and put options.[5] A European call option gives the holder the right to buy the underlying security from the writer at a predetermined price K at a future time T. A European put option gives the holder the right to sell the underlying security at the price K at time T. The pay-off for a call option is[6]

$$f_{call}(S_T) = (S_T - K)^+,$$

and for a put option it is

$$f_{put}(S_T) = (K - S_T)^+.$$

In the special case of the Black-Scholes model, it is easy to derive a closed form expression for the price from (3.5). The price of a call option is[7]

$$V_{call} = S_0 N(d_1) - Ke^{-rT} N(d_2),$$

and the price of a put option is

$$V_{put} = Ke^{-rT} N(-d_2) - S_0 N(-d_1),$$

where $N(\cdot)$ is the cdf for the normal distribution, and

$$\begin{aligned} d_1 &= \frac{\log\left(\frac{S_0}{K}\right) + \left(r + \frac{\sigma^2}{2}\right) T}{\sigma\sqrt{T}}, \\ d_2 &= d_1 - \sigma\sqrt{T}. \end{aligned}$$

[5]A European derivative is a derivative with a fixed exercise date. In contrast, an American derivative gives the holder the right to exercise the derivative at any time of her choosing before maturity.

[6]We use A^+ for the positive part of the function, i.e., $A^+ = \max(A, 0)$.

[7]The formula is derived from (3.5) by first noting that

$$\overline{S}_t = S_0 \exp\left(\sigma\widetilde{W}_t - \frac{\sigma^2}{2} t\right)$$

and substituting it into the expression:

$$\begin{aligned} V_{call} = \mathbb{E}^Q\left[\frac{f_{call}(S_T)}{R_t}\right] &= \int_{-\infty}^{\infty} \left(\frac{S_T}{R_T} - \frac{K}{R_T}\right)^+ dQ \\ = \frac{1}{\sqrt{2\pi T}} \int_{-\left(\log\left(\frac{S_0}{K}\right) + \left(r - \frac{\sigma^2}{2}\right)T\right)/\sigma}^{\infty} & S_0 \exp\left[-\frac{(w - \sigma wT)^2}{2T}\right] - K\exp\left[-rT - \frac{w^2}{2T}\right] dw. \end{aligned}$$

By splitting the integral into two parts and changing the variables, it follows that

$$V_{call} = S_0 \frac{1}{\sqrt{2\pi}} \int_{-\left(\log\left(\frac{S_0}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T\right)/\sigma\sqrt{T}}^{\infty} e^{-\frac{y^2}{2}} dy - Ke^{-rT} \frac{1}{\sqrt{2\pi}} \int_{-\left(\log\left(\frac{S_0}{K}\right) + \left(r - \frac{\sigma^2}{2}\right)T\right)/\sigma\sqrt{T}}^{\infty} e^{-\frac{y^2}{2}} dy.$$

The derivation of the formula for a put option is similar.

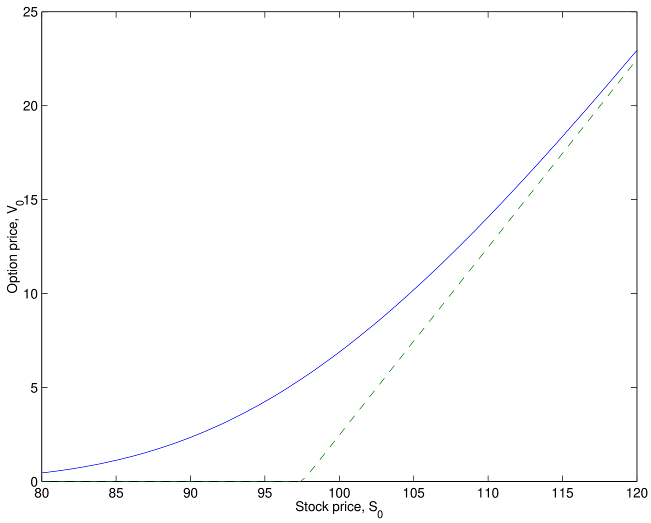


Figure 3.1: The solid line is the price of a European call option in the Black-Scholes model. The dashed line is the function $(S_0 - Ke^{-rT})^+$, the value of a European option as $\sigma \to 0$.

Figure 3.1 shows an example with the price of a European call option as a function of the current price of the asset. In this case, it is an option with strike price $K = 100$, and 6 months to maturity, $T = 0.5$, an interest rate of 5% and volatility $\sigma = 0.2$.

3.2.3 How pricing models are used in market risk models

If the Black-Scholes model were correct, the drift would be the only difference between the stock price processes for the probability spaces with the risk neutral and the real world measures. However, this approach does not reproduce observable prices for traded options. Therefore, instead of using a volatility estimated from stock price data, pricing models typically use parameters implied by option prices, i.e., the pricing model is used as a form of interpolation scheme. In value-at-risk simulation, we are interested in changes in value over a short time period and what is needed is a model that captures the local dynamics of the value. Hence, compared to the models used for derivatives trading where the whole lifespan of the contract must be considered, the quality of the pricing model is less critical. In our examples we use the Black-Scholes model to construct such local approximations, but the ideas could in principle be extended to more complex pricing models.

Two problems that we touch lightly upon in this thesis are volatility risk and mapping of risk factors. Both topics are important in the implementation of market risk models. As mentioned, parameters of option pricing models must be chosen to reproduce prices in the market. Unfortunately, parameters such as the volatility σ in the Black-Scholes models tend to change over time. Therefore, a natural extension is to make volatility stochastic (see for example [59]). This makes potential changes in volatility a source of risk and it can be introduced as a risk factor in a value-at-risk model. In Chapter 4 and Chapter 5, we study a very simple version of such a model and we see that it leads to some interesting qualitative changes to the problem. Mapping of risk factors is the process where some risk factors are replaced by a few general factors. An example is replacing a continuum of interest rates with different maturities by a few representative rates. The dimension reduction problem in Chapter 5 can be viewed as an automatic mapping method.

3.3 Taylor approximation for portfolios

As we have seen the price for a derivative may be a complicated function. Often there is no explicit formula and it must be priced using a separate simulation. Furthermore, the number of different types of securities in a derivatives portfolio will typically be much larger than the number of basic risk factors. In such cases, Taylor's theorem provides a tool to approximate the value (3.1) of a portfolio by a function with a simple mathematical form.

3.3.1 Δ-approximation

Taylor approximations are accurate close to the point of expansion, if the function is sufficiently smooth. Under this assumption and using a first order approximation, we get [8]

$$\Pi_t = \Pi_0 + \left(\sum_{i=1}^{k} \theta_i \frac{\partial V_t^i}{\partial t}\right) t + \sum_{j=1}^{n} \left(\sum_{i=1}^{k} \theta_i \frac{\partial V_t^i}{\partial s_j}\right) S_0^j r_j + O(t^2 + r_1^2 + \cdots + r_n^2).$$

Collecting the coefficients, we get

$$\Pi_t \approx \Pi_0 + \Theta t + \Delta^T r \tag{3.6}$$

where

$$\Theta = \left(\sum_{i=1}^{n} \theta_i \frac{\partial V_t^i}{\partial t}\right) \quad \text{and} \quad \Delta_j = \left(\sum_{i} \theta_i \frac{\partial V_t^i}{\partial s_j}\right) S_0^j.$$

In finance, such a linear approximation is called a Δ-approximation. The Δ is often used to hedge a portfolio. By taking a position, for example by buying the risk factor or future contracts, that offsets the derivative, a portfolio's sensitivity to changes in the underlying can be reduced. As a consequence, in risk models for derivative portfolios a significant component of the risk will be made up of higher-order effects.

Consider the call option from the previous section. Suppose that the current price of the underlying S_0 is equal to the strike price, $K = 100$, i.e, the option is at-the-money. The

[8] Taylor's theorem gives an explicit formula for the error. For the first order approximation the error is

$$\text{error} = \frac{1}{2!}\left(\sum_i \theta_i \frac{\partial^2 V_t^i}{\partial t^2}\right)(t')^2 + \frac{1}{2!}\left(\sum_{i,j} \theta_i \frac{\partial^2 V_t^i}{\partial t \partial s_j} S_0^j r_j'\right) t' + \frac{1}{2!}\sum_{i,j,k} \theta_i \frac{\partial^2 V_t^i}{\partial s_j \partial s_k} S_0^j S_0^k r_j' r_k'$$

where the derivatives are evaluated at some $t' \in (0, t)$ and $r_j' \in (0, r_j)$ for $j = 1, \ldots, n$.

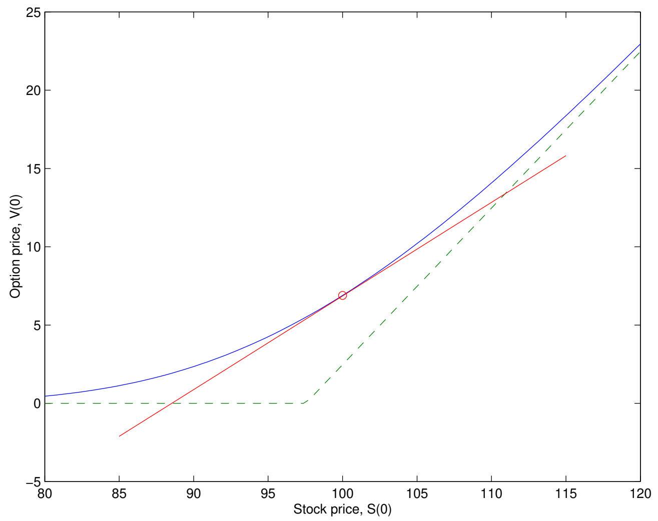


Figure 3.2: The price of a European call option in Black-Scholes model compared to the at-the-money Δ-approximation for the value.

derivatives of the value for an option are[9]

$$\frac{\partial V_t}{\partial s} = N(d_1), \tag{3.7}$$
$$\frac{\partial V_t}{\partial t} = -sN'(d_1)\frac{\sigma}{\sqrt{T-t}} + rKe^{-r(T-t)}N(d_2). \tag{3.8}$$

Hence,

$$\Delta = \frac{\partial V_t}{\partial s} S_0 \approx 59.77 \quad \text{and} \quad \Theta \approx -8.12,$$

and the Δ-approximation of the gain of the call option is

$$\Pi_t \approx \Pi_0 + \Theta t + \Delta r \approx 6.89 - 8.12t + 59.77r \tag{3.9}$$

where $r = (s - S_0)/S_0$. Figure 3.2 shows the Δ-approximation (as a function of the stock price s for one day, $t = 1/250$) compared to the Black-Scholes price. It is accurate for small returns, but quickly deteriorates as $|r|$ increases. Finally, when the return $|r|$ is large enough, the approximation is negative or less than $(s - Ke^{-r(T-t)})^+$, and therefore it violates the basic principle of no arbitrage.

[9] The Δ and Θ are derived by differentiating the Black-Scholes equation. Using $d_2 = d_1 - \sigma\sqrt{T-t}$, it follows that

$$e^{-r(T-t)}N'(d_2) = N'(d_1)e^{d_1\sigma\sqrt{T-t}}e^{-(r+\sigma^2/2)(T-t)} = \frac{s}{K}N'(d_1).$$

Therefore,

$$\frac{\partial V_t}{\partial s} = N(d_1) + sN'(d_1)\frac{\partial d_1}{\partial s} - Ke^{-r(T-t)}N'(d_2)\frac{\partial d_2}{\partial s} = N(d_1)$$

and

$$\begin{aligned}\frac{\partial V_t}{\partial t} &= sN'(d_1(t))\frac{\partial d_1}{\partial t} + rKe^{-r(T-t)}N(d_2) - Ke^{-r(T-t)}N'(d_2)\frac{\partial d_2}{\partial t}\\ &= -sN'(d_1(t))\frac{\sigma}{2\sqrt{T-t}} + rKe^{-r(T-t)}N(d_2).\end{aligned}$$

Wilmott's book [59] has a list of derivatives, or "greeks", for calls, puts and other common derivative securities.

3.3.2 $\Delta\Gamma$-approximation

For non-linear portfolios, the Δ-approximation is useful when the time period considered is relatively small. But, as Figure 3.2 illustrates, ignoring the curvature for an option portfolio may lead to a large truncation error. This feature becomes particularly important if Δ-hedging is used, since the linear term is hedged out leaving a higher order residual for the portfolio value.

The non-linearity can be better approximated by including more terms from the Taylor series. Keeping the first two terms in r gives the quadratic approximation[10]

$$\begin{aligned}\Pi_t = \Pi_0 &+ \left(\sum_{i=1}^{k} \theta_i \frac{\partial V_t^i}{\partial t}\right) t + \sum_{j=1}^{n} \left(\sum_{i=1}^{k} \theta_i \frac{\partial V_t^i}{\partial s_j}\right) S_0^j r_j \\ &+ \frac{1}{2!} \sum_{l=1}^{n} \sum_{j=1}^{n} \left(\sum_{i=1}^{k} \theta_i \frac{\partial^2 V_t^i}{\partial s_j \partial s_l}\right) S_0^j S_0^l r_j r_l + O(t^2 + t \textstyle\sum_i r_i + \sum_i r_i^3)\end{aligned}$$

or

$$\Pi_t \approx \Pi_0 + \Theta t + \Delta^T r + \frac{1}{2} r^T \Gamma r. \tag{3.10}$$

The vector Δ and the scalar Θ are as in (3.6), and

$$\Gamma_{jl} = \left(\sum_{i=1}^{k} \theta_i \frac{\partial^2 V_t^i}{\partial s_j \partial s_l}\right) S_0^j S_0^l, \qquad j, l = 1, 2, \ldots n.$$

In finance, the quadratic approximation (3.10) is called a $\Delta\Gamma$-approximation. Because it approximates the curvature of the value function, it is a more accurate local approximation to the portfolio value than the linear Δ-approximation.

We return to the example with the Δ-approximation (3.9) for a call option. In this case, the second derivative for the value is

$$\frac{\partial^2 V_t}{\partial s^2} = \frac{N'(d_1)}{s\sigma\sqrt{T-t}}.$$

Then,

$$\Gamma = \frac{\partial^2 V_t}{\partial s^2} S_0^2 \approx 273.6$$

which gives the $\Delta\Gamma$-approximation

$$\Pi_t \approx \Pi_0 + \Theta t + \Delta r + \Gamma \frac{r^2}{2} \approx 6.89 - 8.12t + 59.77r + 273.6 \frac{r^2}{2} \tag{3.11}$$

where $r = (s - S_0)/S_0$. The approximation compared to the exact values is shown in Figure 3.3. In comparison to the Δ-approximation, the $\Delta\Gamma$-approximation is much closer to the price in the Black-Scholes model.

[10] The truncation error is

$$\begin{aligned}\text{error} = &\frac{1}{2!}\left(\sum_i \theta_i \frac{\partial^2 V_t^i}{\partial t^2}\right) t^2 + \frac{1}{2!}\left(\sum_{i,j} \theta_i \frac{\partial^2 V_t^i}{\partial t \partial s_j} S_0^j r_j'\right) t' \\ &+ \frac{1}{3!}\sum_{i,j,k}\left(\theta_i \frac{\partial^3 V_t^i}{\partial t \partial s_j \partial s_k} S_0^j S_0^k r_j' r_k'\right) t' + \frac{1}{3!}\sum_{i,j,k,l} \theta_i \frac{\partial^3 V_t^i}{\partial s_j \partial s_k \partial s_l} S_0^j S_0^k S_0^l r_j' r_k' r_l'\end{aligned}$$

for some $t' \in (0, t)$ and $r_j' \in (0, r_j)$ for $j = 1, \ldots, n$.

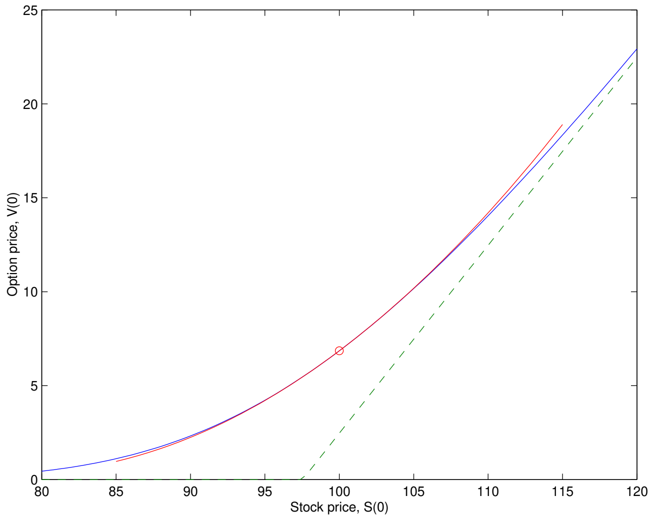


Figure 3.3: The price of a European call option compared to the at-the-money $\Delta\Gamma$-approximation.

3.4 Estimation of parameters and factorization of $\Delta\Gamma$-approximations

In the algorithm that we develop in Chapter 4, we assume that derivatives portfolios are represented as $\Delta\Gamma$-approximations (3.10). The change in value, over the time period from time 0 to time t, is then

$$\Delta\Pi_t = \Pi_t - \Pi_0 \quad \approx \quad \Delta\widetilde{\Pi}_t = \Theta t + \Delta^T r + \frac{1}{2} r^T \Gamma r. \tag{3.12}$$

The return R on the risk factors is a vector of random variables, and, when $\Delta\Pi_t$ is viewed as a function of R, it is a random variable modeling the change in the portfolio's value over the time period. We assume that the stochastic model for returns R is of the type presented in Chapter 2.

The parameters of the risk factor model are estimated, independently from the pricing model, from a time series of historical returns. This choice comes at the cost of ignoring any connections, suggested by the theory of arbitrage-free pricing, about the relation of the real world and risk neutral processes. However, it has the important advantage of more flexibility in choosing the underlying model for price changes. Furthermore, as illustrated by the Black-Scholes model, the conclusion that risk neutral and real world processes have the same variance, properties implied by the pricing model, may not hold up to scrutiny. For these reasons, we believe that this pragmatic approach is justified.

3.4.1 Portfolio decomposition and portfolio dependent estimation

In this section, we present a new method for portfolio dependent estimation of parameters for the risk factor models. The strategy for parameter estimation builds on the observation that a $\Delta\Gamma$-approximation can be decomposed as a sum of one-dimensional quadratic functions. This decomposition has both a long history in both applied mathematics and statistics, and it has been used by other authors in the computation of value-at-risk. Still, the multivariate risk factor model we present is fundamentally different in that the resulting risk factor models are portfolio dependent.

Suppose that $\{r_i\}_{i=1}^d$ is a time series of returns, where r_i is an n-vector. We know, as discussed in Chapter 2, that the mean $\widehat{\mu}$ can be estimated with standard sample estimators (in the case of the asymmetric T model, the mean is replaced by the median $\widehat{m}$). Similarly, the standard sample statistics

$$\widehat{C}_{ij} = \frac{1}{d-1}\sum_{k=1}^{d}(r_{ik} - \widehat{\mu}_i)(r_{jk} - \widehat{\mu}_j), \tag{3.13}$$

can be used to estimate the elements $C_{ij} = \mathrm{cov}(R_i, R_j)$ of the covariance matrix C. For the normal model, $\widehat{\mu}$ and $\widehat{C}$ characterize the model completely. For the other two models, we explain below how to estimate the remaining parameters.

Recall that, as discussed in Section 2.3, we want to approximate the dependence structure of the risk factor returns with a product pdf. The first step to construct such a model is to observe that a $\Delta\Gamma$-approximation can be factored by solving the generalized eigenvalue problem

$$\begin{cases} \Lambda = A^T \Gamma A, \\ \widehat{C} = AA^T. \end{cases} \tag{3.14}$$

The matrix A is nonsingular and

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n).$$

Let

$$X = A^{-1}(R - \widehat{\mu})$$

define a new vector of random variables. Then,

$$\mathbb{E}[X] = 0 \quad \text{and} \quad \mathbb{E}[XX^T] = I. \tag{3.15}$$

Also, we substitute $r = Ax + \widehat{\mu}$ to express $\Delta\widetilde{\Pi}_t$ in terms of our new variables:

$$\Delta\widetilde{\Pi}_t = \left(\Theta t + \widehat{\mu}^T\Delta + \frac{1}{2}\widehat{\mu}^T\Gamma\widehat{\mu}\right) + x^T A^T(\Delta + \Gamma\widehat{\mu}) + \frac{1}{2}x^T\Lambda x.$$

Define

$$\Xi = \Theta t + \widehat{\mu}^T\Delta + \frac{1}{2}\widehat{\mu}^T\Gamma\widehat{\mu}, \tag{3.16}$$

$$\Delta' = A^T(\Delta + \Gamma\widehat{\mu}). \tag{3.17}$$

The $\Delta\Gamma$-approximation to the change in portfolio value is

$$\Delta\widetilde{\Pi}_t = \Xi + \sum_{i=1}^{n}\left(\Delta'_i x_i + \frac{\lambda_i}{2}x_i^2\right) \tag{3.18}$$

This is simply a sum of one-dimensional quadratic functions

$$\pi_i = \Delta'_i x_i + \frac{\lambda_i}{2}x_i^2. \tag{3.19}$$

We postulate that the joint distribution of X is a product pdf,

$$p(x) = p_1(x_1) \cdots p_n(x_n), \tag{3.20}$$

or equivalently that that the components of X are independent. As (3.15) shows, the components of X are uncorrelated by construction, so the approximation with a product pdf extrapolates from this property to a more general assumption. Finally, the remaining parameters for the asymmetric T and Parzen models can be estimated for each component X_i individually. The complete parameter estimation and factorization procedure is summarized in Algorithm 1.

3.4.2 Testing independence

The assumption that the risk factor returns can be modeled by a product pdf is central to what follows in Chapter 4. It is therefore natural to question whether it is a valid assumption. We note that returns tend to be scattered around a central point, as in Figure 2.2. Also, the parameter estimation procedure produce uncorrelated random variables for which the sample correlation is zero by construction. These two points form the basis for our belief that a product pdf is a reasonable model.

For our two dimensional example, returns on BCE and CTRa, it is possible to construct a statistical test.[11] To test the assumption, we use a binomial test, as discussed in for example [37]. As we will show, the experiment suggests that, in this case, the independence assumption is valid.[12]

Consider a portfolio with at-the-money call options on BCE and CTRa. Using the Black-Scholes model to price this portfolio, we get a $\Delta\Gamma$-approximation with

$$\Delta = \begin{bmatrix} 25.4932 \\ 10.9513 \end{bmatrix} \quad \text{and} \quad \Gamma = \begin{bmatrix} 69.4300 & 0 \\ 0 & 29.7313 \end{bmatrix}.$$

The portfolio dependent estimation procedure, with the standard sample estimates and the data in Figure 2.2, gives us

$$\begin{aligned} \lambda_1 &= 0.0064, & \Delta_1' &= 0.1371, \\ \lambda_2 &= 0.0169, & \Delta_2' &= 0.6382. \end{aligned}$$

[11] In fact, the possibilities for statistical tests are infinite, see for example Feuereverger [22] and the references therein. If X and Y are independent, then for any functions h and g

$$\mathbb{E}[h(X)g(Y)] = \mathbb{E}[h(X)]\mathbb{E}[g(y)],$$

provided that the integrals exist. This equality is taken as the null hypothesis for a statistical test. Given a set of samples, it is possible to compare sample estimates for the left and right hand sides. By the law of large numbers the two converge for independent random variables. If it is possible to find confidence intervals for the sample estimates, then the estimates can be tested to accept or reject the null hypothesis.

Statistical tests can not prove independence, they can only *reject* independence through rejection of the null hypothesis. Furthermore, although the same idea could in principle be used in high-dimensions, formulating tests with sufficient power and obtaining a set of samples that is large enough makes such tests practically unfeasible.

