Expansion Methods Applied to Distributions and Risk Measurement in Financial Markets

Kohei Marumo, BE(*Tokyo*), MSc(*Oxon*)

Thesis submitted for the degree of Doctor of Philosophy

School of Economics and Finance
Queensland University of Technology
GPO Box 2434, Brisbane, QLD 4001
Australia

10 October 2007

To Junko

Abstract

Obtaining the distribution of the profit and loss (PL) of a portfolio is a key problem in market risk measurement. However, existing methods, such as those based on the Normal distribution, and historical simulation methods, which use empirical distribution of risk factors, face difficulties in dealing with at least one of the following three problems: describing the distributional properties of risk factors appropriately (description problem); deriving distributions of risk factors with time horizon longer than one day (time aggregation problem); and deriving the distribution of the PL given the distributional properties of the risk factors (risk aggregation problem).

Here, we show that expansion methods can provide reasonable solutions to all three problems. Expansion methods approximate a probability density function by a sum of orthogonal polynomials multiplied by an associated weight function. One of the most important advantages of expansion methods is that they only require moments of the target distribution up to some order to obtain an approximation. Therefore they have the potential to be applied in a wide range of situations, including in attempts to solve the three problems listed above. On the other hand, it is also known that expansions lack robustness: they often exhibit unignorable negative density and their approximation quality can be extremely poor. This limits applications of expansion methods in existing studies.

In this thesis, we firstly develop techniques to provide robustness, with which expansion methods result in a practical approximation quality in a wider range of examples than investigated to date. Specifically, we investigate three techniques: standardisation, use of Laguerre expansion and optimisation. Standardisation applies expansion methods to a variable which is transformed so that its first and second moments are the same as those of the weight function. Use of Laguerre

expansions applies those expansions to a risk factor so that heavy tails can be captured better. Optimisation considers expansions with coefficients of polynomials optimised so that the difference between the approximation and the target distribution is minimised with respect to mean integrated squared error. We show, by numerical examples using data sets of stock index returns and log differences of implied volatility, and GARCH models, that expansions with our techniques are more robust than conventionl expansion methods. As such, marginal distributions of risk factors can be approximated by expansion methods. This solves a part of the description problem: the information on the marginal distributions of risk factors can be summarised by their moments. Then we show that the dependence structure among risk factors can be summarised in terms of their cross-moments. This solves the other part of the description problem. We also use the fact that moments of risk factors can be aggregated using their moments and cross-moments, to show that expansion methods can be applied to both the time and risk aggregation problems. Furthermore, we introduce expansion methods for multivariate distributions, which can also be used to approximate conditional expectations and copula densities by rational functions.

Keywords: Conditional expectations; Copulas; Hermite polynomials; Laguerre polynomials; Orthogonal expansion; Risk aggregation.

Contents

Table of Symbols and Abbreviations

- BS formula: Black and Scholes formula (for European call option premium). (Section 5.2)
- d: A function $d : \mathbf{R}^p \mapsto \mathbf{R}$, which maps the p risk factors to a PL. (Section 2.2.1)
- ES: Expected shortfall. (Section 2.1.2)
- HS method: Historical simulation method. (Section 2.2.2)
- HV: Historical volatility. (Section 5.2)
- i.i.d.: Independently identically distributed.
- ISE: Integrated squared error. (Section 4.2.1)
- IV: Implied volatility. (Section 5.2)
- KS: Kolmogorov-Smirnov (test statistic). (Section 3.2)
- MC: Monte Carlo (method). (Section 2.2.1)
- MISE: Mean integrated squared error. (Section 3.1)
- RMSE: Root mean squared error. (Section 3.2)
- pdf: Probability density function.
- PL: Profit and loss (of a portfolio). Here, losses are expressed as negative profits. (Section 2.1.2)
- VaR: Value at Risk. (Section 2.1.2)
- $\stackrel{\mathcal{D}}{=}$: Random variables on the left and right hand sides have the same distribution.

Statement of Originality

This is to certify that the work contained in this thesis has never previously been submitted for a degree or diploma in any university and that, to the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Date Kohei Marumo

Acknowledgment

The author thanks Professor Rodney Wolff (Queensland University of Technology, Australia) for supervising the project throughout three years, and Professor Stan Hurn (Queensland University of Technology, Australia) for discussions on Monte Carlo simulation. The author also wishes to thank Dr. Steven Li (University of South Australia) for inspiring discussion and suggestions for Chapters 3 and 4, and Professeure Dominique Guégan (Universiteé Paris I la Sorbonne, France) for suggesting the benefit of greater discussion on coherent measures of risk.

Chapter 1

Introduction

Market risk can be considered as potential loss that can happen to a portfolio due to possible changes in market conditions. Market risk consists of two basic contributions: the components of a portfolio and possible changes in market conditions. Measuring market risk might be considered as quantifying these two basic contributions[1].

There are many technical issues associated with measuring market risk. For instance, a risk factor — a market variable which can affect the profit and loss (PL) of a portfolio — can be described as a random variable, and this treatment can involve a number of arguments on modelling techniques. Furthermore, when a portfolio is affected by more than one risk factor, we must deal with a multivariate distribution and aggregate the risk factors, which can be even more troublesome. In spite of being so technical, measuring market risk is one of the most actively discussed subjects by both academia and the financial industry, and several papers on this topic have been published[2]. This particularly suggests the existence of

[1]Measures of market risk which only account for the components of a portfolio do exist. A possible loss of a portfolio under an arbitrarily defined market scenario can be an example. However, we do not deal with such measures in this thesis.

[2]Some of those papers are cited in Chapter 2.

unsolved problems and the need for better practices.

This thesis deals with some of the technical problems which can arise in measuring market risk of a portfolio. It attempts to propose reasonable solutions to these problems — methods to derive measures of market risk which are capable of capturing characteristics both of a portfolio and of risk factors — in the following manner.

In Chapter 2, we firstly overview a general background of market risk measurement practices, and discuss the fact that obtaining the distribution of the PL of a portfolio can be prominent in measuring market risk. Then we state the following three technical problems which can arise when we derive a PL distribution.

The first problem is related to the fact that the PL of a portfolio, in general, is exposed to multiple risk factors. How do we describe the marginal distributions of risk factors and the dependence structure among them and pass these pieces of information to the calculation process of PL distribution? The use of classical parametric multivariate distributions may be capable of dealing with them; however, tractable multivariate distributions, such as the multivariate Normal distribution, are exceptional. Let us call this the description problem.

The second problem arises when we are interested in measuring market risk with a long time horizon such as ten days. Distributional properties or features of risk factors with time horizon of one day may be inferred from daily observations of market variables. However, use of less frequent observations in order to infer the properties of risk factors with a longer time horizon can result in insufficient numbers of observations. Let us call this the time aggregation problem.

The third problem is to determine how to derive the distribution of the PL given the distributional properties of the risk factors. This problem is not trivial, especially when a portfolio is subject to multiple risk factors, or when the effect of risk factors on the PL of a portfolio is non-linear. Let us call this the risk aggregation problem.

Further review of existing methods reveals that they face difficulty in coping with at least one of the three problems. For instance, we often observe heavy tails in distributions of risk factors in real markets. However, the methods which assume that the risk factors have a multivariate Normal distribution fail to capture these heavy tails. We discuss the fact that they also face difficulty in dealing with the risk aggregation problem. On the other hand, the historical simulation (HS) methods, which use empirical distributions of risk factors, do not have a solution to the time aggregation problem.

Then, we demonstrate that expansion methods, which approximate a probability density function (pdf) by the sum of orthogonal polynomials multiplied by an associated weight function, can provide solutions to all of the three problems, and that those solutions are capable of capturing characteristics both of a portfolio and of risk factors.

Expansion methods only require moments of target distributions, and therefore have the potential to be applied to a wide rage of situations, including the three problems described above. However, their biggest shortcoming is a lack of robustness: approximations by expansion methods can often exhibit unignorable negative density.

Here we set our aim of this thesis as solving these three problems using expansion methods, despite their shortcoming. That is, we attempt to make the most of the advantage of expansion methods — expansion methods only require the moments of a target distribution to approximate that distribution — by developing techniques to mitigate their shortcoming.

Estimating moments in various situations may be tricky and can be a vast study area in its own right. In this thesis, we do not deal with such estimation problems; instead, we will show that our methods can be applied as long as moments are available, no matter which method is used for their estimations.

In Chapter 3, we introduce techniques, which combine to provide robustness

to expansion methods, and which are essential for further applications of expansion methods beyond those in existing studies. We also demonstrate with extensive numerical examples that expansion methods with these techniques can approximate the distributions of risk factors fairly well, given the moments of risk factors. This solves half of the description problem: marginal distributions of risk factors can be described in terms of their moments.

In Chapter 4, we discuss applications of expansion methods to the time aggregation problem. We show that expansion methods can be applied to the time aggregation problem when the moments of the risk factors are available. Furthermore, if the autocross-moments are available, expansion methods are capable of capturing serial dependence structure of risk factors.

In Chapter 5, we apply expansion methods to the risk aggregation problem. There are two necessary conditions for expansion methods to be applied to risk aggregation: that moments and cross-moments of the risk factors are available, and that a deterministic function which links the risk factors to the PL distribution can be approximated by its Taylor expansion. We discuss the fact that the dependence structure among the risk factors is described in terms of their cross-moments.

In Chapter 6, we develop expansion methods for joint distributions and discuss further applications to approximating conditional expectations and copula densities. We demonstrate with numerical examples for bivariate cases that, given moments and cross-moments of risk factors, expansion methods can be used to successfully visualise the dependence structure between two risk factors. This, together with the discussion in Chapter 5, solves the remaining half of the description problem: dependence structure among the risk factors can be described in terms of their cross-moments.

Chapter 7 concludes the thesis.

Chapter 2

Market Risk Measurement and Its Technical Aspects

In this chapter, we firstly review the background and existing studies of market risk measurement, and then we discuss some technical issues associated with current practices. We also describe the aim and scope of this thesis.

2.1 Overview of market risk measurement

We discuss in this section the motivation for market risk measurement with relation to liquidity of markets, and review the development and application of measures of market risk.

2.1.1 Liquidity of markets and measuring market risk

There is a close relationship between liquidity of markets and the need for market risk measurement, which can be explained as follows.

Suppose that we have a portfolio subject to market risk. If the liquidity of the market were unlimited and we could cash our portfolio at any instant at

the market price, then measuring market risk might not matter so much. For instance, we could avoid an unexpected loss by the following procedure: we set a limit level to the loss of the portfolio, monitor the net present value of the portfolio, and cash the portfolio as soon as the loss hits the limit level.

Unfortunately, however, the liquidity of real markets is limited, and the size of an average financial institution's portfolio can be unignorable compared to the daily traded volume in markets. Therefore, one must sell a portfolio bit by bit, find a counterpart to buy the portfolio, or construct a hedge position in order to eliminate the market risk. In any way, it is more realistic to assume that it takes a certain period of time, say $T > 0$ days, before one can cash the portfolio and get out of the market. During the period after the decision to cash the portfolio and before the transaction is completed, the portfolio is exposed to market risk, and this can be a chance to incur an unexpected loss. Measuring market risk can be a powerful means to know about what could possibly happen to a portfolio, and therefore it can help with preparing for adverse outcomes.

We also find from this discussion that it makes sense to associate the time horizon for measures of market risk with the expected time needed either to get out of the market, or to change financial decisions[1].

2.1.2 Development of measures of market risk

The development of measures of market risk is summarised by Szegö (2002). According to him, market risk used to be considered as a correcting factor of expected return, and risk-adjusted returns were defined on an *ad hoc* basis, until Markowitz (1959) suggested measuring the market risk of a portfolio using the variance of returns of all the assets involved and the covariance of their pairs. His

[1]This point is discussed, for example, in Sections 1.3.3 and 5.2.2 of Jorion (2001).

risk measure is given by

$$\mathrm{V}(S_1 + \cdots + S_p) = \sum_i \mathrm{V}(S_i) + 2\sum_{i<j} \mathrm{Cov}(S_i, S_j),$$

where S_i, $i = 1, \ldots, p$, are random variables which represent the returns on the p assets, and V and Cov denote variance and covariance respectively. His main innovation is that he included the covariances in the measure of risk, which account partly for the dependence structure among the returns of the assets.

In 1994, the concept of Value at Risk, or VaR was introduced (RiskMetrics 1994)[2]. VaR is defined as a tail quantile of the profit and loss (PL) distribution of a portfolio at a given time horizon, namely,

$$\mathrm{VaR}_\alpha = \max\{-\inf\{z | P(Z \leq z) \geq \alpha\}, 0\},$$

where Z is a random variable which represents the PL at the time horizon, and α is the parameter called probability level. For the market risk measurement purposes, $\alpha = 0.01$ is the most commonly used[3]. Note that the PL of a portfolio at a given time horizon is considered to be the difference between present value of the portfolio and a potential value of the portfolio at the time horizon, provided the composition of the portfolio stays the same, while an actual portfolio can change by transactions. Here, losses are expressed as negative profits.

Since VaR does not assume a specific type of distribution for risk factors and PL, it is capable of capturing more information than variance and covariance, such as skewness and kurtosis of individual risk factors and the PL, or dependence structure other than through covariances. Practically, this point was considered

[2]Rogachev (2007) suggests that VaR first received wide representation in the report by the Group of Thirty (1993), however, Szegö (2002) dates the introduction of VaR to the disclosure of the risk management practice using VaR by JP Morgan, which is known as the first edition of RiskMetrics (1994).

[3]$\alpha = 0.01$ is also suggested by the Basel Committee (1996). Since it is the 0.99 quantile of the loss distribution, VaR with $\alpha = 0.01$ is often called 99%VaR.

to be convenient for measuring the market risk of derivative positions[4], whose PL distribution may not be well characterised by variance and covariance alone. Due to such advantages, VaR soon became a common measure of market risk in the financial industry[5].

However, Artzner, Delbaen, Eber & Heath (1999) introduced the idea of coherent measures of risk, and they showed that VaR may not be coherent. That is, they defined a measure of risk ρ as a mapping from a set of random variables, which represent risk factors that can affect the PL, into the real numbers, and suggested that a coherent measure should satisfy the following four conditions.

- Transition invariance: For all risk factors X and all real numbers α, we have $\rho(X + \alpha(1 + r_f)) = \rho(X) + \alpha$, where r_f is the return on a risk free asset.

- Subadditivity: For all pairs of risk factors X_1 and X_2, $\rho(X_1+X_2) \leq \rho(X_1)+\rho(X_2)$.

- Positive homogeneity: For all risk factors X and real numbers $\lambda \geq 0$, $\rho(\lambda X) = \lambda\rho(X)$.

- Monotonicity: For all pairs of risk factors X and Y with $X \leq Y$, we have $\rho(Y) \leq \rho(X)$.

Then they showed an example where VaR does not satisfy the Subadditivity condition. In fact, VaR is coherent only when the joint distribution of the risk factors is elliptic[6].

[4]See Part III of RiskMetrics (1996).

[5]The use of VaR is also mentioned by the regulators. See Basel Committee (1996).

[6]A random vector $\boldsymbol{X}$ is elliptic if there exists an affine transformation $\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{b}$, where A and $\boldsymbol{b}$ are a real matrix and vector respectively, such that $\boldsymbol{Y}$ is spherical. $\boldsymbol{Y}$ is spherical if, for any orthonormal transformation Q, $\boldsymbol{Y} \stackrel{\mathcal{D}}{=} Q\boldsymbol{Y}$, where '$\stackrel{\mathcal{D}}{=}$' denotes that the both sides have the same distribution. An orthonormal transformation is defined as a transformation which always preserves the length of any vector and the angle between any two vectors.

Instead, it is shown by Acerbi & Tasche (2002) that the expected shortfall (ES), which is defined as an expected loss under the condition that the loss exceeds the VaR, namely

$$\mathrm{ES}_\alpha = -\mathrm{E}(Z|Z \leq -\mathrm{VaR}_\alpha),$$

is coherent as long as the PL distribution is continuous.

Moreover, Rockafellar & Uryasev (2000) showed that the portfolio can be optimised with respect to ES unig a simple algorithm. They proposed a linear programming algorithm to find a portfolio with a minimum ES for a given level of expected return. An advantage of ES over VaR is that VaR may have local optima, and therefore searching for the global optimum is much more difficult for VaR than for ES, which may be optimised using linear programming.

For such reasons, ES is attracting people's attention, although VaR still is the most popular practice in the industry[7].

One aspect of such development of measures of risk is that it stresses the importance of obtaining reliable distributions of risk factors and of the PL. For instance, the definitions of VaR and ES implicitly assume that a reliable PL distribution of a portfolio can be obtained. Furthermore, the problem with coherence of VaR is, in a sense, encouraging the pursuit of sophisticated estimation of distributions of risk factors, since the problem only arises when an elliptic distribution, such as the Normal distribution, turns out to be a poor approximation.

2.1.3 Applications of measures of market risk

Several of applications of market risk measures are proposed. Here we list only some of these mentioned by RiskMetrics (1996), Mina & Xiao (2001), Duffie &

[7]According to a survey by Rogachev (2007) over 57 Swiss private banks, nearly one third of respondents reported that their banks currently implement a VaR concept, while none is reported to use ES.

Pan (1997) and Jorion (2001).

Risk limit

Position limits, or limits to the amount of assets a portfolio can hold set by risk managers, have traditionally been expressed in nominal terms, futures equivalents or other denominators (RiskMetrics 1996). Setting limits in terms of a measure of market risk is more sensible because such measures are associated with possible loss, for which one needs to prepare.

Regulation

Regulators can set a rule on market activities of financial institutions using a measure of market risk. For instance, the Basel Committee (1996) allows banks to calculate their capital requirement using VaR.

Risk analysis and reporting

A measure of market risk based on a PL distribution provides a common ground for different classes of assets such as equity and foreign currency. By measuring the risk of each asset class, one can analyse how much risk one is taking, and in which markets. Reporting such analyses to the senior management, shareholders and regulators can increase the transparency of market business, by which some financial institutions might benefit[8].

[8]See Chapter 7 of Mina & Xiao (2001) for an example of risk analysis and reporting practice. They suggest that reporting PL distributions as well as the risk measures is also informative.

2.2 Technical aspects of market risk measurement

As discussed in Section 2.1, obtaining distributions of risk factors and of a PL is a key problem in market risk measurement.

In this section, we review some technical aspects of current practices for obtaining PL distributions. We firstly state the problems on which we focus, and then review how current practices deal with these problems. Finally, we describe the aim and scope of this thesis.

2.2.1 Obtaining PL distributions and related problems

Current practices of VaR calculation, which almost inevitably involves the estimation of a PL distribution, are well explained in RiskMetrics (1996), Mina & Xiao (2001), Jorion (2001), Duffie & Pan (1997) and Linsmeier & Pearson (1996).

Based on this literature, the general idea of obtaining the PL distribution can be summarised as the following three steps[9].

1. Identify the risk factors $\boldsymbol{x} = (x_1, \ldots, x_p)$ which can affect the PL of the portfolio, and formulate a deterministic function d which expresses the PL z, namely, $z = d(\boldsymbol{x})$.

2. Describe the risk factors as random variables $\boldsymbol{X} = (X_1, \ldots, X_p)$.

3. The PL distribution is given as the distribution of the random variable $Z = d(\boldsymbol{X})$.

We have, however, a number of technical problems when we implement this procedure. The following three are among the most frequently discussed.

[9]See, for instance, Chapter 6 of RiskMetrics (1996) and Linsmeier & Pearson (1996).

Description problem

The first problem arises in the second step above: how do we describe the risk factors as random variables $\boldsymbol{X} = (X_1, \ldots, X_p)$? It involves the description of p marginal distributions and of the dependence structure among them. We also need to pass these pieces of information to the calculation process of the PL distribution. The use of classical parametric multivariate distributions may be capable of dealing with them; however, tractable multivariate distributions, such as the multivariate Normal distribution, are exceptional. Let us call this the description problem.

This problem can be too obvious to be clearly stated in most of academic literature; however, the studies which pursue better models of asset returns and dependence structure among them might be considered as dealing with this problem. We will review some of these in Section 2.2.2.

Time aggregation problem

The second problem can be considered as a special case of the first one: how do we describe the risk factors when the time horizon is long (such as $T = 10$ days as suggested by the Basel Committee (1996))? It can be especially difficult to infer the characteristics of the risk factors from the past records due to unavailability of sufficiently large samples. For example, we need around 20 years' records in order to obtain 500 non-overlapping samples of 10-day increments of markets variables, however, this is not realistic in many cases. Such difficulties in dealing with risk factors with a long time horizon is called the time aggregation problem[10].

We might describe a risk factor with a T-day time horizon X^T using a series

[10]See Section 4.5 of Jorion (2001), for example. He treats this problem as one of the building blocks of calculation of VaR. This problem is also discussed by Daníelsson (2002), however, he only mentions "the difficulty given current technology of creating reliable 10 day VaR" and criticises Basel Committee (1996) for setting a regulation which relies on 10 days VaR.

of the risk factors with a one-day time horizon $X^1(t)$, $t = 1, \ldots, T$ by $X^T = X^1(1) + \cdots + X^1(T)$ or similarly using transformations such as logarithms. Then, the problem becomes that of deriving the distribution of the sum of T random variables.

Risk aggregation problem

The third problem is: given the distributional properties or features of the risk factors, how do we derive the distribution of PL? It is generally non-trivial, given random variables $\boldsymbol{X} = (X_1, \cdots, X_p)$ and a deterministic function d, to obtain the distribution of $Z = d(\boldsymbol{X})$. This problem is called the risk aggregation problem[11].

A naive solution when the joint density function f of $\boldsymbol{X}$ is available is to integrate numerically $P(Z \leq z) = \int_{d(\boldsymbol{x}) \leq z} f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$, perhaps by the Monte Carlo (MC) method, which might be costly in terms of computational load[12].

2.2.2 Current practices

Now we discuss how the current practices deal with these problems. According to RiskMetrics (1996), Mina & Xiao (2001), Jorion (2001), Duffie & Pan (1997) and Linsmeier & Pearson (1996), the most common solutions to these problems seem to be categorised into two groups, with one group based on the Normal distribution and the other on empirical distributions. The methods based on

[11]This problem is discussed widely in the literature, if not called "risk aggregation problem," including RiskMetrics (1996), Mina & Xiao (2001), Jorion (2001) and Duffie & Pan (1997). The discussion by Artzner et al. (1999) also has a close relationship with this problem, since the subadditivity condition concerns the measure of multiple risks.

[12]Since risk measures are mainly concerned with the tails of a distribution, the number of samples needed can be much more than when we are interested in statistics such as mean or variance of a distribution. Moreover, when p is not small, it can be another problem to generate $p-$variate pseudo-random numbers. Section 9.4.3 of Jorion (2001) points out the problems of MC method used in market risk measurement, including these.

empirical distributions are called historical simulation (HS) methods[13].

Normal based methods

Modelling the risk factors using the Normal distribution is one of the most common methods to obtain a PL distribution in practice[14]. These methods are also discussed in, for instance, Chapter 6 of RiskMetrics (1996), Chapter 2 of Mina & Xiao (2001), Section 9.2 of Jorion (2001), and Duffie & Pan (1997).

Here we review how Normal based methods deal with the problems in Section 2.2.1.

Description problem. Among the Normal based methods, the most basic solution to the description problem is to assume that the risk factors have a multivariate Normal distribution[15]. This approach has an advantage in terms of tractability: it only requires the mean vectors and covariance matrix of the risk factors. On the other hand, however, it ignores other characteristics such as skewness and heavy tails of risk factors, and dependence structures among them other than through covariances, which are often observed in the markets[16].

According to Duffie & Pan (1997), jump diffusion and stochastic volatility models are proposed in order to describe the heavy tails which are often observed in asset return distributions. The jump diffusion model can be regarded as adding a random variable, which represents a shock that arrives randomly, to a Normally distributed variable. The shock arrival process can be modelled as a Poisson distribution. Stochastic volatility models assume that the volatility (the square

[13] According to a survey by Rogachev (2007) over 57 Swiss private banks, approximately 75% of the banks using the VaR apply the HS method, and more than 20% apply methods based on the Normal distribution.

[14] See footnote 13.

[15] See, for instance, Chapter 6 of RiskMetrics (1996), Chapter 2 of Mina & Xiao (2001) and Section 9.2 of Jorion (2001) for details of this method.

[16] See, for instance, Rachev (2003) for heavy tails observed in the markets.

root of the variance of the Normal distribution) is also random. They comprise a wide variety of models, including the regime-switching volatility model, which assumes that the volatility process is a finite state Markov chain; and ARCH[17] and GARCH[18] models, which are capable of incorporating the autocorrelation of volatility.

The drawback of using these models in measuring market risk is that they require additional parameter estimation. Since they centre upon quantities such as the jump process and volatility process, which cannot be observed in the market, the calibration of the model to fit the observations can be another tricky problem. Also, the distributions of risk factors under such models are not necessarily Normal, and therefore can be less tractable, especially when faced with the time and risk aggregation problems[19].

Time aggregation problem. The most common solution of the Normal based methods to the time aggregation problem can be to assume further that the risk factors are serially independent[20]. Under this assumption, a risk factor with a T-day time horizon is a sum of T independent Normal risk factors with one-day time horizons and therefore has a Normal distribution. The mean vector and covariance matrix can be obtained by multiplying those of risk factors with a one-day time horizon by T, or equivalently, we have a scaling property: $\boldsymbol{X}^T - T\boldsymbol{\mu}^1 \stackrel{\mathcal{D}}{=} \sqrt{T}(\boldsymbol{X}^1 - \boldsymbol{\mu}^1)$, where $\boldsymbol{\mu}^1$ is the mean vector of risk factors with a one-day time horizon and $\boldsymbol{X}^T$ and $\boldsymbol{X}^1$ denote the risk factors with a T-day and one-day time horizons, respectively.

[17] Autoregressive conditional heteroscedasticity. See Engle (1982).

[18] Generalised autoregressive conditional heteroscedasticity. See Kim, Shephard & Chib (1998). We will discuss GARCH models later in Chapter 3.

[19] See Duffie & Pan (1997) and Kim et al. (1998).

[20] For instance, see Section 4.5 of Jorion (2001). Chapter 2 of Mina & Xiao (2001) implicitly assumes that the risk factors are serially independent.

In order to incorporate the serial dependence of second moments of the risk factors, the GARCH[21] and the regime-switching volatility models are also proposed. The drawbacks are as described above.

Risk aggregation problem. There seem to be two common directions for Normal based methods to solve the risk aggregation problem. One is to apply a Taylor approximation to the function d.

When only the linear terms are included in the approximation, it is called the delta-Normal method[22]. This method is particularly handy, because under this approximation, the PL is a linear combination of risk factors, whose distribution is also Normal. However, since it ignores non-linear effects of the risk factors, it is considered to be unsuitable to measure the market risk of some types of derivative positions such as options.

If the Taylor approximation includes up to the quadratic terms, it is called the delta-gamma-Normal method[23]. It can be considered to be more reliable in the sense that it include some non-linear effects of the risk factors; however, the PL distribution is not Normal and obtaining the PL distribution function requires further approximation. In this topic, Feuerverger & Wong (2000) compared technical aspects of the saddlepoint approximation[24] with those of the Fourier inversion[25] applied to approximating the PL distribution under the delta-

[21]See, for instance, Duffie & Pan (1997), Chapter 5 of RiskMetrics (1996) and Section 8.2.3 of Jorion (2001).

[22]See Section 6.3.2 of RiskMetrics (1996), Section 2.3 of Mina & Xiao (2001), Section 9.1.1 of Jorion (2001) and Duffie & Pan (1997).

[23]See Section 6.3.3 of RiskMetrics (1996), Section 9.1.3 of Jorion (2001).

[24] This approximation method can be regarded as a variation of expansion methods, which approximates a distribution using a non-negative function and polynomials. Such expansion methods are related to the main focus of this thesis, and therefore we will cite this article again in the context of reviewing studies on expansion methods. See, for instance, Kolassa (1997) or Jensen (1995) for details about the saddlepoint approximation.

[25]The characteristic function of the PL distribution under the delta-gamma-Normal method

gamma-Normal method. They concluded that the two are best viewed as being complementary. That is, there are contrasts between these two such as that the saddlepoint approximation requires a cumulant generating function of a tractable form while the Fourier inversion requires characteristic function in a form which can be numerically integrated in a practical computational time.

Jaschke (2002) proposed an application of the Cornish-Fisher expansion[26] to approximate the quantile of PL distribution under the delta-gamma-Normal method. He concluded that the quality of an approximation by the Cornish-Fisher expansion is comparable to those by other methods such as Fourier inversion, saddlepoint expansion and the MC method, if the PL distribution is close to the Normal distribution.

The other direction for solving the risk aggregation problem is to use the MC method, as mentioned in Section 2.2.1. For a mean vector and covariance matrix, either estimated directly from the observations or given from models such as jump diffusion and stochastic volatility models, we generate N pseudo random vectors $\boldsymbol{x}^{(i)}$, $i = 1, \ldots, N$, from a multivariate Normal distribution, then we have N samples of PL, $z^{(i)} = d(\boldsymbol{x}^{(i)})$, $i = 1, \ldots, N$. We use the empirical distribution of $\{z^{(i)}\}$ as an approximation of the PL distribution. The drawbacks of this method are discussed in Section 2.2.1.

Historical simulation methods

HS methods use the empirical distribution of risk factors to derive the PL distribution. Suppose that we observe N samples $\boldsymbol{x}^{(i)}$, $i = 1, \ldots, N$, of risk factors from the historical records, then we have N samples of PL $z^{(i)} = d(\boldsymbol{x}^{(i)})$, $i = 1, \ldots, N$.

is derived. Numerical Fourier-inversion of this can be used to compute an approximation for the distribution function.

[26] The Cornish-Fisher expansion is also one of the expansion methods, and this article will be cited again as well. See footnote 24. See, for instance, Hall (1992) for details about the Cornish-Fisher expansion.

We use the empirical distribution of $\{z^{(i)}\}$ as an estimate of PL distribution[27].

Description problem. When the number of observations is large enough, using the empirical distribution of risk factors can be a reasonable solution to the description problem, since it incorporates, in a natural way, characteristics of the marginal distributions and dependence structures of the risk factors, including skewness, heavy tails and dependence structures other than through covariances, which are ignored in Normal based methods.

The drawback can be that the discreteness may not be favourable especially in the tails where the observations are sparse.