[12] Whether the independence assumption holds is not the central question. The important question is whether the approximation leads to good simulation results. Our experience using different stocks and portfolios is that for four years of data rejections become more common as the scales of the portfolio components become more different. However, a difference in scales indicates that the influence of one component dominates the dynamics making the independence a secondary issue. For shorter time series, the rejection rate decreases and rejections with two years of data appear to be rare.

Parameter estimation and factorization

Input: A $\Delta\Gamma$-approximation. A time series $\{r_i\}_{i=1}^d$ of daily returns.

Output: Parameter estimates for the market model. A factorization of the $\Delta\Gamma$-approximation.

- Compute estimates for the mean $\widehat{\mu}$ and the covariance matrix $\widehat{C}$.
- Solve the eigenvalue problem

$$\begin{cases} \Lambda = A^T \Gamma A, \\ \widehat{C} = A A^T. \end{cases}$$

- Compute

$$\begin{aligned} \Xi &= \Theta t + \widehat{\mu}^T \Delta + \frac{1}{2} \widehat{\mu}^T \Gamma \widehat{\mu}, \\ \Delta' &= A^T (\Delta + \Gamma \widehat{\mu}). \end{aligned}$$

- For each variable X_i, estimate the remaining parameters, from the transformed returns $\{A^{-1}(r_i - \widehat{\mu})\}_{i=1}^d$, using the methods in Chapter 2.

Algorithm 1: Algorithm describing the steps of the portfolio dependent parameter estimation procedure, and the simultaneous factorization of the $\Delta\Gamma$-approximation.

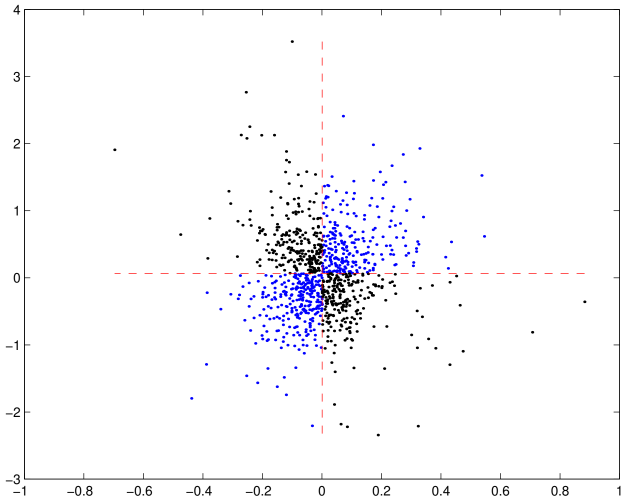


Figure 3.4: Scatter plot for the empirical returns on the two portfolio components π_1 and π_2. The event B corresponds to the bottom-left and top-right quarters of the plane.

We want to test if the two portfolio components π_1 and π_2 are independent. To formulate this question as a binomial test, viewing π_1 and π_2 as random variables, define the events

$$A_1 = \{\pi_1 : \pi_1 > \widehat{\mu}_{\pi_1}\},$$
$$A_2 = \{\pi_2 : \pi_2 > \widehat{\mu}_{\pi_2}\},$$

where $\widehat{\mu}_{\pi_i}$ is the sample mean for π_i. From the time series we estimate the probabilities of the events:

$$\widehat{p}_1 = \frac{\#\text{ samples in } A_1}{\#\text{ samples}} = \frac{476}{1006},$$
$$\widehat{p}_2 = \frac{\#\text{ samples in } A_2}{\#\text{ samples}} = \frac{490}{1006}.$$

Consider the event

$$B = (A_1 \cap A_2) \cup (\overline{A_1} \cap \overline{A_2}),$$

the event that the pair of returns are both either larger or smaller than their estimates for the mean (see Figure 3.4). We then formulate our null hypothesis: the probability $P(B)$ is equal to

$$q = \widehat{p}_1\widehat{p}_2 + (1 - \widehat{p}_1)(1 - \widehat{p}_2).$$

Estimate the probability of B

$$\widehat{q} = \frac{\#\text{ samples in } B}{\#\text{ samples}} = \frac{522}{1006}.$$

Treating each sample as an independent Bernoulli trial and normalizing the random variable of the number of successes leads to

$$\frac{522 - 1006q}{\sqrt{1006q(1-q)}} = 1.15$$

which is within both the confidence intervals for 95%, $[-1.96, 1.96]$, and 90%, $[-1.64, 1.64]$. Therefore, the null hypothesis should be accepted and the portfolio components have passed the independence test.

3.4.3 A few implementation issues

In the generalized eigenvalue problem (3.14), the Hessian Γ is symmetric and the covariance matrix $\widehat{C}$ is non-negative definite. Symmetric-definite eigenvalue problems arise in other applications. The standard method [28] is to compute the Cholesky factorization

$$\widehat{C} = R^T R, \tag{3.21}$$

and then apply the QR algorithm to $R\Gamma R^T$ to compute the Schur form

$$\Lambda = Q^T(R\Gamma R^T)Q$$

which gives $A = R^T Q$. To solve this eigenvalue problem requires $O(n^3)$ floating point operations. Combined with the estimation of the covariance matrix, this gives a total of $O(n^3 + n^2 d)$ floating point operations.

However, if we take advantage of the structure of the problem (3.14), it is possible to improve slightly on this procedure. Let $\{r_i\}_{i=1}^d$ be the time series of returns. Define the $d \times n$ matrix

$$W = \begin{bmatrix} r_1^T - \mu^T \\ \vdots \\ r_d^T - \mu^T \end{bmatrix}.$$

The estimate of the covariance matrix (3.13) can be written as

$$\widehat{C} = \frac{1}{n} W^T W.$$

A variety of estimators for the covariance matrix C have been proposed in the finance literature (see for example [12, 33, 59]). Many of the estimators are of the form

$$\widehat{C} = W^T D W, \tag{3.22}$$

where D is a weight matrix.[13]

Depending on whether the number of dates d is greater or smaller than the number of risk factors n, we get two cases. Suppose that $d > n$. Rather than explicitly forming the covariance matrix $\widehat{C}$, it is preferable to factor it directly by computing the QR factorization

$$D^{\frac{1}{2}} W = QR \tag{3.23}$$

where Q is a $d \times n$ matrix with orthonormal columns and R is a $d \times d$ upper triangular matrix. The QR algorithm (3.23) is then applied to the matrix $R\Gamma R^T$. The QR factorization takes about $2n^2 d$ floating point operations compared to approximately $n^2 d + n^3/3$ for (3.21). The

[13] Two examples are the exponentially weighted moving average with

$$D = \frac{1-\lambda}{1-\lambda^d} \operatorname{diag}(1, \lambda, \ldots, \lambda^{d-1})$$

and the multivariate GARCH(1,1) model [9, 12, 33] $\Sigma_n = (1-\alpha-\beta)V + \alpha r_i r_i^T + \beta \Sigma_{n-1}$, where V is the standard estimate for the long-run average volatility, which has

$$D = \operatorname{diag}(\alpha, \alpha\beta, \alpha\beta^{d-1}) + \frac{1-\beta^d}{1-\beta}(1-\alpha-\beta)I.$$

advantage of the QR algorithm is that it can be shown that the forward error is smaller than for the Cholesky factorization.[14]

On the other hand, if $d < n$ then the matrix $\widehat{C}$ will be singular. The best approach when $d \leq n$ is to use the factorization (3.22) and compute the Schur decomposition

$$\Lambda = Q^T (D^{\frac{1}{2}} W) \Gamma (D^{\frac{1}{2}} W)^T Q.$$

This effectively reduces the size of the eigenvalue problem from order n to order d and therefore the whole step requires only $O(dn^2 + d^2 n + d^3)$ floating point operations. We conclude that performing computation directly on the time series matrix, rather than forming the covariance matrix explicitly, is convenient, effective and numerically sound.

[14] Sun [57, 32] proves an upper bound for the forward error for Cholesky factorization. We adapt Sun's result for our purpose. Suppose that $W^T W$ is positive definite and that ΔW is a sufficiently small perturbation. Consider the Cholesky factorizations $W^T W = R^T R$ and $(W + \Delta W)^T (W + \Delta W) = (R + \Delta R)^T (R + \Delta R)$. Then,

$$\frac{\|\Delta R\|_F}{\|R\|_p} \leq 2^{-1/2} \frac{\kappa_2^2(W)\epsilon}{1 - \kappa_2^2(W)\epsilon}, \qquad \epsilon = 2\frac{\|\Delta W\|_F}{\|W\|_p} + O(\Delta W^2), \quad p = 2, F.$$

This bound is larger than the forward error bound for the QR factorization proved by Stewart [55, 32]. Suppose that W has rank n and that

$$W = QR \qquad \text{and} \qquad W + \Delta W = (Q + \Delta Q)(R + \Delta R)$$

are QR factorizations. Stewart shows that for sufficiently small perturbations

$$\frac{\|\Delta R\|_F}{\|R\|_F} \leq c_n \kappa_F(W) \frac{\|\Delta W\|_F}{\|W\|_F}$$

where c_n is a constant.

Chapter 4

The fast convolution method for value-at-risk

The risk factor model and the pricing model provide the necessary ingredients to make the the equation (1.1) defining value-at-risk meaningful. In this chapter, we present a new Fourier transform algorithm for computing value-at-risk. The method is different from existing Fourier methods (see [42, 15, 27]) in that it does not assume that the characteristic function is explicitly known. Our method has the advantage of greater freedom in choosing the risk factor model. We present an extended example illustrating the performance of the algorithm and the importance of risk factor models with fat-tails. In the second part of the chapter, we extend the method to compute the gradient of value-at-risk. The chapter concludes with two computational examples. The first is a simple linear approximation to the change in value-at-risk with changes in portfolio composition. In the second example we hedge a derivatives portfolio by solving an optimization problem to minimize the value-at-risk.

4.1 Distribution of profits and losses

The local dynamics of changes in portfolio value is approximated by

$$\Delta\widetilde{\Pi} = \Theta t + \Delta^T r + \frac{1}{2} r^T \Gamma r.$$

Therefore, we may approximate the value-at-risk by the solution to

$$\mathrm{P}[\Delta\widetilde{\Pi} \leq -\mathrm{VaR}] = 1 - \alpha. \tag{4.1}$$

Because $\Delta\widetilde{\Pi}$ is a quadratic function, it is easier to solve (4.1) than (1.1). At the same time, since $\Delta\widetilde{\Pi}$ is locally accurate, the solution VaR to (4.1) is a good approximation to value-at-risk, provided that the probability of large changes in the risk factors is relatively small. We return to study accuracy in Chapter 6.

Equation (4.1) is solved in two steps. We find the pdf of $\Delta\widetilde{\Pi}$ and compute the value-at-risk from this distribution. Consider a risk factor model and a factorization of the portfolio of the type produced by Algorithm 1. That is, the joint distribution of the transformed risk factors has a product pdf

$$p(x) = p_1(x_1) \cdots p_n(x_n),$$

and the $\Delta\Gamma$-approximation is a sum of independent quadratic functions

$$\Delta\widetilde{\Pi} = \Xi + \sum_{i=1}^{n} \pi_i(x_i) \qquad \text{where} \qquad \pi_i = \Delta_i' x_i + \frac{\lambda_i}{2} x_i^2.$$

Let $p_{\pi_i}(x)$ be the pdf of π_i for $i = 1, \ldots, n$. Then, the pdf of $\Delta\widetilde{\Pi}$ is[1]

$$p_{\Delta\widetilde{\Pi}} = \tau_{\Xi}(p_{\pi_1} * \cdots * p_{\pi_n}). \tag{4.2}$$

4.1.1 The pdf of a quadratic random variable

Given a single risk factor x_i with the pdf p_i and a quadratic portfolio component

$$\pi_i = \Delta_i' x_i + \frac{\lambda_i}{2} x_i^2.$$

It is easy to derive the pdf p_{π_i} for π_i. Let $x_+(u)$ and $x_-(u)$ be the two roots of $\pi_i(x) - u = 0$,

$$x_{\pm}(u) = \frac{-\Delta_i' \pm \sqrt{(\Delta_i')^2 + 2\lambda_i u}}{\lambda_i}, \qquad (-(\Delta_i')^2/2\lambda_i \leq u). \tag{4.3}$$

The pdf p_{π_i} is the derivative of

$$\mathrm{P}[\pi_i \leq u] = \mathrm{P}[x \in x_+([-\infty, u])] + \mathrm{P}[x \in x_-([-\infty, u])].$$

The sets $x_{\pm}([-\infty, u])$ are empty for $u < -(\Delta_i')^2/2\lambda_i$, and otherwise

$$p_{\pi_i}(u) = p_i(x_+(u))x_+'(u) - p_i(x_-(u))x_-'(u)).$$

It follows that the pdf is

$$p_{\pi_i}(u) = \begin{cases} 0, & \text{if } u < -(\Delta_i')^2/2\lambda_i, \\ \frac{p_i(x_+(u)) + p_i(x_-(u))}{\sqrt{(\Delta_i')^2 + 2\lambda_i u}}, & \text{if } u \geq -(\Delta_i')^2/2\lambda_i. \end{cases}$$

The pdf has a singularity at $u = -(\Delta_i')^2/2\lambda_i$, i.e., at the critical point $x = -\Delta_i'/\lambda_i$ of π_i.

4.1.2 Discretization

Let $[-a, a]$ be an interval. Consider a regular grid $\{\xi_j = -a + jh\}_{j=0}^{N-1}$ where $h = 2a/N$. Because p_{π_i} has a singularity, it is necessary to use a discretization scheme that conserves probability (see Figure 4.1). Therefore, we take

$$p_{\pi_i} \approx p_i^D = h \sum_{j=1}^{N-1} p_i^j \delta_j \quad \text{where} \quad p_i^j = \frac{1}{h} \int_{\xi_j - h/2}^{\xi_j + h/2} p_{\pi_i}(u) du \tag{4.4}$$

where δ is the delta function and $\delta_j = \delta(\xi_j - x)$. For convenience, let $x_0 = -a$ and define $p_i^0 = 0$.

[1]The function $f * g$ is the convolution product:

$$f * g(x) = \int_{-\infty}^{\infty} f(y)g(x - y)\,\mathrm{d}y.$$

The functional τ_a is the shift operator defined by

$$\tau_a f(x) = f(x + a).$$

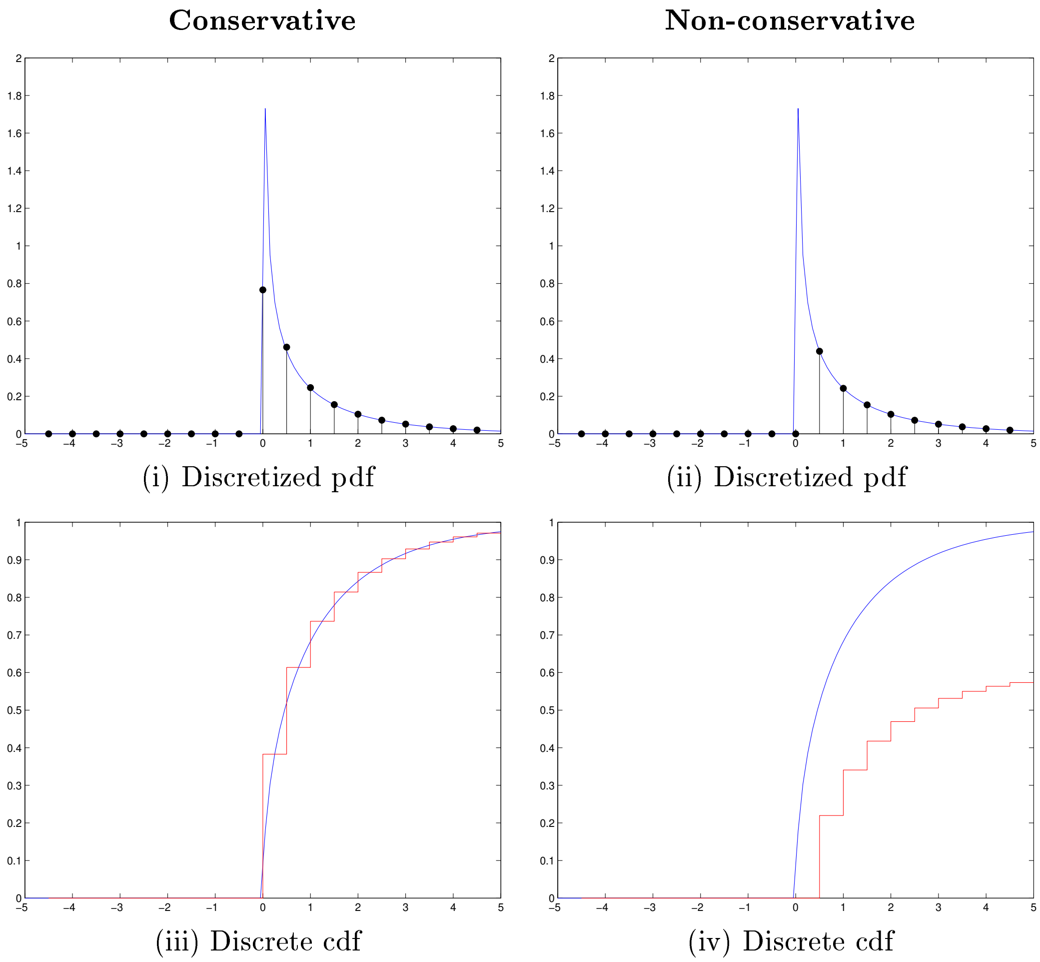


Figure 4.1: An example of the importance of a discretization method that conserves probability density. Consider $\pi = X^2$ where X is normal—a χ^2 random variable with one degree of freedom. For the conservative method, the discrete pdf and cdf are close to the pdf and cdf of the χ^2 random variable. In the non-conservative method, the pdf is sampled at the grid points to get a discrete approximation. As the graph for the cdf shows, the distribution function for the discrete approximation is not close the original cdf.

Accuracy and convergence

The accuracy of the discretized function depends on the density outside the interval $[-a, a]$ and the number of grid points N. Suppose that

$$\mathrm{P}[\pi_i \leq -a + h/2] < \epsilon/2, \qquad \text{and} \qquad \mathrm{P}[\pi_i \geq a - h/2] < \epsilon/2,$$

for $i = 1, \ldots, n$ and some small $\epsilon > 0$. On $[-a, a]$, the discretization converges to the probability density in the weak sense.[2] As a special case of convergence, the approximate cdf converges linearly to the exact cumulative density: if $\xi_k - h/2 \leq y < \xi_k + h/2$ then

$$\left| \int_{\xi_1 - h/2}^{y} p_{\pi_j}(x) - h \sum_{j=0}^{N-1} p_i^j \delta_j(x) dx \right| \quad \leq \quad p_i^k h,$$

and they agree exactly for $y = \xi_k + h/2$. For the cdf, we have

$$\left| \int_{-\infty}^{y} p_{\pi_j}(x) - p_{\pi_j}^D(x) dx \right| \leq \epsilon + h \max_k p_i^k.$$

The computational details

To compute the coefficients in (4.4) is a bit messy. We consider the case $\lambda_i \neq 0$ (the case $\lambda_i = 0$ is straight forward). Considering an interval $[\xi_j - h/2, \xi_j + h/2]$ the computation of p_i^j falls into one of three categories.

(i). The polynomial $\pi_i(x) - u$ does not have any zeros for any $u \in [\xi_j - h/2, \xi_j + h/2]$. This gives
$$p_i^j = 0. \tag{4.5}$$

(ii). The polynomial $\pi_i(x) - u$ has a double zero for some $u \in [\xi_j - h/2, \xi_j + h/2]$. It follows that
$$p_i^j = \begin{cases} \frac{1}{h} \int_{x_-(\xi_j + h/2)}^{x_+(\xi_j + h/2)} p_i(x) dx, & \text{if } \lambda_i > 0, \\ \frac{1}{h} \int_{x_+(\xi_j - h/2)}^{x_-(\xi_j - h/2)} p_i(x) dx, & \text{if } \lambda_i < 0. \end{cases} \tag{4.6}$$

(iii). The polynomial $\pi_i(x) - u$ has two distinct zeros for each $u \in [\xi_j - h/2, \xi_j + h/2]$. This yields
$$p_i^j = \frac{1}{h} \int_{x_+(\xi_j - h/2)}^{x_+(\xi_j + h/2)} p_i(x) dx + \frac{1}{h} \int_{x_-(\xi_j + h/2)}^{x_-(\xi_j - h/2)} p_i(x) dx \tag{4.7}$$

[2] Let $\phi(x)$ be a continuous test function. Then,

$$|\langle \phi, p_{\pi_i} - p_i^D \rangle| \leq \int_{-a}^{-a+h/2} |\phi(y)| p_{\pi_i}(y) dy + \int_{a-h/2}^{a} |\phi(y)| p_{\pi_i}(y) dy$$
$$+ \max_j \max_{x \in I_j} |\phi(x) - \phi(\xi_j)| \left(\int_{-a+h/2}^{a-h/2} p_{\pi_i}(y) dy \right) \quad \text{where} \quad I_j = [\xi_j - h/2, \xi_j + h/2]$$

As $h \to 0$ all three terms vanish, which proves weak convergence.

4.1.3 Convolution with the fast Fourier transform

Because the pdf for the $\Delta\Gamma$-approximation $\Delta\widetilde{\Pi}$ is a convolution product, it can be computed using ideas from Fourier analysis [24]. The convolution product and its Fourier transform[3] satisfies

$$p_{\pi_1} * \cdots * p_{\pi_n} = \left(\prod_{k=1}^{n} \widehat{p_{\pi_k}}\right)^{\vee}.$$

It is possible to compute an approximate pdf for the $\Delta\Gamma$-approximation $\Delta\widetilde{\Pi}$ by multiplying and inverting the discrete Fourier transform of the coefficients of the discretized densities:[4]

$$p_{\pi_1} * \cdots * p_{\pi_n} \approx p^D_{\Delta\widetilde{\Pi}} = h \sum_{j=0}^{N-1} p_j \delta_j \quad \text{where} \quad P^j = (-1)^{j(n-1)} h^{n-1} \prod_{k=1}^{n} P^j_k. \tag{4.8}$$

The sequences $\{P^j\}_{j=0}^{N-1}$ and $\{P^j_k\}_{j=0}^{N-1}$ are the DFT of the sequences $\{p_j\}_{k=0}^{N-1}$ and $\{p^j_k\}_{k=0}^{N-1}$ in $p^D_{\Delta\widetilde{\Pi}}$ and $p^D_j(x)$, respectively. The DFT and the inverse DFT of a sequence with N points can be computed with the fast Fourier transform (FFT) in $O(N \log N)$ floating point operations [10, 24]. To compute the discrete approximation $p^D_{\Delta\widetilde{\Pi}}(\xi_i)$ therefore requires a total of $O(nN \log N)$ floating point operations.

To prove that this method works and that the computed distribution converges linearly as h and ϵ decrease requires a bit of work. Essentially, the proof is an exercise in Fourier analysis and it proceeds in two steps. In the first step, we prove that, for a fixed interval $[-a, a]$, the equation (4.8) converges to the cyclic convolution. In the second step, we show that as a grows the cyclic convolution approximates the standard convolution.

Step 1: Convergence to the cyclic convolution

For the set of integrable functions with compact support in $[-a, a]$, we define the *cyclic convolution* by the integral

$$f \circledast g(x) = \int_{-a}^{a} f^p(x-y) g(y) dy \qquad x \in [-a, a] \tag{4.9}$$

and to be zero elsewhere. The function $f^p(x)$ is the periodic extension of $f(x)$, or in other words $f^p(\cdot) = \sum_k f(\cdot - 2ak)$. It is easy to show that the cyclic convolution is commutative and

[3] The continuous Fourier transform of a function $f(x)$ is

$$\widehat{f}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega x} f(x) dx.$$

The inverse Fourier transform is

$$(\widehat{f})^{\vee}(x) = \int_{-\infty}^{\infty} e^{i\omega x} \widehat{f}(\omega) d\omega.$$

[4] Given a sequence $\{y_k\}_{k=0}^{N-1}$, the discrete Fourier transform (DFT) $\{Y^k\}_{k=0}^{N-1}$ and its inverse are defined by

$$y_k = \frac{1}{N} \sum_{j=0}^{N-1} Y^j \omega^{jk}, \qquad k = 0, 1, \ldots, N-1,$$

$$Y^k = \sum_{j=0}^{N-1} y_j \omega^{-jk}, \qquad k = 0, 1, \ldots, N-1,$$

where $\omega = e^{-i\frac{2\pi}{N}}$.

linear. Furthermore, if the two functions are in L^1, we have

$$\|f \circledast g\|_1 \leq \|f\|_1 \|g\|_1,$$

mimicking the similar property for the standard convolution.