Time aggregation problem. For the time aggregation problem, however, available solutions for HS methods are not so attractive[28]. Even if we assume that the risk factors are serially independent, it is not easy to obtain the distribution of risk factors with a T-day time horizon. For instance, assume that the risk factors with a one-day time horizon $\boldsymbol{X}^1(t)$, $t = 1, \ldots, T$, are independently distributed with an empirical distribution with realisations $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\}$. The risk factor with a T-day time horizon is given by $\boldsymbol{X}^T = \boldsymbol{X}^1(1) + \cdots + \boldsymbol{X}^1(T)$. The number of possible realisations of $\boldsymbol{X}^T$ is of the order of N^T, which can easily exceed the range of computational practicality[29].

Instead, scaling by $\sqrt{T}$, which is used in the Normal based method, is sometimes applied as a rule of thumb (Mina & Xiao 2001).

Sampling T days increments from past records allows for the serial dependence in a natural way; however, non-overlapping sampling can be difficult because of

[27]See Section 2.3 of RiskMetrics (1996), Chapter 3 of Mina & Xiao (2001) and Section 9.3 of Jorion (2001) for details of HS methods.

[28]Finger (2006) discusses problems of the HS method including this point.

[29] For the case with $T = 10$ days time horizon and $N = 500$ samples, $N^T \sim 10^{26}$. Drawing enough, but not too many, samples randomly from $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}\}^T$ may also be an option, however, reliability of such a 'simulation on simulation' is debatable.

the insufficient number of samples, as mentioned in Section 2.2.1. Instead, overlapping sampling is often employed in practice, although this causes very strong dependence among the samples and therefore the estimation can be inefficient[30].

Risk aggregation problem. For the risk aggregation problem, the HS method provides a handy solution: given samples $\boldsymbol{x}^{(i)}$, $i = 1, \ldots, N$, and function d, we can use the empirical distribution of $z^{(i)} = d(\boldsymbol{x}^{(i)})$ as an estimator of the distribution of $Z = d(\boldsymbol{X})$.

The drawbacks are that discreteness, especially in the tails where the observations are sparse, can be unfavourable, and that tails of the empirical distribution might be regarded as too sensitive to a small number of observations. For instance, the 0.01 quantile of the empirical distribution of 500 samples can only depend on five smallest observations, or particularly, on the fifth smallest observation alone[31].

Other methods

Other less common methods in use include the following. Here we also review the use of expansion methods, which is the main focus of the thesis.

Copulas. Copulas[32], which are used to describe dependence structure more generally among the risk factors, can solve a part of the description problem

[30] See Vlaar (2000) for example. He investigated the HS methods applied to a Dutch bond portfolio using overlapping samples, fully aware of the problem, and commented that the use of overlapping samples might have caused an underestimation of VaR.

[31] Schachter & Butler (1996) suggest use of a kernel density estimator to obtain a continuous PL distribution function; however, calculating quantiles of their kernel density estimator can require intensive numerical calculation.

[32] A class of $p-$variate copula functions can be defined as distribution functions on $[0,1]^p$, which have standard uniform marginal distributions. We will further discuss copulas later in Chapter 6.

(Junker & May 2005). A copula is flexible enough to account for the dependence structure other than through covariances, however, it requires specification of the copula function *a priori*[33]. Also, we may have to rely on MC methods or other numerical integration methods to deal with the risk aggregation problem.

EVT based methods. The tails of marginal distributions can be approximated using extreme value theory (EVT), which can solve another part of the description problem which copulas do not. Namely, EVT yields that, under general conditions, the distribution of a random variable X beyond a sufficiently large cutoff point u can be approximated by

$$P(X > x + u | X \geq u) \simeq 1 - G_{\beta,\xi}(x) = \exp\left(- \int_0^{\frac{x}{\beta}} \frac{\mathrm{d}v}{(1+\xi v)_+} \right),$$

where $G_{\beta,\xi}$ is the distribution function of the generalised Pareto distribution, $\beta > 0$ and $\xi \in \mathbf{R}$ are scale and shape parameters, respectively, and $(\cdot)_+$ is defined for $y \in \mathbf{R}$ as $(y)_+ = \max\{0, y\}$. Methods to estimate the parameters β and ξ from observations or models have been proposed[34].

The methods based on EVT also provide a scaling property similar to that of Normal based methods, which can be applied to the time aggregation problem when we assume serial independence. That is, the scale parameter of a risk factor with a T-day time horizon can be approximated by T^{ξ} times the scale parameter of the risk factor with a one day time horizon, while the shape parameter remains ξ the same[35].

The dependence structure among the risk factors, however, has to be dealt

[33]Although a non-parametric copula has been introduced, its calibration is not trivial. See Capéraà, Fougères & Genest (1997).

[34]Embrechts, Klüppelberg & Mikosch (1997) is one of the standard references for EVT and its application to finance. Further theoretical details of EVT are given by Resnick (1987) and Pickands III (1975).

[35]See Daníelsson (1997) or Chapter 10 of Jorion (2001).

with by other methods such as using a copula (Natale 2006)[36].

Also, EVT based methods do not have a handy solution for the risk aggregation problem, except that Longin (2000) proposed a formula to approximate the VaR of a portfolio using a linear combination of quantiles of risk factors.

Use of non-Normal distributions. Instead of the Normal, the multivariate Student's t distribution can also be used in order to describe heavy tails of the risk factors' distributions[37]. Similarly to the Normal based methods, it also ignores the dependence structure other than through covariances. For the time and risk aggregation problems, similar solutions to those of the Normal based method are available. Note that the multivariate Student's t distribution is also elliptic.

Furthermore, the class of stable distributions[38], generalised error distributions[39] and the mixed Normal distribution[40] are among the proposed distributions. They might capture some characteristics of risk factors better than the Normal distribution; however, in general, they can be less tractable than the Normal distribution, especially when faced with the problems described in Section 2.2.1.

Use of expansion methods. Besides the methods discussed so far, approximation techniques called expansion methods are sometimes used to obtain distributions of risk factors and of the PL of a portfolio.

[36]Multivariate extreme value theory has been discussed by Ari Hauksson, Dacorogna, Domenig, Müller & Samorodnitsky (2001). They showed that they can make an inference on tails of multivariate distributions, however, their method requires sufficiently frequent and long records, such as 12 years' hourly observations, which may not be always available.

[37]See Appendix A.2 of Mina & Xiao (2001) or Section 4.2.4 of Jorion (2001).

[38]See Weron (2004).

[39]See Section 4.2.4 of Jorion (2001)

[40]See Appendix A.2 of Mina & Xiao (2001). Mixed Normal distribution can be viewed as a jump diffusion model with Normally distributed shocks.

Expansion methods approximate a probability density function by the sum of orthogonal polynomials multiplied by an associated weight function[41]. One of the advantages of expansion methods over other methods to approximate distributions, such as methods based on EVT, is that expansion methods only require the moments of the target distribution up to some order. They also have a closed form which can be used for calculating ES[42].

Because of such advantages, expansion methods have the potential to be applied in a wide range of situations. In fact, successful examples of their applications to asset return distributions can be found in, for example, Mauleón & Perote (2000*b*). Furthermore, an application to the PL of a portfolio which is described as a linear combination of risk factors is considered by Perote & Del Brío (2001). Apart from applications to market risk measurement, Buckland (1992) considers a density estimator based on expansion methods. That is, he formulates the estimator as a polynomial multiplied by a weight function and estimates the coefficients of the polynomial using maximum likelihood. He points out that this provides a good fit to observations. However, this is a parametric approach, in stark contrast to our methodology, which might limit further applications of Buckland's method[43].

In general, however, expansion methods are subject to a serious problem: they are lacking in robustness. Very often their approximation quality is poor, especially in the tails, and can exhibit unignorable negative probability density[44]. Mainly due to this disadvantage, most applications of expansion methods to risk

[41] We will define orthogonal polynomials and an weight function in Section 2.2.3.

[42]This closed form has not appeared in financial literature, as far as we can determine. Including this, properties of expansion methods are discussed in more deatil in Section 2.2.3.

[43] Other important applications of expansion methods include Aït-Sahalia (2002), who uses expansion methods for constructing the likelihood function to estimate the parameters of stochastic differencial equation.

[44]For instance, Gordy (2002) and Jaschke (2002) point out this problem. This will be discussed later in Section 2.2.3.

measurement so far are limited to dealing with certain classes of models or to some special cases.

For example, an application of the saddlepoint approximation, a refinement of the conventional expansion methods, to the PL under the delta-gamma-Normal method is considered by Feuerverger & Wong (2000)[45]. Note that the saddlepoint approximation can be applied only to the cases where the cumulant generating function of the target distribution is of a tractable form. Jaschke (2002) applied the Cornish-Fisher approximation[46] also to the PL under the delta-gamma-Normal method. He concludes that the approximation by the expansion is competitive if the PL distribution is reasonably close to the Normal distribution. An expansion up to fourth order has been applied to the GARCH model by Lillo & Mantegna (2002). They analysed the market data after a crash and found some evidence of power-law characteristics, which cannot be captured by the GARCH model.

Despite such a disadvantage, we expect that the expansion methods may provide reasonable solutions to all of the description, time aggregation and risk aggregation problems. Therefore, we elaborate expansion methods in Section 2.2.3, and we explain in Section 2.3 our attempt at applying expansion methods to these three problems in wider situations than existing studies.

2.2.3 Orthogonal polynomials and expansion methods

As briefly introduced in Section 2.2.2, expansion methods have the potential to be applied to a wide range of situations, including the three problems described in Section 2.2.1, while they suffer from lacking in robustness.

[45] See footnote 24.

[46] An expansion method which approximates the target distribution by the product of the Normal pdf with the sum of Hermite polynomials. See footnote 26. The Hermite polynomials are defined in Section 2.2.3.

In this section, we firstly review orthogonal polynomials and expansion methods using them. Then we discuss properties of expansion methods. Further theoretical background for orthogonal polynomials and expansion methods is given by Szegö (1975), Jackson (1963) or Freud (1971).

Orthogonal polynomials

Let $\{g_k; k = 0, 1, \ldots\}$ be a series of polynomials, each of which is of kth degree, and w be a non-negative function with support $\mathcal{S} \subseteq \mathbf{R}$ which satisfies $\int_{\mathcal{S}} w(u)\mathrm{d}u = 1$. Then $\{g_k\}$ is said to be a set of orthogonal polynomials with weight function w if

$$\int_{\mathcal{S}} w(x)g_k(x)g_l(x)\mathrm{d}x = \begin{cases} 0 & (k \neq l) \\ e_k > 0 & (k = l) \end{cases}.$$

When $e_k = 1$ for $k = 0, 1, \ldots$, the set is called orthonormal. For any orthogonal polynomials $\{g_k\}$, $\{g_k/\sqrt{e_k}\}$ is orthonormal, and therefore we can deal only with orthonormal polynomials without loss of generality. We call the combination $(w, \{g_k\})$ an orthonormal system.

Expansion methods

Consider a probability density function (pdf) f with support $\mathcal{S}$ belonging to a Hilbert space[47] $\mathcal{H}$, and let $\{g_k; k = 0, 1, \ldots\}$ be an orthonormal polynomial basis of $\mathcal{H}$ with weight function w. That is, assume that f can be expressed as

$$f(x) = w(x)\sum_{k=0}^{\infty} C_k g_k(x), \tag{2.1}$$

for some real coefficients $C_0, C_1, \ldots$. Then, from orthonormality, the coefficients are given by

$$C_k = \int_{\mathcal{S}} g_k(u)f(u)\mathrm{d}u$$

[47] A vector space for which an inner product $\langle \cdot, \cdot \rangle$, such as $\langle \phi, \psi \rangle = \int_{\mathcal{S}} \phi(u)\psi(u)\mathrm{d}u$, is defined and the norm $\|\phi\| = \sqrt{\langle \phi, \phi \rangle}$ makes the space complete. That is, all Cauchy sequences $\{\phi_1, \phi_2, \ldots\}$ that satisfy $\lim_{\min\{m,n\}\to\infty} \|\phi_m - \phi_n\| = 0$, are convergent.

for $k = 0, 1, \ldots$. Let X be a random variable with pdf f. Then, $C_k = \mathrm{E}(g_k(X))$ is a linear combination of moments of X up to kth order. Note that $g_0(x)$ is a constant and $C_0 = g_0(x) \equiv 1$ is required so that $\int_{\mathcal{S}} f(u)\mathrm{d}u = 1$ is satisfied.

Given the moments $\mathrm{E}(X), \ldots, \mathrm{E}(X^n)$, Equation (2.1) can be approximated by a sum up to finite n, and we have the approximation for the density function

$$f(x) \simeq \hat{f}(x) = w(x) \sum_{k=0}^{n} \mathrm{E}(g_k(X)) g_k(x). \tag{2.2}$$

This approximation technique is called the expansion method.

Expansion methods can be distinguished by the orthonormal system employed. The following two expansions are often discussed.

Hermite expansion. For the support $\mathcal{S} = (-\infty, \infty)$, the Hermite system is the most commonly used; that is, g_k is the kth Hermite polynomial given by

$$g_k(x) = \mathrm{He}_k(x) = \frac{(-1)^k}{\sqrt{k!}} \mathrm{e}^{\frac{x^2}{2}} \frac{\mathrm{d}^k}{\mathrm{d}x^k} \mathrm{e}^{-\frac{x^2}{2}}, \tag{2.3}$$

and w is the pdf of the standard Normal distribution[48]. Then, Equation (2.2) is of the form

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \sum_{k=0}^{n} \mathrm{E}(\mathrm{He}_k(X)) \mathrm{He}_k(x). \tag{2.4}$$

Let us call this the Hermite expansion. For the distribution function, by using Equation (2.3), we have

$$\int_{-\infty}^{x} f(u)\mathrm{d}u \simeq \frac{1}{\sqrt{2\pi}} \left\{ \int_{-\infty}^{x} \mathrm{e}^{-\frac{u^2}{2}} \mathrm{d}u - \mathrm{e}^{-\frac{x^2}{2}} \sum_{k=1}^{n} \frac{\mathrm{E}(\mathrm{He}_k(X))}{\sqrt{k}} \mathrm{He}_{k-1}(x) \right\}. \tag{2.5}$$

[48] Actually, orthonormal systems can be constructed using arbitrary weight functions with support $\mathcal{S}$, and approximations using Equation (2.2) can be made using them. See Jackson (1963) and Appendix A.1 for Schmidt's process, which explains a procedure to obtain a set of orthogonal polynomials associated with an arbitrary weight function. However, useful properties such as Equations (2.5), (2.6), (2.8) and (2.9) may not be available in general.

Furthermore, we have

$$\int_{-\infty}^{x} uf(u)\mathrm{d}u \simeq \frac{1}{\sqrt{2\pi}}\left\{-\mathrm{e}^{-\frac{x^2}{2}} + \mathrm{E}(\mathrm{He}_1(X))\left(\int_{-\infty}^{x}\mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u - x\mathrm{e}^{-\frac{x^2}{2}}\right)\right.$$
$$\left. -\mathrm{e}^{-\frac{x^2}{2}}\sum_{k=2}^{n}\mathrm{E}(\mathrm{He}_k(X))\left(\frac{x\mathrm{He}_{k-1}(x)}{\sqrt{k}} + \frac{\mathrm{He}_{k-2}(x)}{\sqrt{k(k-1)}}\right)\right\}, \quad (2.6)$$

which can be used for calculating the ES[49].

Laguerre expansion. For the support $\mathcal{S} = [0, \infty)$, the generalised Laguerre system is often used; that is, g_k is the kth generalised Laguerre polynomial given by

$$g_k(x) = L_k^{(\beta-1)}(x) = \sqrt{\frac{\Gamma(k+1)\Gamma(\beta)}{\Gamma(k+\beta)}}\sum_{l=0}^{k}\binom{k+\beta-1}{k-l}\frac{(-x)^l}{l!},$$

where $\beta > 0$ is a parameter, and w is the pdf of the standard gamma distribution with the parameter β. Then Equation (2.2) is of the form

$$\hat{f}(x) = \left\{\frac{1}{\Gamma(\beta)}x^{\beta-1}\mathrm{e}^{-x}\right\}\sum_{k=0}^{n}\mathrm{E}(L_k^{(\beta-1)}(X))L_k^{(\beta-1)}(x). \quad (2.7)$$

Let us call this the Laguerre expansion. Similarly to the Hermite expansion, using the identity

$$L_k^{(\beta-1)}(x) = \sqrt{\frac{\Gamma(\beta)}{\Gamma(\beta+k)\Gamma(k+1)}}x^{-(\beta-1)}\mathrm{e}^{x}\frac{\mathrm{d}^k}{\mathrm{d}x^k}\left(x^{\beta-1+k}\mathrm{e}^{-x}\right),$$

we have

$$\int_0^x f(u)\mathrm{d}u \simeq \frac{1}{\Gamma(\beta)}\left\{\int_0^x u^{\beta-1}\mathrm{e}^{-u}\mathrm{d}u + x^{\beta}\mathrm{e}^{-x}\sum_{k=1}^{n}\frac{\mathrm{E}(L_k^{(\beta-1)}(X))}{\sqrt{\beta k}}L_{k-1}^{(\beta)}(x)\right\}, \quad (2.8)$$

[49]Neither this formula, nor Equation (2.9), has been used in financial literature as far as we can determine.

and

$$\int_0^x uf(u)\mathrm{d}u \simeq \frac{1}{\Gamma(\beta)}\Bigg\{\left(1-\frac{\mathrm{E}(L_1^{(\beta-1)}(X))}{\sqrt{\beta}}\right)\int_0^x u^\beta \mathrm{e}^{-u}\mathrm{d}u + x^{\beta+1}\mathrm{e}^{-x}\left(\sum_{k=1}^{n}\frac{\mathrm{E}(L_k^{(\beta-1)}(X))}{\sqrt{\beta k}}L_{k-1}^{(\beta)}(x) - \sum_{k=2}^{n}\frac{\mathrm{E}(L_k^{(\beta-1)}(X))}{\sqrt{\beta(\beta+1)(k-1)k}}L_{k-2}^{(\beta+1)}(x)\right)\Bigg\}. \quad (2.9)$$

Properties of expansion methods

Now, let us discuss some properties of approximation by Equation (2.2).

First of all, the approximation by Equation (2.2) only requires the moments of the target distribution up to nth order[50].

Moreover, given the moments $\mathrm{E}(X^1), \ldots, \mathrm{E}(X^n)$, Equation (2.2) assures

$$\int_{\mathcal{S}} u^k \hat{f}(u)\mathrm{d}u = \mathrm{E}(X^k),$$

for $k = 0, 1, \ldots, n$. This suggests that an approximation by expansion methods may preserve up to nth moments of original variable X. For higher order moments, we have

$$\int_{\mathcal{S}} u^k \hat{f}(u)\mathrm{d}u = \int_{\mathcal{S}} u^k w(u)\mathrm{d}u,$$

for $k = n+1, \ldots$, which suggests that the higher order moments of the approximated distribution can be those of the weight function. Therefore we might expect a good approximation quality if the target distribution is close to the weight function.

On the other hand, the approximation by expansion methods is lacking in robustness. We find that Equation (2.2) does not exclude the possibility of negative density: $\hat{f}$ has some region where $\hat{f}(x) \leq 0$ whenever $\sum_{k=0}^{n} E(g_k(X))g_k(x) = 0$

[50]See Sections 3.2.2 and 4.2.2 and Appendix A.6 for examples and discussion on expansion methods applied to distributions for which some moments do not exist.

has real roots. We might try to mitigate the negative density by increasing n in Equation (2.2), however, it can be considered as rather exceptional that $\hat{f}$ in Equation (2.2) converges to f in Equation (2.1) as $n \to \infty$, and therefore increasing n does not necessarily improve the approximation quality[51]. This can be the biggest drawback of the expansion methods.

These facts outline the properties of expansion methods.

2.3 Aim and scope of this thesis

In this thesis, we aim to demonstrate that expansion methods provide reasonable solutions to the three problems described in Section 2.2.1, despite their lack of robustness.

We firstly attempt to develop techniques which can provide robustness for expansion methods, before we deal with the three problems associated with market risk measurement. The three problems, then, will be dealt with in the following manner.

As explained in Section 2.2.3, expansion methods only require the moments of the target distribution. This suggests that the marginal distributions of the risk factors can be expressed in terms of their moments. Also, we show that the dependence structure among risk factors can be expressed adequately in terms of cross-moments. These will be the solution to the description problem.

For the time aggregation problem, where we wish to find the distribution of a risk factor with a T-day time horizon $X^T = X^1(1) + \cdots + X^1(T)$, expansion methods can provide a solution even when we assume some serial dependence, as

[51] Jaschke (2002) points out this convergence problem for Edgeworth expansions, which have similar formulae to those of the Hermite expansion, but derived differently. See Hall (1992) for construction of Edgeworth expansions. See also Freedman (1981) or Szegö (1975) for convergence properties of expansions. We will also discuss the effect of n on approximation quality later in Chapter 3.

long as moments and autocross-moments of $X^1(1), \ldots, X^1(T)$ are available.

The Taylor approximation of the function d is a polynomial in $X_1, \ldots, X_p$, and therefore the moments of $Z = d(X_1, \ldots, X_p)$ can be approximated using the moments and cross-moments of the individual risk factors. This suggests that, given moments and cross-moments of the individual risk factors, expansion methods can be applied to approximate the distribution of Z. This can be a solution for the risk aggregation problem. Note that, unlike the Normal based method, expansion methods can handle higher order terms. Also, the approximations by expansion methods are continuous. This can be an advantage over the HS methods, which suffer from discreteness.

Estimating the moments and cross-moments in various situations can also be a tricky problem, however, we limit the scope of this thesis to discussing expansion methods and their solutions to the description, time aggregation and risk aggregation problems, given the estimates of the moments and cross-moments. Therefore, we use conventional methods for estimating moments and cross-moments, where necessary.

Chapter 3

Expansion Methods Applied to Risk Factors

In this Chapter, we discuss the application of expansion methods to approximating the risk factors' marginal distributions and introduce related techniques which can provide robustness for expansion methods. Numerical examples are also given[1].

This Chapter has two purposes. One is to examine expansion methods in the simplest situations, where expansion methods are applied to univariate distributions. Here we develop techniques to provide robustness for the conventional use of Hermite expansions and demonstrate how the techniques works. The other is to solve a part of the description problem in Section 2.2.1. That is, if the expansion methods successfully approximate the marginal distributions of risk factors, we might be able to describe the marginal distributions in terms of their moments.

Further applications and multivariate cases are discussed in later Chapters.

[1]Parts of this Chapter and of Chapter 4 are summarised in Marumo & Wolff (2007).

3.1 Application and related techniques

Based on the properties of expansion methods explained in Section 2.2.3, we discuss here the applications of these methods to approximating the risk factors' marginal distributions. Here we also develop techniques to provide robustness to the expansion methods.

3.1.1 Application to distributions of risk factors

It is natural that we apply the Hermite expansion, whose support is $(-\infty, \infty)$, to approximating distributions of risk factors and the PL, which can take either positive or negative values. In fact, the expansions employed by existing studies including those reviewed in Section 2.2.2 can be categorised as either the Hermite expansion or a version of it, as far as we can determine. However, as pointed out by Gordy (2002) and Jaschke (2002), the Hermite expansion is lacking in robustness, and a naive use of the Hermite expansion — such as plugging the sample moments of a risk factor into Equation (2.4) directly — can result in a very poor approximation.

Therefore, some improvement which provides robustness to the conventional use of expansion methods might be essential for further applications than those considered in existing studies.

3.1.2 Techniques to provide robustness

Here we introduce three techniques, combinations of which can provide robustness for the naive Hermite expansion.

Standardisation

As discussed in Section 2.2.3, if the target pdf is not so different from the weight function, we expect that the expansion method can provide a good approximation.

One obvious way to make the target pdf closer to the weight function is to standardise the variable so that the first and second moments of the target pdf are equal to those of the weight function. Actually, this technique is used, sometimes implicitly, in most existing applications.

For the Hermite expansion, where we wish to approximate a pdf f of a random variable X, let $\mu = \mathrm{E}(X)$, $\sigma^2 = \mathrm{E}(X^2) - \mu^2$, $X' = (X - \mu)/\sigma$ and $f_{X'}$ be the pdf of X'. We apply the Hermite expansion Equation (2.4) to X' to obtain its approximation $\hat{f}_{X'}$. The pdf of X can be approximated by $\hat{f}(x) = \hat{f}_{X'}\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}$.

For the Laguerre expansion, let

$$\mu = \mathrm{E}(X) \text{ and } \beta = \frac{\{\mathrm{E}(X)\}^2}{\mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2},$$

and apply the Laguerre expansion with parameter β to $X'_L = \beta X/\mu$ to obtain the approximation for the density function $\hat{f}_{X'_L}$ of X'_L. We have $\hat{f}(x) = \hat{f}_{X'_L}\left(\frac{\beta x}{\mu}\right)\frac{\beta}{\mu}$.

Note that $C_1 = C_2 = 0$ is assured by such standardisation.

Use of Laguerre expansions

Very often asset return distributions exhibit heavier tails than those of the Normal distribution, and this can be regarded as one of the reasons why the Hermite expansion can perform poorly, especially at the tails. Since the gamma distribution has a heavier right tail than that of the Normal, we expect that the Laguerre expansion, whose weight function is the gamma density, can approximate the heavy tail better.

In fact, the Laguerre expansion is often used to approximate the distributions of physical quantities, such as light scattered by the turbulent atmosphere (Barakat 1999), intensity of photoelectrons' arrival (Barakat 1996) and the cosmic microwave background (Gaztañaga, Fosalba & Elizalde 2000).

However, since the support of the Laguerre expansion is non-negative, it cannot be applied directly to variables which can take either positive or negative val-

ues. One modification is to shift the variable by some $M > 0$ so that $X_L = X + M$ can be regarded as non-negative and then apply the Laguerre expansion to X_L to obtain the approximation $\hat{f}_L$ of the pdf of X_L. Hence f is approximated by $\hat{f}(x) = \hat{f}_L(x + M)$.

Furthermore, we can apply the Laguerre expansion to $X_{L^*} = (X + M)^2$, so that X_{L^*} is non-negative, and obtain approximation $\hat{f}_{L^*}$ of the pdf of X_{L^*}. Thus f is approximated by $\hat{f}(x) = 2(x+M)\hat{f}_{L^*}((x + M)^2)$. Let us call this the *Laguerre expansion with squaring*. We expect that it works for large enough M, so that $M + X$ can be regarded as positive[2]. Note that, since we deal with the squared variable, Equation (2.9) is not available for obtaining the conditional expectation of X, however, it still can be calculated by a linear combination of incomplete Gamma functions. See Appendix A.4 for details.

Since the left end of the support of these Laguerre expansions is bounded, the approximation quality of the Laguerre expansions at the left tail can be poor. This problem can be dealt with by applying these expansions to $-X$.

Such use of the Laguerre expansions has not been introduced in the literature yet, to the best of our knowledge.

Optimisation

Consider the case where we have unbiased estimators of $\mathrm{E}(g_k(X))$ and $(\mathrm{E}(g_k(X)))^2$, $k = 0, \ldots, n$. A typical situation is the case where we observe i.i.d. observations $X^{(1)}, \ldots, X^{(N)}$ from f. Also assume that f can be expanded as

$$f(x) = w(x) \sum_{k=0}^{\infty} C_k g_k(x),$$

where $C_k = \mathrm{E}(g_k(X))$. We might use $\hat{C}_k = N^{-1} \sum_{i=1}^{N} g_k(X^{(i)})$ as an estimator for $\mathrm{E}(g_k(X))$ and plug this into Equation (2.2) in order to construct an approximation

[2]See Appendix A.5 for discussion about the choice of M.

for f. Alternatively, we can consider another class of estimator

$$\hat{f}(x) = w(x) \sum_{k=0}^{n} \alpha_k \hat{C}_k g_k(x)$$

and choose the coefficients $\alpha_k \geq 0$, $k = 0, 1, \ldots, n$, to minimise the weighted mean integrated square error (MISE)[3]

$$\mathrm{E}\left(\int_{\mathcal{S}} \frac{\left\{ \hat{f}(u) - f(u) \right\}^2}{w(u)} \mathrm{d}u \right). \tag{3.1}$$

Here, the weight function w in the denominator of Equation (3.1), which is identical to the weight function in Equation (2.2), works in two ways. One is to put more importance on the error in the tail than that of the centre of the distribution. The other is that it makes Equation (3.1) tractable. In fact, it can be shown that the MISE in Equation (3.1) is equal to

$$\sum_{k=0}^{n} \alpha_k^2 \mathrm{E}\left(\hat{C}_k^2 \right) - 2 \sum_{k=0}^{n} \alpha_k C_k^2 + \sum_{k=0}^{\infty} C_k^2. \tag{3.2}$$

See the Appendix A.2 for details. An unbiased estimator for this MISE is given by

$$\sum_{k=0}^{n} \alpha_k^2 \hat{C}_k^2 - 2 \sum_{k=0}^{n} \alpha_k \frac{N\hat{C}_k^2 - \hat{B}_k^2}{N-1} + \sum_{k=0}^{\infty} \frac{N\hat{C}_k^2 - \hat{B}_k^2}{N-1}, \tag{3.3}$$

where $\hat{B}_k^2 = N^{-1} \sum_{i=1}^{N} \left\{ g_k(X^{(i)}) \right\}^2$. Now we consider $\{\alpha_k\}$ which minimises Equation (3.3). Firstly, $\alpha_0 = 1$ is required so that $\int_{\mathcal{S}} \hat{f}(u) \mathrm{d}u = 1$ is satisfied. If the variable is already standardised so that the first and second moments are identical to those of w, we have $C_1 = C_2 = 0$, and therefore we might set $\alpha_1 = \alpha_2 = 0$. For $k = 3, \ldots, n$, Equation (3.3) is minimised when

$$\alpha_k = \frac{N\hat{C}_k^2 - \hat{B}_k^2}{(N-1)\hat{C}_k^2},$$

[3] An almost equivalent discussion for approximating a discrete distribution is found in Hall (1983*b*). He considers a case with a different orthogonal system from ours in which MISE is not weighted.

if $N\hat{C}_k^2 - \hat{B}_k^2 > 0$, and $\alpha_k = 0$, otherwise.

This optimisation, as well as the use of Laguerre expansions, has not been introduced in the literature yet, to the best of our knowledge.

3.2 Numerical examples

Now we apply the expansion methods and related techniques to distributions of risk factors and show how they work.

We design our exhibition as follows. We firstly demonstrate in Section 3.2.1 how the expansion methods approximate the empirical distributions, given sample moments. As discussed in Chapter 2, empirical distributions naturally capture the characteristics of risk factors, but they suffer from discreteness especially in the tails. Therefore, continuous approximations for empirical distributions might be particularly desired. Furthermore, if the approximation is successful, it might be justified to summarise the information about the marginal distributions in terms of their moments, which can solve a part of the description problem discussed in Section 2.2.1. In Section 3.2.2, we apply the expansion methods to GARCH model as an example of expansion methods applied to a parametric model.