Theorem 4.1. *Assume f and g are Riemann integrable and have a finite number of discontinuities. Let*

$$f^D = h \sum_{k=0}^{N-1} f^k \delta_k \qquad \text{and} \qquad g^D = h \sum_{k=0}^{N-1} g^k \delta_k.$$

be discretizations (in the manner of Section 4.1.2). Then,

$$f^D \circledast g^D = h \sum_{k=0}^{N-1} a^k \delta_k \qquad \text{where} \qquad A^j = (-1)^j h F_k^j G_k^j,$$

and $f^D \circledast g^D \to f \circledast g$ and $f^D \circledast g^D \to (f \circledast g)^D$ in the weak sense.

Proof. We prove the two statements separately, starting with the first one.

1. The cyclic convolution $f^D \circledast g^D$ is

$$f^D \circledast g^D(x) = h^2 \sum_{k=0}^{N-1} \left(\sum_{l=0}^{N-1} f^{k-l+N/2} g^l \right) \delta_k(x),$$

and

$$a^k = h \sum_{l=0}^{N-1} f^{k-l+N/2} g^l, \qquad k = 0, 1, \ldots, N-1.$$

By applying the DFT with $\omega = e^{-i\frac{2\pi}{N}}$, it follows that

$$\begin{aligned} A^k &= h \sum_{j=0}^{N-1} a^j \omega^{-jk} \\ &= h\omega^{-kN/2} \sum_{l=0}^{N-1} g^l \omega^{-lk} \left(\sum_{j=0}^{N-1} f^{(j-l)-N/2} \omega^{-k(j-l)+kN/2} \right) \\ &= (-1)^k h F^k G^k. \end{aligned}$$

2. It remains to show that $f^D \circledast g^D \to f \circledast g$ and $f^D \circledast g^D \to (f \circledast g)^D$ in the weak sense. Since f and g are piecewise continuous, we have, for all points of continuity,

$$\begin{aligned} &\frac{1}{h} \int_{x-h/2}^{x+h/2} f(z-y)dy = f(z-x) + q(z-x,h)h \\ &\frac{1}{h} \int_{x-h/2}^{x+h/2} g(y)dy = g(x) + r(x,h)h \end{aligned}$$

where $q(\cdot,h)h \to 0$ and $r(\cdot,h)h \to 0$ as $h \to 0$. Let $\phi(x)$ be a smooth test function. Then,

$$\int_{-a}^{a} \phi(x) f^D \circledast g^D(x)dx = \sum_{k=0}^{N-1} \phi(\xi_k) \sum_{l=0}^{N-1} \left(\int_{\xi_l - h/2}^{\xi_l + h/2} f^p(\xi_k - y)dy \right) \left(\int_{\xi_l - h/2}^{\xi_l + h/2} g^p(y)dy \right)$$
$$= \sum_{k=0}^{N-1} h\phi(\xi_k) \sum_{l=0}^{N-1} h \left(f(\xi_k - \xi_l) + hq(\xi_k - \xi_l, h) \right) \left(g(\xi_l) + hr(\xi_l, h) \right).$$

We note that

$$\sum_{l=0}^{N-1} hf(\xi_k - \xi_l) g(\xi_l) \to f \circledast g(\xi_k)$$

and

$$\sum_{k=0}^{N-1} h\phi(\xi_k) f \circledast g(\xi_k) \to \int_{-a}^{a} \phi(x) f \circledast g(x)dx.$$

If necessary, the finite number of points with discontinuities may be excluded without changing the limits. The remaining integrals involving r and q vanish proving that $f^D \circledast g^D \to f \circledast g$.

Since $f \circledast g$ is Riemann integrable and piecewise continuous, it follows that

$$\frac{1}{h} \int_{x-h/2}^{x+h/2} f \circledast g(y)dy = f \circledast g(x) + hs(x,h),$$

at all points of continuity. Hence,

$$\int_{-a}^{a} \phi(x)(f \circledast g)^D(x)dx = \sum_{k=0}^{N-1} h\phi(\xi_k) \frac{1}{h} \int_{\xi_k - h/2}^{\xi_k + h/2} f \circledast g(y)dy \to \int_{-a}^{a} \phi(x) f \circledast g(x)dx,$$

as $h \to 0$.

□

Step 2: The cyclic approximation to the standard convolution

By repeated applications of Theorem 4.1, we conclude that the right hand side of (4.8) converges weakly to the cyclic convolution of the truncated densities. Our remaining obligation is to show that the standard convolution can be approximated by the cyclic convolution. The intuition being that since a pdf decays in the tails, for a large enough interval, the density of the overlapping regions in the cyclic convolution decreases.

In the first lemma, we show that the error from restricting the pdfs, to an interval containing the majority of the density, is small.

Lemma 4.1. *Given the probability density functions p_{π_i} where $i = 1, \dots, n$, let*

$$\widetilde{p}_{\pi_i}(x) = \begin{cases} p_{\pi_i}(x), & \text{if } -a \le x \le a, \\ 0, & \text{otherwise.} \end{cases}$$

where $\|p_{\pi_i} - \widetilde{p}_{\pi_i}\|_1 \le \epsilon$ for some $\epsilon > 0$ and for all $i = 1, \dots, n$. Then,

$$\|p_{\pi_1} * \dots * p_{\pi_n} - \widetilde{p}_{\pi_1} * \dots * \widetilde{p}_{\pi_n}\|_1 \le \epsilon n.$$

Proof. The statement follows immediately from the inequality $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$:

$$\begin{aligned}\|p_{\pi_1} * \cdots * p_{\pi_n} - \widetilde{p}_{\pi_1} * \cdots * \widetilde{p}_{\pi_n}\|_1 &\leq \|(p_{\pi_1} - \widetilde{p}_{\pi_1}) * \cdots * p_{\pi_n}\|_1 \\ &\quad + \cdots + \|\widetilde{p}_{\pi_1} * \cdots * (p_{\pi_n} - \widetilde{p}_{\pi_n})\|_1 \\ &\leq \epsilon n.\end{aligned}$$

□

In the next lemma, we show that if two pdfs are small outside an interval around the origin and if the cyclic convolution is taken over a large enough interval then the error is small.

Lemma 4.2. *Assume that the pdfs f and g satisfy*

$$f(x) \leq C_1 \frac{1}{|x|^{\alpha+1}} \qquad \textit{and} \qquad g(x) \leq C_2 \frac{1}{|x|^{\alpha+1}}, \tag{4.10}$$

for some $\alpha > 0$ for all x outside some interval $[-b, b]$. Suppose that $a > 2b$. Let $\widetilde{f}$ and $\widetilde{g}$ be the restrictions of f and g, respectively, to the interval $[-a, a]$. Then,

$$0 \leq \widetilde{f} \circledast \widetilde{g}(x) - \widetilde{f} * \widetilde{g}(x) \leq \frac{1}{(a - |x|/2)^{\alpha+1}} \big(C_1 D_2(x) + C_2 D_1(x)\big)$$

where

$$D_1(x) = \begin{cases} \int_{-a+x/2}^{-a+x} f(y)dy, & \textit{if } x > 0, \\ \int_{a+x}^{a+x/2} f(y)dy, & \textit{if } x \leq 0, \end{cases} \qquad \textit{and} \qquad D_2(x) = \begin{cases} \int_{-a+x/2}^{-a+x} g(y)dy, & \textit{if } x > 0, \\ \int_{a+x}^{a+x/2} g(y)dy, & \textit{if } x \leq 0. \end{cases}$$

Proof. We prove the bound in two steps. Consider a point x in the interval $[-a, a]$. Suppose that $0 < x \leq a$.

1. Then,

$$\widetilde{f} * \widetilde{g}(x) = \int_{-a+x}^{a} f(x-y)g(y)dy$$

and

$$\widetilde{f} \circledast \widetilde{g}(x) = \int_{-a}^{-a+x} f(x-y-2a)g(y)dy + \int_{-a+x}^{a} f(y-x)g(y)dy.$$

Since f and g are positive, the error satisfies

$$0 \leq \widetilde{f} \circledast \widetilde{g}(x) - \widetilde{f} * \widetilde{g}(x) = \int_{-a}^{-a+x} f(x-y-2a)g(y)dy.$$

2. Consider the interval $[-a, -a+x]$. The functions $g(y)$ and $f(y)$ satisfy (4.10) in $[-a, -a+x/2]$ and $[-a+x/2, -a+x]$, respectively. Hence,

$$\begin{aligned}\int_{-a}^{-a+x/2} f(x-y-2a)g(y)dy &\leq C_2 \int_{-a}^{-a+x/2} \frac{1}{|y|^{\alpha+1}} f(x-y-2a)dy, \\ \int_{-a+x/2}^{-a+x} f(x-y-2a)g(y)dy &\leq C_1 \int_{-a}^{-a+x/2} \frac{1}{|y|^{\alpha+1}} g(x-y-2a)dy.\end{aligned}$$

Since $|y|^{\alpha+1} < (a-|x|/2)^{\alpha+1}$, by combing the two equalities we find that

$$0 \leq \int_{-a}^{-a+x} f(x-y-2a)g(y)dy \leq \frac{1}{(a-|x|/2)^{\alpha+1}} \big(C_2 D_1(x) + C_1 D_2(x)\big),$$

as required.

The proof for negative x is similar. The case $x \notin [-a, a]$ is trivial since both functions are zero outside the interval. $\square$

With the help of the two lemmas, we are in a position to relate the cyclic approximation to the standard convolution.

Theorem 4.2. *Let p_{π_i} for $i = 1, \dots, n$ be probability density functions that satisfy the assumptions in Lemma 4.1 and Lemma 4.2. Then,*

$$\|p_{\pi_1} * \cdots * p_{\pi_n} - \widetilde{p}_{\pi_1} \circledast \cdots \circledast \widetilde{p}_{\pi_n}\|_1 = O(\epsilon + a^{-\alpha}).$$

Proof. Use Lemma 4.2 to bound the error from the cyclic convolution. Since the result is asymptotic, we may assume that $a > 2b$ so the the lemma can be applied. Note that

$$\begin{aligned}\widetilde{p}_{\pi_1} * \cdots * \widetilde{p}_{\pi_n} - \widetilde{p}_{\pi_1} \circledast \cdots \circledast \widetilde{p}_{\pi_n} &= (\widetilde{p}_{\pi_1} * \widetilde{p}_{\pi_2} - \widetilde{p}_{\pi_1} \circledast \widetilde{p}_{\pi_2}) * \cdots * \widetilde{p}_{\pi_n} \\ &\quad + \cdots + \widetilde{p}_{\pi_1} \circledast \cdots \circledast (\widetilde{p}_{\pi_{n-1}} * \widetilde{p}_{\pi_n} - \widetilde{p}_{\pi_{n-1}} \circledast \widetilde{p}_{\pi_n})\end{aligned}$$

By the lemma and because $|D(x)| \leq 1$, it follows that

$$\begin{aligned}\|\widetilde{p}_{\pi_k} * \widetilde{p}_{\pi_{k+1}} - \widetilde{p}_{\pi_k} \circledast \widetilde{p}_{\pi_{k+1}}\|_1 &\leq 2(C_k + C_{k+1}) \int_0^a \frac{1}{(a - x/2)^{\alpha+1}} dx \\ &= (C_k + C_{k+1}) \frac{4}{\alpha} \left(\left(\frac{2}{a} \right)^{\alpha} - \frac{1}{a^{\alpha}} \right).\end{aligned}$$

Hence,

$$\|\widetilde{p}_{\pi_1} * \cdots * \widetilde{p}_{\pi_n} - \widetilde{p}_{\pi_1} \circledast \cdots \circledast \widetilde{p}_{\pi_n}\| = O(a^{-\alpha}).$$

By Lemma 4.1 and the triangle inequality, it follows that

$$\|p_{\pi_1} * \cdots * p_{\pi_n} - \widetilde{p}_{\pi_1} \circledast \cdots \circledast \widetilde{p}_{\pi_n}\|_1 = O(\epsilon + a^{-\alpha}).$$

$\square$

Theorem 4.1 and Theorem 4.2 show that (4.8) works. As the later examples demonstrate, the rate of convergence for numerical experiments appear to be linear (see Figure 4.2). It is therefore reasonable to guess that a stronger version of Theorem 4.1 may be true.

4.2 Computing value-at-risk

In this section, we discuss how to compute value-at-risk from the discrete approximation (4.8) to the pdf of the $\Delta\Gamma$-approximation. Recall that the value-at-risk is defined by a non-linear equation (1.1), and we approximate it by the solution to (4.1). Given the pdf

$$p^D_{\Delta\widetilde{\Pi}} = h \sum_{j=0}^{N-1} p_j \delta_j,$$

we simply add the coefficients to get the cdf at the grid points

$$P_{\Delta\widetilde{\Pi}}(\xi_k) = h\frac{p_0}{2} + h \sum_{j \leq k} p_j \qquad \text{and} \qquad P_{\Delta\widetilde{\Pi}}(\xi_N) = P_{\Delta\widetilde{\Pi}}(a) = hp_0 + h \sum_{j=1}^{N-1} p_j. \tag{4.11}$$

Since the first point of the grid corresponds to the interval $[-a, -a + h/2] \cup [a - h/2, a]$, we choose to assign half of the density to the right-end grid point a and half to the left-end grid point $-a$.

Value-at-risk can then be computed by interpolating the cdf. Since the cdf is an increasing function, we search for the index k such that

$$P_{\Delta\widetilde{\Pi}}(\xi_k) \leq 1 - \alpha \leq P_{\Delta\widetilde{\Pi}}(\xi_{k+1})$$

and therefore $-\mathrm{VaR} - \Xi$ is in the interval $[\xi_k, \xi_{k+1}]$. The linear interpolant to the inverse cdf is

$$L(p) = \xi_k + \left(\frac{p - P_{\Delta\widetilde{\Pi}}(\xi_k)}{P_{\Delta\widetilde{\Pi}}(\xi_{k+1}) - P_{\Delta\widetilde{\Pi}}(\xi_k)}\right) h, \qquad p \in [P_{\Delta\widetilde{\Pi}}(\xi_k), P_{\Delta\widetilde{\Pi}}(\xi_{k+1})].$$

The desired approximation to value-at-risk is

$$\mathrm{VaR} = -L(1 - \alpha) - \Xi.$$

4.2.1 Richardson's extrapolation improves accuracy

In practice, we have found that the observed rate of convergence can be improved with a step of Richardson's extrapolation. In our implementation, we compute two solutions VaR_f and VaR_c on a fine grid with step size h and a coarse grid with step size $2h$, respectively. Richardson's extrapolation then gives the solution[5]

$$\mathrm{VaR} = 2\mathrm{VaR}_f - \mathrm{VaR}_c.$$

Although Richardson's extrapolation can be extended to eliminate higher-order errors, we have found that additional levels of extrapolation do not lead to further improvements. The steps of the fast convolution method for value-at-risk are summarized in Algorithm 2.

The advantage of Richardson's extrapolation is clearly illustrated by the computational example in Figure 4.2. The four graphs show the error as a function of the steps size h. The functions are

$$\begin{aligned}
(i) &\qquad \Delta\widetilde{\Pi} = X_1 + X_2, \\
(ii) &\qquad \Delta\widetilde{\Pi} = X_1^2 + X_2^2, \\
(iii) &\qquad \Delta\widetilde{\Pi} = -(X_1^2 + X_2^2), \\
(iv) &\qquad \Delta\widetilde{\Pi} = X_1^2 + X_2^2 - X_3^2 - X_4^2 + X_5.
\end{aligned}$$

and the random variables X_i are normal. The graphs show the error for confidence level α equal to 1%, 5%, 10% and 20%; and for value-at-risk computed with and without Richardson's extrapolation. The error is estimated as the difference between the value-at-risk for consecutive

[5]Suppose that we want to compute y. If we have two approximations satisfying

$$\begin{aligned}
y_1 &= y + ch + o(h), \\
y_2 &= y + c(2h) + o(h),
\end{aligned}$$

then the linear error term can be eliminated with Richardson's extrapolation

$$2y_1 - y_2 = y + o(h).$$

It is easy to derive similar formulas for higher-order terms.

Fast convolution method for value-at-risk

Input: A $\Delta\Gamma$-approximation. A time series of daily returns. A confidence level $0 < \alpha < 1$. A number of grid points N and a bound for the grid interval a.

Output: The value-at-risk for one day with confidence level α.

- Estimate the parameters and factorize the $\Delta\Gamma$-approximation with Algorithm 1.
- Create the grid. Define $\{\xi_k = -a + hk\}_{k=0}^{N-1}$ where $h = 2a/N$.
- Discretize the densities,
$$p_i^D = h \sum_{j=1}^{N-1} p_i^j \delta_j \quad \text{where} \quad p_i^j = \frac{1}{h} \int_{\xi_j - h/2}^{\xi_j + h/2} p_\pi(u) du.$$
- Compute the density for the $\Delta\Gamma$-approximation with the FFT:
$$p_{\Delta\tilde{\Pi}}^D = h \sum_{j=0}^{N-1} p_j \delta_j \quad \text{where} \quad P^j = (-1)^{j(n-1)} h^{n-1} \prod_{k=1}^{n} P_k^j.$$
- Compute the discrete cdf $P_{\Delta\tilde{\Pi}}$ and find k such that
$$P_{\Delta\tilde{\Pi}}(\xi_k) \leq 1 - \alpha \leq P_{\Delta\tilde{\Pi}}(\xi_{k+1}).$$
- Compute the linear interpolant $L(\cdot)$ for the inverse over $[\xi_k, \xi_{k+1}]$ and let
$$\mathrm{VaR} = -L(1-\alpha) - \Xi.$$
- Repeat from step 2 for a grid with step size $2h$. Extrapolate to get a more accurate approximation.

Algorithm 2: The fast convolution method for value-at-risk.

grid sizes. We computed the rate of convergence with linear regression. Without extrapolation, the observed rate of convergence is very close to 1, for all four problems and all confidence levels. When extrapolation is used, all problems show a faster rate of convergence, but the systematic relationship is less clear. The estimated rates of convergence range from approximately 1.5 for problem (ii) and (iii) to approximately 2 for problem (i).

4.2.2 Computational complexity

The number of floating point operations for Algorithm 2 is $O(n^2 \min(d, n) + nN \log N)$, where n is the number of risk factors, d is the number of dates in the time series, and N is the grid size. Figure 4.3 shows the computation time for portfolios for increasing n, the remaining parameters are fixed with $d = 1000$ and $N = 4096$. The figure shows that the time is essentially linear in n. In addition, we note that a large portion of the computation time is spent in the parameter estimation step. As expected, the parameter estimation for the Parzen model is much slower than for asymmetric T model, which in turn is slower than the normal model.

4.3 Example: fat-tails and value-at-risk

We illustrate the performance of the algorithm with an example. The example also demonstrates the importance of a return model that incorporates fat-tails and that includes information about the curvature of the function for the portfolio value. Consider a portfolio containing one European call option on each of BCE and Canadian Tire. The options are at-the-money and have 3 months to maturity. The Black-Scholes price of the options are \$3.23 and \$1.39. The Hessian in the $\Delta\Gamma$-approximation is

$$\Gamma = \begin{bmatrix} 99.0967 & 0 \\ 0 & 42.6138 \end{bmatrix}.$$

The portfolio is similar to the example in Section 3.4.2, but has shorter time to maturity which increases the curvature.

An investor who has sold this portfolio will see her holdings decrease in value if the stock prices increase. To hedge the portfolio she might take a linear position that offsets the Δ of the portfolio. To examine how the value-at-risk changes with the Δ, we computed the 95% and 99% value-at-risk on a grid with $-10 \leq \Delta_i \leq 10$ where $i = 1, 2$. Figures 4.4 and 4.5 show the level sets for value-at-risk. Each graph is computed with a 30×30 grid for Δ. The first two graphs in Figure 4.4 show the surface for the method from Section 1.2, the dynamics of portfolio value are approximated by a linear function and the returns on the risk factors are modeled by a multivariate normal with zero mean. The remaining six graphs in Figures 4.4 and 4.5 were computed with the fast convolution method with $N = 4096$ grid points. The dynamics of portfolio value are approximated by the same $\Delta\Gamma$-approximation in all six simulations, and the risk factors are modeled using the three models introduced in Chapter 2.

We see that the linear model over simplifies the problem. In particular, it suggests that the risk is eliminated completely by the hedging strategy, which is not true. From the graph it is clear that it also underestimates the risk away from the origin.

Although the remaining six graphs are more in agreement with each other, we can see some interesting differences. Again the estimate for the value-at-risk is smaller for the normal model compared to the asymmetric T and Parzen models. For $\Delta = 0$, the relative differences are approximately 16% (38%) and 80% (92%) for the asymmetric T model (Parzen model) and for

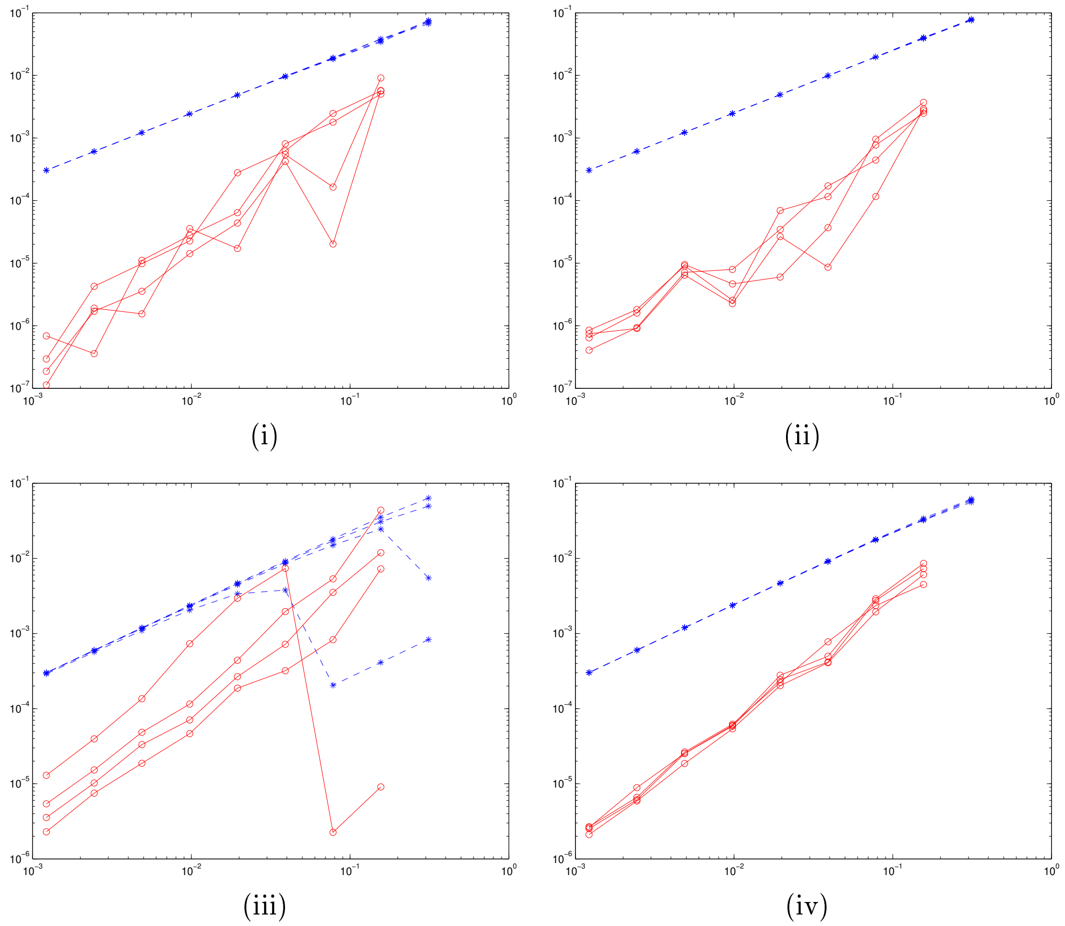


Figure 4.2: The error in value-at-risk as a function of step size for four abstract problems. Each graph shows the error for the confidence levels 1%, 5%, 10% and 20%, and for value-at-risk computed with and without Richardson's extrapolation. Extrapolation is superior since the errors are smaller and the observed rate of convergence improves from 1 to between 1.5 and 2, depending on the problem.