We have found from preliminary investigation that the approximation quality of expansions without standardisation is far from practical and therefore it will not be discussed in this thesis.

When we discuss the adequacy of approximations, we focus on the following outputs.

Tail plots. Measures of market risk such as VaR and ES concern tails of distribution functions. Here we focus on distribution functions, rather than density functions. The first reason for this is that it makes it possible to make a comparison between approximations and observations. That is, by comparing the

distribution functions obtained by expansions and other approximations with empirical distribution of observations, we can evaluate visually the quality of approximations in the tails. Secondly, distribution functions are more relevant to quantiles, which are necessary for calculation of VaR. Furthermore, we can use the closed form in Equation (2.5) or (2.8) for deriving distribution functions by expansion methods. Note that non-monotonicities in these plots indicate existence of negative density.

Measures of difference. In order to compare the overall fit of an approximated distribution function to an empirical distribution, we show the Kolmogorov-Smirnov (KS) test statistics[4] and root mean squared error (RMSE), which are defined as

$$\mathrm{KS} = \sup_{x \in \mathcal{S}} \{|\hat{F}(x) - S(x)|\} \tag{3.4}$$

and

$$\mathrm{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{F}(x^{(i)}) - S(x^{(i)}))^2}, \tag{3.5}$$

where $\hat{F}$ is the approximation for the distribution function, S is the empirical distribution function, and $x^{(i)}$ are the observations. KS and RMSE can be considered as measures of difference between $\hat{F}$ and S. Therefore, we might consider that smaller KS and RMSE indicate that $\hat{F}$ is fits better to the empirical distribution.

Total area of negative density. As discussed in Section 2.2.3, the approximations by expansion methods may have some regions where the density is negative. We define the total area of negative density of an approximation $\hat{f}$ for a pdf by

$$-\int_{\hat{f}(x)<0} \hat{f}(x)\mathrm{d}x = \frac{1}{2}\left(\int_{\mathcal{S}} |\hat{f}(x)|\mathrm{d}x - 1\right). \tag{3.6}$$

[4]See Massey (1951).

3.2.1 Approximating empirical distributions

Here we apply the expansion methods to the empirical distribution. We firstly apply the Hermite expansions to the risk factors' marginal distributions. The purpose of these examples is to investigate the conventional use of the Hermite expansion, which has already been studied by literature including those reviewed in Section 2.2.2. Then we apply our techniques introduced in Section 3.1 in order to show how they improve the conventional use of the Hermite expansion.

Four market variables, Nikkei 225 stock index daily log-returns, the daily log-differences of Black and Scholes implied volatility (IV)[5] of a call option (three month, at-the-money) on it, S&P 500 stock index daily log-returns, and the daily log-differences of IV of call option (three month, at-the-money) on it are chosen as examples of risk factors. See Table 3.1 for summaries of the data.

	N225	N225 IV	SP500	SP500 IV
Observation period	From 25/10/2004 to 6/11/2006		From 25/10/2004 to 18/10/2006	
Number of observations	500		500	
Mean ($\times 10^{-4}$)	8.574	−2.931	−4.425	−5.006
Std. dev. ($\times 10^{-2}$)	1.081	3.987	0.654	5.018
Skewness ($\times 10^{-1}$)	−2.269	3.125	0.715	−2.510
Kurtosis	4.013	4.957	3.376	4.782
Min ($\times 10^{-1}$)	−0.423	−1.375	−0.185	−2.275
Max ($\times 10^{-1}$)	0.352	1.996	0.213	1.910
Correl. coef.	−0.3612		−0.7443	

Table 3.1: Summary statistics of the data sets. Correlation coefficients are shown for later reference.

[5]See Section 5.2 for discussion on IV.

Approximation by the Hermite expansion

Here we investigate the approximation quality of the standardised Hermite expansions.

At the same time, we also show how the degree of expansion, n in Equation (2.4) changes the approximations. In existing studies, for instance, $n = 4$ (Gordy 2002, Feuerverger & Wong 2000, Gaztañaga et al. 2000), $n = 5$ (Barakat 1999) and $n = 8$ (Mauleón & Perote 2000*b*) are employed. Although criteria for the choice of n is not given in existing studies as far as we can determine, it seems that all employ $n = 4$ or larger so that the approximation captures important features such as skewness and kurtosis. Here, we compare expansions with degrees $n = 2$, 4, 8 and 16, similarly to Jaschke (2002). Note that the Hermite expansion with $n = 2$ is equivalent to an approximation by the Normal distribution.

In Figures 3.1 to 3.4 we show the tails of empirical distributions and approximations obtained by Hermite expansions. Tables 3.2 and 3.3 show the measures of overall difference, and Table 3.4 shows the total area of negative density.

We can summarise our findings as follows.

- Non-monotonicities are found in the Hermite expansion with a high degree such as $n = 8$ and 16 in Figure 3.2 and $n = 8$ in Figure 3.4. These demonstrate visually recognisable existence of negative probability density. In fact, Table 3.4 shows that the these three approximations have the largest total area of negative density. This typically suggests naive use of the Hermite expansion can perform poorly.

- For other examples, expansions resulted in better fits to the empirical distribution than approximation by the Normal distribution (Tables 3.2 and 3.3).

- A Smaller measure of difference does not necessarily mean a smaller total area of negative density (Tables 3.2 to 3.4).

We might consider searching for an optimal n, however, such optimisation may have the following drawbacks. Firstly, the optimisation has to combine a measure of difference and a total area of negative density; however, such criteria may be determined arbitrarily. For instance, in the Nikkei 225 example, $n = 16$ fits better to the empirical distribution while $n = 8$ has smaller total area of negative density, and determining which is the better can be a tricky problem.

Secondly, the optimisation has to be done for each example. For instance, the optimal n for Nikkei 225 might be around 8 or even larger, while the optimal n for Nikkei 225 IV might be around 4, and expansion with $n = 8$ can be useless. This particularly indicates that the optimal n can vary largely depending on the data considered, and therefore n may not have a convenient value that works fairly well for a wide range of examples.

Instead, our approach is to develop techniques which can provide robustness, with which approximations work fairly well for a wide range of situations, rather than to choose an optimal n.

In the next set of examples, we work with cases with $n = 8$. The Hermite expansion with $n = 8$ exhibits visually recognisable negative densities (Figures 3.2 and 3.4). Therefore cases with $n = 8$ allow us to see if our techniques introduced in Section 3.1 provide robustness. Also, expanding with $n = 8$ covers the examples studied by literature mentioned above, except for $n = 16$ by Jaschke (2002), who compared the results from $n = 2$, 4, 8 and 16, and commented that the approximation with $n = 16$ is the least reliable. In the next set of examples, we show how our techniques work using the same data sets.

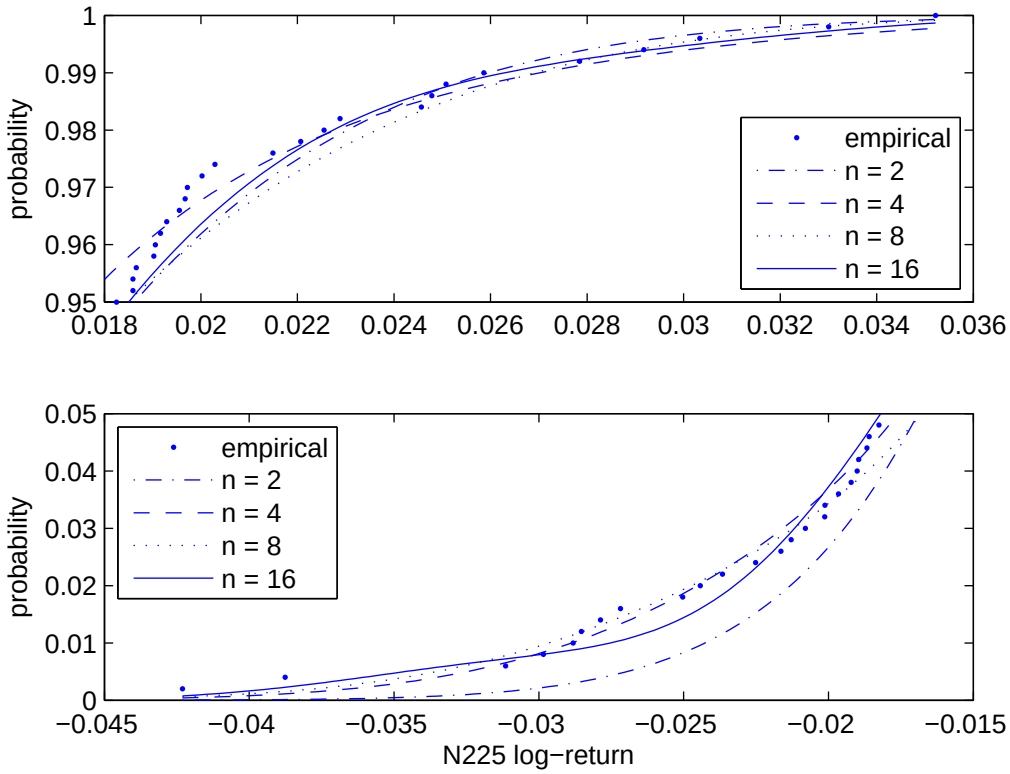

Figure 3.1: The empirical distribution function of daily log-returns of Nikkei 225 and Hermite expansions applied to it: n corresponds to the degree of expansion in Equation (2.4). Upper and lower plots show the right tails and left tails, respectively. From the lower plots, we can see that expansions capture the heavy tail of the empirical distribution better than the approximation by the Normal distribution ($n = 2$). See Tables 3.2 and 3.3 for the overall difference, and Table 3.4 for the total area of negative density.

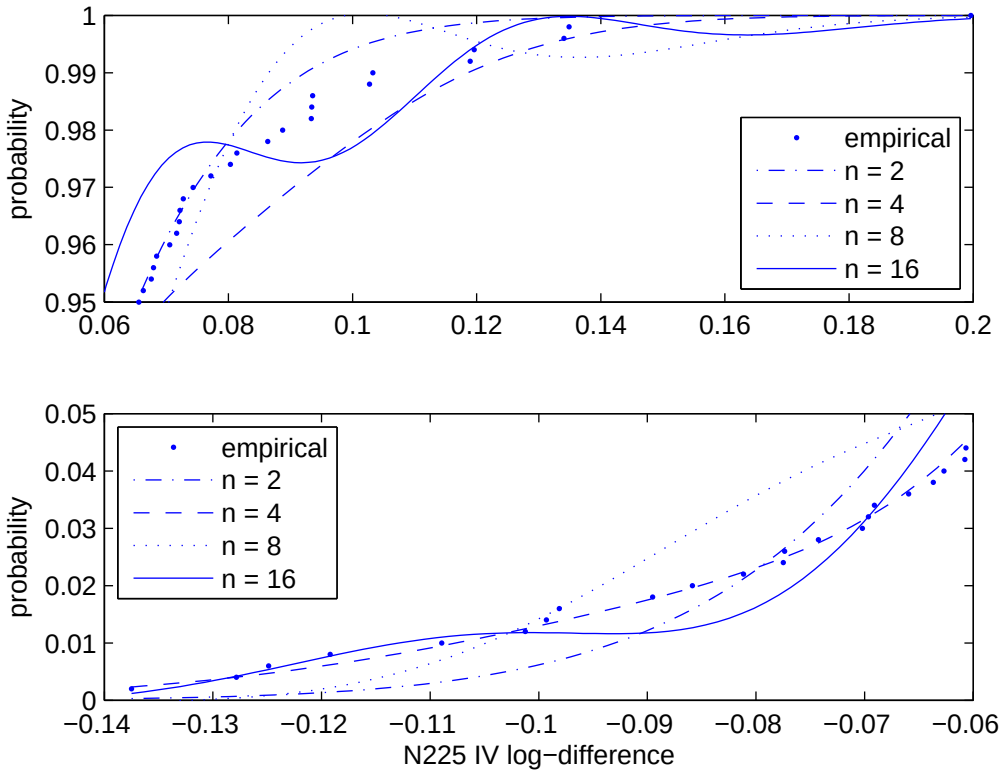

Figure 3.2: The empirical distribution function of daily log-differences of Black and Scholes implied volatility of Nikkei 225 call option (three months, at-the-money) and Hermite expansions applied to it: n corresponds to the degree of expansion in Equation (2.4). Upper and lower plots show the right tails and left tails, respectively. Upper plots show that expansions with $n = 8$ and 16 can be very poor approximations. See Tables 3.2 and 3.3 for the overall difference, and Table 3.4 for the total area of negative density.

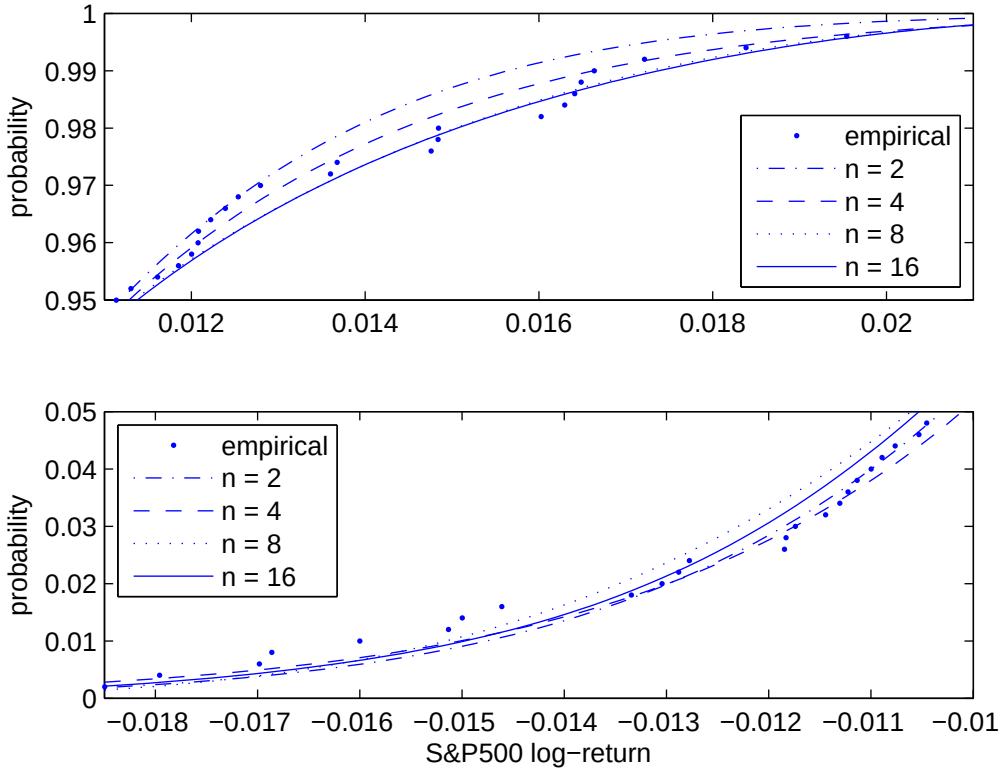

Figure 3.3: The empirical distribution function of daily log-returns of S&P 500 and Hermite expansions applied to it: n corresponds to the degree of expansion in Equation (2.4). Upper and lower plots show the right tails and left tails, respectively. Upper plots show that expansions capture the heavy tail of empirical distribution better than the approximation by the Normal distribution ($n = 2$). See Tables 3.2 and 3.3 for the overall difference, and Table 3.4 for the total area of negative density.

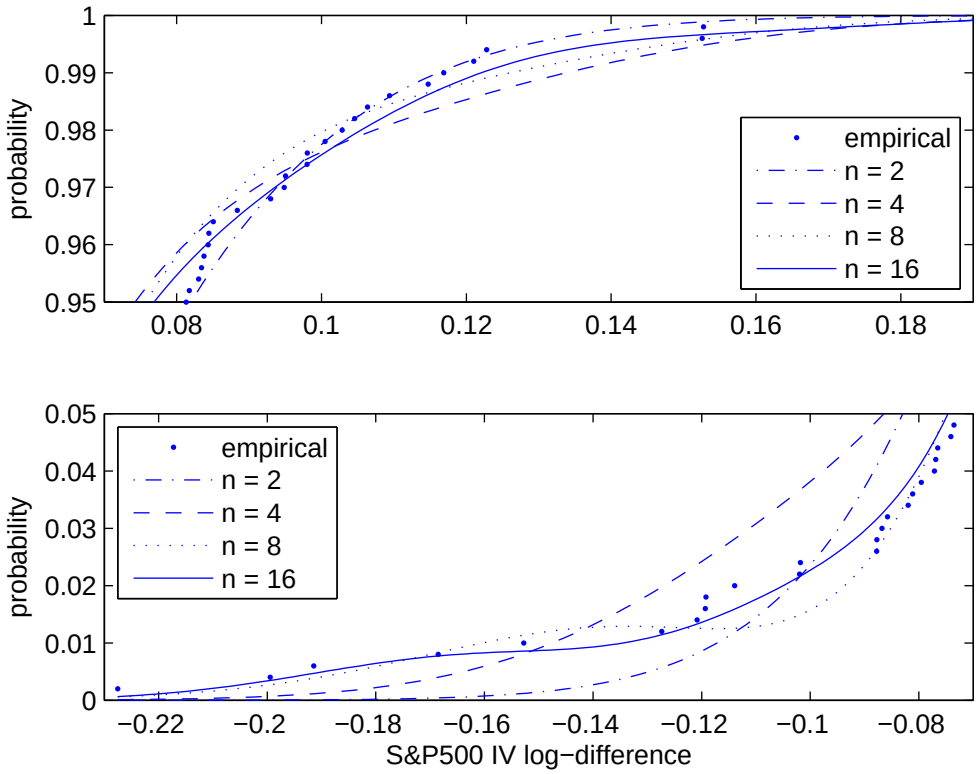

Figure 3.4: The empirical distribution function of daily log-differences of Black and Scholes implied volatility of S&P 500 call option (three months, at-the-money) and Hermite expansions applied to it: n corresponds to the degree of expansion in Equation (2.4). Upper and lower plots show the right tails and left tails, respectively. Lower plots show that the expansion with $n = 8$ shows slight non-monotonicity. See Tables 3.2 and 3.3 for the overall difference, and Table 3.4 for the total area of negative density.

$(\times 10^{-2})$	Nikkei 225	Nikkei 225 IV	S&P 500	S&P 500 IV
$n = 2$ (Normal)	6.336	4.445	3.944	3.640
$n = 4$	4.709	3.273	3.543	6.081
$n = 8$	3.895	5.067	2.807	2.697
$n = 16$	2.511	3.839	2.709	2.214

Table 3.2: KS test statistics defined by Equation (3.4): n denotes the degree of expansion in Equation (2.4). we use this test statistic as a measure of overall difference between an approximation and the empirical distribution function. See Figures 3.1 to 3.4 for the tail fit. In the statistical testing sense, the null hypothesis that the distribution of observations is identical to the approximation is rejected with 10% significance only for $n = 2$ for Nikkei 225 and $n = 4$ for S&P 500 IV.

$(\times 10^{-2})$	Nikkei 225	Nikkei 225 IV	S&P 500	S&P 500 IV
$n = 2$ (Normal)	2.916	2.271	1.885	1.604
$n = 4$	2.082	1.513	1.605	2.335
$n = 8$	1.643	2.965	1.065	0.830
$n = 16$	0.887	1.810	0.883	0.726

Table 3.3: RMSE defined by Equation (3.5): n denotes the degree of expansion in Equation (2.4). This also can be used as a measure of overall difference of an approximation to the empirical distribution function. See Figures 3.1 to 3.4 for the tail fit.

	Nikkei 225	Nikkei 225 IV	S&P 500	S&P 500 IV
$n = 2$ (Normal)	0	0	0	0
$n = 4$	0	0	0	0
$n = 8$	1.106×10^{-5}	8.010×10^{-3}	1.376×10^{-4}	4.107×10^{-4}
$n = 16$	1.430×10^{-4}	7.561×10^{-3}	2.437×10^{-4}	4.999×10^{-5}

Table 3.4: Total area of negative density defined by Equation (3.6): n denotes the degree of expansion in Equation (2.4). See Figures 3.1 to 3.4 for the tail.

Comparison of techniques

In this next set of examples, we compare optimised Hermite expansions and Laguerre expansions with squaring with Hermite expansions. Since we saw in the previous set of examples that Hermite expansions performed poorly for Nikkei 225 IV and S&P 500 IV, it is especially interesting to see if expansions with techniques introduced in Section 3.1 perform better for these two cases.

Figures 3.5 to 3.8 show tails of the empirical distributions and approximations using the Hermite expansion, the optimised Hermite expansion, the Laguerre expansion with squaring, and the Normal distribution. Tables 3.5 to 3.7 compare their measures of difference and total area of negative density. The Hermite expansions are the same as those in the previous set of examples. For all expansions, $n = 8$, as mentioned above.

From these Figures and Tables, we can point out the following:

- The optimised Hermite expansion and Laguerre expansion with squaring are stable in all cases, and the total area of negative density is smaller for those than for the Hermite expansion. This suggests that the expansions with the techniques introduced in Section 3.1 are more robust, at least for these examples.

- Except in the cases where negative density is observed, the Hermite expansion is almost identical to the Laguerre expansion with squaring (lower plots of Figure 3.5 and both plots of Figure 3.7) or its approximation quality is comparable to that of the Laguerre expansion with squaring.

- From Tables 3.5 and 3.6, we can see that approximation quality of expansions is much better than the Normal distribution, in the sense that it is closer to the empirical distribution, except for the Hermite expansion for Nikkei 225 log-return.

- Figures 3.5 to 3.8, show that the approximations by expansion have heavier tails than the Normal distribution, except in the upper plots in Figure 3.5.

We might conclude that the optimised Hermite expansion and the Laguerre expansion with squaring are more robust than naive Hermite expansion, and that the overall fit of these expansions is better than that of a Normal approximation, at least for our four examples.

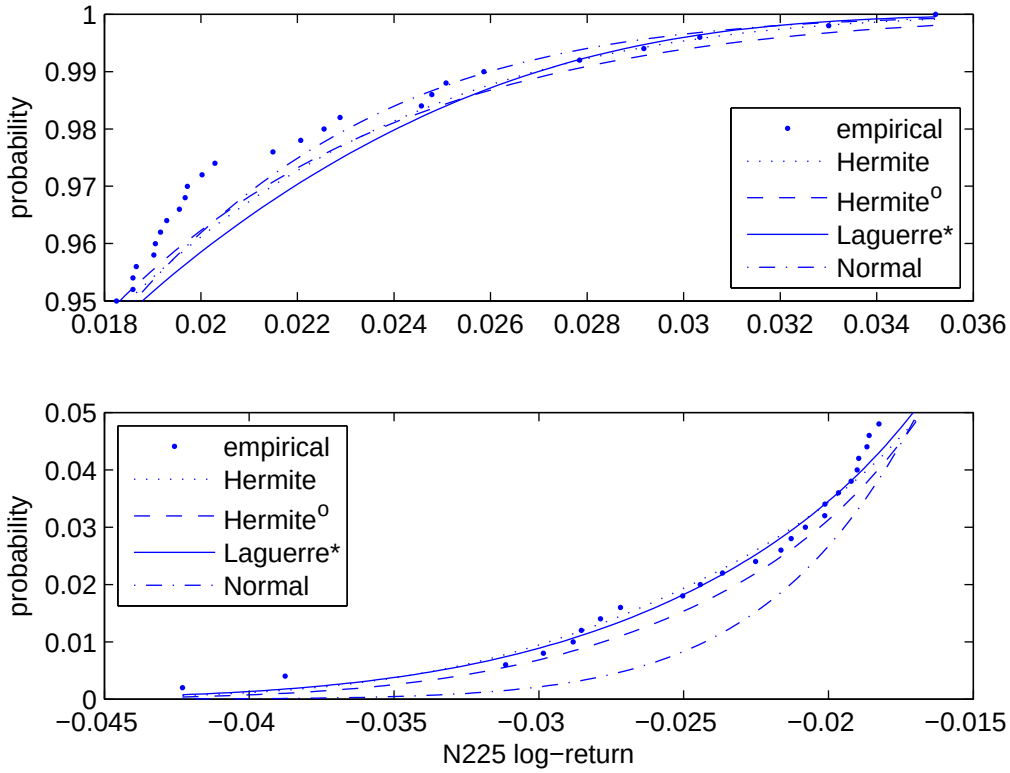

Figure 3.5: The empirical distribution function of daily log-returns of Nikkei 225 and three expansions applied to it. 'Hermite^{o}' and 'Laguerre*' correspond to the optimised Hermite expansion and to the Laguerre expansion with squaring respectively: $n = 8$ in Equation (2.2) for all expansions. Upper and lower plots show the right tails and left tails respectively. Lower plots show that expansions capture the heavy tail of empirical distribution better than the approximation by the Normal distribution. See Tables 3.5 and 3.6 for the overall difference.

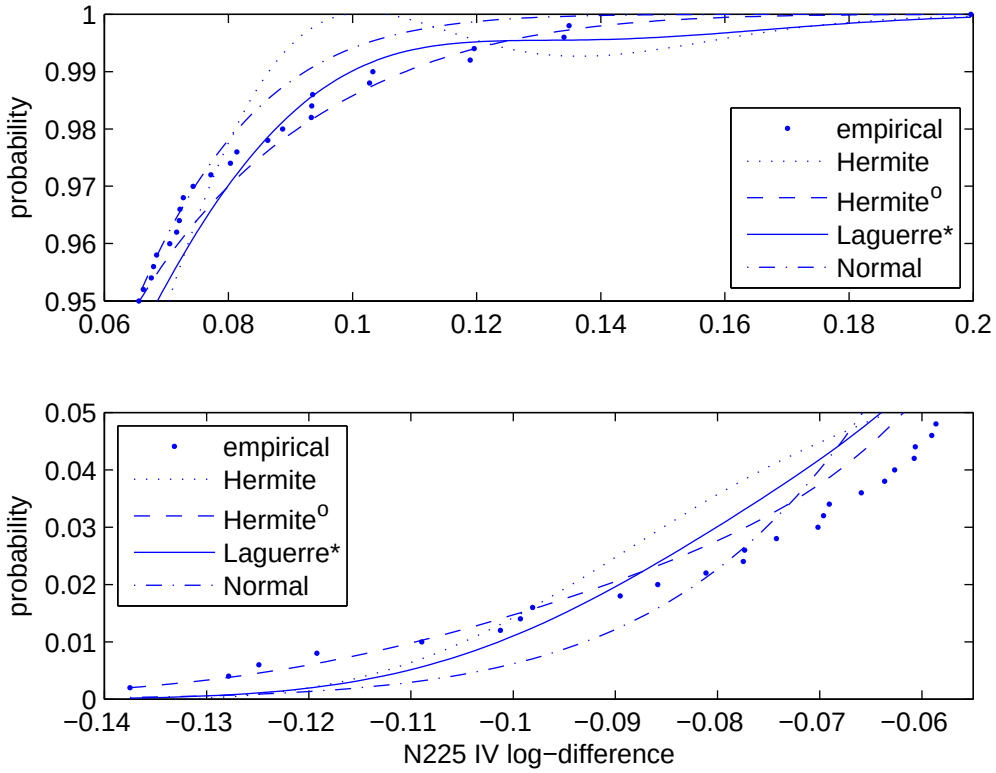

Figure 3.6: The empirical distribution function of daily log-differences of Black and Scholes implied volatility of Nikkei 225 call option (three months, at-the-money) and three expansions applied to it. 'Hermite°' and 'Laguerre*' correspond to the optimised Hermite expansion and to the Laguerre expansion with squaring respectively: $n = 8$ in Equation (2.2) for all expansions. Upper and lower plots show the right tails and left tails respectively. We can see that the optimised Hermite expansion and the Laguerre expansion with squaring are more stable than the Hermite expansion. See Tables 3.5 and 3.6 for the overall difference.

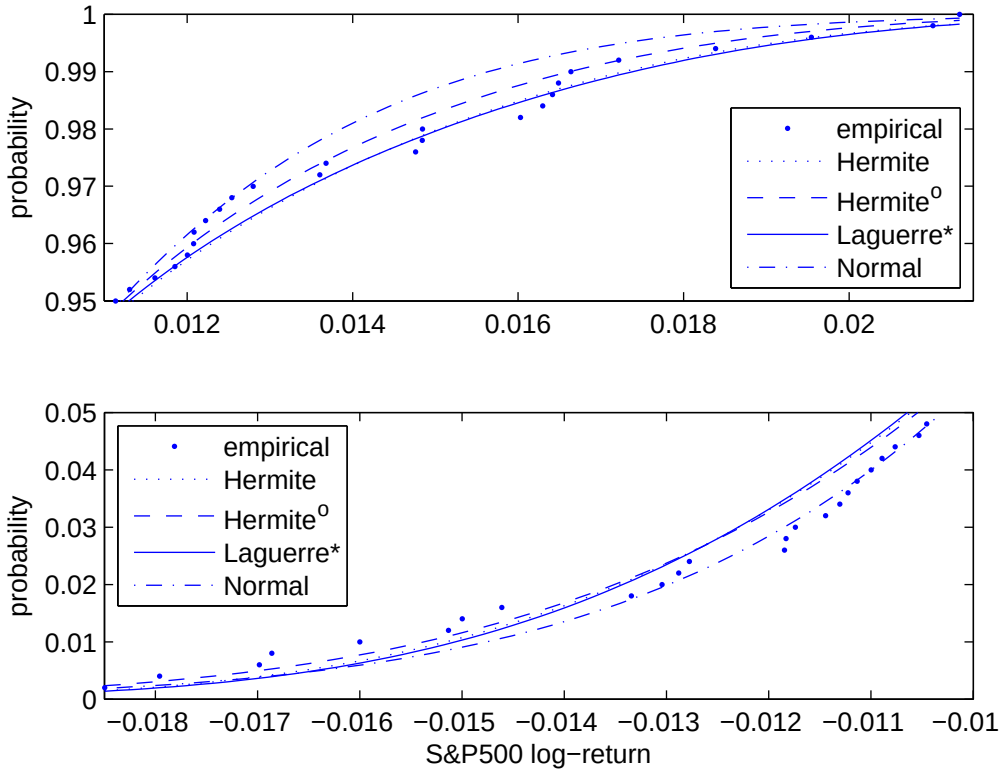

Figure 3.7: The empirical distribution function of daily log-returns of S&P 500 and three expansions applied to it. 'Hermite^{o}' and 'Laguerre*' correspond to the optimised Hermite expansion and to the Laguerre expansion with squaring respectively: $n = 8$ in Equation (2.2) for all expansions. Upper and lower plots show the right tails and left tails respectively. We can see that the Hermite expansion and the Laguerre expansion with squaring are almost identical to each other. See Tables 3.5 and 3.6 for the overall difference.