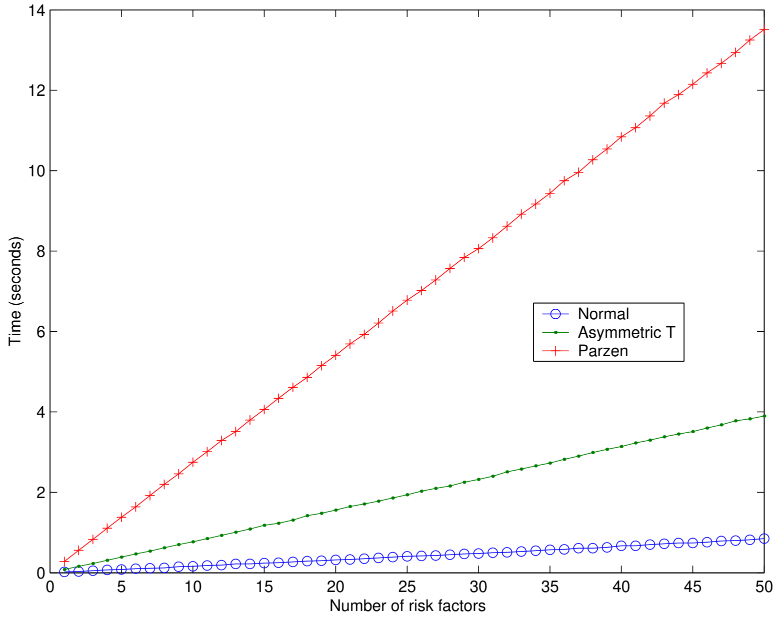


Figure 4.3: Computation time versus the number risk factors for the three return models. The parameter estimation takes a large portion of the total time.

the 95% and 99% value-at-risk, respectively. The differences between the asymmetric T and Parzen model are about 19% and 7% for the 95% and 99% value-at-risk, respectively. The Parzen model gives a larger value-at-risk compared to both the other models. The level sets of the normal and asymmetric T models are elliptical whereas the level sets for the Parzen model display less symmetry.

The fat-tails are primarily a concern for portfolios with negative curvature. Figure 4.6 shows the 95% value-at-risk for a long position in the example portfolio. The differences between the four graphs are much smaller, and the last three are almost identical close to the origin. Of course, this just confirms that buying options is much less risky than selling them.

4.3.1 So which result can we trust?

To better understand the simulations results, we decided to take a closer look at a long position in the Δ-hedged portfolio, the portfolio with $\Delta = 0$. Recall that if the value-at-risk is correct, we expect to have approximately 5 (25) losses exceeding the 99% (95%) value-at-risk for a sample of 500 returns. We repeated the value-at-risk simulation for 500 consecutive days and computed the number of losses greater than value-at-risk over the 500 returns used in the calculation. Figure 4.7 shows histograms for the normal, asymmetric T and Parzen models.

From the examples in Chapter 2, we know that the normal model does not produce an accurate model for the tails. This is confirmed by the simulation since the number of losses greater than value-at-risk deviates from the expected values. The histograms for the asymmetric T model are centered close to the expected values. The Parzen model produces a good estimate for the 99% value-at-risk, but it seems to overestimate the smaller confidence level. We conclude that the asymmetric T and Parzen models are preferable to the normal model. Furthermore, for this example, both models give acceptable results for the 99% value-at-risk,

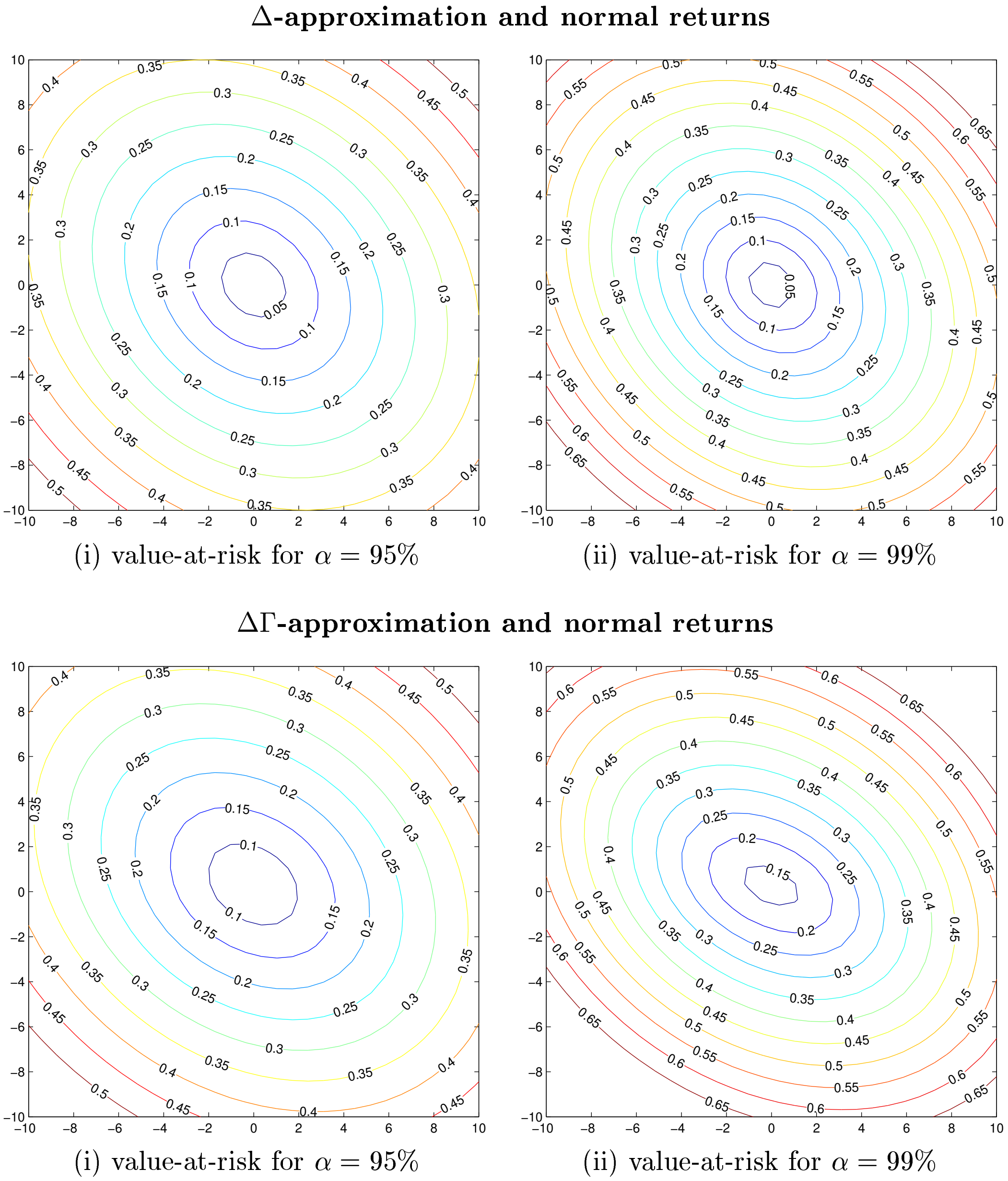
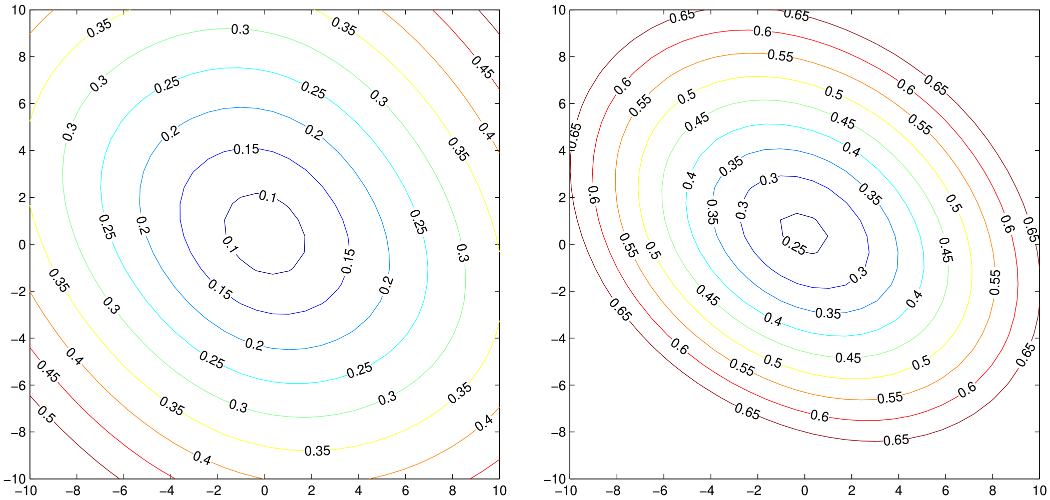


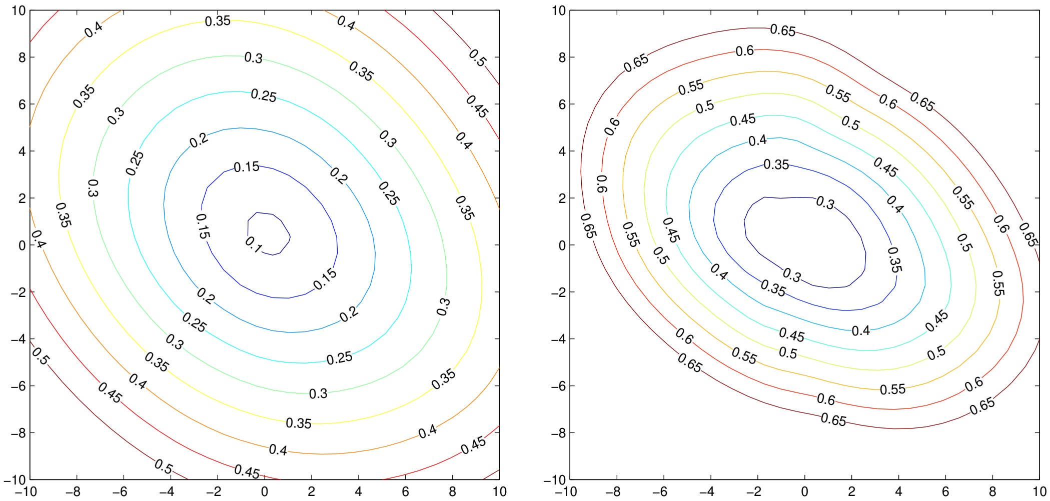
Figure 4.4: Part I. Value-at-risk for a short position in an option portfolio as a function of Δ_1 and Δ_2. The horizontal axis is Δ_1, the linear position in BCE; and the vertical axis is Δ_2, the linear position in CTRa.

ΔΓ-approximation and asymmetric-T returns



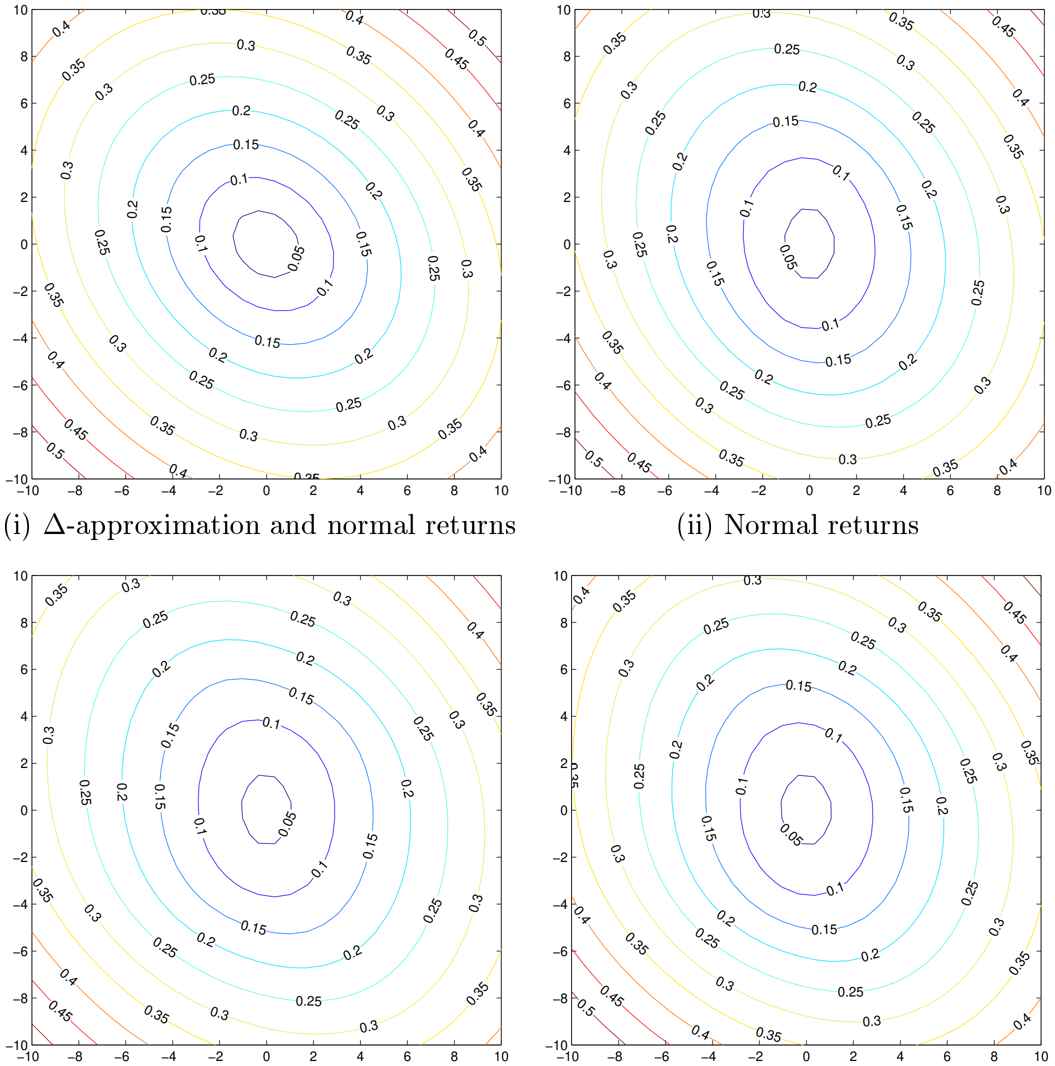
(i) value-at-risk for $\alpha = 95\%$ (ii) value-at-risk for $\alpha = 99\%$

ΔΓ-approximation and Parzen estimate for returns



(i) value-at-risk for $\alpha = 95\%$ (ii) value-at-risk for $\alpha = 99\%$

Figure 4.5: Part II. Value-at-risk for a short position in an option portfolio as a function of Δ_1 and Δ_2. The horizontal axis is Δ_1, the linear position in BCE; and the vertical axis is Δ_2, the linear position in CTRa.



(i) Δ-approximation and normal returns

(ii) Normal returns

(iii) Asymmetric T returns

(iv) Parzen estimate for returns

Figure 4.6: Part III. The 95% value-at-risk for long position in an option portfolio as a function of Δ_1 and Δ_2. The horizontal axis is Δ_1, the linear position in BCE; and the vertical axis is Δ_2, the linear position in CTRa.

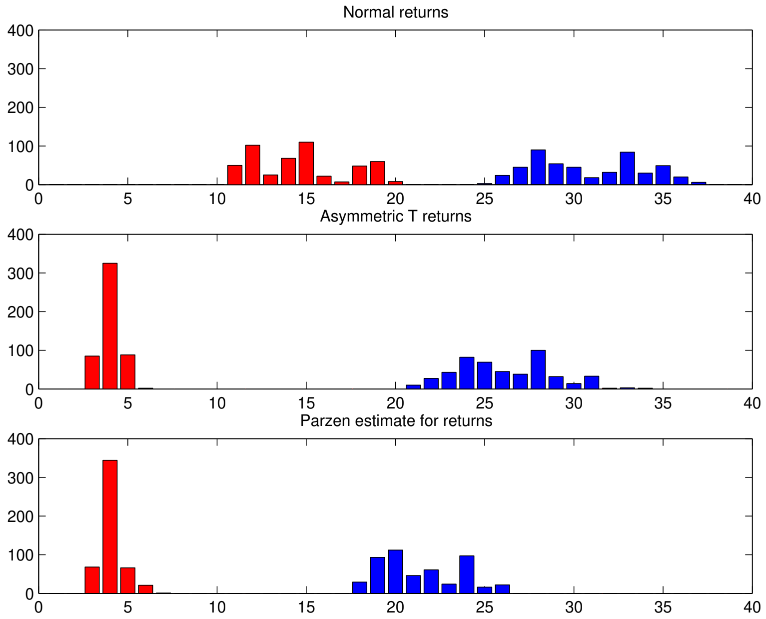


Figure 4.7: Histograms of the number of losses larger than value-at-risk for $\alpha = 95\%$ and $\alpha = 99\%$. Each graph shows two superimposed histograms where the left "bump" is the result for 99% value-at-risk and the right one is the result for the 95% value-at-risk.

and the asymmetric T model produces a better estimate for the 95% value-at-risk.

4.4 Computing the gradient of value-at-risk

For small portfolios, such as the one in our example, it is possible to understand how value-at-risk changes with portfolio composition by computing it many times, varying the parameters, and visualizing the result. For large portfolios and for increasing number of parameters, this is a time consuming strategy and the result becomes harder to interpret. In this section, we extend the fast convolution method to compute the gradient of value-at-risk. It is an interesting problem because the gradient gives local information about how the value-at-risk changes with portfolio composition. Gradient information is important to understand and evaluate decisions about changes in the portfolio. In [41] Mausser and Rosen review applications and methods for value-at-risk gradients. Monte Carlo methods to compute value-at-risk have been proposed independently by several authors, see for example the methods developed by Paun [47] and Mausser and Rosen [41]. For the linear model with normal returns from Section 1.2, the gradient may be interpreted in terms of "risk contributions" from instruments or risk factors, see for example the paper by Hallerbach [29].

Recall that value-at-risk is defined by the non-linear equation (1.1). In the fast convolution method, it is approximated by the solution to (4.1). We can reformulate (4.1) as a one-dimensional integral over the pdf of $\Delta\widetilde{\Pi}$,

$$\int_{-\infty}^{-\mathrm{VaR}} p_{\Delta\widetilde{\Pi}}(x)dx = 1 - \alpha.$$

Then, the gradient can be computed by implicit differentiation:

$$\frac{\partial \mathrm{VaR}}{\partial \Theta} = \frac{1}{p_{\Delta\widetilde{\Pi}}(-\mathrm{VaR})} \int_{-\infty}^{-\mathrm{VaR}} \frac{\partial p_{\Delta\widetilde{\Pi}}}{\partial \Theta} dx, \tag{4.12}$$

$$\frac{\partial \mathrm{VaR}}{\partial \Delta_i'} = \frac{1}{p_{\Delta\widetilde{\Pi}}(-\mathrm{VaR})} \int_{-\infty}^{-\mathrm{VaR}} \frac{\partial p_{\Delta\widetilde{\Pi}}}{\partial \Delta_i'} dx, \tag{4.13}$$

$$\frac{\partial \mathrm{VaR}}{\partial \Lambda_{ij}} = \frac{1}{p_{\Delta\widetilde{\Pi}}(-\mathrm{VaR})} \int_{-\infty}^{-\mathrm{VaR}} \frac{\partial p_{\Delta\widetilde{\Pi}}}{\partial \Lambda_{ij}} dx. \tag{4.14}$$

Of course, rather than the derivatives for the parameters in the portfolio factorization, (3.14) and (3.16), we want the derivatives for the parameters in the original $\Delta\Gamma$-approximation or the portfolio positions θ_i. Fortunately, these can be computed from (4.12)-(4.14)[6].

4.4.1 The value-at-risk gradient and portfolio composition

Suppose that the VaR gradient with respect to the parameters in the $\Delta\Gamma$-approximation are know. The gradient with respect to θ_i, the quantity invested in the ith security, follows from the chain rule:

$$\frac{\partial \mathrm{VaR}}{\partial \theta_i} = \frac{\partial \mathrm{VaR}}{\partial \Theta}\frac{\partial \Theta}{\partial \theta_i} + \sum_{k=1}^{n} \frac{\partial \mathrm{VaR}}{\partial \Delta_k}\frac{\partial \Delta_k}{\partial \theta_i} + \sum_{k=1}^{n}\sum_{j=1}^{n} \frac{\partial \mathrm{VaR}}{\partial \Gamma_{kj}}\frac{\partial \Gamma_{kj}}{\partial \theta_i}. \tag{4.15}$$

The derivatives for the parameters of the $\Delta\Gamma$-approximation can in turn be computed from the derivatives with respect the parameters in the portfolio factorization (3.14), (3.16) and (3.17). The Θ derivative (4.12) immediately gives

$$\frac{\partial \mathrm{VaR}}{\partial \Theta} = -t.$$

For the remaining two derivatives, it follows after some calculation that the gradient vector for Δ_i is

$$\left[\frac{\partial \mathrm{VaR}}{\partial \Delta_i}\right]_{i=1}^{n} = A\left(\left[\frac{\partial \mathrm{VaR}}{\partial \Delta_i'}\right]_{i=1}^{n} - \widehat{\mu}\right) \tag{4.16}$$

and the Jacobian matrix for Γ_{ij} is

$$\left[\frac{\partial \mathrm{VaR}}{\partial \Gamma_{ij}}\right]_{i,j=1}^{n} = \frac{1}{2}A\left(\left[\frac{\partial \mathrm{VaR}}{\partial \Lambda_{ij}}\right]_{i,j=1}^{n} + \left[\widehat{\mu}_i \frac{\partial \mathrm{VaR}}{\partial \Delta_j'}\right]_{i,j=1}^{n} + \left[\frac{\partial \mathrm{VaR}}{\partial \Delta_i'}\widehat{\mu}_j\right]_{i,j=1}^{n} - \widehat{\mu}\widehat{\mu}^T\right)A^T. \tag{4.17}$$

4.4.2 Computing the gradient

Consider the integrals (4.12)-(4.14). We note that the density $p_{\Delta\widetilde{\Pi}}(-\mathrm{VaR})$ is directly available in Algorithm 2; it can be computed by interpolation. Therefore, the task that remains is to approximate the two integrands in (4.13) and (4.14).

[6] In the gradient computation, we have left out the contribution to the gradient from the parameter estimates.

Partial derivatives of the pdf

The derivative with respect to Δ_i' is

$$\frac{\partial p_{\Delta\tilde{\Pi}}}{\partial \Delta_i'} = \tau_\Xi\big(p_{\pi_1} * \cdots * \frac{\partial p_{\pi_i}}{\partial \Delta_i'} * \cdots * p_{\pi_n}\big).$$

Similarly, the derivative with respect to the diagonal element Λ_{ii} is

$$\frac{\partial p_{\Delta\tilde{\Pi}}}{\partial \Lambda_{ii}} = \tau_\Xi\big(p_{\pi_1} * \cdots * \frac{\partial p_{\pi_i}}{\partial \Lambda_{ii}} * \cdots * p_{\pi_n}\big).$$

To find the derivatives for Λ_{ij} where $i \neq j$, we have to resort to a slightly different technique[7]. The matrix Λ is a Hessian and hence symmetric. Therefore, we only have to consider derivatives in directions that preserve symmetry. Consider a perturbation in the direction $E_{ij}+E_{ji}$ $(i \neq j)$. Differentiating the curve

$$\begin{pmatrix} 1 & -t \\ t & 1 \end{pmatrix} \begin{pmatrix} \Lambda_{ii} & \\ & \Lambda_{jj} \end{pmatrix} \begin{pmatrix} 1 & t \\ -t & 1 \end{pmatrix},$$

we get

$$\left.\frac{d}{dt}\right|_{t=0} \begin{pmatrix} 1 & -t \\ t & 1 \end{pmatrix} \begin{pmatrix} \Lambda_{ii} & \\ & \Lambda_{jj} \end{pmatrix} \begin{pmatrix} 1 & t \\ -t & 1 \end{pmatrix} = (\Lambda_{ii} - \Lambda_{jj}) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Since the curve is tangent to $E_{ij} + E_{ji}$, we consider $p_{\Delta\tilde{\Pi}}$ along this curve as a function of t

$$p_{\Delta\tilde{\Pi}} = p_{\pi_1} * \ldots * p_{\pi_i}(\Lambda_{ii}(t), \Delta_i'(t)) * \ldots * p_{\pi_j}(\Lambda_{jj}(t), \Delta_j'(t)) * \ldots * p_{\pi_n}.$$

For this curve, we get

$$\left.\frac{d}{dt}\right|_{t=0} \Delta'(t) = \left.\frac{d}{dt}\right|_{t=0} \begin{pmatrix} 1 & t \\ -t & 1 \end{pmatrix}^{-1} \Delta' = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \Delta'$$

The derivative of $p_{\Delta\tilde{\Pi}}$ in the direction of $T = \frac{1}{2}(E_{ij} + E_{ji})$ is the same as the derivative of $p_{\Delta\tilde{\Pi}}$ along the curve renormalized to compensate for the curve not having unit speed:

$$\nabla_T p_{\Delta\tilde{\Pi}} = \left.\frac{d}{dt}\right|_{t=0} \frac{\tau_\Xi\big(p_{\pi_1} * \ldots * p_{\pi_i}(\Lambda_{ii}(t), \Delta_i' + t\Delta_j') * \ldots * p_{\pi_j}(\Lambda_{jj}(t), \Delta_j' - t\Delta_i') * \ldots * p_{\pi_n}\big)}{2(\Lambda_{ii} - \Lambda_{jj})}.$$

Hence, it follows that

$$\frac{\partial p_{\Delta\tilde{\Pi}}}{\partial \Lambda_{ij}} = \frac{1}{2(\Lambda_{ii} - \Lambda_{jj})} \left(\Delta_j' \frac{\partial p_{\Delta\tilde{\Pi}}}{\partial \Delta_i'} - \Delta_i' \frac{\partial p_{\Delta\tilde{\Pi}}}{\partial \Delta_j'} \right)$$

Putting it together

The computation boils down to finding derivatives of the pdf p_{π_i} with respect to Δ_i' and Λ_{ii}. In our implementation, we approximate the derivatives by differentiating the coefficients of the

[7]Differentiating a function $f : \mathbb{R}^n \to R$ is often simplified if the derivation is carried out in a convenient basis. The derivative in the direction of $\boldsymbol{v}$ at $\boldsymbol{x}$ satisfies

$$\nabla_{\boldsymbol{v}} f = \boldsymbol{v}^T \nabla f(\boldsymbol{x})$$

where $\nabla f(\boldsymbol{x})$ is the gradient of f at $\boldsymbol{x}$. Hence, if $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ is a basis for $\mathbb{R}^n$, then the gradient can be computed from the derivatives $\nabla_{\boldsymbol{b}_i} f$ by

$$\nabla f(\boldsymbol{x}) = B^{-1} \left[\nabla_{\boldsymbol{b}_i} f\right]$$

where the rows of the matrix B are the basis vectors.

Fast convolution method for value-at-risk gradient

Input: A $\Delta\Gamma$-approximation. A time series of daily returns. A confidence level $0 < \alpha < 1$. A number of grid points N and a bound for the grid interval a.

Output: The value-at-risk for one day with confidence level α. The gradient of value-at-risk for the parameters of the $\Delta\Gamma$-approximation.

- Compute value-at-risk with Algorithm 2.
- Compute the discretized partial derivative for p_{π_i}

$$\frac{\partial p^D_{\pi_i}}{\partial \Delta'_i} = h \sum_{j=0}^{N-1} q_i^j \delta_j .$$

- Convolve the functions with the FFT

$$\left(\frac{\partial p_{\Delta\tilde{\Pi}}}{\partial \Delta'_i}\right)^D = h \sum_{j=0}^{N-1} r_j \delta_j \quad \text{where} \quad R^j = (-1)^{j(n-1)} h^{n-1} Q_i^j \prod_{k \neq i} P_k^j$$

- Integrate over $[-a, -\mathrm{VaR}]$ by adding the coefficients $\{r_j\}_{j=0}^{N-1}$ and linear interpolation. Let I be this approximation.
- Compute $p_{\Delta\tilde{\Pi}}(-\mathrm{VaR})$ with linear interpolation for $p^D_{\Delta\tilde{\Pi}}$.
- Set

$$\left[\frac{\partial \mathrm{VaR}}{\partial \Delta'_i}\right]_i = \frac{1}{p_{\Delta\tilde{\Pi}}(-\mathrm{VaR})} I$$

- Repeat from step 2 for each Δ'_i and for Λ_{ii}.
- Repeat for a grid with step size $2h$. Extrapolate both value-at-risk and the gradients to get more accurate answers.
- Perform the change of coordinates using (4.16) and (4.17).

Algorithm 3: The fast convolution method for the gradient of value-at-risk.

discretized pdf $p^D_{\Pi_i}$, i.e., by differentiating (4.5)-(4.7) with respect to Δ'_i and Λ_{ii}. This procedure gives two discretized functions:

$$\frac{\partial p^D_{\pi_i}}{\partial \Delta'_i} = h \sum_{j=0}^{N-1} q^j_i \delta_j \quad \text{and} \quad \frac{\partial p^D_{\pi_i}}{\partial \Lambda_{ii}} = h \sum_{j=0}^{N-1} r^j_i \delta_j .$$

Hence, we can compute discrete approximations to the partial derivatives with the DFT. The algorithm is identical to the corresponding step (4.8) in the value-at-risk algorithm. Finally, the integrals, (4.13) and (4.14), are approximated by summing the coefficients and with linear interpolation over the final interval. The integration step is similar to (4.11). The complete computational procedure is summarized in Algorithm 3.

4.5 Example: sensitivity analysis and linear approximation

Consider a portfolio with two European call options on each of BCE and Canadian Tire. Both options are at-the-money and mature in 3 months. Let θ_1 and θ_2 be the number of contracts held for each of the two options. When priced using the Black-Scholes model, the value of the portfolio is

$$\Pi = 3.23 \cdot \theta_1 + 1.39 \cdot \theta_2 .$$

The $\Delta\Gamma$-approximation is given by

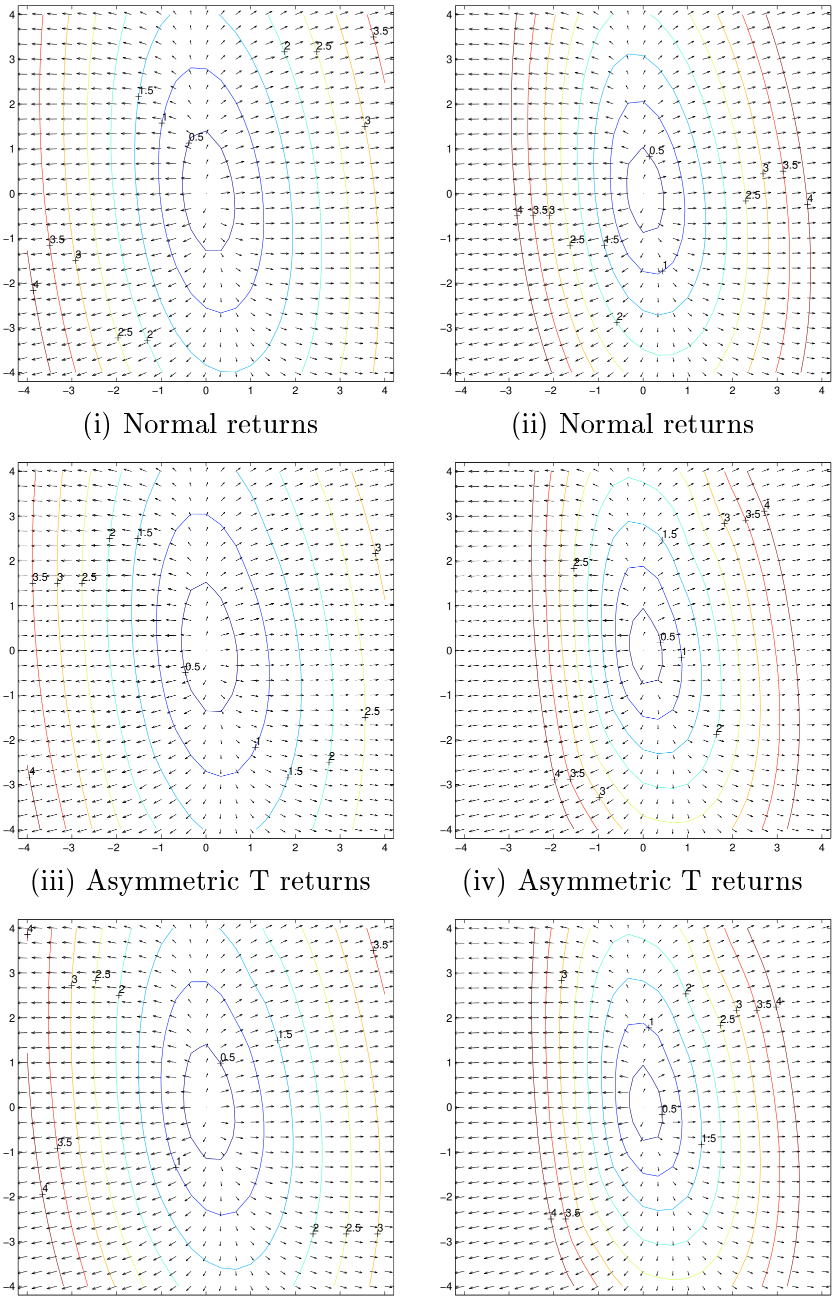
$$\begin{aligned} \Theta &= -0.0277 \cdot \theta_1 - 0.0119 \cdot \theta_2 \\ \Delta &= \begin{bmatrix} 24.3797 \cdot \theta_1 \\ 10.4725 \cdot \theta_2 \end{bmatrix} \\ \Gamma &= \begin{bmatrix} 99.0967 \cdot \theta_1 & 0 \\ 0 & 42.6138 \cdot \theta_2 \end{bmatrix} \end{aligned}$$

As an example of a gradient calculation, we computed value-at-risk for a 25×25 grid of (θ_1, θ_2) portfolios. To understand the sensitivities of value-at-risk to the portfolio parameters, we also computed the gradient at each grid point. Figure 4.8 shows the computed gradient field superimposed over the level curves of value-at-risk. The vectors show the direction of largest sensitivity to changes in the portfolio. The six graphs show the results for the 95% and 99% value-at-risk and for the three return models introduced in Chapter 2. From the figure we make two observations. First, for all plots the computed vector field agrees with the level sets and shows that changes in θ_1, the position in the BCE option, has the most impact on the risk. Second, similar to the previous example, the asymmetric T and Parzen models for return give a larger estimate of the risk compared to the normal model.

Third, the level sets for the Parzen model are less smooth, but the computed gradient still seem to agree quite well with the macro scale of change for the function. For value-at-risk simulations the lack of smoothness of computed value-at-risk is a minor concern, but when applied to optimization problems this is a serious short coming.

Consider portfolios where $\theta_2 = 1$ is fixed and θ_1 varies, i.e, the portfolios on a horizontal line in each of the graphs in Figure 4.8. For value-at-risk as a function of θ_1, the derivative of value-at-risk gives a linear approximation. Figure 4.9 shows the value-at-risk and the linear approximation computed for $\theta_1 = -0.2$. In all six cases, the linear approximations accurately describe the local dynamics of the value-at-risk. Neglecting to differentiate the market model does not lead to an inaccurate derivative.

95% **value-at-risk** 99% **value-at-risk**



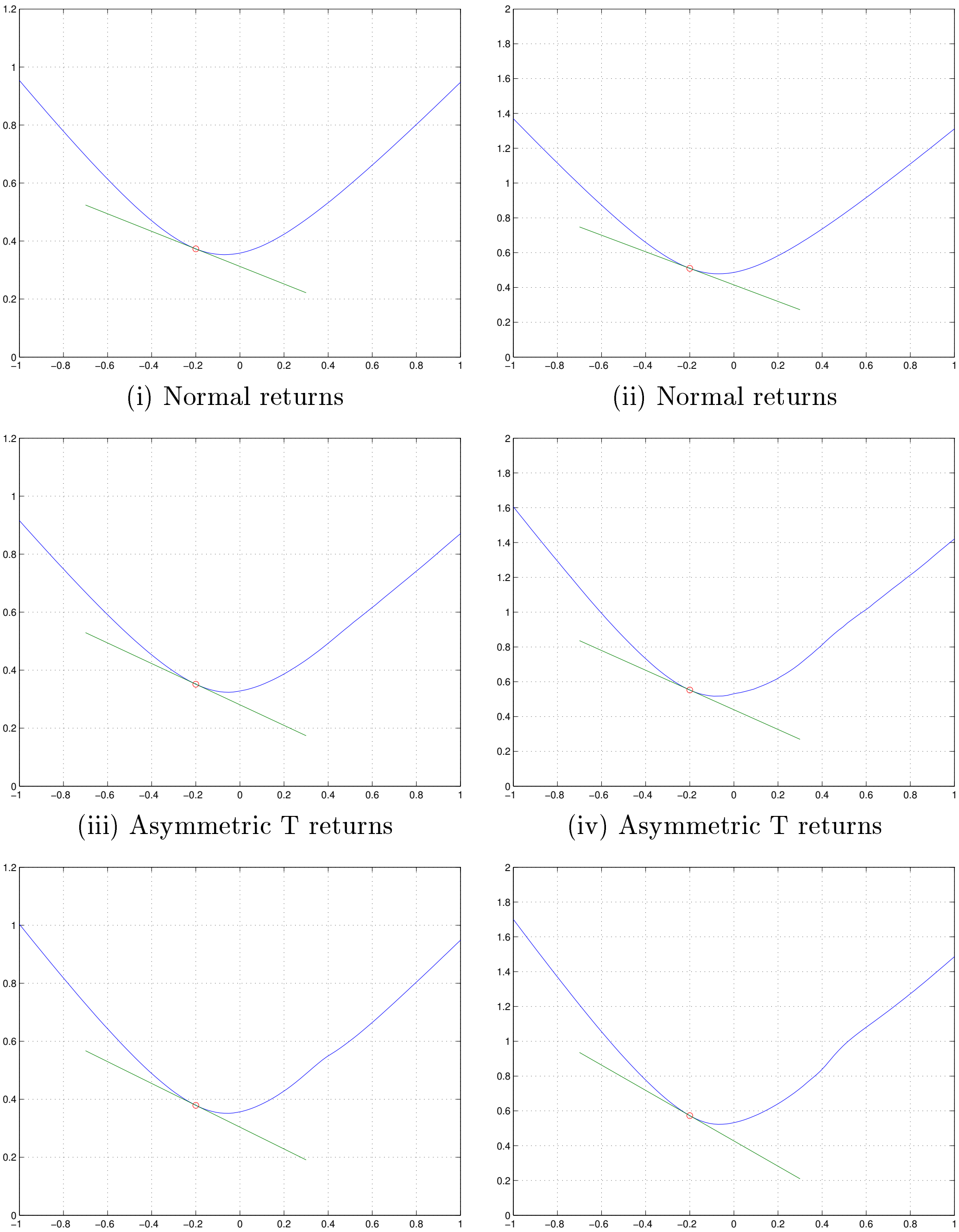
(i) Normal returns (ii) Normal returns

(iii) Asymmetric T returns (iv) Asymmetric T returns

(v) Parzen estimate for returns (vi) Parzen estimate for returns

Figure 4.8: Level sets for value-at-risk and the gradient field for value-at-risk. The horizontal axis is θ_1, the position in the BCE call option, and the vertical axis is θ_2, the position in the Canadian Tire option.

95% value-at-risk **99% value-at-risk**



(i) Normal returns
(ii) Normal returns
(iii) Asymmetric T returns
(iv) Asymmetric T returns
(v) Parzen estimate for returns
(vi) Parzen estimate for returns

Figure 4.9: Linear approximations to the 95% and 99% value-at-risk as a function of θ_1, the position in the call option on BCE. The linear approximation is computed for $\theta_1 = -0.2$.

4.6 Example: hedging with value-at-risk

We conclude this chapter with an optimization example. Similar to the previous examples we consider a portfolio with a long position in one European call option each on BCE and Canadian Tire. The two options both mature in 3 months and are at-the-money. The Toronto35 is an index with 35 stocks trading on the Toronto Stock Exchange. Extend the model with two more securities: the index itself and an at-the-money call option on the index. Since the Toronto35 includes both BCE and Canadian Tire and there is significant positive correlation of the returns, it is possible to find a position that reduces the value-at-risk of the original portfolio. Let θ_1 and θ_2 be the number of index and call options on the index in the portfolio. The $\Delta\Gamma$-approximation for this portfolio is given by

$$\Theta = 0.0396 - 0.2354 \cdot \theta_2$$

$$\Delta = \begin{bmatrix} -24.3797 \\ -10.4725 \\ 564.75 \cdot \theta_1 + 321.71 \cdot \theta_2 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} -99.0967 & 0 & 0 \\ 0 & -42.6138 & 0 \\ 0 & 0 & 2233.2 \cdot \theta_2 \end{bmatrix}$$

This leads to the following optimization problem

$$\min_{\theta_1, \theta_2} \mathrm{VaR}(\theta_1, \theta_2).$$

The value-at-risk surface has a flat fold, see Figure 4.10. The portfolios along the fold all have approximately the same value-at-risk and corresponds to portfolios where Δ_3 is constant. To study the performance of the fast convolution method when used in combination with optimization software, we computed the minimum for each of the three return models.[8] The computed solutions are marked in the plots in Figure 4.10. From studying the surface and inspecting the iterations, we note that the computed solutions are different and that the solutions for all three problems are very close to degenerate. Hence, hedging the portfolio using both index and index options does not lead to any significant reduction in the value-at-risk compared to using the index alone

4.6.1 Adding stochastic volatility

The picture changes when the model is extended by making volatility a risk factor. We chose to consider a simple extension. Suppose that the volatility for all three risk factors is the same, i.e, the changes in volatilities satisfy

$$\frac{\Delta\sigma^1}{\sigma_0^1} = \frac{\Delta\sigma^2}{\sigma_0^2} = \frac{\Delta\sigma^3}{\sigma_0^3}$$

and we use the Black-Scholes equation to extract a $\Delta\Gamma$-approximation. Although this model is too simple to be of use in practise, it captures the main property of a stochastic volatility model, it introduces a risk factor that can only be hedged using the option. We therefore believe that the qualitative properties of the example are correct, and for the purpose of exploring value-at-risk optimization it is therefore an appropriate model problem.

[8] We used the quasi-Newton method `fminunc` in Matlab's optimization toolbox.

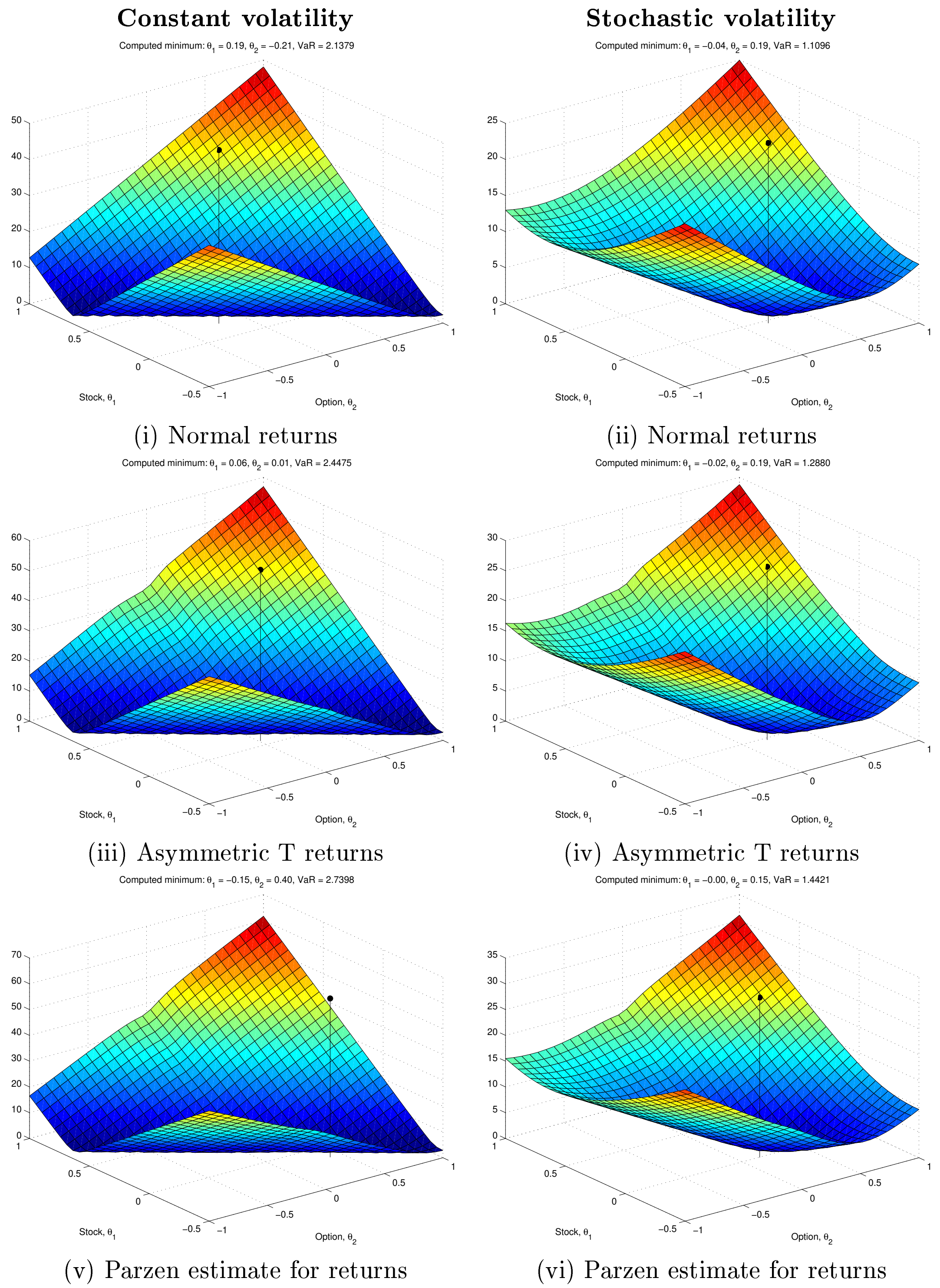


Figure 4.10: The 99% value-at-risk surface as a function of θ_1, the position in Toronto35, and θ_2, the position in call options on Toronto35. The point marked is the computed solution to the optimization problem. Volatility as a risk factor changes the shape of the value-at-risk surface and makes the minimum well defined.