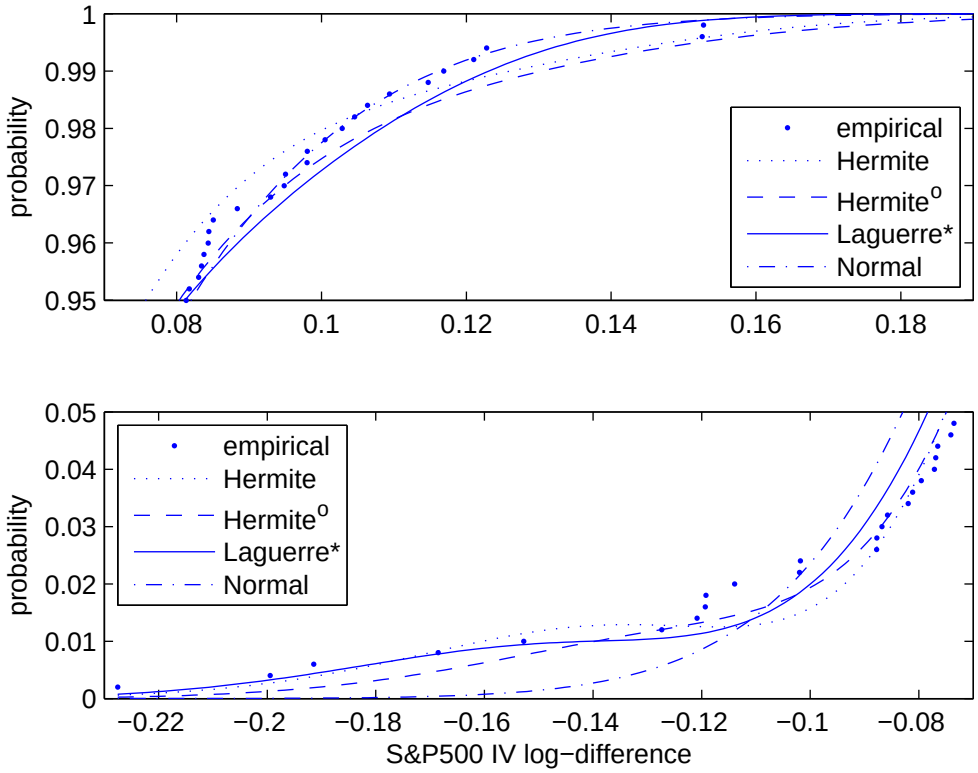

Figure 3.8: The empirical distribution function of daily log-differences of Black and Scholes implied volatility of S&P 500 call option (three months, at-the-money) and three expansions applied to it. 'Hermite°' and 'Laguerre*' correspond to the optimised Hermite expansion and to the Laguerre expansion with squaring respectively: $n = 8$ in Equation (2.2) for all expansions. Upper and lower plots show the right tails and left tails respectively. Lower plots show that the optimised Hermite expansion and the Laguerre expansion with squaring are more stable than the Hermite expansion. See Tables 3.5 and 3.6 for the overall difference.

($\times 10^{-2}$)	Nikkei 225	Nikkei 225 IV	S&P 500	S&P 500 IV
Hermite	3.895	5.067	2.807	2.696
Hermite^{o}	4.602	3.037	3.288	3.560
Laguerre*	3.558	3.417	2.824	3.901
Normal	6.335	4.445	3.943	3.639

Table 3.5: KS test statistics defined by Equation (3.4). We use this test statistic as a measure of overall difference between an approximation and the empirical distribution function. See Figures 3.5 to 3.8 for the tail fit. In the statistical testing sense, the null hypothesis that the distribution of observations is identical to the approximation is rejected with 10% significance only for the Normal approximation for Nikkei 225.

($\times 10^{-2}$)	Nikkei 225	Nikkei 225 IV	S&P 500	S&P 500 IV
Hermite	1.642	2.965	1.065	0.830
Hermite^{o}	1.964	0.931	1.249	1.460
Laguerre*	1.579	1.359	1.062	1.276
Normal	2.916	2.271	1.885	1.604

Table 3.6: RMSE defined by Equation (3.5). This also can be used as a measure of overall difference of an approximation to the empirical distribution function. See Figures 3.5 to 3.8 for the tail fit.

	Nikkei 225	Nikkei 225 IV	S&P 500	S&P 500 IV
Hermite	1.106×10^{-5}	8.010×10^{-3}	1.376×10^{-4}	4.107×10^{-4}
Hermite^{o}	6.504×10^{-9}	1.046×10^{-6}	9.831×10^{-5}	0
Laguerre*	9.451×10^{-6}	1.931×10^{-6}	1.128×10^{-4}	2.414×10^{-5}
Normal	0	0	0	0

Table 3.7: Total area of negative density defined by Equation (3.6). See Figures 3.5 to 3.8 for the tail.

For completeness, we also subjected data from the S&P 500, in the period 1 September 2000 to 4 September 2002, to the same analyses. The index returns exhibited much greater volatility than the sample period in the previous examples (see Tables 3.1 and 3.8). In brief, our empirical findings were not challenged.

	SP500 (2)
Observation period	From 1/9/2000 to 4/9/2002
Number of observations	500
Mean ($\times 10^{-4}$)	-10.639
Std. dev. ($\times 10^{-2}$)	1.446
Skewness ($\times 10^{-1}$)	2.716
Kurtosis	4.265
Min ($\times 10^{-1}$)	-0.505
Max ($\times 10^{-1}$)	0.557

Table 3.8: Summary statistics of the more volatile data set.

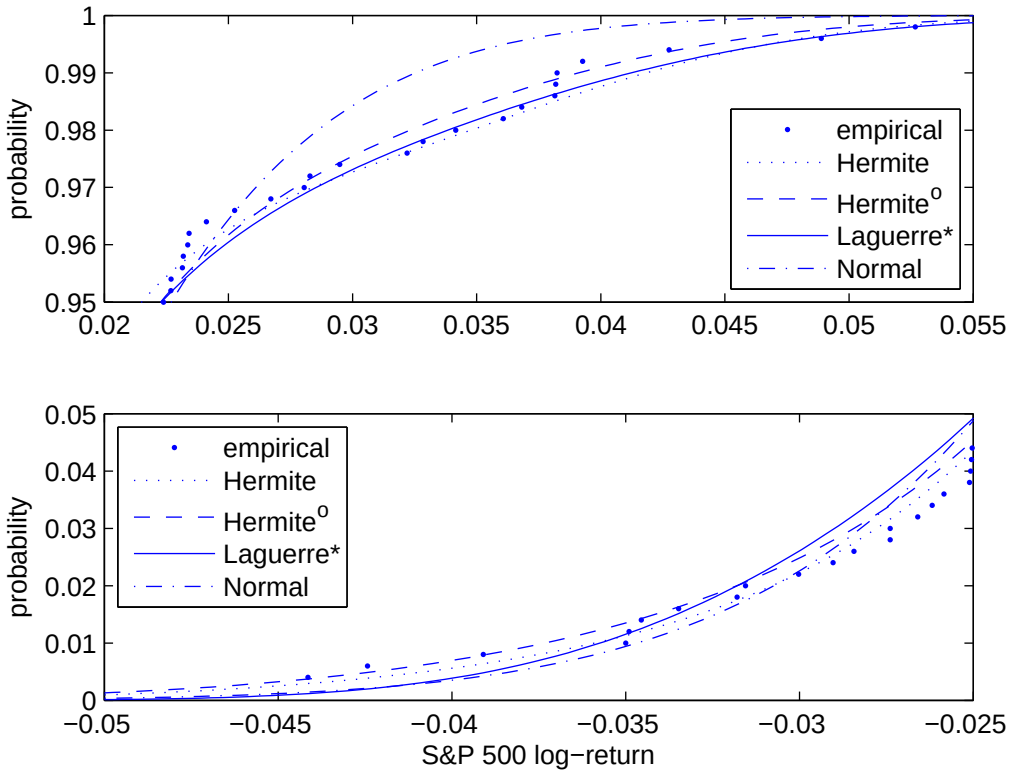

Figure 3.9: The empirical distribution function of daily log-returns of S&P 500 and three expansions applied to it. The observation period is different from that in Figure 3.7 (see Table 3.8). 'Hermite°' and 'Laguerre*' correspond to the optimised Hermite expansion and to the Laguerre expansion with squaring respectively: $n = 8$ in Equation (2.2) for all expansions. Upper and lower plots show the right tails and left tails respectively. We can see that the expansions capture the heavy right tail better than the approximation by Normal. See Table 3.9 for the overall difference.

	f^-	KS ($\times 10^{-2}$)	RMSE ($\times 10^{-2}$)
Hermite	1.960×10^{-5}	2.468	1.022
Hermite$^{\text{o}}$	5.138×10^{-7}	2.321	1.037
Laguerre*	2.389×10^{-5}	2.530	0.908
Normal	0	4.751	2.285

Table 3.9: Total area of negative density (f^-) defined by Equation (3.6), KS test statistics defined by Equation (3.4) and RMSE defined by Equation (3.5), of daily log-return of S&P 500. The observation period is different from that in Tables 3.5 to 3.7. See Table 3.8 for summary statistics and Figure 3.9 for the tail.

3.2.2 Application to GARCH models

GARCH models are widely used in studies and practices in finance. As discussed in Section 2.2.2, they are capable of describing the autocorrelation of volatility and of realising heavier tails than the Normal distribution. However, the closed form for associated distributions may not be available. On the other hand, expansion methods only require the moments of the variable, which can be calculated for GARCH models, as long as they exist and are finite. In this section, as an example of the application of expansion methods to a parametric model, we discuss the GARCH models.

We present two sets of numerical examples: a case where the variable has heavy tails, and a case where the variable has light tails. Note that the optimisation is not available here, since the unbiased estimator for MISE is not available.

GARCH models

Let $\{X(t) : t = 1, 2, \ldots\}$ be a process adapted to a filtration[6] $\{\mathcal{F}(t)\}$. Then $\{X(t)\}$ is said to be a GARCH(p, q) model if it satisfies

$$\mathrm{E}(X(t)|\mathcal{F}(t-1)) = 0$$

and

$$\mathrm{E}(X(t)^2|\mathcal{F}(t-1)) = \sigma(t)^2 = c + \sum_{i=1}^{p} a_i X(t-i)^2 + \sum_{j=1}^{q} b_j \sigma(t-j)^2, \qquad (3.7)$$

where c, $a_1, \ldots, a_p$, $b_1, \ldots, b_q$ are positive parameters. Note that $\{\sigma(t)\}$ is $\{\mathcal{F}(t-1)\}-$adapted. We focus on the GARCH model with Gaussian innovations, $X(t)|\mathcal{F}(t-1) \sim N(0, \sigma(t)^2)$ where $\sigma(t)^2$ is given in Equation (3.7), which is the most frequently discussed. It is convenient to denote $X(t) = \sigma(t)W(t)$, where $\{W(t)\}$ is an i.i.d. Normal sequence adapted to $\{\mathcal{F}(t)\}$.

Unconditional distribution in the stationary state

From Equation (3.7), we can see that the distribution of $X(t)$ depends on the time t. However, $X(t)$ may converge to a state in which the distribution f of $X(t)$ does not depend on time as $t \to \infty$. Let us call it the stationary state[7] and the distribution f the unconditional distribution in the stationary state, or simply the unconditional distribution.

In this set of examples, we aim to approximate the unconditional distribution in the stationary state, for which a closed form is not available[8].

[6] See Appendix A.3 for definitions of terms related with stochastic processes.

[7] For conditions for existence of a stationary state see, for example, Gouriéroux (1997). For GARCH models, the stationary state is unique, if it exists.

[8] The conditional distribution of $X(t)$ given $\mathcal{F}(t-1)$ is the Normal distribution with known variance and therefore dealing with such conditional distributions can be trivial.

Unconditional moments of the GARCH variables

Even when the GARCH process is in a stationary state, the unconditional moments of the variable may not be finite[9]. However, as long as they exist and are finite, we can use the following procedure to obtain the unconditional moments analytically.

The unconditional variance of $X(t)$ can be derived by solving

$$m_2 = c + \sum_{i=1}^{p} a_i m_2 + \sum_{j=1}^{q} b_j m_2$$

for m_2.

For $p = q = 1$ in Equation (3.7), the higher order moments can be similarly derived by solving

$$m_{2k} = \sum_{i=0}^{k} \sum_{j=0}^{i} \frac{k!}{(k-i)!(i-j)!j!} c^{k-i} a_1^j b_1^{i-j} \frac{\mu_{2j}^{\mathrm{N}} \mu_{2k}^{\mathrm{N}}}{\mu_{2i}^{\mathrm{N}}} m_{2i} \tag{3.8}$$

for m_{2k}, where

$$\mu_{2k}^{\mathrm{N}} = \frac{(2k)!}{2^k k!}$$

is the $2k$th moment of the Normal distribution, and $m_{2k-1} = 0$ from symmetry.

In this case, the autocross-moments $\mathrm{E}(X(t)^{2k} X(t-s)^{2l})$ can be derived by recurrence. From Equation (3.7) we have

$$\sigma(t)^2 = c + (a_1 W(t-1)^2 + b_1)\sigma(t-1)^2.$$

By applying this s times, $X(t)^{2k} X(t-s)^{2l}$ can be expressed in terms of $\sigma(t-s)$ and $W(t), \ldots, W(t-s)$ only. Since $W(t), \ldots, W(t-s)$ are i.i.d. Normal, the autocross-moment can be calculated from up to the $2(k+l)$th moments of $X(t)$. Also from symmetry, we have $\mathrm{E}(X(t)^{2k-1} X(t-s)^{2l}) = \mathrm{E}(X(t)^{2k} X(t-s)^{2l-1}) = \mathrm{E}(X(t)^{2k-1} X(t-s)^{2l-1}) = 0$.

[9] See, for example, Carnero, Peña & Ruiz (2004).

For general cases with $p \geq 2$ or $q \geq 2$ in Equation (3.7), solving a system of linear equations is required in order to obtain the higher order moments and autocross-moments[10].

Heavy tailed case

Here we show application of the Hermite expansion and the Laguerre expansion with squaring to a GARCH(1, 1) model with heavy tails. A set of parameters $a_1 = 0.13$, $b_1 = 0.82$ and $c = 0.1$ in Equation (3.7) realises an unconditional kurtosis of 4.592, which is within the range of sample kurtosis values found in Table 3.1.

For this example, only the moments up to 6th order exist, and therefore, the degree of expansion employed is $n = 6$ for the Hermite expansion, and $n = 3$ for the Laguerre expansion with squaring. This can be an interesting example, since it exhibits how expansion methods perform when higher moments do not exist. From the properties of expansion methods discussed in Section 2.2.3, we find that the higher order moments, which do not exist for the 'true' distribution, are replaced by those of the weight functions.

Figure 3.10 shows the approximations for the distribution function of the GARCH model by the Hermite expansion, the Laguerre expansion with squaring and the Normal distribution. Since the closed form for the 'true' distribution is not available, we show in Figure 3.10 the empirical distribution of 5,000 samples from the MC method instead. The measures of difference in Table 3.10 are for

[10]As an example, we describe the procedure of deriving unconditional 4th moment of the GARCH(2,2). Consider $\mathrm{E}(X(t)^4) = \mathrm{E}(\sigma(t)^4)\mu_4^{\mathrm{N}}$. Expanding the right hand side using Equation (3.7) and the stationarity show that it involves the terms associated with $\mathrm{E}(\sigma(t)^4)$, $\mathrm{E}(\sigma(t)^2\sigma(t-1)^2)$, $\mathrm{E}(\sigma(t)^2X(t-1)^2)$, $\mathrm{E}(\sigma(t)^2)$ and a constant. By expanding $\mathrm{E}(X(t)^2X(t-1)^2)$ and $\mathrm{E}(X(t)^2X(t-2)^2)$ in similar ways, we have two other equations with terms associated with the same moments only. The 4th moment, as well as the autocross-moments $\mathrm{E}(\sigma(t)^2\sigma(t-1)^2)$ and $\mathrm{E}(\sigma(t)^2X(t-1)^2)$, can be derived by solving this system of three linear equations.

the difference from the MC sample empirical distribution and not from the 'true' distribution.

From Figure 3.10 and Table 3.10, we find that the Hermite expansion can be a very poor approximation, while the Laguerre expansion with squaring is stable. The tail plots suggest that the Laguerre expansion with squaring has heavier tails than the Normal distribution, and that the Laguerre expansion with squaring agrees with the empirical distribution of the MC samples in the tail. By comparing the measures of difference, we might say that the approximation quality of the Laguerre expansion with squaring can be slightly better than that of the Normal distribution; however, it also must be noted that they are the measures of difference from the MC sample empirical distribution, not from the true distribution.

With respect to the point that we apply expansion methods to a distribution which does not have higher order moments, the results from this particular example do not exhibit obvious evidence that the Laguerre expansion with squaring is an inappropriate approximation.

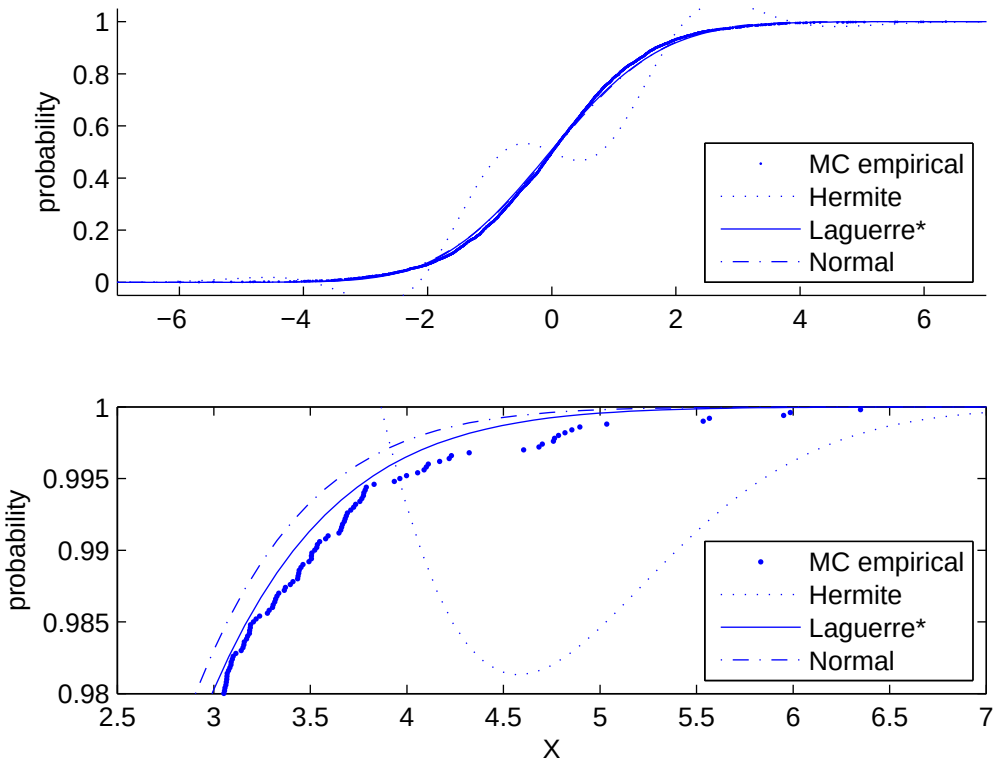

Figure 3.10: Approximations for the unconditional distribution function of the GARCH(1, 1) model with $a_1 = 0.13$, $b_1 = 0.82$ and $c = 0.1$ in Equation (3.7). The empirical distribution function of 5,000 samples from MC method is also shown. 'Laguerre*' corresponds to the Laguerre expansion with squaring: $n = 6$ in Equation (2.2) for the Hermite expansion and $n = 3$ for the Laguerre expansion with squaring. Upper plots show the overall fit and lower plots magnify the right tails. They show that the Hermite expansion is a very poor approximation, while the Laguerre expansion with squaring is stable. We can also see that the Laguerre expansion with squaring captures the heavy tail better than the Normal approximation. See Table 3.10 also.

	f^-	KS ($\times 10^{-2}$)	RMSE ($\times 10^{-2}$)
Hermite	0.253	23.760	15.651
Laguerre*	0	2.734	1.521
Normal	0	3.061	1.601

Table 3.10: The total area of negative density (f^-), KS test statistics and RMSE, defined in Equations (3.6), (3.4) and (3.5), respectively, of the three approximations shown in Figure 3.10. KS and RMSE measure the difference from the MC sample empirical distribution.

Light tailed case

Here we show applications of the Hermite expansion and the Laguerre expansion with squaring to a GARCH(1, 1) model with light tails. A set of parameters $a_1 = 0.05$, $b_1 = 0.9$ and $c = 0.1$ in Equation (3.7) realises an unconditional kurtosis 3.1622, which is closer to the Normal distribution than the previous example. For this example, the moments up to 16th order exist, and therefore, we employ $n = 8$ as the degree of expansion for both expansions. We also show the empirical distribution of 5,000 samples from MC method. Again, the measures of difference in Table 3.11 are the differences from the MC sample empirical distribution.

From Figure 3.11 and Table 3.11, we find that all approximations perform fairly well, including the approximation by the Normal distribution. We also find that expansion methods have slightly heavier tails than the Normal (lower plots of Figure 3.11), which might be reflecting a slightly larger kurtosis than that of the Normal distribution. By comparing the measures of difference, we might say that the approximation quality of the Hermite expansion can be the best of the three, however, again it also must be noted that they are the measures of difference from the MC sample empirical distribution, not from the true distribution.

For completeness, we also applied the analysis to a GARCH(1, 1) model with a large persistence parameter b_1; that is, we used $a_1 = 0.01$, $b_1 = 0.95$ and $c = 1.0 \times 10^{-5}$. It resulted in a distribution very close to the Normal distribution, with a kurtosis value of 3.008, and all the approximations were almost identical to each other. Figures are omitted for brevity.

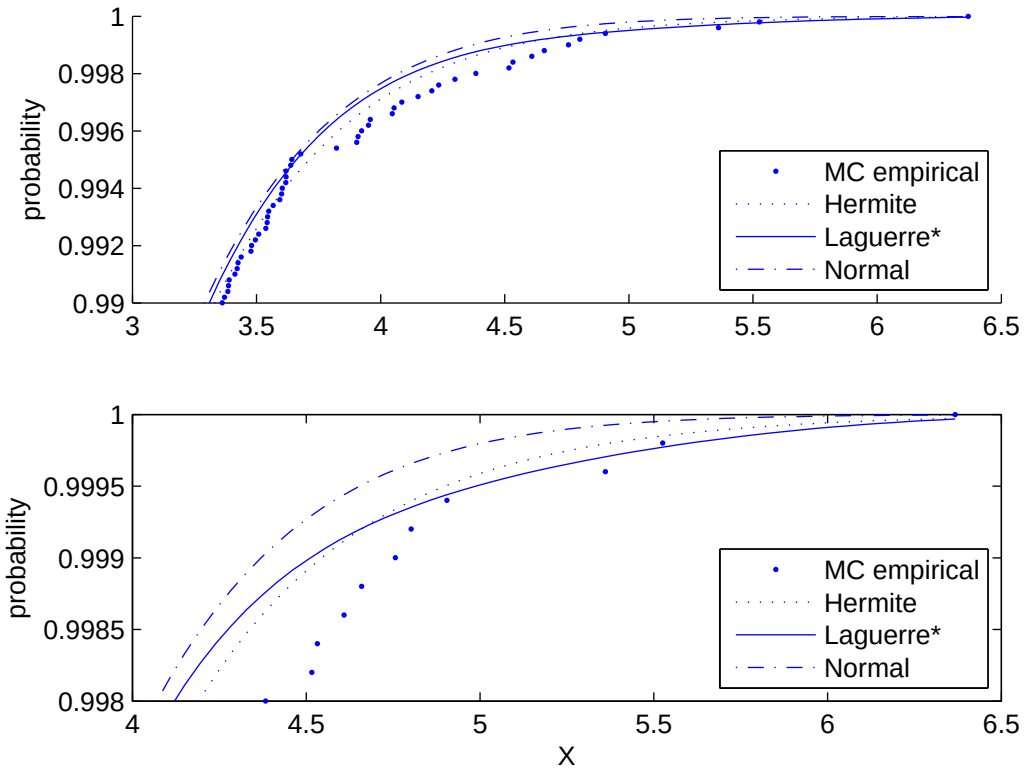

Figure 3.11: Approximations for the unconditional distribution function of GARCH(1, 1) model with $a_1 = 0.05$, $b_1 = 0.9$ and $c = 0.1$ in Equation (3.7). The empirical distribution function of 5,000 samples from MC method is also shown. 'Laguerre*' corresponds to the Laguerre expansion with squaring: $n = 8$ in Equation (2.2) for all expansions. Upper plots show the right tails and lower plots magnify far right tails. We can see that all approximations are performing well, while lower plots show that expansions capture the heavy tails slightly better than the approximation by the Normal distribution. See Table 3.11 also.

	f^-	KS ($\times 10^{-2}$)	RMSE ($\times 10^{-3}$)
Hermite	0	1.108	4.404
Laguerre*	5.726×10^{-6}	1.443	5.636
Normal	0	1.208	4.895

Table 3.11: The total area of negative density (f^-), KS test statistics and RMSE, defined in Equations (3.6), (3.4) and (3.5), respectively, of the three approximations in Figure 3.11. KS and RMSE measure the difference from the MC sample empirical distribution.

3.3 Summary of this Chapter

In this Chapter, we considered applications of expansion methods to marginal distributions of risk factors. We firstly introduced techniques to provide robustness to the conventional use of the Hermite expansion. Note that the two techniques, the use of Laguerre expansions and optimisation, have not appeared in literature to the best of our knowledge.

Then we applied expansions to the empirical distributions of risk factors using real markets' observations, and to GARCH models, as an example of parametric models.

We found cases where the Hermite expansion performs poorly, from both empirical and GARCH examples. Then we also showed that expansions with our techniques can provide fairly good approximations for those examples in which the Hermite expansions performs poorly.

For the GARCH example with heavy tails, where moments higher than 6th order do not exist, the approximation by the Laguerre expansion with squaring agreed with the empirical distribution of MC samples. This suggests the possibility that expansion methods can perform well even though the target distribution does not have some higher order moments.

If our purpose is only to obtain a continuous approximation which fits well to a given empirical distribution, some smoothing technique such kernel[11] might perform better. Further work would usefully involve comparisons of expansion methods with such competing methods. However, expansion methods can also be applied to a parametric model as shown in Section 3.2.2, where an empirical distribution is not available.

At least for our examples in this Chapter, expansion methods successfully approximated the distributions given the moments of the risk factors. Considering the properties of expansion methods discussed in Section 2.2.3 and the construction of related techniques discussed in Section 3.1, we expect that expansion methods can be successfully applied more widely to examples other than those in this Chapter.

This can solve a part of description problem stated in Chapter 2: the marginal distributions of risk factors can be described in terms of their moments.

Furthermore, since expansion methods are now shown to have some robustness, they might be applied to a wider range of situations than ever. Therefore, expansion methods might be able to deal with other problems discussed in Chapter 2. This will be the purpose of the following Chapters.

[11] See Hall (1983*a*), for instance.

Chapter 4

Application to Time Aggregation

As discussed in Chapter 2, we are often interested in distributions of risk factors and PL with a time horizon longer than one day, such as $T = 10$ days. Such time horizons might be determined with relation to the time needed to eliminate market risk. Also, the Basel Committee (1996) suggests the use of VaR with a ten-day time horizon.

We also discussed in Section 2.2.1 that it can be particularly difficult to infer distributional properties or features of risk factors with a long time horizon, due to unavailability of sufficiently large samples. Instead, we might describe a risk factor with a T-day time horizon X^T using a series of the risk factor with a one-day time horizon $X^1(t)$, $t = 1, \ldots, T$, by

$$X^T = X^1(1) + \cdots + X^1(T), \tag{4.1}$$

or similarly using transformations such as logarithms. Then the time aggregation problem is that of deriving the distribution of the sum of T random variables. However, as reviewed in Section 2.2.2, current practices often face difficulty in obtaining the distribution of X^T even using Equation (4.1).

In this Chapter, we discuss applications of expansion methods to the time aggregation problem, or time-aggregated expansions. We also show example anal-

yses using time-aggregated expansions. Here, we only deal with univariate cases. Discussion on multivariate cases is given in Section 5.1.2 and Chapter 6.

4.1 Solutions for the time aggregation problem

In this section, we firstly elaborate in Section 4.1.1 on some issues associated with current practices to deal with the time aggregation problem, which were reviewed briefly in Section 2.2.2. Then we discuss application of expansion methods to the time aggregation problem in Section 4.1.2. Some ideas for further applications using these time-aggregated expansions are discussed in Section 4.1.3.

4.1.1 Current practices

Normal based methods

As we reviewed in Section 2.2.2, solutions to the time aggregation problem which are available for the Normal based methods can be categorised into the following two kinds.

One is to assume that risk factors are from a serially independent and identical Normal distribution, namely, $X^1(t)$, $t = 1, \ldots, T$, in Equation (4.1) are from independent Normal distributions with the same mean and variance. Then the time aggregation is simply obtaining a sum of i.i.d. Normal variables and we have the scaling rule

$$X^T - T\mu^1 \stackrel{\mathcal{D}}{=} \sqrt{T}(X^1 - \mu^1), \tag{4.2}$$

where $\mu^1 = \mathrm{E}(X^1(1))$.

The other kinds of methods use parametric models, such as GARCH models, in order to incorporate serial dependence. These methods can be, however, less tractable than cases with the i.i.d. Normal assumption and closed form for the

distribution of X^T may not be available. We will discuss in Section 4.2.2 the application of expansion methods to time aggregation with GARCH models.

Historical simulation methods

For HS methods, which use empirical distributions of risk factors, we do not have a reasonable solution to the time aggregation problem. This is due to the difficulty of obtaining an empirical distribution of X^T from observations.

For instance, consider the time aggregation of HS methods where we assume serial independence of a risk factor. From an analogy with the Normal based methods, we can deduce that $X^1(t)$, $t = 1, \ldots, T$, in Equation (4.1) are independent and their marginal distributions are common and identical to the empirical distribution of the risk factor with a one-day time horizon. A naive solution to obtain the distribution of such X^T would be to consider a 'virtual' risk factor with a T-day time horizon $\tilde{X}^T$ as follows. Assume that we observe N realisations $\{x^{1,(1)}, \ldots, x^{1,(N)}\}$ of a risk factor with a one-day time horizon. Then, virtual realisations of $\tilde{X}^T = X^1(1) + \cdots + X^1(T)$ can be obtained by sampling each $X^1(t)$, $t = 1, \ldots, T$, from $\{x^{1,(1)}, \ldots, x^{1,(N)}\}$, for all possible combinations. Thus, we have N^T virtual observations of $\tilde{X}^T$. However, N^T can easily exceed the range of computational practicality[1], and therefore the empirical distribution of $\tilde{X}^T$ may not be obtained by this naive method.

Instead, as a rule of thumb, scaling by $\sqrt{T}$ in Equation (4.2) is often applied to approximating the distribution of $\tilde{X}^T$. This scaling rule is an exact result when random variables are i.i.d. Normal. However, more generally, including empirical distributions, scaling only adjusts the first and second moments and does not account for higher order moments, and therefore its reliability is not very clear. In Sections 4.1.2 and 4.1.3, we will show that expansion methods also can provide approximations for the distribution of $\tilde{X}^T$, and that they account for higher order

[1]See footnote 29 in Section 2.2.2.

moments.

4.1.2 Time aggregation problem and expansion methods

As we showed in Chapter 3, expansion methods successfully approximated the empirical distributions of risk factors given their moments, at least for the considered examples. We might deduce that the distributions of risk factors can be approximated fairly well given their moments, even for the cases where we cannot observe the empirical distributions of risk factors. We notice that time aggregation can be one such case, since, as discussed in Section 4.1.1, empirical distributions of risk factors or the PL with a T-day time horizon are not observable, while their moments can be available by the following procedure.

Time aggregation of moments

Now we show that the moments of X^T, a risk factor with a T-day time horizon, can be calculated from moments and autocross-moments of $X^1(t)$, $t = 1, \ldots, T$, a time series of a risk factor with a one-day time horizon. Let m_k^T be the kth moment of X^T. Then, by definition, we have

$$\begin{aligned} m_k^T &= \mathrm{E}((X^1(1) + \cdots + X^1(T))^k) \\ &= \sum_{k_1+\cdots+k_T=k} \frac{k!}{k_1! \cdots k_T!} \mathrm{E}(X^1(1)^{k_1} \cdots X^1(T)^{k_T}). \end{aligned} \tag{4.3}$$

and therefore m^T is expressed in terms of moments and autocross-moments of $X^1(t)$.

If we further assume that $X^1(t)$ are i.i.d., we have

$$m_k^T = \sum_{k_1+\cdots+k_T=k} \frac{k!}{k_1! \cdots k_T!} m_{k_1}^1 \cdots m_{k_T}^1, \tag{4.4}$$

where m_k^1 is the kth moment common to $X^1(t)$, $t = 1, \ldots, T$.

Time-aggregated expansions

Equations (4.3) and (4.4) suggest that expansion methods can be applied to the distribution of X^T as long as moments and autocross-moments of a risk factor with a one-day time horizon are available. Let us call these expansion methods which use time-aggregated moments, *time-aggregated expansions*.

This is the solution to the time aggregation problem available for expansion methods.

Since the empirical distribution of X^T is not available, here we do not have measures of fit to evaluate how well time-aggregated expansions approximate the distribution of X^T, unlike the cases with a one-day time horizon with which we dealt in Chapter 3. However, it seems reasonable to assume that time-aggregated expansions are as reliable as ordinary expansions in Chapter 3. That is, we assume that time-aggregated expansions approximate the distribution of X^T fairly well, unless they exhibit unignorable negative density[2].

4.1.3 Applications of time-aggregated expansions

By using sample moments of a risk factor with a one-day time horizon in Equation (4.4), expansions can approximate the distribution of $\tilde{X}^T$ discussed in Section 4.1.1, whose exact distribution is not available. We might use a time-aggregated expansion as a proxy for this distribution.

Furthermore, we might be able to evaluate the adequacy of the current practices, such as the use of i.i.d. Normal distribution and scaling in Equation (4.2),

[2]Needless to say, we are aware that additional assumptions introduced when we deal with the time aggregation, such as serial independence of risk factors or specific type of models applied including GARCH models, can reduce the reliability of approximations. However, this problem depends on how appropriately we model and estimate the moments and autocross-moments of risk factors. As discussed in Section 2.3, this is beyond the scope of this thesis and therefore we do not deal with this problem.

by comparing them with this proxy. Note that the use of the Normal distribution and scaling only account for the first and second moments, while expansions are capable of incorporating the effect of time aggregation on higher order moments.

Similarly, applying time-aggregated expansions to parametric models, we can analyse the effect of serial dependence on time aggregation. That is, we can compare the expansions using Equation (4.3), which account for serial dependence, with those using Equation (4.4), which assumes serial independence.

Such analyses using time-aggregated expansions have not been done elsewhere as far as we can determine. Numerical examples of these analyses are shown in Section 4.2

4.2 Example analyses

In this Section, we demonstrate numerical examples of analyses introduced in Section 4.1.3.

In Section 4.2.1, we investigate the adequacy of approximation by the Normal distribution and scaling using Equation (4.2), using the same data sets as in Section 3.2.1. Then, in Section 4.2.2, we investigate the effect of serial dependence on time aggregation using GARCH models with the same parameter sets as in Section 3.2.2.

We consider cases with $T = 10$ days, as suggested by the Basel Committee (1996).

Total area of negative density will not be shown, since it was at most of the order of 10^{-5} and considered to be negligible, except for one obvious example in Section 4.2.2.

4.2.1 Adequacy of current practices

In this Section, we assume that risk factors are serially independent and consider the distribution of $\tilde{X}^T$ discussed in Section 4.1.1. Since the exact distribution of $\tilde{X}^T$ is not available, here we use approximations by time-aggregated expansion as proxies for the distribution of $\tilde{X}^T$. Then we discuss the adequacy of existing approximations, such as those using the Normal distribution and using scaling in Equation (4.2), by comparing them with time-aggregated expansions.

We use the same risk factors as in Section 3.2.1. Table 4.1 shows the summary statistics of $\tilde{X}^T$ obtained by using time aggregation in Equation (4.4) and obvious properties of $\tilde{X}^T$ such as $\max\{\tilde{X}^T\} = \max\{X^1(1)+\cdots+X^1(T)\} = \max\{X^1\}\times T$.

	N225	N225 IV	SP500	SP500 IV
Number of observations	9.765×10^{26}		9.765×10^{26}	
Mean ($\times 10^{-3}$)	8.574	−2.931	−4.426	−5.006
Std. dev. ($\times 10^{-2}$)	3.414	12.598	2.066	1.585
Skewness ($\times 10^{-2}$)	−7.177	9.884	2.262	−7.938
Kurtosis	3.101	3.196	3.037	3.178
Min	−0.423	−1.375	−0.185	−2.275
Max	0.352	1.996	0.213	1.910

Table 4.1: Summary statistics of $\tilde{X}^T$, the virtual data sets of risk factors with 10 days time horizon. See Section 4.1.1 for definition of $\tilde{X}^T$. See Table 3.1 to compare with those with a one-day time horizon.

Comparing the statistics in Table 4.1 with those in Table 3.1, the skewness values are reduced and the kurtosis values are closer to 3, due to the central limit theorem (CLT). This suggests that the distributions of risk factors with a 10-day time horizon can be closer to the Normal than those with one day time horizon.

Adequacy of approximations by Normal distribution

Figures 4.1 to 4.4 show the tails of approximations for distribution function of risk factors with a 10-day time horizon, obtained by time-aggregated expansions and using the Normal distribution.

From these Figures, we can see that the approximations by expansions are almost identical to each other. This may allow us to use only one of the expansions as a proxy for the distribution of $\tilde{X}^T$.

The Normal approximations are closer to the expansions, than they were for the cases with a one-day time horizon in Section 3.2.1. Although this is rather obvious from the CLT, we investigate this point further by introducing the weighted integrated squared error (ISE) for the Normal distribution as a measure of difference between approximations for the density function by the Normal distribution and by expansion. Note that here we measure the difference between two density functions, not distribution functions, for computational convenience. That is, we define

$$\mathrm{ISE} = \int_{\mathcal{S}} \frac{\left\{ \hat{f}(u) - \phi\left(\frac{u-\mu}{\sigma}\right)\frac{1}{\sigma} \right\}^2}{\phi\left(\frac{u-\mu}{\sigma}\right)\frac{1}{\sigma}} \mathrm{d}u, \tag{4.5}$$

where

$$\hat{f}(x) = w(x)\sum_{k=0}^{n} C_k g_k(x)$$

is an approximation for the pdf of a random variable X, $\phi(x) = \exp(-x^2/2)/\sqrt{2\pi}$ is the pdf of the standard Normal distribution, $\mu = \mathrm{E}(X)$ and $\sigma^2 = E(X^2) - \{\mathrm{E}(X)\}^2$. If we use an Hermite expansion as $\hat{f}$ in Equation (4.5), we have $\mathrm{ISE} = \sum_{k=1}^{n} C_k^2$.

Table 4.2 compares the ISE of 10-day time-aggregated approximations with those for a one-day time horizon. Here, an optimised Hermite expansion with $n = 8$ in Equation (2.2) is used as $\hat{f}$ in Equation (4.5). We can confirm that the Normal approximations with a 10-day time horizon are much closer to the

expansions than those with a one-day time horizon.

We also find from Figures 4.1 to 4.4 that distribution functions by expansions have heavier tails than those of the Normal distribution on one side lighter tails on the other side. That is, in Figures 4.1 and 4.4, the expansions have heavier left tails and lighter right tails while in Figures 4.2 and 4.3, they have lighter left tails and heavier right tails. We might assume that these systematic differences between expansions and the Normal distributions can be due largely to skewness of $\tilde{X}^T$, rather than to its kurtosis. For instance, we can find from Table 4.1 that they are consistent with the fact that the distributions with a positive skewness value have a light left tail and heavy right tail.

From these results, we might say that the Normal distribution can approximate better the distribution of $\tilde{X}^T$ than it can approximate the distribution of X^1, which was considered in Section 3.2.1.

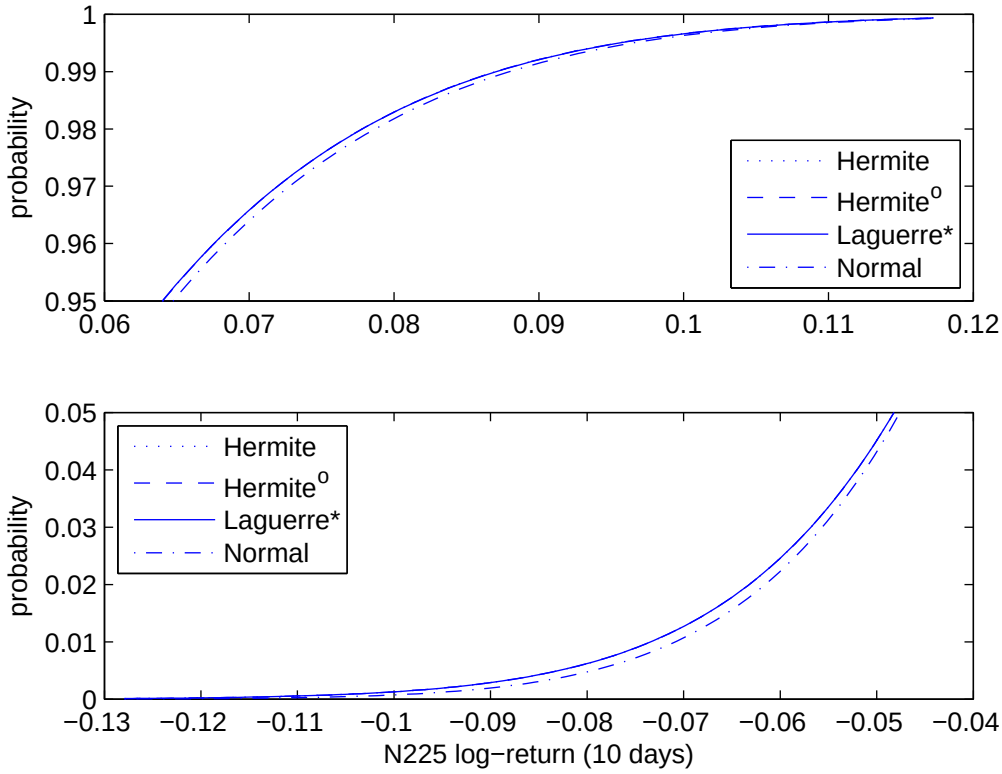

Figure 4.1: Approximations for distribution function of Nikkei 225 log-return with a 10-day time horizon by three expansions and the Normal distribution. 'Hermite^{o}' and 'Laguerre*' correspond to the optimised Hermite expansion and to the Laguerre expansion with squaring respectively. All expansions are time-aggregated using Equation (4.4): $n = 8$ in Equation (2.2) for all expansions. Upper and lower plots show the right tails and left tails respectively. We can see that expansions are almost identical to each other, while the Normal approximation is slightly different.

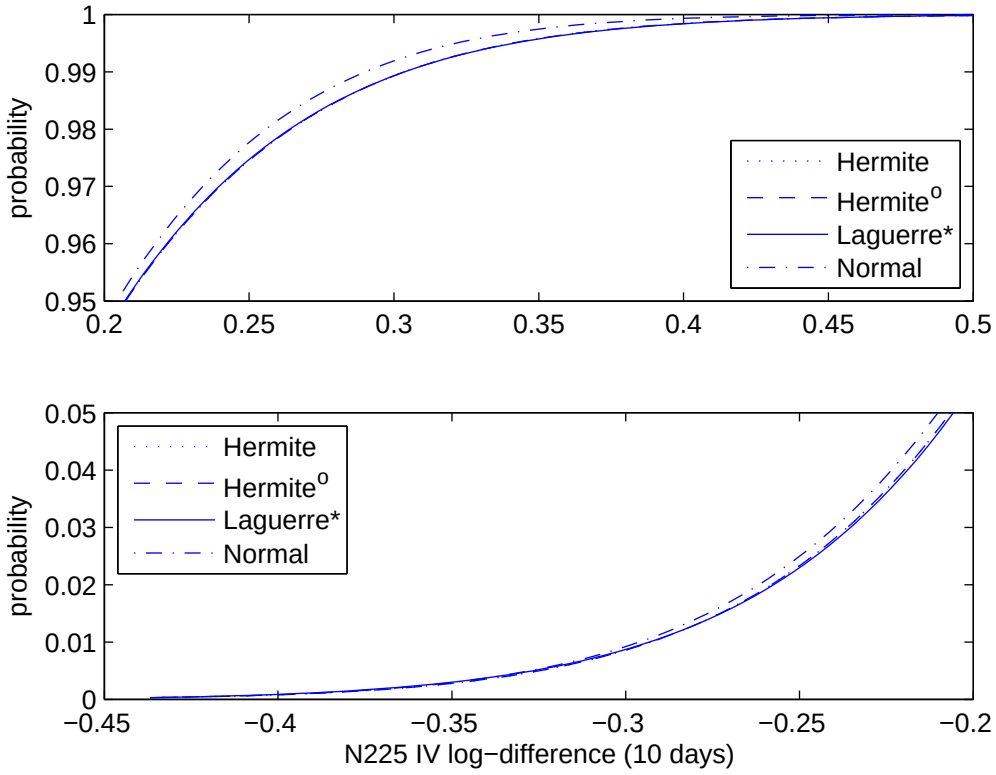

Figure 4.2: Approximations for distribution function of log-differences of Black and Scholes implied volatility of Nikkei 225 call option (three months, at-the-money), with a 10-day time horizon by three expansions and the Normal distribution. 'Hermite°' and 'Laguerre*' correspond to the optimised Hermite expansion and to the Laguerre expansion with squaring respectively. All expansions are time-aggregated using Equation (4.4): $n = 8$ in Equation (2.2) for all expansions. Upper and lower plots show the right tails and left tails respectively. We can see that expansions are almost identical to each other, while the Normal approximation is slightly different.

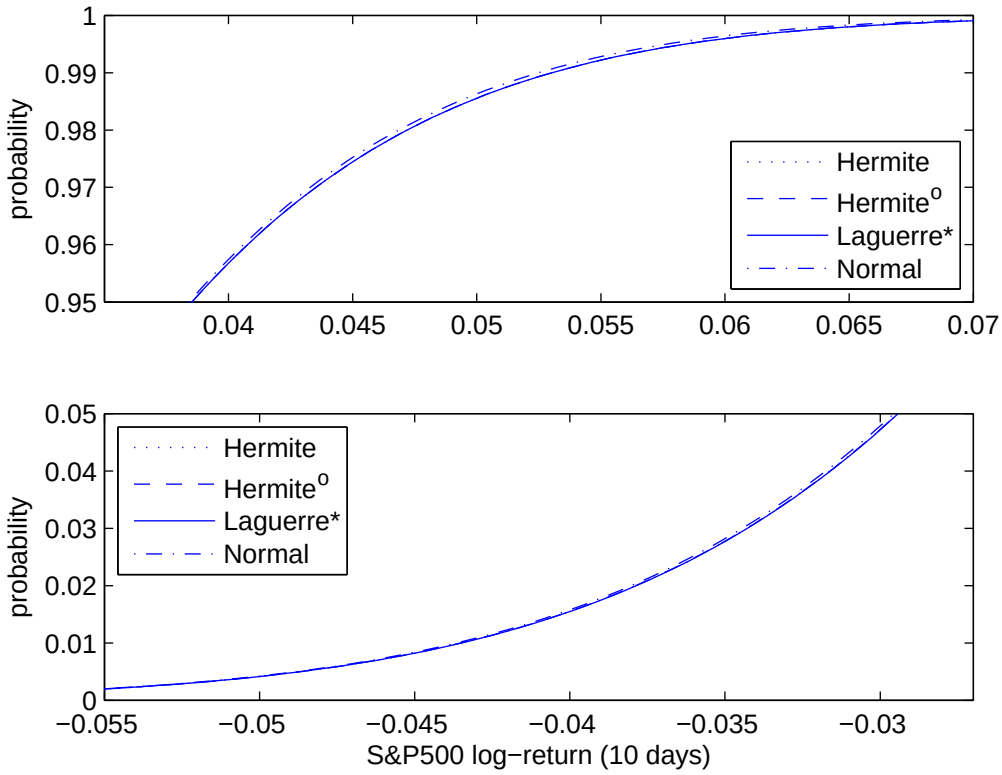

Figure 4.3: Approximations for distribution function of S&P 500 log-return with a 10-day time horizon by three expansions and the Normal distribution. 'Hermite^{o}' and 'Laguerre*' correspond to the optimised Hermite expansion and to the Laguerre expansion with squaring respectively. All expansions are time-aggregated using Equation (4.4): $n = 8$ in Equation (2.2) for all expansions. Upper and lower plots show the right tails and left tails respectively. We can see that all approximations are almost identical to each other.

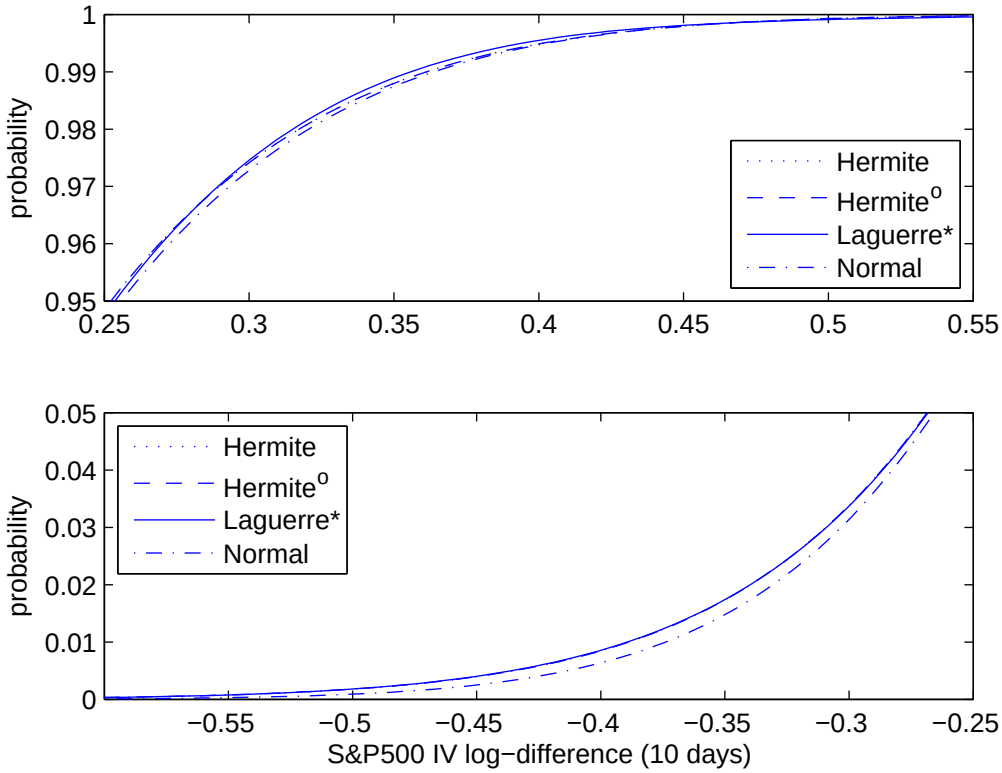

Figure 4.4: Approximations for distribution function of log-differences of Black and Scholes implied volatility of S&P 500 call option (three months, at-the-money), with a 10-day time horizon by three expansions and the Normal distribution. 'Hermite°' and 'Laguerre*' correspond to the optimised Hermite expansion and to the Laguerre expansion with squaring respectively. All expansions are time-aggregated using Equation (4.4): $n = 8$ in Equation (2.2) for all expansions. Upper and lower plots show the right tails and left tails respectively. We can see that expansions are almost identical to each other, while the Normal approximation is slightly different.

$(\times 10^{-2})$	N225	N225 IV	SP500	SP500 IV
$T = 10$ days	0.132	0.361	0.015	0.259
$T = 1$ day	3.193	9.018	0.976	13.805

Table 4.2: Comparison of ISE defined in Equation (4.5), which shows how different the approximation for the pdf of a risk factor by the Normal distribution is from that by a Hermite expansion. Here, the optimised Hermite expansion with $n = 8$ in Equation (2.2) is used as $\hat{f}$ in Equation (4.5). '$T = 10$' denotes weighted ISE of $\tilde{X}^T$, a virtual risk factor with a 10-day time horizon defined in Section 4.1.1, and '$T = 1$' denotes that of a risk factor with a one-day time horizon, which we dealt with in Section 3.2.1.

Adequacy of approximations by scaling

As discussed in Sections 2.2.2 and 4.1.1, we sometimes apply scaling by Equation (4.2) to the empirical distributions of a one-day time horizon in order to make up for the absence of such a distribution for a risk factor with a T-day time horizon in HS methods. However, its adequacy is not very clear. From Equation (4.2), we find that a scaled risk factor has the same mean and variance as the time-aggregated risk factor $\tilde{X}^T$. On the other hand, scaling does not account for the effect of the CLT, and skewness and kurtosis of a scaled risk factor are the same as those of a risk factor with a one-day time horizon. By comparing skewness and kurtosis of scaled risk factors with those of $\tilde{X}^T$ in Tables 3.1 and 4.1, we can see that scaled risk factors are more skewed and have larger kurtosis.

Here we investigate how well scaling in Equation (4.2) can approximate the distribution of $\tilde{X}^T$. That is, we compare an expansion applied to a risk factor scaled by using Equation (4.2) to a time aggregated expansion using Equation (4.4) as a proxy for the distribution of $\tilde{X}^T$. If the approximation by scaling is adequate, we might expect that a scaled expansion can be close to a corresponding

time-aggregated expansion.

Since the time-aggregated expansions in the previous set of examples were almost identical to each other, in this set of examples, we only use the time-aggregated and optimised Hermite expansion with $n = 8$ in Equation (2.2) as a proxy of the distribution of $\tilde{X}^T$.

Figures 4.5 to 4.8 compare the tails of the scaled and optimised Hermite expansion using Equation (4.2) with those of the time-aggregated and optimised Hermite expansion. Approximations by the Normal distribution are also shown for comparison.

From these Figures, we find that the approximations by scaling are not as close to the time-aggregated expansions as approximations by the Normal distribution. Approximations by scaling have heavier tails than other approximations in all examples, possibly reflecting larger kurtosis discussed above. Thus, approximation by scaling can result in overestimation of measures of market risk such as VaR or ES.

Similarly to the previous set of examples, we can introduce a measure of difference between two approximations by Hermite expansions for a pdf as follows. We define a weighted ISE for two Hermite expansions,

$$\hat{f}_a(x) = \phi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}\sum_{k=0}^{n} C_k^a \mathrm{He}_k\left(\frac{x-\mu}{\sigma}\right)$$

and

$$\hat{f}_b(x) = \phi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}\sum_{k=0}^{n} C_k^b \mathrm{He}_k\left(\frac{x-\mu}{\sigma}\right),$$

by

$$\mathrm{ISE} = \int_{\mathcal{S}} \frac{\left\{\hat{f}_a(u) - \hat{f}_b(u)\right\}^2}{\phi\left(\frac{u-\mu}{\sigma}\right)\frac{1}{\sigma}}\mathrm{d}u = \sum_{k=1}^{n}(C_k^a - C_k^b)^2. \tag{4.6}$$

Table 4.3 shows the ISE between a time-aggregated optimised Hermite expansion and a scaled and optimised Hermite expansion. The ISE for an approximation by the Normal distribution, which is the same as '$T = 10$' in Table 4.2,

is also shown for comparison. Here, we also see that the approximations by the Normal distribution are much closer to time-aggregated expansions than are the approximations by scaling.

From these results, we might say that approximations by the Normal distribution can be more appropriate for time aggregation under the assumption of serial independence, than those by scaling. This conclusion is valid at least for our four examples, however, considering that time aggregation under i.i.d. assumption in general can be strongly affected by the CLT, we might expect to meet the same conclusion in a wider range of examples, especially when $T = 10$ days or longer.

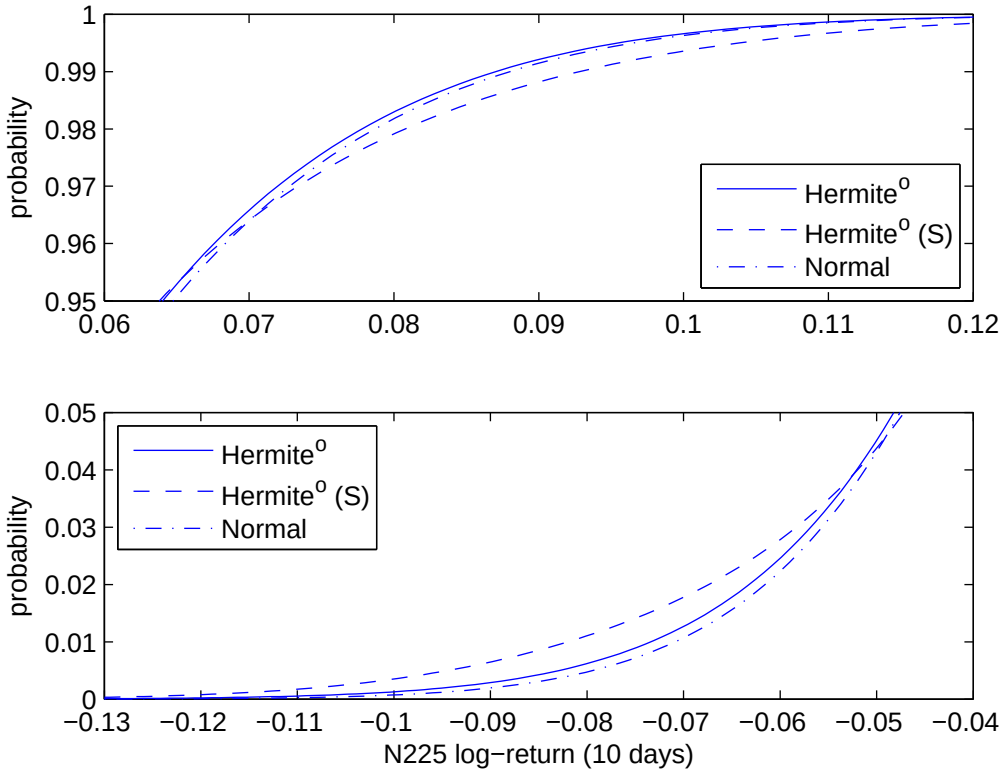

Figure 4.5: The optimised Hermite expansions and the Normal approximation applied to Nikkei 225 log-return with a 10-day time horizon. 'Hermite^{o}' denotes the time-aggregated and optimised Hermite expansion, while 'Hermite^{o} (S)' corresponds to the scaled and optimised Hermite expansion using Equation (4.2): $n = 8$ in Equation (2.2) for all expansions. Upper and lower plots show the right tails and left tails respectively. The scaled expansion does not necessarily agree with the time-aggregated expansion and has heavier tails than the others, while the Normal approximation is close to the time-aggregated expansion.