If we include first order volatility risk this gives a new $\Delta\Gamma$-approximation for the portfolio:

$$\Theta = 0.0396 - 0.2354 \cdot \theta_2$$

$$\Delta = \begin{bmatrix} -24.3797 \\ -10.4725 \\ 564.75 \cdot \theta_1 + 321.71 \cdot \theta_2 \\ -4.1963 + 22.0409 \cdot \theta_2 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} -99.0967 & 0 & 0 & 0 \\ 0 & -42.6138 & 0 & 0 \\ 0 & 0 & 2233.2 \cdot \theta_2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

In Figure 4.10, we see that the value-at-risk surfaces for the stochastic volatility model is different from the model without volatility risk. In the stochastic volatility model, the degenerate minimum is replaced by a well defined minimum. Furthermore, the portfolio with optimal value-at-risk combines a position in the index and the index option to reduce the risk, and using only the index would give a less efficient hedge. Finally, the minima computed using optimization software are close to each other for all three return models.

From the example we can draw several interesting conclusions. First, the example shows that value-at-risk leads to non-trivial optimization problems. The fast convolution method provides an efficient basis for solving hedging problems with value-at-risk as the risk measure. Second, the shape of the value-at-risk surface changes when volatility risk is included in the model. In the stochastic volatility model options are important to reduce the value-at-risk. Third, our experience is that optimization with the Parzen model is less reliable than the other two return models. The reason is that the surface for the Parzen model has small scale variations caused by the non-parameteric density estimator. When it is used in an optimization algorithm, the lack of smoothness makes the use of finite difference approximations to the derivatives less stable. To some extent this can be handled by varying the tolerances used in these computations, but it does not change the fact that the the steps and stopping criteria are less reliable and special care must be taken to check the accuracy of the computed solution.

Chapter 5

Dimension reduction

For many portfolios, a few main directions determine the dominant behavior of the dynamics of the portfolio value. This is a combined effect of correlation of risk factor returns and the quantities of each security held in the portfolio. Therefore, it is natural to search for a simpler approximation depending on fewer factors that is close to the original model. In this chapter, we develop two portfolio dependent methods for dimension reduction (see also [2]).

This problem has been examined by other authors, e.g., Hull [33] shows how to use principal component analysis in an interest rate model; Kreinin et al. [36] present a principal component based method for linear portfolios with normally distributed risk factors; and Reimers and Zerbs [49] develop a reduction method where asset blocks are represented by their dominant principal components and cross block covariances by the covariance for the largest principal component of each block. Both methods that we have developed take a $\Delta\Gamma$-approximation and compute a new approximation of lower dimensionality which is close to the original function. The chapter concludes with two examples. In the first example, we compare the performance of the methods for a sample portfolio. We conclude that the method based on mean square error is more accurate, easier to implement and slightly more efficient. In the second example, we apply the dimension reduction method to an optimization problem. This experiment shows that dimension reduction can be effective in reducing computation time while maintaining good accuracy. Parts of these results have appeared in Albanese, Jackson and Wiberg [2], but the numerical experiments presented here are more extensive and the conclusions about non-normal models are new.

5.1 The dimension reduction problem

A $\Delta\Gamma$-approximation can be written as

$$\Delta\widetilde{\Pi} = \Xi + x^T\Delta' + \frac{1}{2}x^T\Lambda x = \Xi + \sum_{i=1}^{n}\big(\Delta_i' x_i + \frac{\lambda_i}{2}x_i^2\big). \tag{5.1}$$

The vector x is an affine transformation of the original risk factor returns r (see Section 3.4). The vector of risk factor returns R is a random variable and the the transformed random variable X satisfies

$$\mathbb{E}[X] = 0 \quad \text{and} \quad \mathbb{E}[XX^T] = I.$$

If n is large, the time to compute value-at-risk is large. The objective in dimension reduction is to find a $\Delta\Gamma$-approximation $\Delta\widetilde{\Pi}_1$ that captures the main properties of $\Delta\widetilde{\Pi}$, but $\Delta\widetilde{\Pi}_1$ depends on

$k \ll n$ dimensions. Successful dimension reduction reduces the time to compute value-at-risk. This is particularly important for problems where value-at-risk must be computed repeatedly, for example, to solve optimization problems.

The strategy that we propose is to restrict $\Delta\widetilde{\Pi}$ to a subspace such that $\Delta\widetilde{\Pi}$ and the restriction $\Delta\widetilde{\Pi}_1$ are close. Let $\mathcal{P}$ be a projection[1] onto the subspace spanned by the orthonormal columns of an $n \times k$ matrix Q_1, and let $\mathcal{P}^{\perp}$ be the projection onto the complementary subspace spanned by the orthonormal columns of Q_2. Let $z = (z_1, z_2)$ be defined by

$$x = \mathcal{P}x + \mathcal{P}^{\perp}x = Q_1 z_1 + Q_2 z_2. \tag{5.2}$$

Based on this factorization of the risk factor space, the we conclude that the reduced approximation has the form

$$\Delta\widetilde{\Pi}(x) \approx \Delta\widetilde{\Pi}(\mathcal{P}x) = \Delta\widetilde{\Pi}_1(z_1). \tag{5.3}$$

In the sections below, we present two methods for finding a projection so that $\Delta\widetilde{\Pi}$ and $\Delta\widetilde{\Pi}_1$ are close. The methods are motivated by two different views of the meaning of closeness. The first method solves the problem by finding a lower dimension approximation with a small mean square error. The second method uses the identification of quadratic forms with matrices and solves the problem, after re-scaling the variables, by finding a lower rank matrix close to the original $\Delta\Gamma$-approximation.

5.2 Method 1: reduction with small mean square error

In Method 1, we find $\Delta\widetilde{\Pi}_1$ in (5.3) by insisting that the mean square error $\mathbb{E}[(\Delta\widetilde{\Pi} - \Delta\widetilde{\Pi}_1)^2]$ is small. To motivate the algorithm for dimension reduction, we need the following lemma:

Lemma 5.1. *Let A be an $n \times n$ matrix. Suppose that*

$$\max_{\|a\|_2=1} \max_{\|b\|_2=1} \mathbb{E}[(a^T x)^2 (b^T x)^2] \leq \beta.$$

Then,

$$\mathbb{E}[(x^T A x)^2] \leq n\beta \|A\|_F^2.$$

Proof. From the Cauchy-Schwartz inequality, it follows that

$$\mathbb{E}[(x^T A x)^2] \leq \mathbb{E}[(x^T x)(x^T A^T A x)].$$

Since $A^T A$ is symmetric, there is an orthogonal matrix Q such that $A^T A = Q^T \Sigma^2 Q$ where Σ is diagonal. Hence,

$$\mathbb{E}[(x^T x)(x^T A^T A x)] = \mathbb{E}[(x^T Q^T Q x)(x^T Q^T \Sigma^2 Q x)].$$

Define $y_i = q_i^T x$ where q_i^T is the ith row vector of Q. By assumption, we have $\mathbb{E}[y_i^2 y_j^2] \leq \beta$ for all i and j, and it follows that

$$\begin{aligned}
\mathbb{E}[(y^T y)(y^T \Sigma^2 y)] &= \mathbb{E}\left[(\sum_k y_k^2)(\sum_l \sigma_l^2 y_l^2) \right] \\
&= \sum_l \sigma_l^2 \sum_k \mathbb{E}[y_k^2 y_l^2] \\
&\leq n\beta \sum \sigma_j^2.
\end{aligned}$$

Since $\sum \sigma_j^2 = \|A\|_F^2$ the lemma follows. □

[1] A *projection* is a Hermitian matrix $\mathcal{P}$ such that $\mathcal{P}^2 = \mathcal{P}$.

Suppose that the 4th order moments in the lemma are bounded, and that we have an estimate β for this bound. Consider a tolerance $\epsilon > 0$. By re-ordering the components of x, we can order the eigenvalues, the diagonal elements of Λ, such that $|\lambda_i| \geq |\lambda_{i+1}|$ for all i. Let k be the smallest index for which

$$\sum_{i=k+1}^{n} \lambda_k^2 \leq \epsilon. \tag{5.4}$$

Partition $x = (x_1, x_2)$ with x_1 containing the first k components. We can then write (5.1) as

$$\begin{aligned} \Delta\widetilde{\Pi} = \Xi + & \left[x_1^T, 0\right] \begin{bmatrix} \Delta_1' \\ 0 \end{bmatrix} + \frac{1}{2}\left[x_1^T, 0\right] \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ 0 \end{bmatrix} \\ & + \left[0, x_2^T\right] \begin{bmatrix} 0 \\ \Delta_2' \end{bmatrix} + \frac{1}{2}\left[0, x_2^T\right] \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \begin{bmatrix} 0 \\ x_2 \end{bmatrix}. \end{aligned}$$

The contribution to the mean square error from Λ_2 is small relative to ϵ, but the effect on the gain from Δ_2' may still be large. With a simple trick we can keep all the information in Δ_2'. Let $V = [v_1, V_2]$ be an orthogonal matrix with the first column $v_1 = \Delta_2' / \|\Delta_2'\|_2$. Since Δ_2' is orthogonal to the column vectors in V_2, we define the $(k+1)$-vector z_1 and the $(n-k-1)$-vector z_2 in (5.2) by

$$z_1 = \begin{bmatrix} x_1 \\ v_1^T x_2 \end{bmatrix} \quad \text{and} \quad z_2 = V_2^T x_2.$$

Hence, the reduced $\Delta\Gamma$-approximation is

$$\Delta\widetilde{\Pi}_1(z_1) = \Xi + z_1^T \begin{bmatrix} \Delta_1' \\ \|\Delta_2'\|_2 \end{bmatrix} + \frac{1}{2} z_1^T \begin{bmatrix} \Lambda_1 & 0 \\ 0 & v_1^T \Lambda_2 v_1 \end{bmatrix} z_1.$$

This dimension reduction method is summarized in Algorithm 4.

To relate the reduced $\Delta\Gamma$-approximation to the mean square error, we observe that the residual is a pure quadratic term:

$$\Delta\widetilde{\Pi} - \Delta\widetilde{\Pi}_1 = \frac{1}{2}[z_1^T, z_2^T] \left[\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & 0 & v_1^T \Lambda_2 V_2 \\ \hline 0 & V_2^T \Lambda_2 v_1 & V_2^T \Lambda_2 V_2 \end{array}\right] \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}.$$

It is easy to show that

$$\left\| \begin{bmatrix} 0 & v_1^T \Lambda_2 V_2 \\ V_2^T \Lambda_2 v_1 & V_2^T \Lambda_2 V_2 \end{bmatrix} \right\|_F \leq \left\| V^T \Lambda_2 V \right\|_F.$$

By the unitary invariance of the Frobenius norm and the criterion (5.4), we can apply the lemma to prove the bound for the mean square error summarized in the following theorem.

Theorem 5.1. *Given $\epsilon > 0$ and $\beta > 0$ such that*

$$\max_{\|a\|_2=1} \max_{\|b\|_2=1} \mathbb{E}[(a^T x)^2 (b^T x)^2] \leq \beta.$$

Then the $\Delta\Gamma$-approximation $\Delta\widetilde{\Pi}_1$ produced by Algorithm 4 satisfies

$$\mathbb{E}[(\Delta\widetilde{\Pi} - \Delta\widetilde{\Pi}_1)^2] \leq \frac{\beta \epsilon n}{4}.$$

Dimension reduction, Method 1

Input: A $\Delta\Gamma$-approximation. A time series $\{r_i\}_{i=1}^d$ of daily returns. A tolerance $\epsilon > 0$.

Output: Parameter estimates for the market model. A factorization of the reduced $\Delta\Gamma$-approximation.

- Compute estimates for the mean $\widehat{\mu}$ and the covariance matrix $\widehat{C}$.
- Solve the eigenvalue problem

$$\begin{cases} \Lambda = A^T \Gamma A, \\ \widehat{C} = AA^T. \end{cases}$$

  Order the eigenvalues so that $|\lambda_i| \geq |\lambda_{i+1}|$.
- Compute

$$\begin{aligned} \Xi &= \Theta t + \widehat{\mu}^T \Delta + \frac{1}{2} \widehat{\mu}^T \Gamma \widehat{\mu}, \\ \Delta' &= A^T (\Delta + \Gamma \widehat{\mu}). \end{aligned}$$

- Find the smallest integer k such that

$$\sum_{i=k+1}^{n} \lambda_k^2 \leq \epsilon.$$

- Compute

$$\begin{aligned} \Delta_1 &= [\Delta(1\colon k), \|\Delta(k+1\colon n)\|_2], \\ v_1 &= \Delta(k+1\colon n)/\|\Delta(k+1\colon n)\|_2, \\ \Lambda_1 &= \begin{bmatrix} \Lambda(1\colon k, 1\colon k) & 0 \\ 0 & v_1^T \Lambda(k+1\colon n, k+1\colon n) v_1 \end{bmatrix}. \end{aligned}$$

- For each Z_i, estimate the remaining parameters, from the transformed returns $\{A^{-1}(r_i - \widehat{\mu})\}_{i=1}^d$ and $\{v_1^T A^{-1}(r_i - \widehat{\mu})\}_{i=1}^d$, using the methods in Chapter 2.

Algorithm 4: Dimension reduction with Method 1. The algorithm combines parameter estimation and factorization with dimension reduction and replaces Algorithm 1 in this case.

5.3 Method 2: reduction by low-rank approximation

The random variables X in (5.1) have zero mean, are uncorrelated, and have variance one. Therefore, the impact of each random variable is approximately the same. The $\Delta\Gamma$-approximation can be written as a quadratic form[2]

$$\Delta\widetilde{\Pi} = \Xi + \frac{1}{2}[x^T, 1] \begin{bmatrix} \Lambda & \Delta' \\ (\Delta')^T & 0 \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}.$$

Therefore, we can define the distance between two $\Delta\Gamma$-approximations, with the same constant term Ξ, as

$$\|\Delta\widetilde{\Pi} - \Delta\widetilde{\Pi}_1\| = \left\| \begin{bmatrix} \Lambda & \Delta' \\ (\Delta')^T & 0 \end{bmatrix} - \begin{bmatrix} \Lambda_1 & \Delta_1' \\ (\Delta_1')^T & \Psi \end{bmatrix} \right\|_2. \tag{5.5}$$

In Method 2, this is the definition of closeness that we use for dimension reduction.

Consider the Schur decomposition for $\Delta\widetilde{\Pi}$,

$$\begin{bmatrix} \Lambda & \Delta' \\ (\Delta')^T & 0 \end{bmatrix} = U\Sigma U^T,$$

where U is an orthogonal matrix and Σ is a diagonal matrix. The diagonal elements of Σ are the eigenvalues and we may assume that they are ordered in decreasing absolute value $|\sigma_1| \geq \ldots \geq |\sigma_{N+1}|$. Let

$$\Sigma_k = \mathrm{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0).$$

The Schmidt-Mirsky theorem says that $U\Sigma_k U^T$ solves the minimization problem

$$\min_{\mathrm{rank}(B)=k} \|\Delta\widetilde{\Pi} - B\|_2 = \|\Delta\widetilde{\Pi} - U\Sigma_k U^T\|_2 = |\sigma_{k+1}|$$

(for a proof see for example [56, p.208] or [28, p.71]).

Given a tolerance $\epsilon > 0$, the Schmidt-Mirsky gives a tool to find the optimal function $\Delta\widetilde{\Pi}_1$. Let k be the smallest k such that

$$|\sigma_{k+1}| \leq \epsilon.$$

Then,

$$\|\Delta\widetilde{\Pi} - U\Sigma_k U^T\|_2 \leq \epsilon.$$

The terms of the reduced function $\Delta\widetilde{\Pi}_1$ can be computed from the matrix $U\Sigma_k U^T$. Partition the orthogonal matrix U as

$$U = \begin{bmatrix} U_{11} & U_{12} \\ u_{21}^T & u_{22}^T \end{bmatrix} = \left[\begin{array}{ccc|ccc} u_{11} & \cdots & u_{1k} & u_{1k+1} & \cdots & u_{1n+1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ u_{n1} & \cdots & u_{nk} & u_{nk+1} & \cdots & u_{nn+1} \\ \hline u_{n+11} & \cdots & u_{n+1k} & u_{n+1k+1} & \cdots & u_{n+1n+1} \end{array} \right].$$

The matrix Q_1 in (5.2) can be taken as the $n \times k$ matrix with orthonormal columns in the QR factorization $Q_1 R = U_{11}$. For this choice, we get the reduced $\Delta\Gamma$-approximation

$$\Delta\widetilde{\Pi}_1 = \Xi + u_{21}^T \Sigma_k u_{21} + z_1^T R \Sigma_k u_{21} + \frac{1}{2} z_1^T R \Sigma_k R^T z_1,$$

[2]In practice, it is not necessary to transform the problem all the way to the form (5.1). It is sufficient to compute the Cholesky factor of the covariance matrix and transform the $\Delta\Gamma$-approximation accordingly. This way the number of eigenvalue problems solved in Algorithm 5 can be reduced from three to two.

which can be factorized again. The steps of Method 2 are summarized in Algorithm 5.

In some special cases, the matrix U_{11} may be rank deficient.[3] This is not a serious problem since it is easy to show that this leads to an approximation of size $(k-1) \times (k-1)$—it is a lucky break that earns us the additional reduction of one dimension.

5.4 Absolute versus relative value-at-risk

Instead of computing the value-at-risk directly from the value-at-risk equation (1.1) with $\Delta\Pi$ replaced by $\Delta\widetilde{\Pi}_1$, we have found that it is more accurate to compute the *relative value-at-risk* defined by

$$\mathrm{VaR}_{rel}(\Delta\widetilde{\Pi}) = \mathrm{VaR}(\Delta\widetilde{\Pi}) - \mathbb{E}[\Delta\widetilde{\Pi}].$$

Instead of computing the value-at-risk from the reduced approximation as

$$\mathrm{VaR}(\Delta\widetilde{\Pi}) \approx \mathrm{VaR}(\Delta\widetilde{\Pi}_1), \tag{5.6}$$

we use the formula

$$\mathrm{VaR}(\Delta\widetilde{\Pi}) \approx \mathrm{VaR}_{rel}(\Delta\widetilde{\Pi}_1) + \mathbb{E}[\Delta\widetilde{\Pi}]. \tag{5.7}$$

The motivation for this choice becomes more clear when we consider the errors in the two formulas. The error for (5.6) is

$$\mathrm{VaR}(\Delta\widetilde{\Pi}) - \mathrm{VaR}(\Delta\widetilde{\Pi}_1) = \mathrm{VaR}_{rel}(\Delta\widetilde{\Pi}) - \mathrm{VaR}_{rel}(\Delta\widetilde{\Pi}_1) + \mathbb{E}[\Delta\widetilde{\Pi} - \Delta\widetilde{\Pi}_1], \tag{5.8}$$

and for (5.7) it is

$$\mathrm{VaR}(\Delta\widetilde{\Pi}) - \mathrm{VaR}(\Delta\widetilde{\Pi}_1) - \mathbb{E}[\Delta\widetilde{\Pi}] = \mathrm{VaR}_{rel}(\Delta\widetilde{\Pi}) - \mathrm{VaR}_{rel}(\Delta\widetilde{\Pi}_1).$$

The extra term $\mathbb{E}[\Delta\widetilde{\Pi} - \Delta\widetilde{\Pi}_1]$ in (5.8) can be large while the difference between the relative value-at-risk term still is small. Since computing the expectation for the full portfolio is easy, separating the terms, and using the reduction $\Delta\widetilde{\Pi}_1$ only to compute the relative value-at-risk, is both easy to do and leads to better accuracy.

5.5 Which method is better?

The ideas used to develop the two methods for dimension reduction are very different. Method 1 has a direct connection to probability theory, it finds a reduced model with small mean square error. Method 2 is based on linear algebra argument, computing reduced model using a low rank approximation of the matrix for the quadratic form. Method 1 has several advantages over Method 2. First, Method 1 is easier to implement and slightly more efficient. Second, in Method 1 the structure of the $\Delta\Gamma$-approximation is preserved in the new quadratic form, except for the final risk factor which captures the residual linear term. Finally, the numerical experiment in the section below shows that Method 1 gives a more accurate reduced model in practice.

[3] Consider for example

$$\begin{bmatrix} 1 & 0 & \sqrt{2} \\ 0 & 0.1 & 0 \\ \sqrt{2} & 0 & 0 \end{bmatrix} = \left[\begin{array}{cc|c} \frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & 1 \\ \hline \frac{1}{\sqrt{3}} & -\frac{\sqrt{2}}{\sqrt{3}} & 0 \end{array}\right] \left[\begin{array}{cc|c} 2 & 0 & 0 \\ 0 & -1 & 0 \\ \hline 0 & 0 & -0.1 \end{array}\right] \left[\begin{array}{cc|c} \frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & 1 \\ \hline \frac{1}{\sqrt{3}} & -\frac{\sqrt{2}}{\sqrt{3}} & 0 \end{array}\right]^T.$$

Dimension reduction, Method 2

Input: A $\Delta\Gamma$-approximation. A time series $\{r_i\}_{i=1}^{d}$ of daily returns. A tolerance $\epsilon > 0$.

Output: Parameter estimates for the market model. A factorization of the reduced $\Delta\Gamma$-approximation.

- Compute estimates for the mean $\widehat{\mu}$ and the covariance matrix $\widehat{C}$.
- Solve the eigenvalue problem

$$\begin{cases} \Lambda = A^T \Gamma A, \\ \widehat{C} = A A^T. \end{cases}$$

- Compute

$$\begin{aligned} \Xi &= \Theta t + \widehat{\mu}^T \Delta + \frac{1}{2}\widehat{\mu}^T \Gamma \widehat{\mu}, \\ \Delta' &= A^T(\Delta + \Gamma\widehat{\mu}). \end{aligned}$$

- Compute the Schur decomposition

$$\begin{bmatrix} \Lambda & \Delta' \\ (\Delta')^T & 0 \end{bmatrix} = U \Sigma U^T.$$

  Order the eigenvalues so that $|\sigma_i| \geq |\sigma_{i+1}|$.
- Find the smallest k such that

$$|\sigma_{k+1}| \leq \epsilon.$$

- Compute the QR factorization

$$Q_1 R = U_{11}.$$

- Compute

$$\begin{aligned} \Gamma_1 &= R \Sigma_k R^T, \\ \Delta_1 &= R \Sigma_k U(n, 1\!:\!k), \\ \Xi_1 &= \Xi + u_{21}^T \Sigma_k u_{21}. \end{aligned}$$

- Factorize $\Delta\widetilde{\Pi}_1$ and estimate the remaining parameters from the transformed returns, using the methods in Chapter 2.

Algorithm 5: Dimension reduction with Method 2. The algorithm combines parameter estimation and factorization with dimension reduction and replaces Algorithm 1 in this case.

5.6 Example: a comparative experiment

In this section, we present the first of two computational examples. The purpose of the first example is to study the performance of dimension reduction and to compare the results from the two methods.

The portfolios in our previous examples have few dimensions. The advantage of dimension reduction is to reduce computation time for portfolios with many risk factors. So, we consider a portfolio with options on each of the stocks in the Toronto35 index. The returns for the 35 stocks have significant correlation and we expect that dimension reduction will produce accurate simulation results for relatively few dimensions. The portfolio consists of short positions with one call option and one put option on each of the stocks. As before, the options are European, at-the-money and have 3 months to maturity. Furthermore, we include one independent risk factor for changes in volatility which is shared by all options, see Section 4.6 for details. Experiments with a similar problem, without volatility risk factor and with normal returns, is reported on by Albanese, Jackson and Wiberg in [2].

To study the effect of dimension reduction, we compute the daily value-at-risk with dimensions $k = 1, \dots, 36$ (see Figure 5.1). Each graph show the 99% value-at-risk and 95% value-at-risk. There is one graph for each of Method 1 and Method 2 applied to the three risk factor models from Chapter 2. Our expectation, that a few dimensions characterize the risk of the portfolio, is confirmed by the simulations.