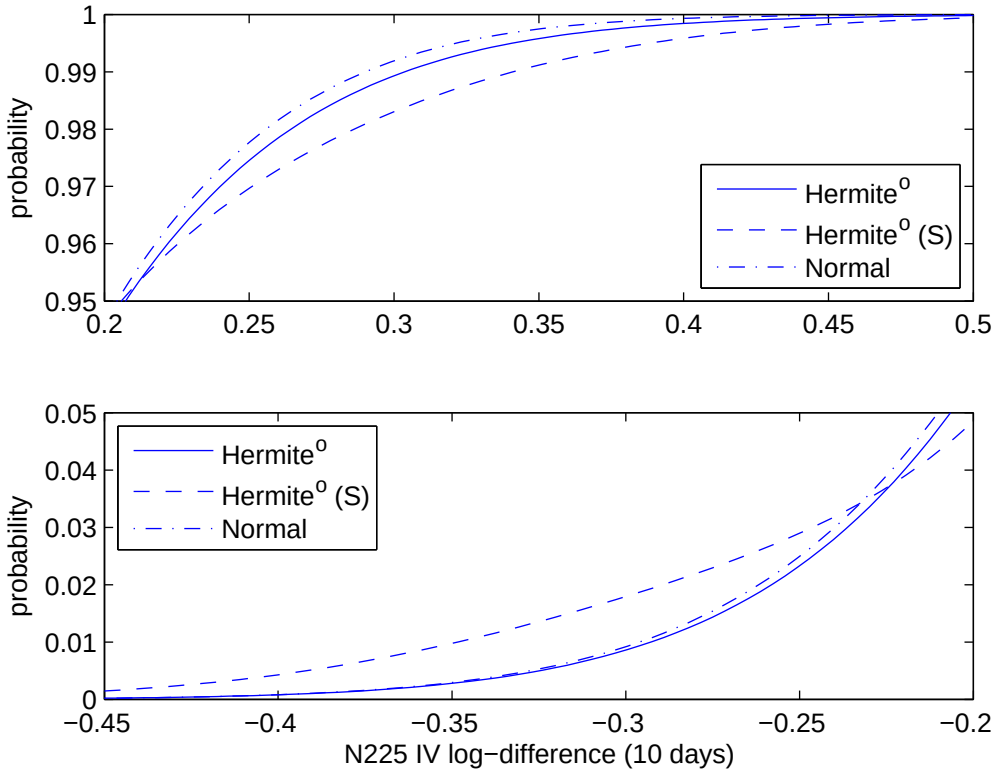

Figure 4.6: The optimised Hermite expansions and the Normal approximation applied to log-differences of Black and Scholes implied volatility of Nikkei 225 call option (three months, at-the-money), with a 10-day time horizon. 'Hermite^{o}' denotes the time-aggregated and optimised Hermite expansion, while 'Hermite^{o} (S)' corresponds to the scaled and optimised Hermite expansion using Equation (4.2): $n = 8$ in Equation (2.2) for all expansions. Upper and lower plots show the right tails and left tails respectively. The scaled expansion does not necessarily agree with the time-aggregated expansion and has heavier tails than the others, while the Normal approximation is close to the time-aggregated expansion.

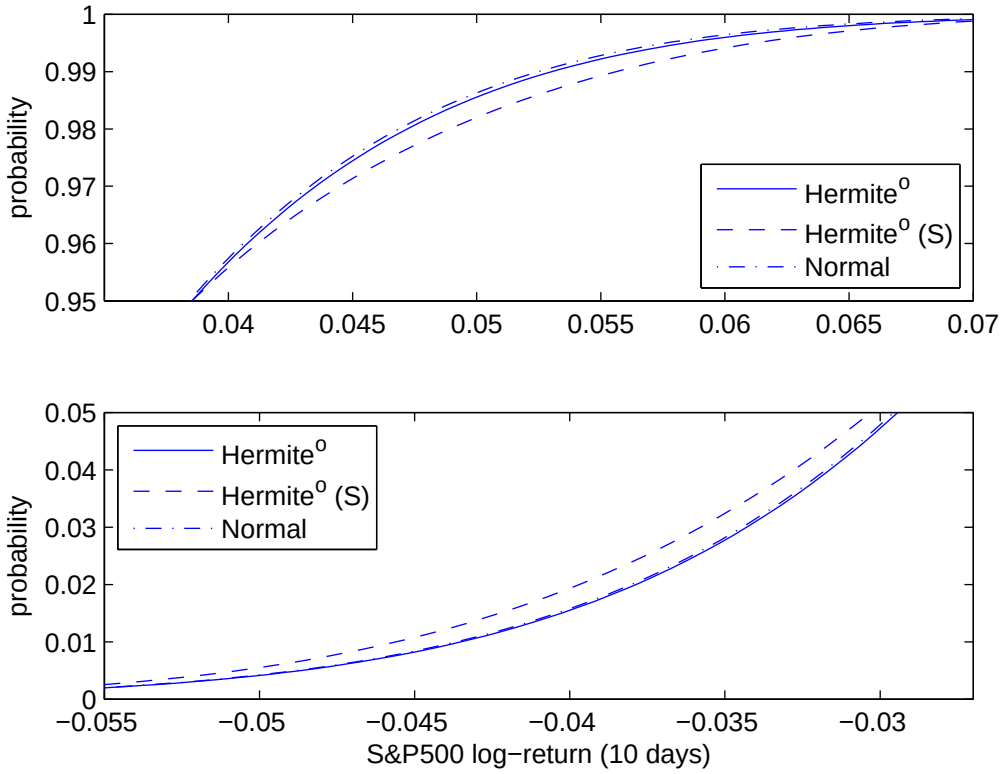

Figure 4.7: The optimised Hermite expansions and the Normal approximation applied to S&P 500 log-return with a 10-day time horizon. 'Hermite^{o}' denotes the time-aggregated and optimised Hermite expansion, while 'Hermite^{o} (S)' corresponds to the scaled and optimised Hermite expansion using Equation (4.2): $n = 8$ in Equation (2.2) for all expansions. Upper and lower plots show the right tails and left tails respectively. The scaled expansion does not necessarily agree with the time-aggregated expansion and has heavier tails than the others, while the Normal approximation is close to the time-aggregated expansion.

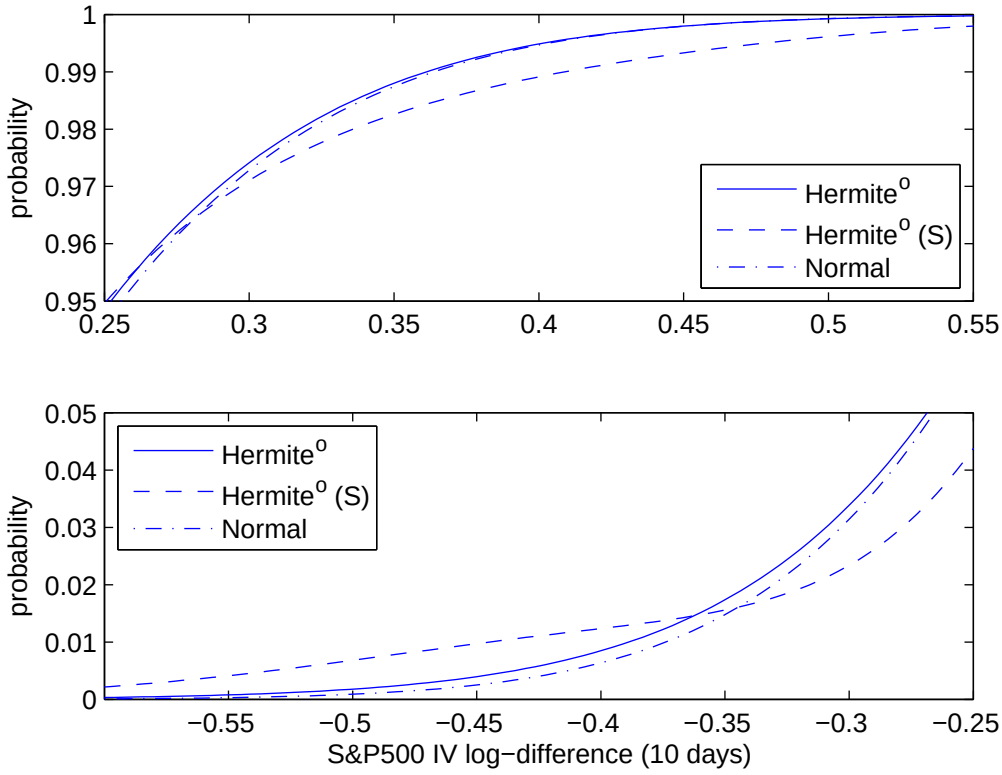

Figure 4.8: The optimised Hermite expansions and the Normal approximation applied to log-differences of Black and Scholes implied volatility of S&P 500 call option (three months, at-the-money), with a 10-day time horizon. 'Hermite^{o}' denotes the time-aggregated and optimised Hermite expansion, while 'Hermite^{o} (S)' corresponds to the scaled and optimised Hermite expansion using Equation (4.2): $n = 8$ in Equation (2.2) for all expansions. Upper and lower plots show the right tails and left tails respectively. The scaled expansion does not necessarily agree with the time-aggregated expansion and has heavier tails than the others, while the Normal approximation is close to the time-aggregated expansion.

$(\times 10^{-2})$	N225	N225 IV	SP500	SP500 IV
scaling	2.360	6.656	0.894	11.444
Normal	0.132	0.361	0.015	0.259

Table 4.3: The row 'scaling' shows ISE as defined in Equation (4.6), which shows how different the approximation for the pdf of a risk factor by a scaled and optimised Hermite expansion is from that by a 10 days time-aggregated and optimised Hermite expansion. Here, $n = 8$ in Equation (2.2). For comparison, we show in the row 'Normal' ISE as defined in Equation (4.5), which is the same as '$T = 10$' in Table 4.2.

4.2.2 Analysing effect of serial dependence

As discussed in Section 4.1.2, time-aggregated expansions are capable of incorporating serial dependence as long as the moments and autocross-moments are available. Therefore, as discussed in Section 4.1.3, time-aggregated expansions allow us to see the effect of serial dependence on the distribution function by comparing the expansions using Equation (4.3), which account for the serial dependence, with those using Equation (4.4), which assumes that a time series of a risk factor is serially independent.

Such analysis, for instance, can give us some idea about the effect of ignoring serial dependence when it possibly exists.

Here, we investigate the effect of serial dependence of GARCH models on a time-aggregated distribution. As we discussed in Section 3.2.2, GARCH models in general have serial dependence in variance; also moments and autocross-moments for the GARCH models are available, as long as they exist and are finite, while exact distributions of random variables from GARCH models are not available. We assume that $X^1(t)$, $t = 1, \ldots, T$ are from a GARCH model and consider the distribution of $X^T = X^1(1) + \cdots + X^1(T)$. Since $X^1(t)$, $t = 1, \ldots, T$ are

serially dependent, Equation (4.3) should be used for time aggregation. Then the distribution of X^T which accounts for the serial dependence is approximated. In this example analysis, however, we also consider another time series $X^1_{\text{iid}}(t)$, $t = 1, \ldots, T$, which is serially independent and each $X^1_{\text{iid}}(t)$ has the same distribution as the marginal distribution of $X^1(t)$. We define $X^T_{\text{iid}} = X^1_{\text{iid}}(1)+\cdots+X^1_{\text{iid}}(T)$. Then, each component of X^T and X^T_{iid} has the same marginal distribution. The only difference is that the components of X^T are serially dependent while those of X^T_{iid} are independent of each other. Therefore, the difference between the distribution of X^T and that of X^T_{iid} can be considered as the effect of serial dependence on a time-aggregated distribution. The distribution of X^T_{iid} can be approximated similarly to that of X^T, but using Equation (4.4) for time aggregation. We exhibit two examples in Figures 4.9 and 4.10. Figure 4.9 compares distribution functions of X^T with those of X^T_{iid} of GARCH(1, 1) model with parameters $a_1 = 0.13$, $b_1 = 0.82$ and $c = 0.1$, which are the same as those in 'Heavy tailed case' in Section 3.2.2. Two expansions, Hermite and Laguerre with squaring are applied to each of X^T and X^T_{iid}. The expansions applied to X^T_{iid} are indicated by '(iid)'. Figure 4.10 is obtained similarly, but using the parameters $a_1 = 0.05$, $b_1 = 0.9$ and $c = 0.1$ for GARCH(1, 1) model, which are the same as those in 'light tailed case' in Section 3.2.2.

From Figure 4.9, we find that the Hermite expansion with serial dependence can be a poor approximation, while others are stable. We can see from the tail plots that the two expansions with serial independence agree with each other, but are clearly different from the Laguerre expansion with squaring and serial dependence. This difference can be considered as the effect of serial dependence structure on the time-aggregated distribution, as discussed above.

From Figure 4.10, we can also see that the two expansions with serial independence agree with each other. Also, they are different from expansions with serial dependence, if not as large a difference as in Figure 4.9, showing the effect

of serial dependence. The two expansions with serial dependence show some difference from each other; however, inferring which can be a better approximation is not easy due to unavailability of the 'true' distribution.

Except for the Hermite expansion in Figure (4.9), which is obviously a poor approximation, expansions which incorporate serial dependence have heavier tails than expansions with serial independence. This suggests that we might underestimate measures of market risk such as VaR and ES, if we ignore serial dependence structure, when a risk factor is from a GARCH model.

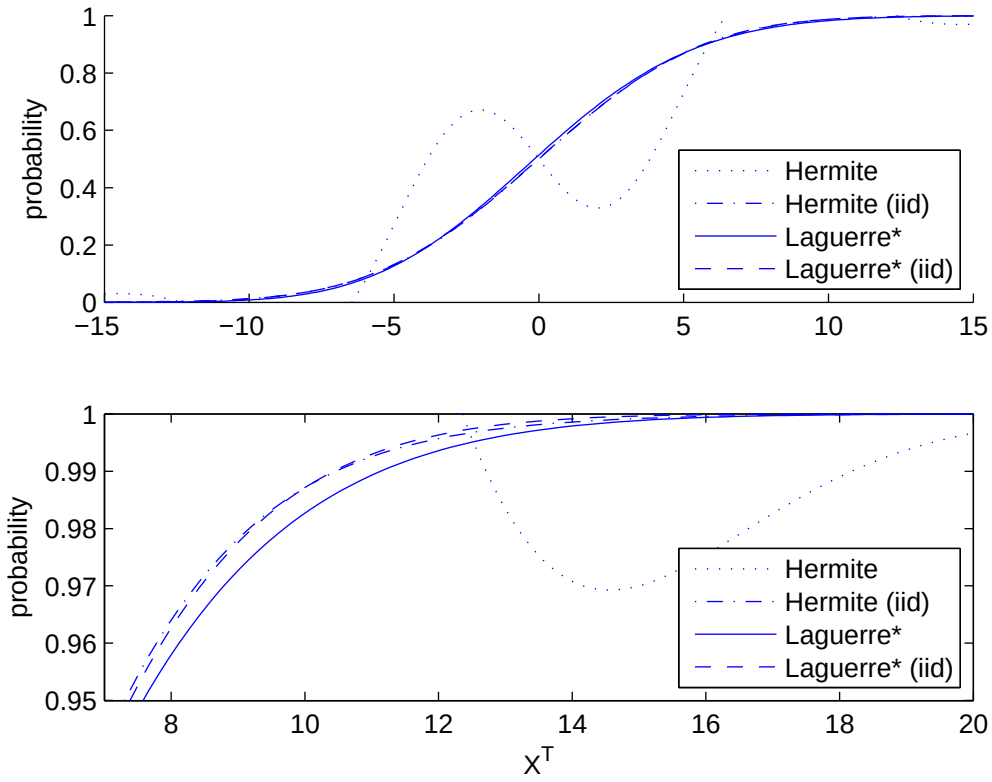

Figure 4.9: Approximations for the 10-day time-aggregated unconditional distribution function of GARCH(1, 1) with $(a_1, b_1, c) = (0.13, 0.82, 0.1)$ in Equation (3.7), as in 'Heavy tailed case' in Section 3.2.2. 'Laguerre*' corresponds to the Laguerre expansion with squaring: $n = 6$ and 3 in Equation (2.2) for Hermite and 'Laguerre*', respectively, similarly to Figure 3.10. Expansions with '(iid)' use Equation (4.4), which assumes serial independence, while others use Equation (4.3), which captures serial dependence of the GARCH model. Upper and lower plots show the overall shapes and the right tails, respectively. The expansions with '(iid)' are close to each other, but different from the time-aggregated 'Laguerre*'. The time-aggregated Hermite performs poorly.

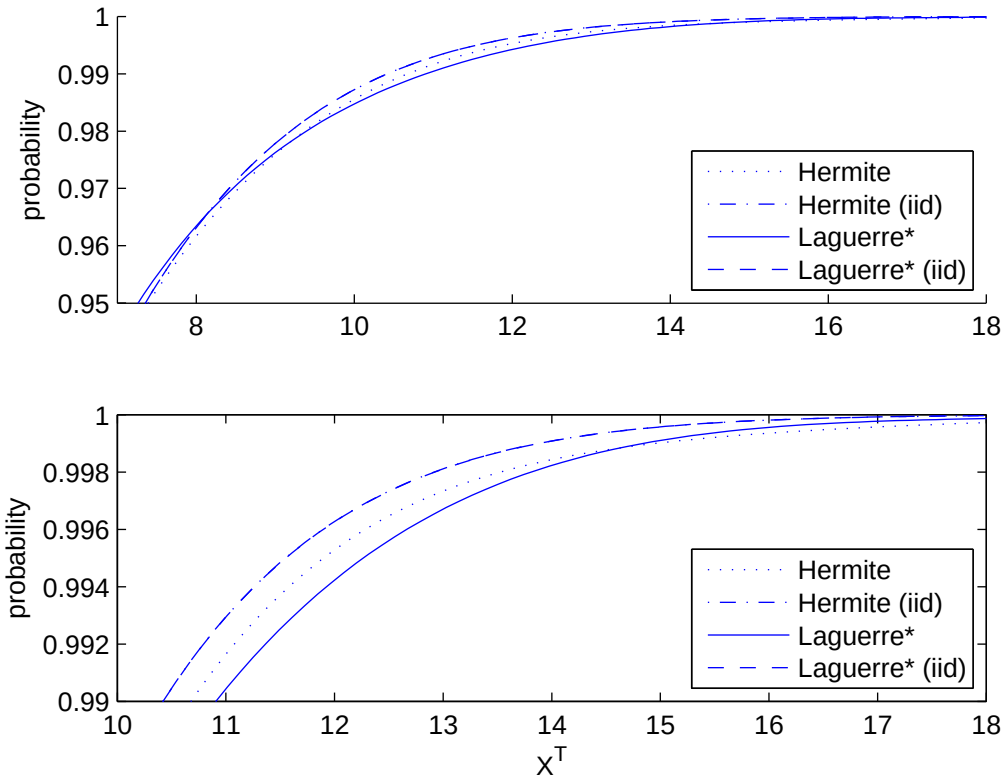

Figure 4.10: Approximations for the 10-day time-aggregated unconditional distribution function of GARCH(1, 1) model with $(a_1, b_1, c) = (0.05, 0.9, 0.1)$ in Equation (3.7), as in 'Light tailed case' in Section 3.2.2. 'Laguerre*' corresponds to the Laguerre expansion with squaring. We set $n = 8$ in Equation (2.2). The expansions with '(iid)' are those using Equation (4.4), which assumes serial independence, while others use Equation (4.3), which captures serial dependence of the GARCH model. Upper plots show the right tails and lower plots magnify the far right tails. We can see that expansions with '(iid)' are almost identical to each other, but different from the time-aggregated expansions.

4.3 Summary of this Chapter

In Section 4.1.2, we proposed a solution to the time aggregation problem using expansion methods. That is, we proposed expansion methods to approximate a distribution of a risk factor with a T-day time horizon, when T is large such as $T = 10$ days. These time-aggregated expansions are available where the moments and autocross-moments are available.

We discussed the possibility that time-aggregated expansions can be used as proxies for time aggregation of empirical distributions. As far as we could determine, reasonable solutions to obtain time-aggregated empirical distributions are not known, but are necessary for the HS methods with a longer time horizon than one day. Our methods can incorporate the effects of time aggregation on high moments such as skewness, kurtosis and even higher, while exising methods only adjust moments up to second order.

Example analyses using the market observations as in Section 3.2.1 suggested that the Normal distribution might better approximate distributions of time-aggregated risk factors than approximation by scaling, due to the CLT. This result is valid for our examples under the assumptions that the risk factors are serially independent and that time-aggregated expansions approximate exact distributions of time-aggregated risk factors fairly well.

When the autocross-moments are available, time-aggregated expansions are capable of accounting for serial dependence. We compared numerical examples of expansions with serial dependence with those assuming serial independence, but with the same marginal distributions, and demonstrated that expansion methods can capture the effect of serial dependence of GARCH models on time-aggregated risk factors' distributions. By the construction described in Section 4.1.2, the application of time-aggregated expansions is not limited to the examples dealt with in this Chapter.

Chapter 5

Application to Risk Aggregation

As discussed in Chapter 2, obtaining a PL distribution is one of the key problems in market risk measurement. In general, risk factors are described as random variables $\boldsymbol{X} = (X_1, \ldots, X_p)$, and they are linked to a PL Z of a portfolio using a deterministic function d as $Z = d(\boldsymbol{X})$. Obtaining the distribution of Z, or risk aggregation, is not trivial, especially when the number of risk factors p is larger than one and when d is not a linear function.

Review of current practices in Section 2.2.2 showed that HS methods, which use empirical distributions of risk factors, can give reasonable solutions to this problem, but they suffer from discreteness.

In this Chapter, we propose continuous approximations for PL distributions using expansion methods.

5.1 Solutions for the risk aggregation problem

In Section 5.1.1, we firstly summarise the current practices for dealing with the risk aggregation problem, which we discussed in Section 2.2.2. Then, in Section 5.1.2, we discuss the application of expansion methods to the risk aggregation problem.

5.1.1 Current practices

Existing solutions for the risk aggregation problem reviewed in Section 2.2.2 can be categorised into the following two directions.

One is to apply a Taylor expansion to the function d, which links the risk factors to the PL of a portfolio, and consider only the linear terms and, for some tractable cases, quadratic terms. This Taylor approximation can provide a continuous distribution as long as we assume the risk factors' distribution is continuous, such as the Normal distribution. However, even when we assume that the risk factors are Normally distributed, including second order moments can require further approximations, as discussed in Section 2.2.2.

The other direction is that of the HS methods, which use the empirical distribution of risk factors; that is, given N observations $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)}$ of risk factors, we use the empirical distribution of $d(\boldsymbol{x}^{(1)}), \ldots, d(\boldsymbol{x}^{(N)})$, as the approximation for the distribution of a PL. Since the HS methods implicitly use the joint empirical distribution of the risk factors, it incorporates the characteristics of the risk factors such as heavy tails or dependence structures among the risk factors in a natural way. On the other hand, the distribution function provided by the HS methods is discrete and this can be inconvenient, especially when we are interested in the tails of the distribution where the observations are sparse. Also, tails of the empirical distribution might be regarded as too sensitive to a small number of observations. For instance, the 0.01 quantile of the empirical distribution of 500 samples will only depend on the five smallest observations.

5.1.2 Expansion methods and risk aggregation

Here, we expect that expansion methods can provide a continuous approximation for the PL distribution which incorporates the characteristics of the market in a natural way. Furthermore, by using Equations (2.5), (2.6), (2.8) and (2.9),

we might also avoid numerical integration when we calculate the quantiles and conditional expectations, which reduces the computational load of calculating VaR and ES.

Let us explain our basic concept using a simple example with two risk factors X_1 and X_2. The extension to the case with three or more risk factors is straightforward.

Approximating the PL of a portfolio

Let $Z = d(X_1, X_2)$ be the PL of the portfolio. Assume that d can be approximated by a finite sum of the form

$$\begin{aligned} d(x_1, x_2) &\simeq \sum_{i,j} \frac{1}{i!j!} \frac{\partial^{i+j}}{\partial x_1^i \partial x_2^j} d(0,0) x_1^i x_2^j \\ &= \sum_{i,j} D_{ij} x_1^i x_2^j, \end{aligned}$$

where

$$D_{ij} = \frac{1}{i!j!} \frac{\partial^{i+j}}{\partial x_1^i \partial x_2^j} d(0,0)$$

are known constants. Then the PL is approximated by

$$Z = \sum_{i,j} D_{ij} X_1^i X_2^j. \tag{5.1}$$

Aggregating the moments

From Equation (5.1), we notice that moments of the PL Z can be derived from the moments and cross-moments of X_1 and X_2. That is, the kth moment m_k^Z of the PL Z can be approximated by

$$\hat{m}_k^Z = \mathrm{E}\left(\left(\sum_{i,j} D_{ij} X_1^i X_2^j \right)^k \right),$$

which can be expanded so that it is expressed as

$$\hat{m}_k^Z = \sum_{i,j} D'_{ij} \mathrm{E}(X_1^i X_2^j), \tag{5.2}$$

a linear combination of moments and cross-moments of X_1 and X_2, with known coefficients D'_{ij}.

Expansions with risk aggregation

Equation (5.2) suggests that expansion methods can be applied to approximating the distribution of PL with non-linear and multiple risk factors, as long as the moments and cross-moments of risk factors are available.

This is the solution to the risk aggregation problem available for expansion methods.

Here, we notice that the dependence structure between X_1 and X_2 is expressed in terms of cross-moments[1].

We might view that our method at once summarises the information of the risk factors by their moments and cross-moments, and then approximates the moments of the PL. Needless to say, if our purpose is only to obtain the distribution of the PL for situations where enough observations of all the risk factors are available, it is much simpler to use the sample moments of $d(\boldsymbol{x}^{(1)}), \ldots, d(\boldsymbol{x}^{(N)})$ directly. However, we notice that the procedure described above is applicable to more general situations. For example, it can be applied to the case where some of the risk factors are given as models, as long as the moments of the models are available, or to the case where some of the risk factors have shorter records than others. Also, our method allows combination with time aggregation as introduced in Section 4.1.2. That is, for instance, if we assume serial independence, we can apply Equation (4.4) to $\hat{m}_k^T$, and, if autocross-moments of risk factors are available, we can use Equation (4.3), which incorporates serial dependence.

Note that this method, which does not assume a certain class of models for risk factor dynamics or portfolios, can be applied to a very wide range of situations. There are only two necessary conditions: that moments and cross-moments of the

[1]This point will be further discussed later in Chapter 6.

risk factors up to some order are available, and that the function d, which links the risk factors to the PL of the portfolio, can be approximated by its Taylor expansion. The main shortcoming of the expansion methods, lack of robustness, can be mitigated by the techniques introduced in Section 3.1. Such applications of expansion methods have not been introduced in the literature to the best of our knowledge.

5.2 Option premium and the Black and Scholes formula

Before we move on to the numerical examples to see how our method explained in Section 5.1.2 works, we review the Black and Scholes formula for pricing a European call option and consider applications of the discussion in Section 5.1.2 to measuring the market risk associated with a European call option.

A European call option is a security whose payoff at maturity depends upon its underlying asset price: it pays $S - K$ units of currency if $S > K$ and 0 otherwise, where K is the prespecified strike price and S is the underlying asset price at maturity. Naturally, the premium for the option traded in the markets before maturity has a strong correlation with the underlying asset price at the corresponding time. It is often easier to treat the innovation of the underlying asset price as a risk factor and then evaluate the option premium using a pricing formula for the European call option than to deal with the observed option premium directly.

In this section we review the Black and Scholes (BS) formula for the European call option premium, discuss risk involved in the option premium, and consider approximations for the BS formula, which are used for applying the expansion methods.

5.2.1 The Black and Scholes formula

The BS formula states that the European call option premium c with time to maturity $\tau > 0$ is given by

$$c = S\Phi(d_1) - K\mathrm{e}^{-r\tau}\Phi(d_1 - \sigma\sqrt{\tau}), \tag{5.3}$$

where Φ is the distribution function of the standard Normal distribution,

$$d_1 = \frac{\log(S/K) + r\tau}{\sigma\sqrt{\tau}} + \frac{\sigma\sqrt{\tau}}{2},$$

S is the price of the underlying asset, $K \geq 0$ is the strike price (constant), $r \geq 0$ is the risk free continuously compounded interest rate (constant), and σ is the volatility of the underlying asset price (constant)[2].

BS delta and gamma

Although the relationship between the option premium and the underlying asset price in Equation (5.3) is non-linear, it is useful to consider a linear component. The first partial derivative of a portfolio value in terms of the underlying asset price is called "delta" and denoted using the Greek letter δ. For the case where the portfolio consists only of the call option described in Equation (5.3), the delta is given by

$$\delta = \frac{\partial c}{\partial S} = \Phi(d_1).$$

Similarly, the second partial derivative is called "gamma" and for the case of the call option, it is given by

$$\gamma = \frac{\partial^2 c}{\partial S^2} = \phi(d_1)\frac{1}{\sigma S\sqrt{\tau}},$$

where ϕ is the pdf of the standard Normal distribution.

[2]See Black & Scholes (1973).

BS implied volatility and vega risk

The BS formula assumes that the volatility of the underlying asset process is constant, which is not realistic in many cases[3]. As a result, the empirical volatility obtained from observations of underlying asset price innovations (the historical volatility or HV) and σ in Equation (5.3) calculated from observed option premium (the Black and Scholes implied volatility or IV) can be significantly different[4], and the IV can fluctuate over time (see Figure 5.1)[5]. Therefore the option position is in fact exposed to the risk of change in the IV as well as that of change in the underlying asset price[6]. This is called the vega risk.

The (linear) sensitivity of a portfolio to the IV change is called "vega" and often denoted by κ. For a portfolio consisting only of the call option in Equation (5.3), the vega is given by

$$\kappa = \frac{\partial c}{\partial \sigma} = S_t \sqrt{\tau} \phi(d_1).$$

5.2.2 Risk factors

Now we apply the discussion in Section 5.1.2 to the portfolio consisting only of the at-the-money call option.

Let $t = 0$ be the present time. We are interested in the distribution of the change in option premium from $t = 0$ to $t = \Delta t > 0$. Since we identify the risk

[3]In fact, pricing formulae which incorporate time varying volatility exist. For example, Heston & Nandi (2000) study an option pricing formula where the underlying asset price return is from the GARCH model.

[4]See, for example, Fabozzi (2007) or Duffie & Pan (1997).

[5]In fact, IV can be different for each transaction. Some studies model IV as a function of moneyness K/S and time to maturity τ. This function is called the implied volatility surface. See, for example, Malz (2000) or Cont & Fonseca (2002). However, in this example, we only focus on the at-the-money three months option premium.

[6] Strictly speaking, it is also exposed to the risk of change in the risk free interest rate; however, we ignore it in this example.

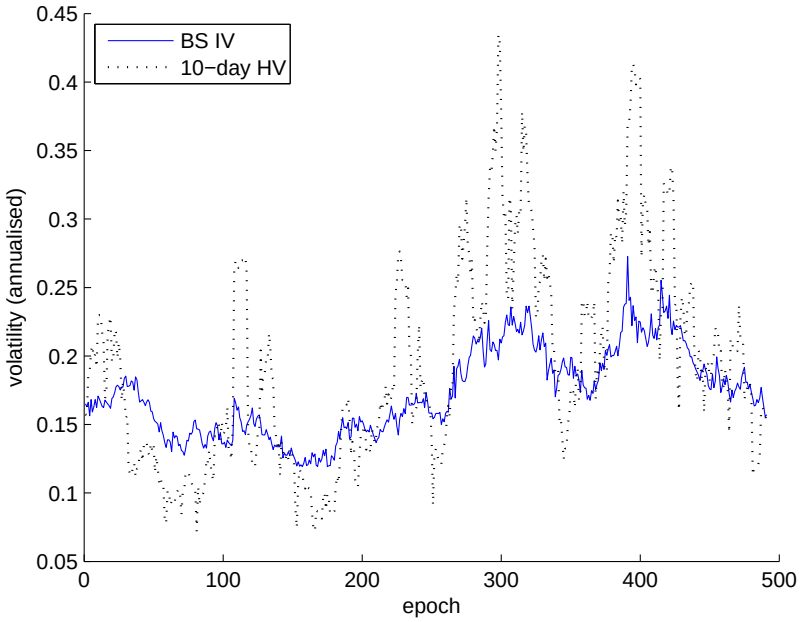

Figure 5.1: The daily innovations of Nikkei 225 index call option implied volatility (three months, at-the-money, IV) and ten days historical volatility (HV) of Nikkei 225. The observation period is from 9 November 2004 to 6 November 2006.

factors as change in the underlying asset price and change in the at-the-money IV, the distribution of the PL is determined by the joint distribution of these risk factors at $t = \Delta t$.[7]

As for the underlying asset price, it is almost a standard practice to work with the log-return[8]; that is, we define the log-return of the underlying asset as

$$X = \log S_{\Delta t} - \log S_0,$$

where S_0 is the underlying asset price at the present time and $S_{\Delta t}$ is the underlying

[7] We can see from Equation (5.3) that the time to maturity τ is reduced by Δt as time passes from $t = 0$ to $t = \Delta t$ and this also affects the option premium. This is called "time decay". However, when Δt is small enough compared to τ, the difference is small and therefore we ignore it in our example.

[8] See, for example, RiskMetrics (1996), Duffie & Pan (1997) and Mina & Xiao (2001).

asset price at time Δt, and then consider the distribution of X. This standard practice is related to the fact that, under many models including the Black and Scholes model, X is assumed to be Normally distributed[9].

Although there seems to be no standard treatment for IV, here we consider the distribution of the log-difference, similarly to Malz (2000),

$$\nu = \log \sigma_{\Delta t} - \log \sigma_0,$$

where σ_0 is the IV at the present time and $\sigma_{\Delta t}$ is the IV at time Δt, similarly to the underlying asset price[10].

Now the VaR of the portfolio consisting only of the call option is given by the quantile of the random variable

$$Z = d(X, \nu) = c(S_0 \mathrm{e}^X, \sigma_0 \mathrm{e}^\nu) - c(S_0, \sigma_0).$$

5.2.3 Approximation for the BS formula

In order to apply expansion methods, we consider the following approximations for function d.

i) *Delta-gamma approximation*

$$d_{\delta\gamma}(X, \nu) = \delta S_0 X + \frac{1}{2}(\delta S_0 + \gamma S_0^2) X^2;$$

ii) *Delta-vega approximation*

$$d_{\delta\kappa}(X, \nu) = \delta S_0 X + \kappa \sigma_0 \nu;$$

[9] See Chapter 4 of Jorion (2001).

[10] We are aware that estimating the moments of X and ν can be tricky especially when asset returns and log differences of the IV cannot be regarded as being stationary. Estimating moments under such circumstances may require some parametric models which incorporate time-inhomogeneity and serial dependence; however, as mentioned in Section 2.3, this is beyond the scope of this thesis. Instead, we notice that the expansion methods are available as long as the moments and cross-moments are available, regardless of which methods or models we use to estimate them.

iii) *Delta-gamma-vega approximation*

$$d_{\delta\gamma\kappa}(X,\nu) = \delta S_0 X + \frac{1}{2}(\delta S_0 + \gamma S_0^2)X^2 + \kappa\sigma_0\nu.$$

These allow us to approximate the moments of Z using the moments and cross-moments of X and ν.

Adequacy of BS approximations

Let us discuss the adequacy of these three approximations for the BS formula using two numerical examples, the call options on Nikkei 225 and on S&P 500 (three months, at-the-money). See Table 3.1 in Section 3.2.1 for a summary of the data sets.

Table 5.1 shows the root mean square differences between the BS formula d and the approximations d_*, defined by

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(d(X^{(i)},\nu^{(i)}) - d_*(X^{(i)},\nu^{(i)})\right)^2}, \tag{5.4}$$

where $(X^{(i)},\nu^{(i)})$ are the observed samples. Figures 5.2 and 5.3 compare the effects of these approximations on the tails of the one day PL distribution. In Figures 5.2 and 5.3, "full evaluation" shows the empirical distribution of $d(X^{(i)},\nu^{(i)})$, $i = 1,\ldots,N$. The other plots are obtained similarly, but using approximations for the function d.

They show that the delta-gamma-vega is a good approximation for the BS formula and incorporating higher order terms may not make a large difference. It is notable that the delta-gamma approximation and the delta-gamma-vega approximation are significantly different. This suggests that vega risk is worth considering.

Needless to say, we could incorporate higher order terms if necessary.

	Nikkei 225 Call option	S&P 500 Call option
$\delta\gamma$	20.02	1.567
$\delta\kappa$	10.71	0.507
$\delta\gamma\kappa$	0.93	0.155

Table 5.1: Root mean square differences between the BS formula and the approximations defined by Equation (5.4). The Greek letters show the components included in the approximations.

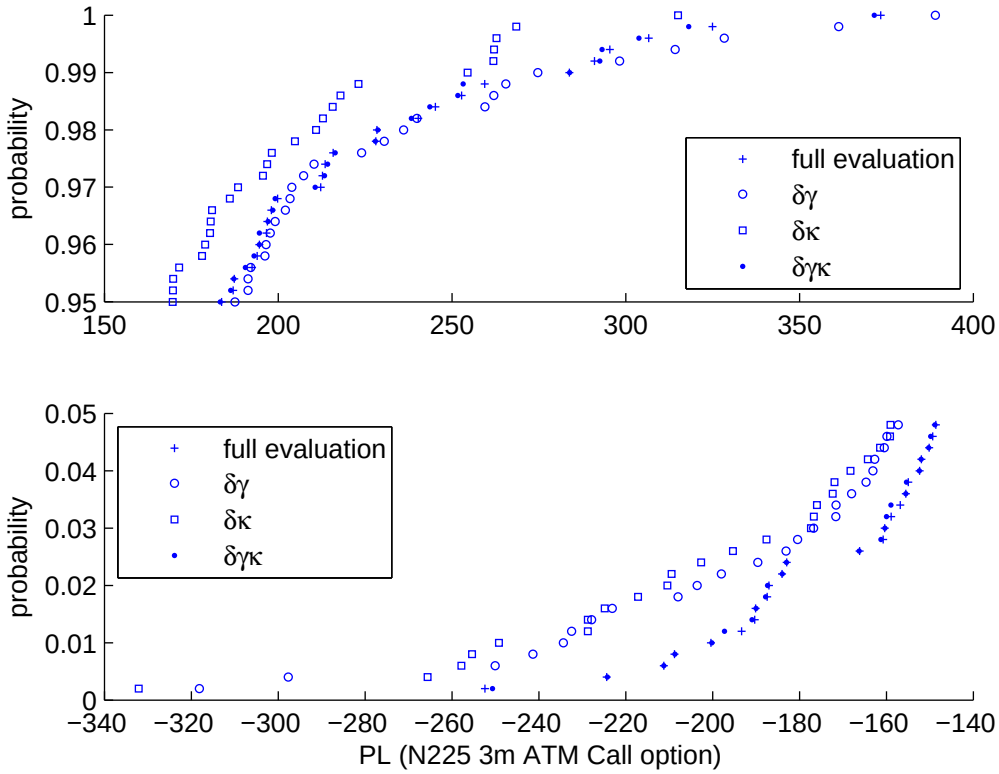

Figure 5.2: The empirical distributions of one day PL of Nikkei 225 call option (three months, at-the-money) by evaluating the BS formula and three approximations for the BS formula. The Greek letters show the components included in the approximations. Upper plots are the right tail and the lower plots are the left tail. We can see that the $\delta\gamma\kappa$ approximation is almost identical to 'full evaluation'. See also Table 5.1 for overall approximation quality.

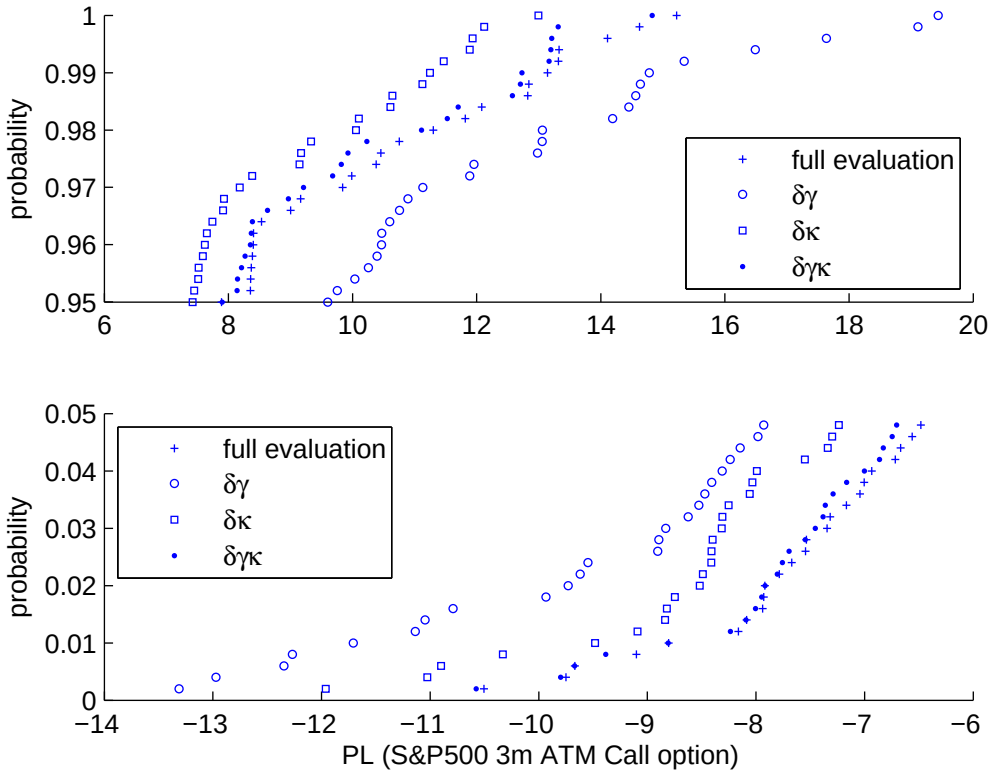

Figure 5.3: The empirical distributions of one day PL of S&P 500 call option (three months, at-the-money) by evaluating the BS formula and three approximations for the BS formula. The Greek letters show the components included in the approximations. Upper plots are the right tail and the lower plots are the left tail. We can see that the $\delta\gamma\kappa$ approximation is very close to 'full evaluation', while others show a large difference. See also Table 5.1 for overall approximation quality.

5.3 Numerical examples

We apply our method described in Section 5.1.2 to the distribution of the PL of call options, using the result of Section 5.2. Here, we use empirical moments and cross-moments as inputs to expansion methods, similarly to the examples in Sections 3.2.1 and 4.2.1. We might use some parametric models such as GARCH models to estimate the moments and cross-moments of risk factors. However, as discussed in Section 2.3, estimating risk factors' moments and cross-moments is beyond the scope of thesis[11]. Instead, using empirical moments and cross-moments enables us to make some judgement about how well the expansion methods are approximating the target distribution, by comparing expansions with empirical distributions. We also notice that the empirical distribution of a PL is equivalent to the PL distribution derived by the HS methods, which is widely used in the financial industry.

Figures 5.4 and 5.5 provide two examples, PLs of the call options on Nikkei 225 and on S&P 500 (three months and at-the-money) respectively. The data set used is the same as the one used in Sections 3.1 and 5.2.3. See Table 3.1 for a summary of the data sets.

[11] See also footnote 10.

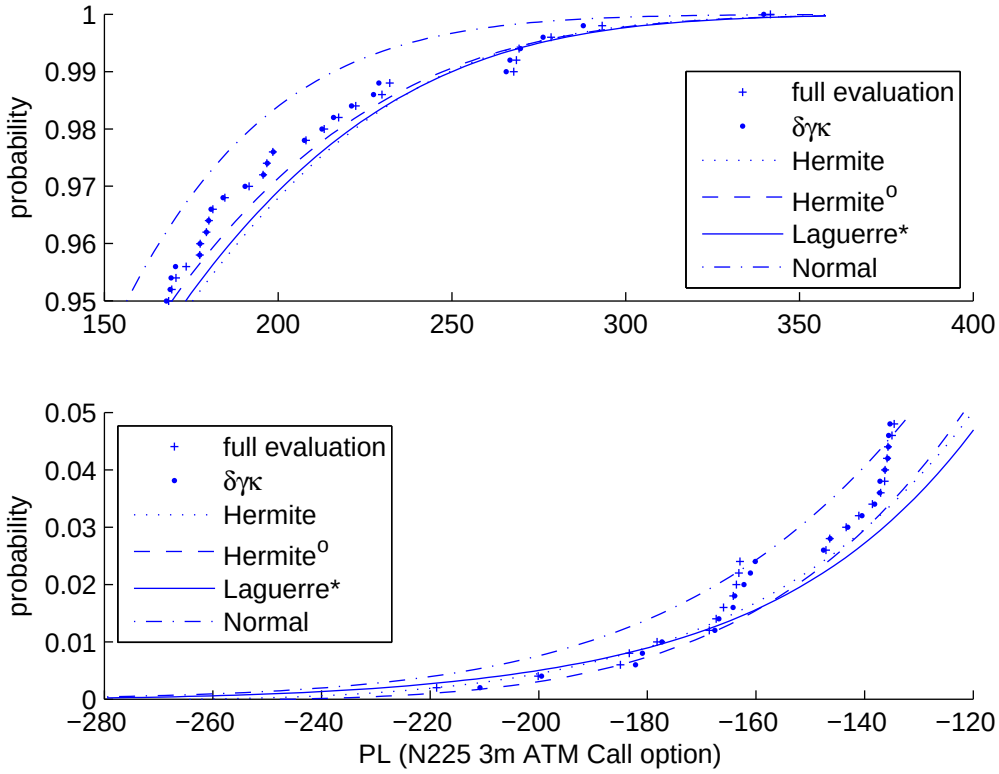

Figure 5.4: Two empirical distributions of one day PL of Nikkei 225 call option (three months, at-the-money), the approximations by the Hermite expansion, the optimised Hermite expansion ('Hermite°'), the Laguerre expansion with squaring ('Laguerre*') and the Normal distribution. The empirical distributions are the same as in Figure 5.2: $n = 8$ in Equation (2.2) for expansions. The upper and lower plots are the right and left tails. We can see that expansions are close to each other and capture the tail shape of 'full evaluation' better than the Normal approximation.

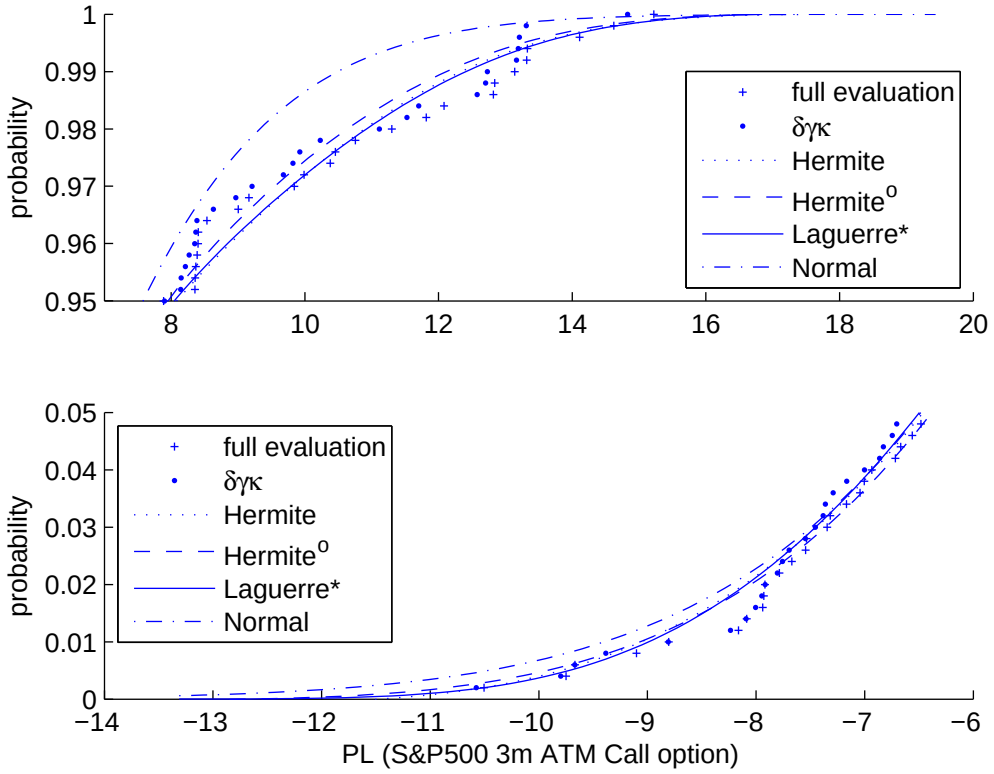

Figure 5.5: Two empirical distributions of one day PL of S&P 500 call option (three months, at-the-money), the approximations by the Hermite expansion, the optimised Hermite expansion ('Hermite°'), the Laguerre expansion with squaring ('Laguerre*') and the Normal distribution. The empirical distributions are the same as in Figure 5.3: $n = 8$ in Equation (2.2) for expansions. The upper and lower plots are the right and left tails. We can see that expansions are close to each other and capture the tail shape of 'full evaluation' better than the Normal approximation.

Tables 5.2 and 5.3 show the tail quantiles (Q) and conditional expectations (CE) of these distributions which correspond to 99% VaR and to ES. They show that the approximation by the Normal distribution has the largest difference from the 'full evaluation' which corresponds to the VaR and ES by HS method.

Probability	0.01		0.99	
	Q	CE	Q	CE
full evaluation	−178.2	−193.1	267.8	286.5
$\delta\gamma\kappa$ approx.	−177.3	−190.1	265.6	284.3
Hermite	−176.9	−200.8	249.5	281.7
Hermite$^{\circ}$	−172.0	−192.1	247.5	281.5
Laguerre*	−176.0	−207.7	250.1	284.3
Normal	−191.0	−220.6	216.1	245.7

Table 5.2: Tail quantiles (Q) of the PL of Nikkei 225 call option and the expectation with the condition (CE) that the PL exceeds the quantiles. The distributions are the same as in Figure 5.4. For the empirical distributions, the 5th smallest and largest observations are used as the 0.01 and 0.99 quantiles respectively: $n = 8$ in Equation (2.2) for expansions. CE for Laguerre* is obtained by using incomplete gamma functions (See Appendix A.4 for details), while those for Hermite expansions were obtained by using Equation (2.6).

Probability	0.01		0.99	
	Q	CE	Q	CE
full evaluation	−8.804	−9.566	13.14	13.96
$\delta\gamma\kappa$ approx.	−8.807	−9.647	12.73	13.41
Hermite	−9.054	−9.757	12.34	13.71
Hermite^{o}	−9.069	−10.106	12.12	13.48
Laguerre*	−8.988	−9.912	12.41	13.86
Normal	−9.396	−10.084	10.49	11.94

Table 5.3: Tail quantiles (Q) of the PL of S&P 500 call option and the expectation with the condition (CE) that the PL exceeds the quantiles. The distributions are the same as in Figure 5.5. For the empirical distributions, the 5th smallest and largest observations are used as the 0.01 and 0.99 quantiles respectively: $n = 8$ in Equation (2.2) for expansions. CE for Laguerre* is obtained by using incomplete gamma function (See Appendix A.4 for details), while those for Hermite expansions were obtained by using Equation (2.6).

	Nikkei 225 Call option			S&P 500 Call option		
	KS	RMSE	f^-	KS	RMSE	f^-
	($\times 10^{-2}$)	($\times 10^{-2}$)	($\times 10^{-4}$)	($\times 10^{-2}$)	($\times 10^{-2}$)	($\times 10^{-4}$)
Hermite	3.501	1.721	1.671	2.847	1.296	7.455
Hermite^{o}	4.503	2.255	2.048	3.171	1.493	2.524
Laguerre*	3.984	1.841	0.022	2.843	1.250	1.521
Normal	7.100	3.425	0	5.041	2.376	0

Table 5.4: The KS statistics, RMSE and the total area of negative density (f^-) of approximations of PLs. The distributions are the same as in Figures 5.4 and 5.5: $n = 8$ in Equation (2.2) for expansions.

From Figures 5.4 and 5.5, we can see that the three expansions agree with each other, while the Normal approximations are distinctly different from them.

More importantly, the expansions seem to capture the tail behaviour of the empirical distributions better than the Normal approximations. This can confirmed from the tail statistics in Tables 5.2 and 5.3. The tail quantiles of the expansions are much closer to those of the empirical distributions.

The measures of difference from the empirical distributions of expansions in Table 5.4 are much smaller than those of the Normal approximations.

The total area of negative density of the Hermite expansion for S&P 500 Call option is relatively large; however, non-monotonicity is not visually recognisable in Figure 5.5. The Laguerre expansion with squaring has the smallest measures of difference and total area of negative density of the three expansions for both examples. This might suggest greater robustness for the Laguerre expansion with squaring.

These results suggest that approximations to the distribution by the expansion methods are significantly different from approximations by the Normal distribution and are closer to the empirical distribution. From these findings, we might conclude that the expansion methods can capture some non-Normal characteristics of the risk.

5.4 Summary of this Chapter

We proposed a method to apply the expansion methods to risk aggregation, which can be used for the calculation of measures of market risk such as VaR and ES of a portfolio. We showed that our methods can be applied to the cases where a portfolio has non-linear and multiple risk factors. The numerical examples which deal with the market risk associated with the option premium showed that the expansion methods capture some characteristics of the market observations bet-

ter than methods based on the Normal distribution. It is also notable that the expansion methods have analytical solutions for the cumulative density function and conditional expectation, which can help us to avoid numerical integration. Further work could usefully compare our methods with those requiring such numerical integration for accuracy.

Since our methods only require the moments and cross-moments of the risk factors (and their unbiased estimators, for optimisation), they can also be applied to a wider range of cases than the examples in this thesis, such as those where some of the risk factors are given as models, as long as the moments and cross-moments are available.

Chapter 6

Multivariate Expansions

As discussed in Chapter 2, when a portfolio involves multiple risk factors, we sometimes wish to deal with a joint distribution of multiple risk factors; however, in reality, not many methods are available[1]. Use of the multivariate Normal distribution is one of the most common methods. This only requires the mean vector and the covariance matrix; however, it ignores heavy tails or dependence structure other than covariance, both of which are often observed in the market. Alternatively, empirical distributions incorporate market observations such as heavy tails or non-linear dependence structures among the risk factors. However, discreteness, especially in the tail where the observations are sparse, can be unfavourable[2]. Also, the empirical distributions do not have joint density functions, which give us a comprehensive visual image of dependence structures. By using copulas, we can describe the dependence structure among risk factors, and non-parametric copula estimation is proposed[3]; however, many other copula-based methods require specification of the copula function *a priori*.

[1]See, for instance, Mina & Xiao (2001) and Jorion (2001) for use of Normal distribution and empirical distribution. For the discussion on the use of other distributions, see Section 2.2.2.

[2]See Finger (2006) for example.

[3]See Section 2.2.2 for the use of copulas.

Besides these methods, we show in this Chapter that expansion methods can be applied to approximating the joint distribution of risk factors. In Chapter 3, we showed that marginal distributions of risk factors can be approximated by expansion methods, given their moments. Here, we show that expansion methods can approximate the joint distribution of risk factors, given moments and cross-moments. In addition to the discussion on the PL distribution with multiple risk factors in Chapter 5, deriving joint distributions from the moments and cross-moments further supports our assumption that the dependence structure among risk factors may be described in terms of their cross-moments.

In this topic, Mauleón & Perote (2000*a*) studied a bivariate expansion which incorporates linear correlation of risk factors. However, the extension of expansion methods more generally to multivariate distributions has not been studied in the literature as far as we can determine.

We also show that conditional moments and copula density functions, which also give a comprehensive view on dependence structures, can be approximated by rational functions using our expansion methods.

6.1 Expanding a bivariate joint density function

We firstly introduce in Section 6.1.1 a natural extension of expansions introduced in Section 2.2.3, and then discuss in Section 6.1.2 bivariate versions of techniques introduced in Section 3.1. Numerical examples are shown in Section 6.1.3.

6.1.1 Basic formula

Let X and Y have a joint density function $f(x, y)$ with support $\mathcal{S} \times \mathcal{S}$. Assume that the conditional distribution, which is defined as

$$f(x|Y = y) = \frac{f(x, y)}{\int_{\mathcal{S}} f(u, y)\mathrm{d}u},$$

can be expanded using Equation (2.1) as

$$f(x|Y=y) = w(x)\sum_{k=0}^{\infty} g_k(x)\int_{\mathcal{S}} f(u|Y=y)g_k(u)\mathrm{d}u.$$

By multiplying both sides by $\int_{\mathcal{S}} f(u,y)\mathrm{d}u$, we have

$$f(x,y) = w(x)\sum_{k=0}^{\infty} g_k(x)\int_{\mathcal{S}} f(u,y)g_k(u)\mathrm{d}u. \tag{6.1}$$

Similarly, we can expand the density function with respect to y,

$$f(x,y) = w(y)\sum_{l=0}^{\infty} g_l(y)\int_{\mathcal{S}} f(x,v)g_l(v)\mathrm{d}v. \tag{6.2}$$

By plugging Equation (6.2) into the right hand side of Equation (6.1), we have the expansion for a joint density function,

$$\begin{aligned} f(x,y) &= w(x)\sum_{k=0}^{\infty}\int_{u\in\mathcal{S}} w(y)\sum_{l=0}^{\infty}\int_{v\in\mathcal{S}} f(u,v)g_l(v)\mathrm{d}v\; g_l(y)g_k(u)\mathrm{d}u\; g_k(x) \\ &= w(x)w(y)\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\mathrm{E}(g_k(X)g_l(Y))g_k(x)g_l(y). \end{aligned} \tag{6.3}$$

Since g_k are polynomials, $\mathrm{E}(g_k(X)g_l(Y))$ are linear combinations of moments and cross-moments of X and Y, therefore we notice that all the information from X and Y is summarised in terms of the moments and cross-moments.

Based on Equation (6.3), given the moments and cross-moments of X and Y, $\mathrm{E}(X^pY^q)$, $p=0,\ldots,n$, $q=0,\ldots,n$, the joint density function of X and Y can be approximated by

$$\hat{f}(x,y) = w(x)w(y)\sum_{k=0}^{n}\sum_{l=0}^{n}\mathrm{E}(g_k(X)g_l(Y))g_k(x)g_l(y). \tag{6.4}$$

By construction, we also notice that the expansions applied to X and Y do not have to be the same. Also, the degree of expansions n does not have to be the same for X and Y, however, we use a common n for simplicity.

Properties of bivariate expansions

Here, a similar discussion to Section 2.2.3, where we discussed the properties of univariate expansions, can be made. That is, the approximation by expansions may preserve the moments and cross-moments of the original variables (X, Y) up to the nth order, $\mathrm{E}(X^k Y^l)$, $k = 0, \ldots, n$, $l = 0, \ldots, n$, while, for $k = n+1, \ldots$, and $l = n+1, \ldots$, the moments and cross-moments of the approximation are replaced by those of the product of the weight functions $w(x)w(y)$. Therefore, similarly to univariate expansions, we might expect a good approximation quality if the target distribution is close to $w(x)w(y)$, which, in the bivariate case, is a joint density function of two independent variables.

When some moments or cross-moments do not exist, they are also replaced by the moments or cross-moments of $w(x)w(y)$.

By construction, these bivariate expansions do not exclude the possibility of negative density.

The marginal densities f_X and f_Y of X and Y, respectively, can be expanded as

$$f_X(x) = w(x) \sum_{k=0}^{\infty} \mathrm{E}(g_k(X)) g_k(x) \tag{6.5}$$

$$f_Y(y) = w(y) \sum_{l=0}^{\infty} \mathrm{E}(g_l(Y)) g_l(y), \tag{6.6}$$

so it is easy to see that when X and Y are independent and $\mathrm{E}(g_k(X)g_l(Y)) = \mathrm{E}(g_k(X))\mathrm{E}(g_l(Y))$, we have $f(x, y) = f_X(x) f_Y(y)$.

6.1.2 Application and related techniques

Here, similarly to the discussion in Section 3.1, we expect that a naive application of the Hermite expansion can result in a poor approximation, and that the same techniques — standardisation, use of the Laguerre expansions and optimisation

— may provide expansion methods with greater robustness. We discuss the application of these techniques in the bivariate case and show numerical examples. Use of the Laguerre expansion is explained together with standardisation.

Standardisation

Similarly to the univariate case, we expect that matching first and second moments of the variables to those of the weight function by standardisation can improve the approximation quality. The difference from the univariate case is that we have the cross-moment term $\mathrm{E}(XY)$ here. Our bivariate weight function is $w(x)w(y)$, which is the joint density function of two independent variables. This suggests that the variables should be transformed so that their covariance is 0. For example, let $Z = \gamma X + Y$, where $\gamma = -(\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y))/(\mathrm{E}(X^2) - (\mathrm{E}(X))^2)$, then the covariance of X and Z is 0.

Other moments can be matched to those of the weight function in the same way as the univariate case; that is, for the Hermite expansion, let $X' = (X - \mu_X)/\sigma_X$ and $Z' = (Z - \mu_Z)/\sigma_Z$, where $\mu_X = \mathrm{E}(X)$, $\sigma_X^2 = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$, $\mu_Z = \mathrm{E}(Z)$ and $\sigma_Z^2 = \mathrm{E}(Z^2) - (E(Z))^2$, and apply the expansion to (X', Z') to obtain $\hat{f}_{X'Z'}$. The approximation of f is given by

$$\hat{f}(x, y) = \hat{f}_{X'Z'}\left(\frac{x - \mu_X}{\sigma_X}, \frac{\gamma x + y - \mu_Z}{\sigma_Z}\right)\frac{1}{\sigma_X \sigma_Z}.$$

For the Laguerre expansion, assume that $X_L = X + M_X$ and $Z_L = Z + M_Z$ can be regarded as non-negative for some $M_X > 0$ and $M_Z > 0$ and let $X'_L = \beta_{X_L} X_L / \mu_{X_L}$ and $Z'_L = \beta_{Z_L} Z_L / \mu_{Z_L}$, where

$$\mu_{X_L} = \mathrm{E}(X_L), \quad \beta_{X_L} = \frac{\mathrm{E}(X_L)}{\mathrm{E}(X_L^2) - (\mathrm{E}(X_L))^2},$$
$$\mu_{Z_L} = \mathrm{E}(Z_L), \quad \text{and} \quad \beta_{Z_L} = \frac{\mathrm{E}(Z_L)}{\mathrm{E}(Z_L^2) - (\mathrm{E}(X_L))^2}.$$

We obtain the approximation $\hat{f}_{X'_L Z'_L}$ of the joint density function of (X'_L, Z'_L) by

applying the Laguerre expansion with parameters β_X and β_Z and we have

$$\hat{f}(x,y) = \hat{f}_{X'_L Z'_L}\left(\frac{\beta_{X_L}(x+M_X)}{\mu_{X_L}}, \frac{\beta_{Z_L}(\gamma x + y + M_Z)}{\mu_{Z_L}}\right)\frac{\beta_{X_L}\beta_{Z_L}}{\mu_{X_L}\mu_{Z_L}}.$$

Note that the moments and cross-moments required for these expansions with standardisation are $\mathrm{E}(X^p Y^q)$, $p = 0, \ldots, 2n$, $q = 0, \ldots, n$.