Figure 5.2 shows the relative errors for the simulations in Figure 5.1. The figures lead to several interesting observations. Dimension reduction produces results that resemble the result for the full model, but the quality of the result differs for Method 1 and Method 2. In general, the result for the 95% value-at-risk seems to be better than for the 99% value-at-risk. In the case of normal returns, the error is very small for both methods (see graph (i) and (ii)). The error for Method 1 is small for both the asymmetric T and the Parzen models (see graph (iii) and (v)). Method 2 does not produce accurate results for the last two return models (see graph (iv) and (vi)), and more seriously it is not clear that the results improve when more dimensions are included.

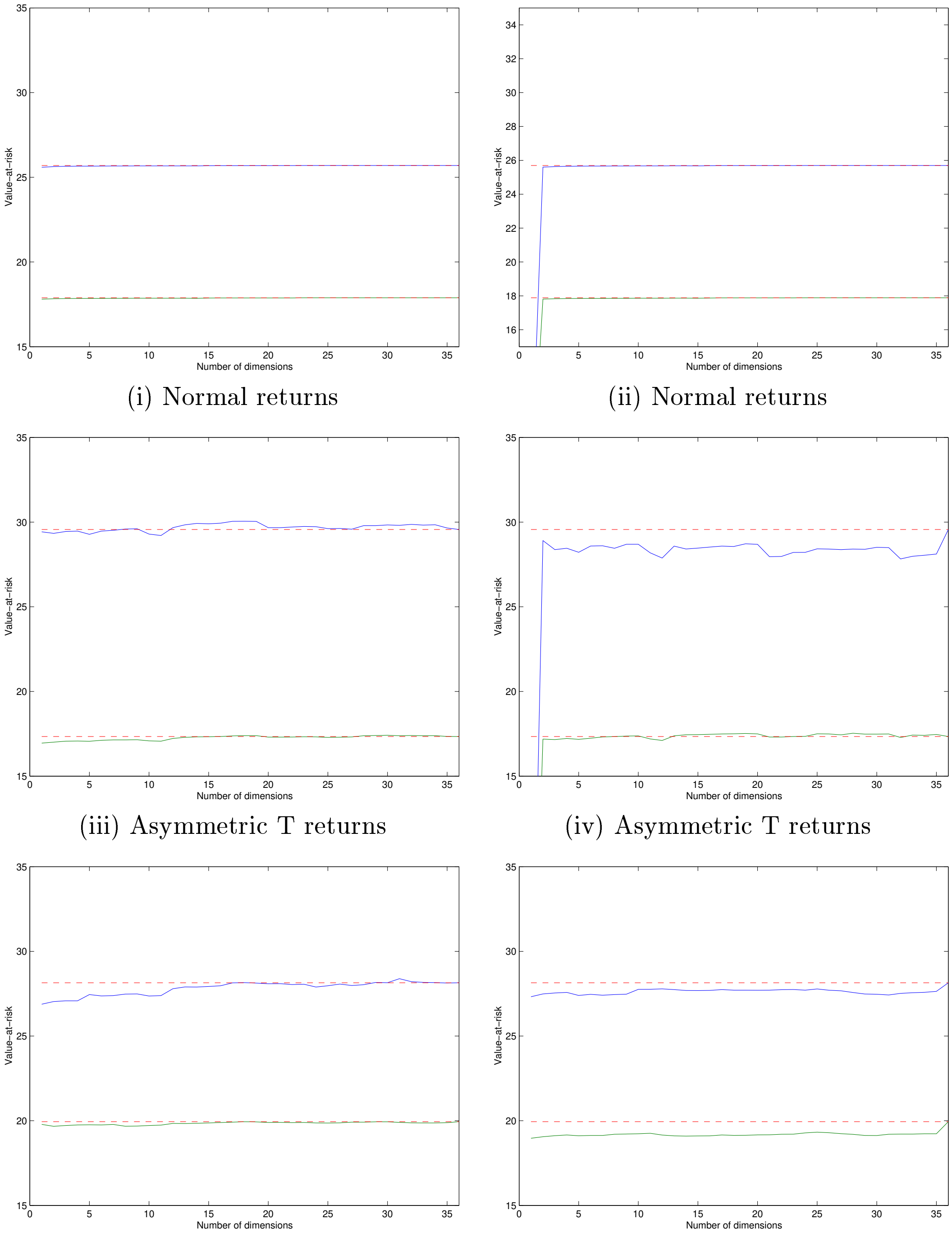
To try to understand the failure of Method 2, we examined the intermediate results produced during the execution. We believe that the problem arises from a large Δ component that is split over two dimensions in the reduced model, one with positive and one with negative curvature. In the normal model, where the model is characterized by the mean and variance of the returns, this does not lead to a deterioration of the simulation result. In the other two models, the estimation error destroys the balance between the two components and this corrupts the result. This is a serious drawback for Method 2 and one that is not shared by Method 1.

To conclude, the value-at-risk computed with Method 1 has a small relative error. Although the error does not decrease monotonically, the trend is clear—more dimension give more accurate results. As demonstrated by this example, Method 1 produces better results than Method 2.

5.7 Example: dimension reduction and optimization

We conclude with an optimization example similar to the one in Section 4.6. The example shows that dimension reduction leads to significant savings in computation time and that the accuracy is preserved despite the reduction. Since Method 1 has clear advantages over Method 2, we restrict our attention to the first method.

Method 1 **Method 2**



(i) Normal returns

(ii) Normal returns

(iii) Asymmetric T returns

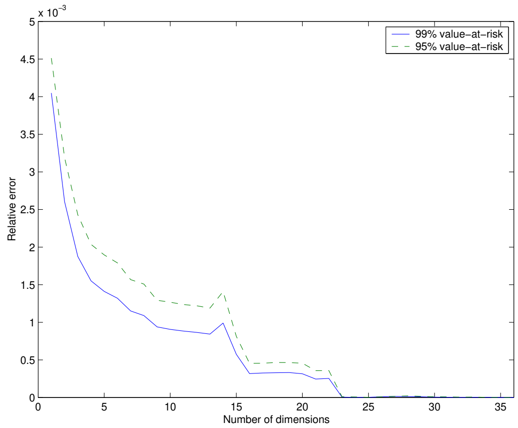
(iv) Asymmetric T returns

(v) Parzen estimate for returns

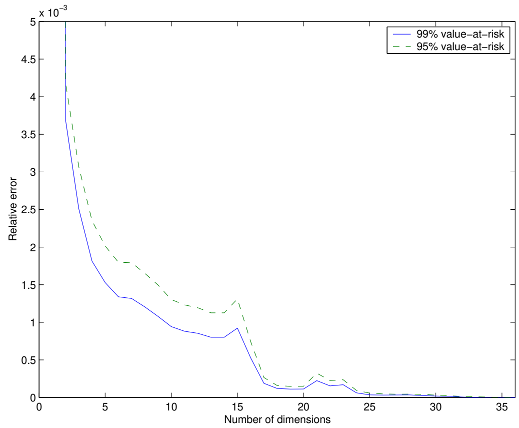
(vi) Parzen estimate for returns

Figure 5.1: Value-at-risk with dimension reduction. The graphs show the 95% and 99% value-at-risk as a function of the number of dimensions used in the simulation.

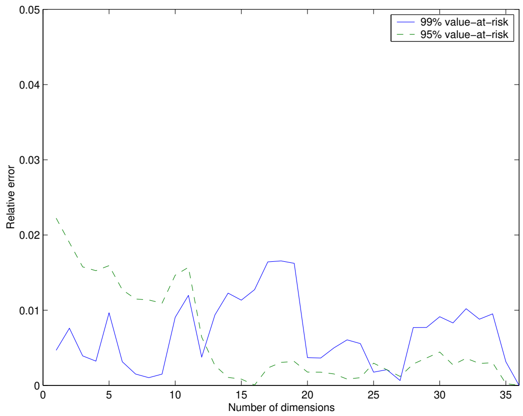
Method 1 **Method 2**



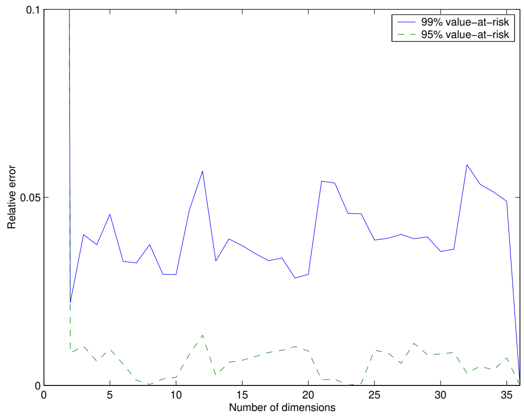
(i) Normal returns



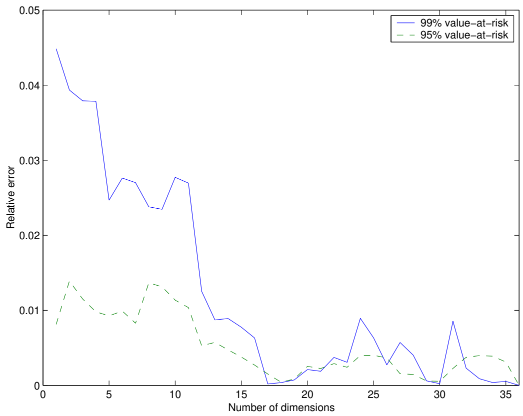
(ii) Normal returns



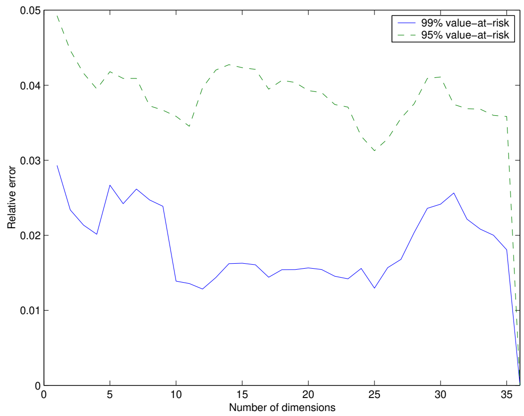
(iii) Asymmetric T returns



(iv) Asymmetric T returns

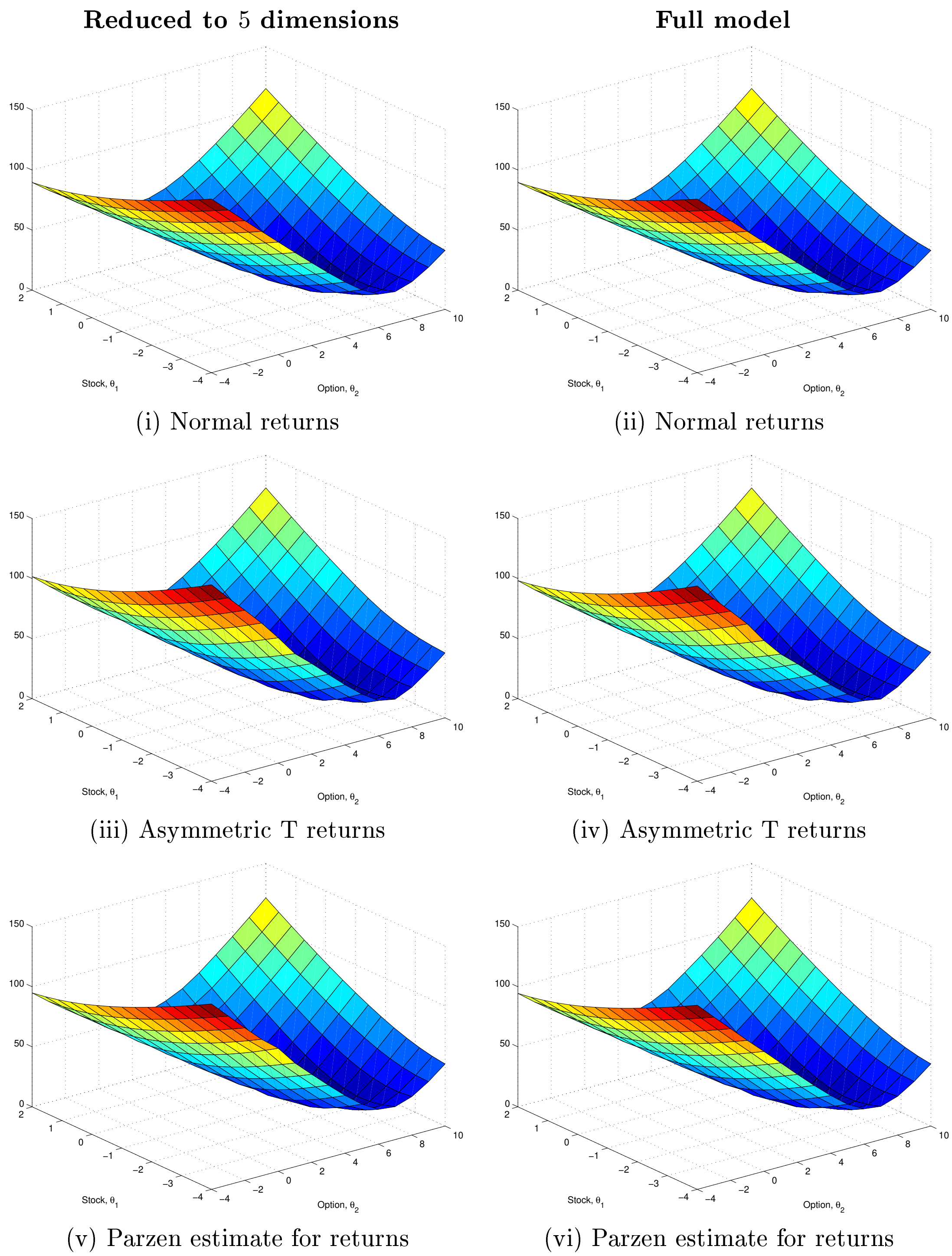


(v) Parzen estimate for returns



(vi) Parzen estimate for returns

Figure 5.2: Relative errors for the simulations in Figure 5.1. Observe that the scales for graph (iii) and (iv) are different.



(i) Normal returns (ii) Normal returns

(iii) Asymmetric T returns (iv) Asymmetric T returns

(v) Parzen estimate for returns (vi) Parzen estimate for returns

Figure 5.3: The 99% value-at-risk surfaces computed with dimension reduction to 5 dimensions and for the full model.

Consider a portfolio with a short position in one call option on each of the stocks in the Toronto35 index. All options are European, at-the-money and have 3 months to maturity. Suppose that we want to minimize the value-at-risk by buying or selling a combination of a linear position in the index, such as the index itself or a future, and a position in a call option on the index. Let θ_1 and θ_2 be the number of index units and call options in the portfolio. To see how dimension reduction affects the shape of the value-at-risk surface, we computed the 99% value-at-risk for the full portfolio and for a reduced model with 5 dimensions. As Figure 5.3 shows, the differences between the surfaces are small. Of course, we are therefore led to believe that the same applies to the optimization problem

$$\min_{\theta_1,\theta_2} \mathrm{VaR}(\theta_1, \theta_2).$$

Table 5.1 show the computed solutions to the optimization problem and statistics about time and number of value-at-risk computations required by the numerical procedure.[4] The variations in value-at-risk at the computed minima are small for the normal and asymmetric T models, and the changes in the location of the minima are relatively small. Similar to the observations in Section 4.6, we note that the performance of the minimization procedure is less reliable for the Parzen model, the denisity estimator is introduces small fluctuations in the value-at-risk. In all three cases, we see that dimension reduction is very effective in reducing the computation time per function evaluation. The total time for the optimization is also reduced. Although the total time is important, it is a not a good indicator since it mostly depends on the success of the stopping criteria used by the optimization algorithm.

[4]We used the quasi-Newton method `fminunc` in Matlab's optimization toolbox.

# Dimensions	θ_1	θ_2	VaR	Time	Func. calls	Time/func. call
1	-1.89	5.31	11.93	8.70	101	0.0861
2	-1.87	5.30	12.15	12.87	129	0.0998
3	-1.87	5.31	12.16	14.79	129	0.1147
4	-1.88	5.32	12.16	15.53	122	0.1273
5	-1.87	5.31	12.16	18.54	130	0.1426
6	-1.87	5.31	12.16	13.41	84	0.1596
8	-1.87	5.29	12.16	24.50	127	0.1929
10	-1.86	5.27	12.17	29.60	131	0.2260
12	-1.89	5.34	12.16	30.40	118	0.2576
17	-1.86	5.27	12.16	33.06	102	0.3241
22	-1.86	5.29	12.17	31.18	79	0.3947
27	-1.92	5.28	12.16	38.70	83	0.4663
32	-1.82	5.19	12.21	50.85	95	0.5353
37	-1.86	5.27	12.17	77.33	127	0.6089

(i) Normal returns

# Dimensions	θ_1	θ_2	VaR	Time	Func. calls	Time/func. call
1	-1.91	5.46	13.15	18.83	85	0.2215
2	-1.91	5.43	13.30	29.10	102	0.2853
3	-1.91	5.40	13.55	36.36	100	0.3636
4	-1.82	5.26	13.61	55.51	128	0.4337
5	-1.80	5.23	13.62	41.45	80	0.5181
6	-1.87	5.30	13.56	65.75	111	0.5923
8	-1.81	5.28	13.59	29.70	39	0.7615
10	-1.81	5.23	13.58	114.8	127	0.9038
12	-1.83	5.31	13.57	146.5	142	1.0316
17	-1.86	5.28	13.57	115.3	83	1.3889
22	-1.80	5.22	13.60	151.0	87	1.7356
27	-1.81	5.23	13.62	167.2	80	2.0905
32	-1.80	5.22	13.60	316.5	129	2.4537
37	-1.84	5.28	13.60	252.9	89	2.8412

(ii) Asymmetrical T returns

# Dimensions	θ_1	θ_2	VaR	Time	Func. calls	Time/func. call
1	-2.05	5.26	14.21	33.50	128	0.2617
2	-1.85	5.27	14.51	50.69	110	0.4608
3	-1.86	5.32	14.70	90.81	140	0.6486
4	-1.82	5.18	13.01	69.43	82	0.8467
5	-1.91	5.27	13.03	73.63	70	1.0519
6	-2.00	5.00	13.32	153.0	122	1.2543
8	-1.88	4.96	13.28	153.4	93	1.6495
10	-1.86	5.16	12.98	108.9	53	2.0555
12	-1.86	5.15	13.00	365.4	150	2.4357
17	-1.83	5.16	13.11	210.1	61	3.4434
22	-1.64	4.98	13.34	725.2	165	4.3950
27	-1.77	5.10	13.22	749.9	139	5.3947
32	-1.85	5.18	12.89	562.6	88	6.3928
37	-1.89	5.27	12.90	601.8	81	7.4291

(iii) Parzen estimate for returns

Table 5.1: The computed solutions to the optimization problem. The quoted times are in seconds of CPU time.

Chapter 6

Perturbation theory

The experiments in Chapter 4 show that the different models for risk factor returns can lead to large differences in the estimate for value-at-risk. Similar observations have been made by several other authors, see Beder [7] and Jorion [34]. In this chapter, we derive a perturbation result that describes how value-at-risk changes with perturbations to the model for risk factor returns. It shows that value-at-risk become increasingly sensitive as the confidence level increases, the tail is more sensitive than the center of the distribution. We present a computable condition number for value-at-risk and illustrate the theory with a numerical example.

6.1 When is value-at-risk well-posed?

Value-at-risk is defined as the solution to the non-linear equation (1.1). If $p(r)$ is the pdf for the risk factor returns then (1.1) is equivalent to the integral equation

$$\int_{\Delta\Pi(r)\leq -\mathrm{VaR}} p(r)dr = 1-\alpha. \tag{6.1}$$

The domain of integration is a set in the n-dimensional space of risk factor returns. The coarea formula [1] gives

$$\int_{-\infty}^{-\mathrm{VaR}} p_{\Delta\Pi}(\rho)d\rho = 1-\alpha. \tag{6.2}$$

where

$$p_{\Delta\Pi}(\rho) = \int_{\{\Delta\Pi(r)=\rho\}} \frac{p(r)}{|D\Delta\Pi(r)|^{1/2}}dA \tag{6.3}$$

and

$$|D\Delta\Pi(r)| = \sum_{i=1}^{n}\left(\frac{\partial\Delta\Pi}{\partial r_i}\right)^2.$$

So, the coarea formula transforms the integral (6.1) into a 1-dimensional integral (6.2) over a new pdf (6.3). In other words, (6.2) is an integral over the pdf for $\Delta\Pi$, which is defined by the surface integral over the level sets of $\Delta\Pi$. The equations (6.1) and (6.2) are equivalent, but we find the second form more convenient in our analysis.

Hadamard's classical definition says that a problem is well-posed if it has a unique solution that depends continuously on the initial data. The properties of the pdf (6.3) determine if

[1] We refer to Evans and Gariepy [19] for a proof and discussion of the coarea formula. The formula can be applied assuming that $\Delta\Pi$ is Lipschitz, differentiable and ess inf $|Df|>0$.

value-at-risk is a well-posed problem. The first condition, existence of a solution, holds without additional assumptions. The cdf

$$\mathrm{P}[\Delta\Pi \leq x] = \int_{-\infty}^{x} p_{\Delta\Pi}(\rho)d\rho \tag{6.4}$$

has the range $(0, 1)$. Hence, since the cdf is a continuous function, the intermediate value theorem implies that (6.2) has a solution for all $\alpha \in (0, 1)$.

For uniqueness and continuity to hold, (6.4) must be strictly increasing in a neighborhood around the solution VaR. Equivalently, it is unique if (6.3) is positive almost everywhere in a neighborhood of VaR. Uniqueness follows by observing that, if there were to be two solutions $\mathrm{VaR}_1 > \mathrm{VaR}_2$, then since (6.4) is an increasing function any x with $-\mathrm{VaR}_1 < x < -\mathrm{VaR}_2$ must be a solution, too. Hence, (6.4) is not strictly increasing, or equivalently the pdf is not positive almost everywhere, in any neighborhood of a solution.

Continuity, the third condition for (6.2) to be well-posed, requires that the solution depends continuously on the data. We show that value-at-risk is continuous for changes in the pdf $p(r)$. Suppose that (6.3) is positive almost everywhere in some interval $(-\mathrm{VaR} - \epsilon, -\mathrm{VaR} + \epsilon)$. Let $\{p_i\}_{i=1}^{\infty}$ be a sequence of pdfs that converges to p, that is $\|p - p_i\|_1 \to 0$ as $i \to \infty$. Moreover, let VaR_i be the solutions to (6.1) corresponding to p_i. Since

$$0 = \int_{-\infty}^{-\mathrm{VaR}} d\rho \int_{\{\Delta\Pi(r)=\rho\}} \frac{p(r)}{|D\Delta\Pi|^{1/2}} dA - \int_{-\infty}^{-\mathrm{VaR}_i} d\rho \int_{\{\Delta\Pi(r)=\rho\}} \frac{p_i(r)}{|D\Delta\Pi|^{1/2}} dA,$$

we get

$$\left| \int_{-\mathrm{VaR}_i}^{-\mathrm{VaR}} p_{\Delta\Pi}(\rho)d\rho \right| \leq \int_{\Delta\Pi(r) \leq -\mathrm{VaR}_i} |p(r) - p_i| dr \leq \|p - p_i\|_1.$$

Furthermore, since $p_{\Delta\Pi}(\rho)$ is positive almost everywhere in a neighborhood of $-\mathrm{VaR}$ and the left hand size of the inequality goes to zero, we must have $\mathrm{VaR}_i \to \mathrm{VaR}$. So, value-at-risk is continuous with respect to the return model p.

Hence, value-at-risk is a well-posed problem, given that the cdf (6.4) is increasing close to the solution. This condition holds for the market risk models we have considered in this thesis. However, for credit risk models, this condition is often violated and value-at-risk is a questionable measure of risk. For a detailed discussion and an axiomatic system of desirable properties of general risk measures, see Artzner et al. [3]. In the analysis above, we examined perturbations of the model for risk factor returns. In the future, it would be interesting to extend the analysis to perturbations of $\Delta\Pi$.

6.2 Perturbations of the return model

The analysis above shows that value-at-risk, defined by (6.2), is well-posed, if the cdf is strictly increasing close to the solution. We argued that value-at-risk is continuous for changes in $p(r)$, but the analysis does not indicate the size of the resulting perturbations. In this section, we quantify the change in value-at-risk for a perturbation of the pdf $p(r)$.

6.2.1 Proof of a first-order perturbation property

We shall now derive a variational, first order perturbation property for value-at-risk. Consider a pdf p such that

$$p_{\Delta\Pi}(\rho) = \int_{\{\Delta\Pi(r)=\rho\}} \frac{p(r)}{|D\Delta\Pi(r)|^{1/2}} dA \tag{6.5}$$

is differentiable. The set of probability density functions is the subset of functions in L^1 that integrate to one.[2] Furthermore, if p and q are pdfs then $hp + (1-h)q$ is a pdf for all $h \in [0,1]$, i.e., the pdfs are a convex set. Consider a variation v where

$$v_{\Delta\Pi}(\rho) = \int_{\{\Delta\Pi(r)=\rho\}} \frac{v(r)}{|D\Delta\Pi(r)|^{1/2}} dA \tag{6.6}$$

is continuous. In addition, for v to be an admissible variation, the function $p + hv$ must be a pdf for all h in some interval $[0,\epsilon)$.

For an admissible variation, the value-at-risk is a function of h that satisfies the equation

$$\int_{-\infty}^{-\mathrm{VaR}(h)} d\rho \int_{\{\Delta\Pi(r)=\rho\}} \frac{p(r) + hv(r)}{|D\Delta\Pi(r)|^{1/2}} dA = 1 - \alpha.$$

The function on the left side,

$$F(x,h) = \int_{-\infty}^{x} d\rho \int_{\{\Delta\Pi(r)=\rho\}} \frac{p(r) + hv(r)}{|D\Delta\Pi(r)|^{1/2}} dA,$$

has the first partial derivatives

$$\frac{\partial F}{\partial x} = \int_{\{\Delta\Pi(r)=x\}} \frac{p(r) + hv(r)}{|D\Delta\Pi(r)|^{1/2}} dA \tag{6.7}$$

and

$$\frac{\partial F}{\partial h} = \int_{-\infty}^{x} d\rho \int_{\{\Delta\Pi(r)=\rho\}} \frac{v(r)}{|D\Delta\Pi(r)|^{1/2}} dA.$$

Since (6.5) is continuous, value-at-risk is well-posed for $h = 0$ if and only if (6.7) is positive for $h = 0$,

$$\left.\frac{\partial F}{\partial x}\right|_{h=0} = p_{\Delta\Pi}(\rho) > 0.$$

Assuming that value-at-risk is well-posed, the implicit function theorem guarantees that the solution $\mathrm{VaR}(h)$ to

$$F(-\mathrm{VaR}(h), h) = 1 - \alpha$$

is continuously differentiable for h in some interval $[0,\delta)$ where $\delta \leq \epsilon$ (see Rudin [53]). The derivative of $\mathrm{VaR}(h)$ at $h = 0$ is

$$\mathrm{VaR}'(0) = \frac{1}{p_{\Delta\Pi}(-\mathrm{VaR})} \int_{\Delta\Pi(r)\leq -\mathrm{VaR}} v(r)dr. \tag{6.8}$$

In the words of variational calculus, $\mathrm{VaR}'(0)$ is the Gateaux variation in the direction of v. Taylor's theorem gives the linear approximation

$$\mathrm{VaR}(h) = \mathrm{VaR}(0) + \mathrm{VaR}'(0)h + O(h^2). \tag{6.9}$$

[2]A function f is in the function space L^1 if it is measurable and has finite L^1 norm, i.e., if

$$\|f\|_1 = \int_{\mathbb{R}^n} |f(x)| dx \leq \infty.$$

6.2.2 Error bounds and the condition number

Taking the absolute value of the linear approximation (6.9), we get an estimate of the absolute error

$$|\mathrm{VaR}(h) - \mathrm{VaR}(0)| \leq |\mathrm{VaR}'(0)| \cdot |h| + O(|h|^2)$$

Note that the derivative $\mathrm{VaR}'(0)$ depends on the variation v.[3] So, in general different variations give different derivatives, and we write $\mathrm{VaR}'_v(0)$ for the Gateaux variation in the direction v to emphasize this dependency. Also, since $\mathrm{VaR}_v(0)$ is independent of v, we write VaR instead.

There are many possible metrics that could be used to measure the distance between two pdfs. We choose to consider the metric induced by the L^1-norm,

$$d(p,q) = \|p - q\|_1,$$

mostly because it leads to an elegant result. We may without loss of generality assume that the variations have unit length, since we may rescale v and h simultaneously to achieve this. Remember that the admissible variations satisfy the constraint that $p + hv$ is a pdf for all h in some interval $[0, \epsilon)$. Since

$$\int_{\mathbb{R}^n} p(r) + hv(r)dr = 1, \qquad h \in [0, \epsilon),$$

it follows that

$$\int_{\mathbb{R}^n} v(r)dr = 0. \tag{6.10}$$

The function $v = v_+ + v_-$ can be separated in its positive v_+ and negative v_- parts. Since v has unit length and satisfies (6.10), we get

$$\int_{\mathbb{R}^n} v_+(r)dr = -\int_{\mathbb{R}^n} v_-(r)dr = \frac{1}{2}. \tag{6.11}$$

The property (6.11) can be used to bound the integral term in (6.8),

$$\left| \int_{\Delta\Pi(r) \leq -\mathrm{VaR}} v(r)dr \right| \leq \frac{1}{2}.$$

Therefore, the derivative is bounded by

$$\left|\mathrm{VaR}'_v(0)\right| \leq \frac{1}{2p_{\Delta\Pi}(-\mathrm{VaR})}$$

for all admissible variations v with unit length. The absolute error is bounded by

$$|\mathrm{VaR}_v(h) - \mathrm{VaR}| \leq \frac{|h|}{2p_{\Delta\Pi}(-\mathrm{VaR})} + O(|h|^2), \tag{6.12}$$

and the relative error is bounded by

$$\frac{|\mathrm{VaR}_v(h) - \mathrm{VaR}|}{\mathrm{VaR}} \leq \frac{|h|}{2\mathrm{VaR}p_{\Delta\Pi}(-\mathrm{VaR})} + O(|h|^2). \tag{6.13}$$

[3]Similarly, the constant in the asymptotic term $O(h^2)$ depends on v. Recall that $O(h^2)$ denotes a function $f(h)$ such that

$$\lim_{h\to 0} \left|\frac{f(h)}{h^2}\right| = C < \infty.$$

For our error bound, C depends on the variation v.

The condition number of value-at-risk is[4]

$$\kappa = \frac{1}{2\mathrm{VaR}p_{\Delta\Pi}(-\mathrm{VaR})}. \tag{6.14}$$

The relative error bound (6.13) provides the first-order error estimate

$$\frac{|\mathrm{VaR}_v(h) - \mathrm{VaR}|}{\mathrm{VaR}} \lesssim \kappa \|p - \tilde{p}\|_1 \tag{6.15}$$

where p is the original pdf, $\tilde{p} = p + hv$ is a perturbed pdf and $v = \frac{p-\tilde{p}}{h}$ is the direction of the perturbation.

6.3 Example: mixture model

We consider two portfolios that depend on a single risk factor, the stock price for BCE. The first portfolio consists of a single stock. The second portfolio is a short position in a European call option. The option is hedged with a position in the stock, i.e, there is a stock position that has been chosen so that the Δ is zero. The option is at-the-money and has three months to maturity. The returns are assumed to be normal,

$$p(r) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(r-\mu)^2}{2\sigma^2}}$$

Although the theory is valid for a large class of return models, we chose this example for its simplicity and for accurate computations.

Figure 6.1 shows four plots with the result of our experiment. There is one plot for each portfolio and for the value-at-risk with $\alpha = 95\%$ and $\alpha = 99\%$. The continuous line shows (6.15) as a function of the size of the perturbation $\|p - \tilde{p}\|_1$. We see that the relative error in value-at-risk grows rapidly for small errors in p. For the stock portfolio, κ can be computed directly. For the option portfolio, we used a Monte Carlo method to compute value-at-risk and a Parzen estimator for $p_{\Delta\Pi}(-\mathrm{VaR})$. The Monte Carlo used $20,000$ random normal samples, and variance reduction with antithetic variables. We computed the option price for each sample with the Black-Scholes formula.

To compare (6.15) and actual relative errors for perturbations of the model, we generated random densities of the form

$$\tilde{p} = H_1 p_1 + H_2 p_2 + (1 - H_1 - H_2)p$$

where p_1 and p_2 are normal pdfs. The random variables H_1 and H_2 are uniform with ranges $[0, 0.1]$ and $[0, 0.01]$, respectively. The parameters of p_1 and p_2 were also generated at random.

[4]Note that this is a variational condition number. The standard condition number (see Rice [50]) for the problem $y = f(x)$ is defined as

$$\kappa = \frac{xf'(x)}{f(x)}.$$

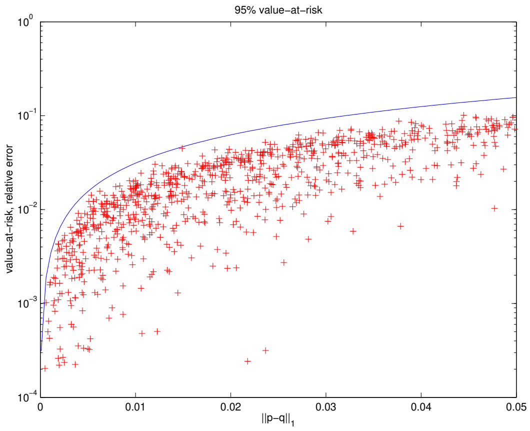
This can be interpreted in a directional sense. Consider $f : \mathbb{R}^n \to R$. The absolute error is

$$f(x + \Delta x) - f(x) \approx \nabla f(x) \cdot \Delta x.$$

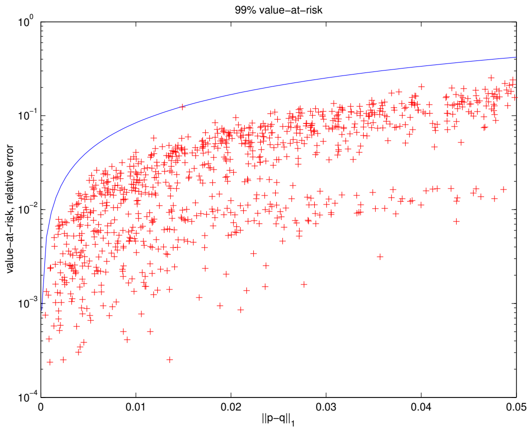
in directional sense

$$f(x + \Delta x) - f(x) \approx \left[\frac{d}{d\epsilon} f(x + \epsilon \Delta x)\right]_{\epsilon=0}$$

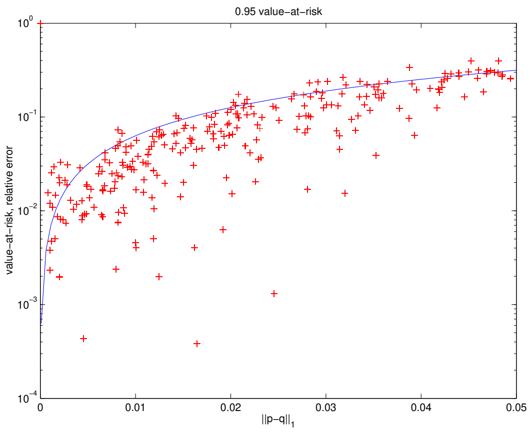
95% value-at-risk **99% value-at-risk**



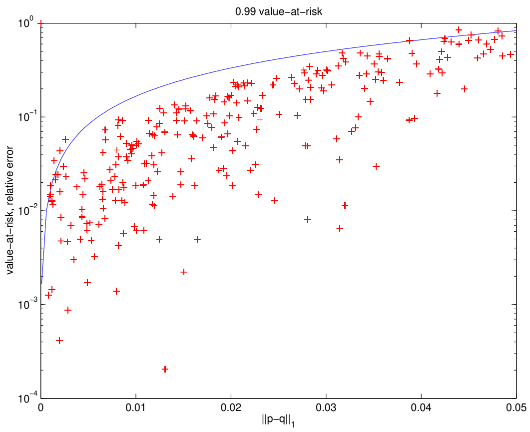
(i) Stock



(ii) Stock



(iii) Call option



(iv) Call option

Figure 6.1: The plots show the relative error in value-at-risk versus the size of the perturbation $\|p - \tilde{p}\|_1$. The continuous line is the error bound (6.15). The plus signs corresponed to random perturbations.

The parameters were generated as

$$\mu_1 = \mu(1 + 2M_1), \qquad \sigma_1 = \sigma|0.5 + 2U_1|$$
$$\mu_2 = \mu(1 + 10M_2), \qquad \sigma_2 = \sigma|0.5 + U_2|$$

where M_1 and M_2 are standard normal random variables, and U_1 and U_2 are standard uniform random variables.

For each random mixture, we computed the relative error in value-at-risk and the size of the perturbation, $\|p - \tilde{p}\|_1$. Each plus sign in the plots in Figure 6.1 marks the result for a randomly perturbed problem. For the stock portfolio, we used 1000 randomly perturbed densities. The relative error in value-at-risk is indeed smaller than the first order, worst case estimate (6.15). For the option portfolio, we computed the value-at-risk and the norm $\|p - \tilde{p}\|_1$ with the Monte Carlo method. Since this procedure is much more time consuming than for the stock portfolio, we had to limit the experiment to 300 random densities. Although some samples are larger than the approximate bound, we see that (6.15) is a good estimate of the relative error. The accuracy of the Monte Carlo method is limited and we see that these plots contain some simulation noise.

6.4 A few final thoughts

In this chapter, we have discussed the properties of the value-at-risk equation (6.2). In our analysis, we argue that in most cases value-at-risk is a well-posed problem. The requirement for well-posedness is that the cdf (6.4) is strictly increasing close to $-\mathrm{VaR}$. An equivalent condition is that (6.3) is positive almost everywhere in a neighborhood of the solution. Credit risk is one important exception where these assumptions will typically not hold, but it is a reasonable assumption for standard value-at-risk models.

Nevertheless, well-posedness alone does not guarantee that a small error in the model for returns translates into a small relative error in value-at-risk. To understand how such errors affect the simulation, we developed a variational perturbation theory. The theory applies to problems that are sufficiently smooth (6.5), and for smooth variations (6.6). The advantage of the variational approach is a theory that is model independent, and it can therefore be used to quantify model risk. The theory provides the estimate (6.15) for the relative error where the condition number can be computed. The stumbling block is to find $p_{\Delta\Pi}(-\mathrm{VaR})$. In some methods, such as the fast convolution method, $p_{\Delta\Pi}(-\mathrm{VaR})$ is computed. In other methods, for example a Monte Carlo method, $p_{\Delta\Pi}(-\mathrm{VaR})$ must be computed with a density estimator, see for example [58].

The condition number (6.14) controls the size of the relative error. The problem becomes increasingly ill-conditioned as $\mathrm{VaR}\,p_{\Delta\Pi}(-\mathrm{VaR})$ decreases to zero. This confirms that the empirical observations of small perturbations to the return model—caused either by changes in the model, in the data or in the estimation procedure—that cause large changes in the simulation result is an intrinsic property for large α. Value-at-risk is ill-conditioned for extreme levels of confidence.

Chapter 7

Conclusions and future work

In this thesis, we have studied the computation of value-at-risk and it contains three main contributions. First, we developed a new Fourier method for computing value-at-risk, the fast convolution method. The innovative part of this method is that it does not assume that the characteristic function for the portfolio returns is known explicitly. The method can be used for non-linear portfolios and for return models with fat tails. Furthermore, we showed that the method can be extended to compute the value-at-risk gradient. Finally, we illustrated the performance of the method with examples and discussed applications to sensitivity analysis and optimization. Second, we developed two dimension reduction algorithms. We concluded that the method based on reduction with small mean square error was the better of the two methods. The examples illustrated that dimension reduction may reduce computation time while retaining good accuracy for the value-at-risk simulation. Third, we developed a variational perturbation theory for value-at-risk. It provides first-order estimates for the absolute and relative errors caused by perturbations of the return model. We also derived a condition number and showed that value-at-risk becomes increasingly ill-conditioned for large confidence levels.

In the future we would like to extend the perturbation theory to errors in the pricing model. It would be interesting to analyze competing measures of risk, such as conditional value-at-risk, to compare their stability properties.

Bibliography

[1] C. Albanese, K. Jackson, and P. Wiberg. Hedging with value-at-risk. Technical report, University of Toronto, 2001.

[2] C. Albanese, K. Jackson, and P. Wiberg. Dimension reduction in the computation of value-at-risk. *To appear in the Journal of Risk Finance*, 2002.

[3] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

[4] Basel Committee on Banking Supervision. International convergence of capital measurement and capital standards. Technical report, Bank for International Settlements, 1988.

[5] Basel Committee on Banking Supervision. Amendment to the Capital Accord to incorporate market risks. Technical report, Bank for International Settlements, 1996.

[6] Basel Committee on Banking Supervision. Overview of the amendment to the Capital Accord to incorporate market risks. Technical report, Bank for International Settlements, 1996.

[7] T.S. Beder. VAR: Seductive but dangerous. *Financial Analysts Journal*, 51(5):12–24, 1995.

[8] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, May-June 1973.

[9] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[10] W.L. Briggs and V.E. Henson. *The DFT: An Owner's Manual for the Discrete Fourier Transform.* SIAM, 1995.

[11] M. Britten-Jones and S.M. Schaefer. Non-linear value-at-risk. *European Finance Review*, 2(2):161–187, 1998.

[12] J.Y. Campbell, A.W. Lo, and A.C. MacKinlay. *The Econometrics of Financial Markets.* Princeton University Press, 1997.

[13] D. Duffie. *Dynamic Asset Pricing Theory.* Princeton University Press, 2001.

[14] D. Duffie and J. Pan. An overview of value-at-risk. *Journal of Derivatives*, 4(3):7–49, 1997.

[15] D. Duffie and J. Pan. Analytic value-at-risk with jumps and credit risk. *Finance and Stochastics*, 5(2):155–180, 2001.

[16] L. El-Jahel, W. Perraudin, and P. Sellin. Value at risk for derivatives. *Journal of Derivatives*, 6(3):7–26, 1999.

[17] R.J. Elliott and P.E. Kopp. *Mathematics of Financial Markets*. Springer-Verlag, 1999.

[18] R.F Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, July 1982.

[19] L.C. Evans and R.F. Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, 1992.

[20] E.F. Fama. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2):383–417, May 1970.

[21] E.F. Fama. Efficient capital markets ii. *Journal of Finance*, 46(5):1575–1617, December 1991.

[22] A. Feuerverger. A consistent test for bivariate dependence. *International Statistical Review*, 61(3):419–433, 1993.

[23] A. Feuerverger and A.C.M. Wong. Computation of value-at-risk for non-linear portfolios. *Journal of Risk*, 3(3):37–55, Fall 2000.

[24] C. Gasquet and P. Witomski. *Fourier Analysis and Applications: Filtering, Numerical Computation, Wavelets*. Springer-Verlag, 1998.

[25] G.D. Gay, J. Kim, and J. Nam. The case of the SK Securities and J.P. Morgan swap: Lessons in VaR frailty. *Derivatives Quarterly*, 5(3):13–26, 1999.

[26] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Variance reduction techniques for estimating value-at-risk. *Management Science*, 46(10):1349–1364, October 2000.

[27] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance*, 12(3):239–269, July 2002.

[28] G.H. Golub and C.F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, USA, second edition, 1989.

[29] W.G. Hallerbach. Decomposing portfolio value-at-risk: A general analysis. Technical Report TI 99-034/2 / ECFR report 9907, Tinbergen Institute, 1999.

[30] J.M. Harrison and D.M. Kreps. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20:381–408, 1979.

[31] J.M. Harrison and S.R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications*, 11:215–260, 1981.

[32] N.J. Higham. *Accuracy and Stability of Numerical Algorithms*. SIAM, 1996.

[33] J.C. Hull. *Options, Futures and Other Derivatives*. Prentice Hall, 4th edition edition, 2000.

[34] P. Jorion. Risk^2: Measuring the risk in value at risk. *Financial Analysts Journal*, 52(6):47–56, 1996.

[35] I. Karatzas and S.E. Shreve. *Methods of Mathematical Finance*. Springer-Verlag, 1999.

[36] A. Kreinin, L. Merkoulovitch, D. Rosen, and M. Zerbs. Principal component analysis in quasi monte carlo simulation. *Algo Research Quarterly*, 1(2):21–30, December 1998.

[37] R.J. Larsen and M.L. Marx. *An Introduction to Mathematical Statistics and Its Applications*. Prentice-Hall, second edition edition, 1986.

[38] R. Lowenstein. *When Genius Failed: The Rise and Fall of Long-Term Capital Management*. Random House, 2000.

[39] Risk Magazine. A billion dollars just isn't what it used to be: Lessons from two decades in the risk management morgue. *Risk*, pages 64–66, May 1999.

[40] B.G. Malkiel. *A Random Walk Down Wall Street*. W.W. Norton, 2000.

[41] H. Mausser and D. Rosen. Beyond VaR: From measuring risk to managing risk. *Algorithmics Research Quarterly*, 1(2):5–20, December 1998.

[42] J. Mina and A. Ulmer. Delta-Gamma four ways. Technical report, RiskMetrics Group, 1999.

[43] J.P. Morgan/Reuters. *RiskMetrics—Technical Document*. J.P. Morgan, 4th edition edition, 1996.

[44] D.B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–369, March 1991.

[45] B. Øksendal. *Stochastic Differential Equations*. Springer-Verlag, 5th edition, 1998.

[46] E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3):1065–1076, September 1961.

[47] S. Păun. Sensitivities of value at risk for non-linear portfolios. Master's thesis, Department of Mathematics, University of Toronto, Toronto, Canada, 1999.

[48] P. Protter. A partial introduction to financial asset pricing theory. *Stochastic Processes and Their Applications*, 91:169–203, 2001.

[49] M. Reimers and M. Zerbs. Dimension reduction by asset blocks. *Algo Research Quarterly*, 1(2):43–58, December 1998.

[50] J.R. Rice. A theory of condition. *SIAM Journal on Numerical Analysis*, 3(2):287–310, June 1966.

[51] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27(3):832–837, September 1956.

[52] S.M. Ross. *Introduction to Probability Models*. Harcourt Academic Press, 7th edition, 2000.

[53] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 3rd edition, 1976.

[54] F. Stambaugh. Risk and value at risk. *European Management Journal*, 14(6):612–621, 1996.

[55] G.W. Stewart. Perturbation bounds for the QR factorization of a matrix. *SIAM Journal on Numerical Analysis*, 14(3):509–518, June 1977.

[56] G.W. Stewart and J.-G. Sun. *Matrix Perturbation Theory.* Academic Press, London, UK, 1990.

[57] J.-G. Sun. Rounding-error and perturbation bounds for the Cholesky and LDL^T factorizations. *Linear Algebra and Its Applications*, 173:77–97, 1992.

[58] J.R. Thompson and R.A. Tapia. *Nonparametric Function Estimation, Modeling, and Simulation.* SIAM, 1990.

[59] P. Wilmott. *Paul Wilmott on Quantitative Finance.* John Wiley & Sons, 2000.

[60] P. Wilmott, S. Howison, and J. Dewynne. *The Mathematics of Financial Derivatives: A Student Introduction.* Cambridge University Press, 1995.