For the Laguerre expansion with squaring, where we consider $X_{L^*} = (X + M_X)^2$ and $Y_{L^*} = (Y + M_Y)^2$, $Z_{L^*} = \gamma^* X_{L^*} + Y_{L^*}$ is defined so that the covariance between X_{L^*} and Z_{L^*} is 0; that is, $\gamma^* = (\mathrm{E}(X_{L^*}Z_{L^*}) - \mathrm{E}(X_{L^*})\mathrm{E}(Z_{L^*}))/(\mathrm{E}(X_{L^*}^2) - (\mathrm{E}(X_{L^*}))^2)$. We may need to consider $Z_{L^*} + M_Z$, $M_Z > 0$ when Z_{L^*} cannot be regarded as non-negative. Then, $X'_{L^*} = \beta_X^* X_{L^*}/\mu_X^*$ and $Z'_{L^*} = \beta_Z^*(Z_{L^*} + M_Z)/\mu_Z^*$ are the standardised variables, where the parameters are defined the same way as those for the Laguerre expansion. From the approximation $\hat{f}_{X'_{L^*}Z'_{L^*}}$ of the joint density function of (X'_{L^*}, Z'_{L^*}), we obtain

$$\begin{aligned}\hat{f}(x,y) &= 4(x+M_X)(y+M_Y) \\ &\times \hat{f}_{X'_{L^*}Z'_{L^*}}\left(\frac{\beta_X^*(x+M_X)^2}{\mu_X^*}, \frac{\beta_Z^*(\gamma^*(x+M_X)^2 + (y+M_Y)^2 + M_Z)}{\mu_Z^*}\right) \\ &\times \frac{\beta_X^*\beta_Z^*}{\mu_X^*\mu_Z^*}.\end{aligned}$$

This expansion requires $\mathrm{E}(X^p Y^q)$, $p = 0, \ldots, 4n$, $q = 0, \ldots, 2n$.

Optimisation

We can make a parallel discussion to the univariate case in Section 3.1 as follows. Consider the case where we observe i.i.d. samples $(X^{(i)}, Y^{(i)})$, $i = 1, \ldots, N$ from f, and f can be expanded as

$$f(x,y) = w(x)w(y)\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} C_{kl}^{XY} g_k(x) g_l(y),$$

where $C_{kl}^{XY} = \mathrm{E}(g_k(X)g_l(Y))$. Let $\hat{C}_{kl}^{XY} = N^{-1}\sum_{i=1}^{N} g_k(X^{(i)})g_l(Y^{(i)})$ then $\mathrm{E}(\hat{C}_{kl}^{XY}) = C_{kl}^{XY}$. We consider a class of estimator

$$\hat{f}(x,y) = w(x)w(y)\sum_{k=0}^{n}\sum_{l=0}^{n} \alpha_{kl}\hat{C}_{kl}^{XY} g_k(x) g_l(y)$$

and optimise the coefficients $\alpha_{kl} \geq 0$, $k = 0, 1, \ldots, n$, $l = 0, 1, \ldots, n$, so that the weighted mean integrated square error (MISE)

$$\mathrm{E}\left(\int_{\mathcal{S}} \int_{\mathcal{S}} \frac{\left\{ \hat{f}(u,v) - f(u,v) \right\}^2}{w(u)w(v)} \mathrm{d}u\mathrm{d}v \right) \tag{6.7}$$

is minimised. Thus Equation (6.7) is equal to

$$\sum_{k=0}^{n} \sum_{l=0}^{n} \alpha_{kl}^2 \mathrm{E}\left((\hat{C}_{kl}^{XY})^2 \right) - 2 \sum_{k=0}^{n} \sum_{l=0}^{n} \alpha_{kl} (C_{kl}^{XY})^2 + \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} (C_{kl}^{XY})^2. \tag{6.8}$$

An unbiased estimator for Equation (6.8) is given by

$$\sum_{k=0}^{n} \sum_{l=0}^{n} \alpha_{kl}^2 (\hat{C}_{kl}^{XY})^2 - 2 \sum_{k=0}^{n} \sum_{l=0}^{n} \alpha_{kl} \frac{N(\hat{C}_{kl}^{XY})^2 - (\hat{B}_{kl}^{XY})^2}{N-1} + \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{N(\hat{C}_{kl}^{XY})^2 - (\hat{B}_{kl}^{XY})^2}{N-1}, \tag{6.9}$$

where $(\hat{B}_{kl}^{XY})^2 = N^{-1} \sum_{i=1}^{N} \left(g_k(X^{(i)}) g_l(Y^{(i)}) \right)^2$. Again, $\alpha_{00} = 1$ is required so that $\int_{\mathcal{S}} \int_{\mathcal{S}} \hat{f}(u,v) \mathrm{d}u\mathrm{d}v = 1$ is satisfied. If the variables are standardised, we can set $\alpha_{kl} = 0$ for $k + l = 1, 2$. For $k + l \geq 3$, Equation (6.9) is minimised when

$$\alpha_{kl} = \frac{N(\hat{C}_{kl}^{XY})^2 - (\hat{B}_{kl}^{XY})^2}{(N-1)(\hat{C}_{kl}^{XY})^2}, \tag{6.10}$$

if $N(\hat{C}_{kl}^{XY})^2 - (\hat{B}_{kl}^{XY})^2 > 0$, and $\alpha_{kl} = 0$, otherwise.

6.1.3 Numerical examples

We show numerical examples using two data sets the same as in Chapter 5: one consists of daily log-returns of the Nikkei 225 stock index and daily log-differences of the European call option implied volatility (three month, at-the-money) on the Nikkei 225 and the other consists of daily log-returns of the S&P 500 stock index and daily log-differences of the European call option implied volatility (three month, at-the-money) on the S&P 500. Summary statistics are given in Table 3.1.

Figures 6.1 to 6.3 are numerical examples of bivariate expansions. Each example approximates the joint density of the log-return of an index and the log-difference of its implied volatility (IV).

We can see that the optimised Hermite expansion and the optimised Laguerre expansion with squaring applied to the Nikkei 225 data set in Figures 6.1 and 6.2 agree with each other: they both capture a high concentration of the observations around the origin with only a few observations around them. On the other hand, the optimised Hermite expansion applied to the S&P 500 data set in Figure 6.3 is clearly different from them: it captures a high concentration of the observations along the line from rear-left to front-right, which also can be seen as a high correlation coefficient in Table 3.1.

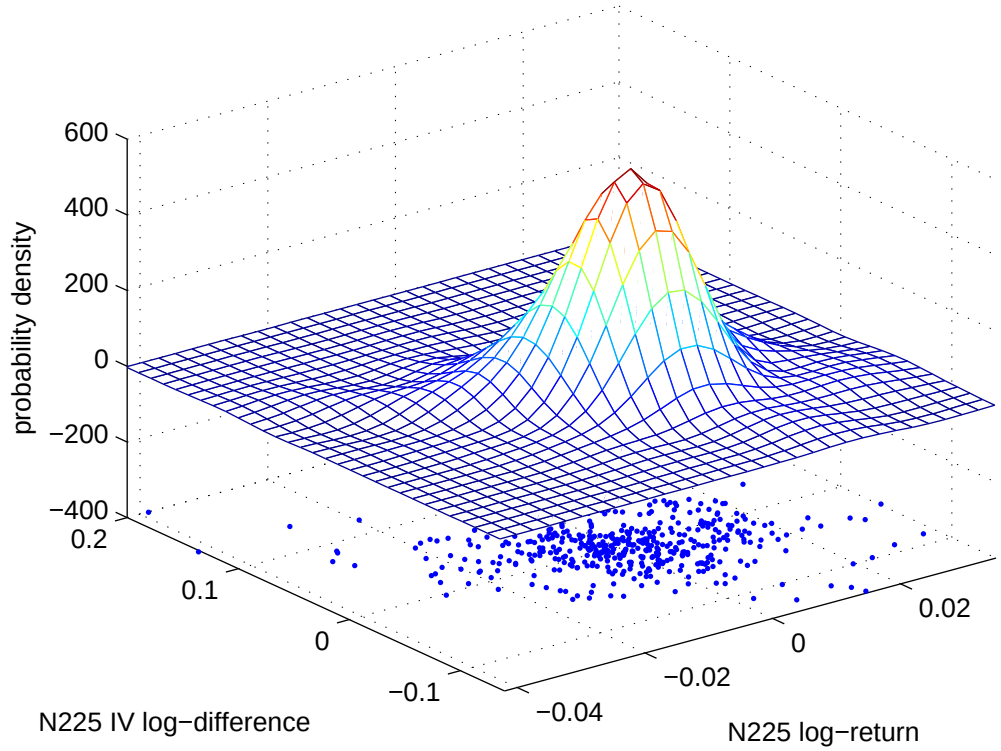

Figure 6.1: Two dimensional standardised and optimised Hermite expansion applied to Nikkei 225 log-return and the IV log-difference (3 month, at-the-money European call option). The observation period is from 25 Oct 2004 to 6 Nov 2006 (500 samples). The sums in Equation (6.4) are taken up to $n = 8$. The observations are scattered on the plane $z = -400$. We can see that it captures some characteristics of original observations, like skewness.

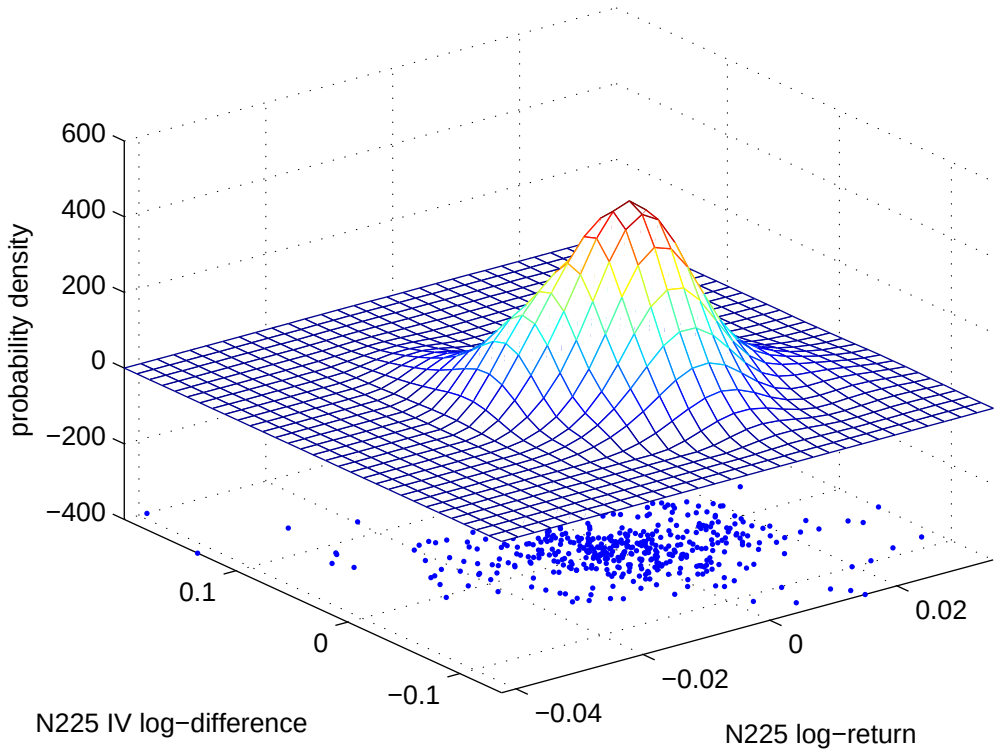

Figure 6.2: Two dimensional standardised and optimised Laguerre expansion with squaring applied to Nikkei 225 log-return and the IV log-difference (3 month, at-the-money European call option). The observation period is from 25 Oct 2004 to 6 Nov 2006 (500 samples). The sums in Equation (6.4) are taken up to $n = 8$. The observations are scattered on the plane $z = -400$.

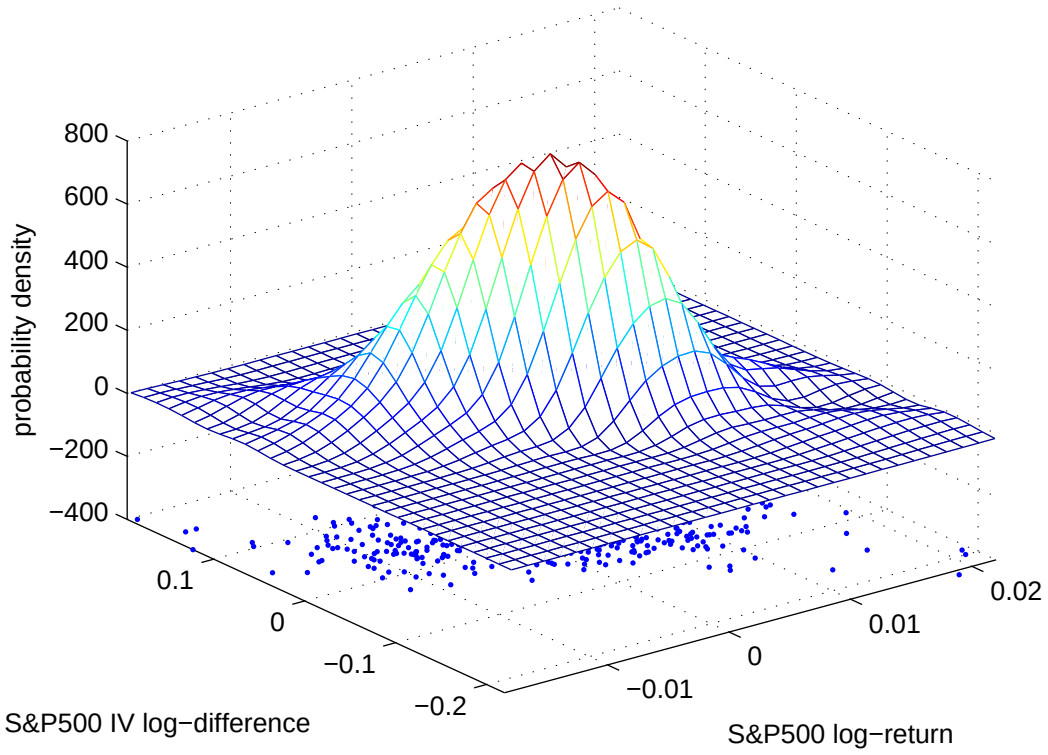

Figure 6.3: Two dimensional standardised and optimised Hermite expansion applied to S&P 500 log-return and the IV log-difference (3 month, at-the-money European call option). The observation period is from 25 Oct 2004 to 18 Oct 2006 (500 samples). The sums in Equation (6.4) are taken up to $n = 8$. The observations are scattered on the plane $z = -400$.

6.2 Conditional expectations and expansion

We show that, using expansion methods, conditional expectations can be approximated by rational functions.

6.2.1 Basic formula and related techniques

For $q = 0, 1, \ldots$, consider $\mathrm{E}(g_q(Y)|X = x)$, the expectation of $g_q(Y)$ under the condition $X = x$, defined as

$$\mathrm{E}(g_q(Y)|X = x) = \frac{\int_{\mathcal{S}} g_q(y) f(x, y) \mathrm{d}y}{\int_{\mathcal{S}} f(x, y) \mathrm{d}y}.$$

By applying expansions to f, we have

$$\begin{aligned} \mathrm{E}(g_q(Y)|X = x) &= \frac{\int_{\mathcal{S}} g_q(y) w(x) w(y) \sum_{k,l=0}^{\infty} C_{kl}^{XY} g_k(x) g_l(y) \mathrm{d}y}{w(x) \sum_{k=0}^{\infty} C_k^X g_k(x)} \\ &= \frac{\sum_{k=0}^{\infty} C_{kq}^{XY} g_k(x)}{\sum_{k=0}^{\infty} C_k^X g_k(x)}, \end{aligned} \tag{6.11}$$

where $C_{kl}^{XY} = \mathrm{E}(g_k(X) g_l(Y))$ and $C_k^X = \mathrm{E}(g_k(X))$. We can approximate the conditional expectation by taking sums in Equation (6.11) up to some finite n.

The conditional moments such as $\mathrm{E}(Y^q|X = x)$ can be obtained by solving a system of linear equations in $\mathrm{E}(g_0(Y)|X = x), \ldots, \mathrm{E}(g_q(Y)|X = x)$.

Standardisation

Conceptually, Equation (6.11) gives an expression for the conditional expectation as a rational function; however, we might consider the standardisation introduced in Section 6.1 in order to improve the approximation quality, that is, we use Equation (6.11) for the standardised variables and then obtain the conditional moments of the original variables.

For the Hermite expansion, we apply Equation (6.11) to (X', Z') where $X' = (X - \mu_X)/\sigma_X$ and $Z' = (\gamma X + Y - \mu_Z)/\sigma_Z$, and we obtain the rational approximations of $\mathrm{E}(g_i(Z')|X' = x')$, $i = 0, \ldots, q$. From these we derive $\mathrm{E}(Z'^i|X = x)$,

$i = 0, \ldots, q$, where $x' = (x - \mu_x)/\sigma_X$. Since $\mathrm{E}(Z'^i|X = x)$ is expanded as

$$\mathrm{E}(Z'^i|X = x) = \frac{1}{\sigma_Z^i} \sum_{j=0}^{i} \binom{i}{j} \mathrm{E}(Y^j|X = x)(x - \mu_X)^{i-j},$$

$\mathrm{E}(Y^i|X = x)$, $i = 0, \ldots, q$, can be calculated.

For the Laguerre expansion, the procedure is almost the same as that of the Hermite expansion, except that the standardised variables are given by $X_L = \beta_{X_L}(X + M_X)/\mu_{X_L}$ and $Z_L = \beta_{Z_L}(\gamma X + Y + M_Z)/\mu_{Z_L}$.

For the Laguerre expansion with squaring, however, a similar procedure only provides $\mathrm{E}\left((Y + M_Y)^{2i}|X = x\right)$, $i = 0, \ldots, q$, and the conditional moments of Y cannot be obtained.

Optimisation

Consider again the case where we observe i.i.d. samples $(X^{(1)}, Y^{(1)}), \ldots, (X^{(N)}, Y^{(N)})$ from f, and f can be expanded as

$$f(x, y) = w(x)w(y) \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} C_{kl}^{XY} g_k(x) g_l(y).$$

Define $h_q(x) \equiv \int_{\mathcal{S}} g_q(v) f(x, v) \mathrm{d}v = w(x) \sum_{k=0}^{\infty} C_{kq}^{XY} g_k(x)$, then $E(g_q(Y)|X = x) = h_q(x)/f_X(x)$. By considering the estimator

$$\hat{h}_q(x) = w(x) \sum_{k=0}^{n} \alpha_{kq} \hat{C}_{kq}^{XY},$$

which minimises the weighted MISE

$$\mathrm{E}\left(\int_{\mathcal{S}} \frac{\left\{ \hat{h}_q(u) - h_q(u) \right\}^2}{w(u)} \mathrm{d}u \right), \tag{6.12}$$

we find that exactly the same α_{kq} as those in Section 6.1 are optimal.

Now $\hat{f}_X$, the estimator for the marginal distribution of X, can be optimised as the univariate distribution.

6.2.2 Numerical examples

Figures 6.4 and 6.5 are numerical examples using the same data sets as in Figures 6.1 to 6.3. Both plots show the expected value of IV log-difference conditioned on stock index log-return (solid line) and ±conditional standard deviation (dotted lines). Both are obtained by applying the standardised and optimised Hermite expansion whose sums in Equation (6.11) are taken up to $n = 8$.

From Figure 6.4, we can see that the conditional standard deviation vanishes at the tail parts. This is because the approximation for the variance in these regions is negative and the standard deviation is imaginary. This suggests that the expansion is not successful in these regions. In fact a close look at these parts in Figure 6.6 reveals that the approximation for the probability density is slightly negative.

On the other hand, the expansion applied to the S&P 500 data set in Figure 6.5 does not suffer from negative density. This is most likely to be due to the fact that the conditional expectation of S&P 500 data set is closer to being linear than that of Nikkei 225 data set, and therefore the standardisation by a linear transformation more successfully reduces dependency between the variables. As discussed in Section 6.1, this can result in a better approximation quality.

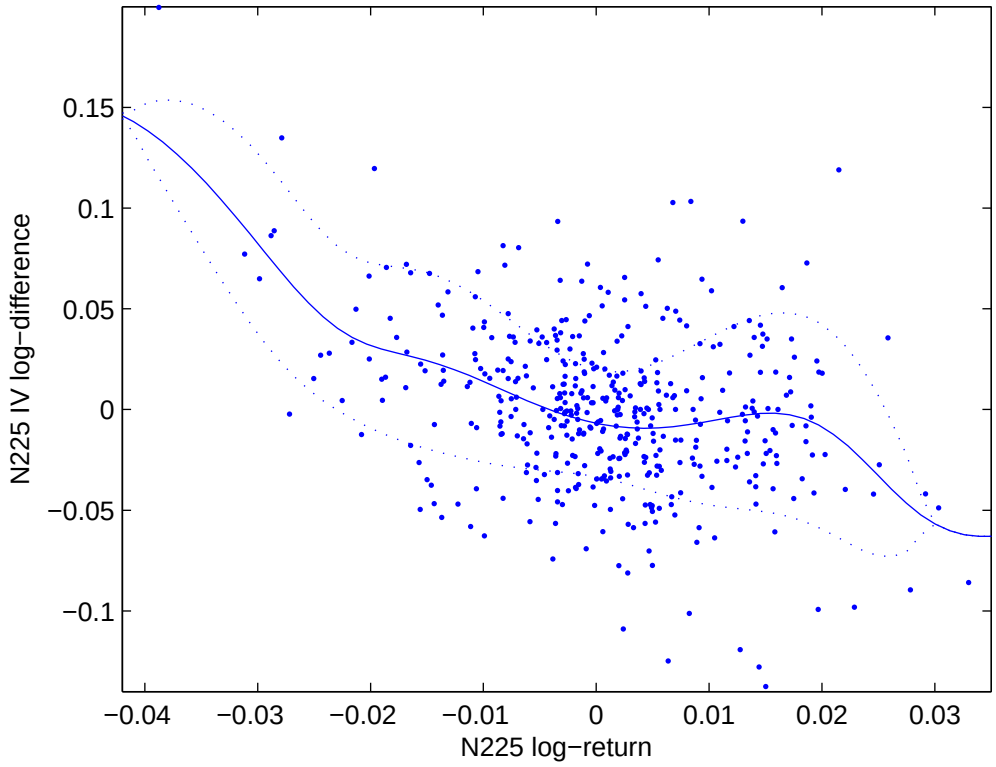

Figure 6.4: Expected value of Nikkei 225 IV log-difference conditioned on Nikkei 225 log-return (solid line) and ±conditional standard deviation (dotted lines). Both are obtained by applying the standardised and optimised Hermite expansion whose sums in Equation (6.11) are taken up to $n = 8$. The data used are the same as in Figures 6.1 and 6.2. Scattered dots represent observations. We can see that the conditional standard deviation vanishes at the tail parts. This is because the approximation for the variance in these regions is negative and the standard deviation is imaginary. This suggests that the expansion is not successful in these regions. In fact a close look at these parts reveals that the approximation for the probability density is slightly negative. See Figure 6.6.

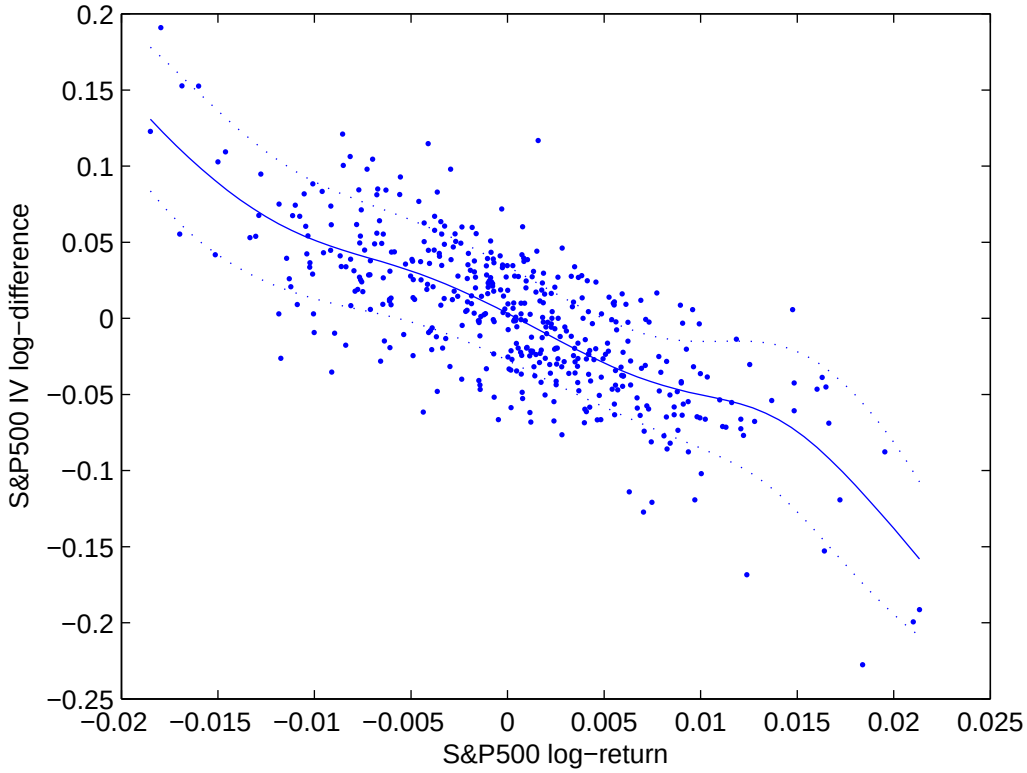

Figure 6.5: Expected value of S&P 500 IV log-difference conditioned on S&P 500 log-return (solid line) and ±conditional standard deviation (dotted lines). Both are obtained by applying the standardised and optimised Hermite expansion whose sums in Equation (6.11) are taken up to $n = 8$. The data used are the same as in Figure 6.3. Scattered dots represent the observations.

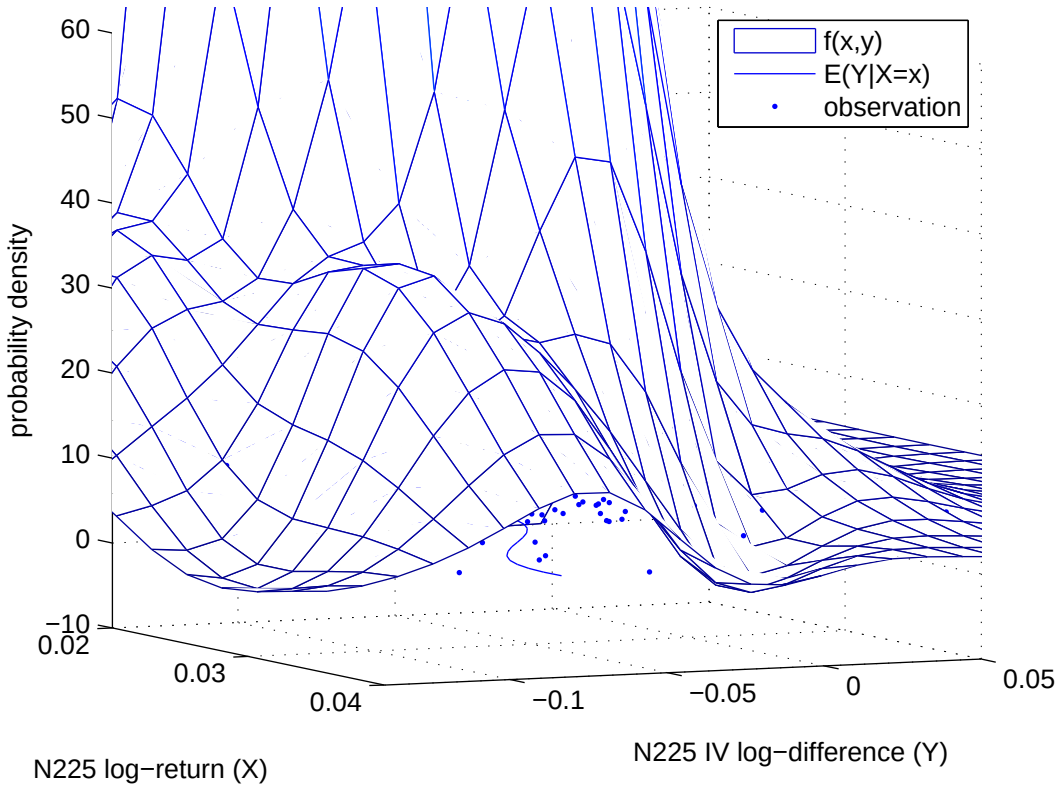

Figure 6.6: A close look at the tail of the approximated density of Nikkei 225 log-return and Nikkei 225 IV log-difference. See Figures 6.1 and 6.4. We see negative density around $(0.035, -0.05)$. The numerically obtained minimum value of the density is -8.48, which is about 1.4% of the density maximum 595.67 in absolute value.

6.3 Copula functions and expansion

Here we show that, using expansion methods, copula densities can be approximated by rational functions.

6.3.1 Copula functions and densities

The two dimensional copula is a function $C : [0,1]^2 \mapsto [0,1]$ which satisfies

$$C(t,0) = C(0,t) = 0, \tag{6.13}$$

$$C(1,t) = C(t,1) = t, \tag{6.14}$$

for all $0 \leq t \leq 1$ and

$$C(u_1,v_1) - C(u_2,v_1) - C(u_1,v_2) + C(u_2,v_2) \geq 0, \tag{6.15}$$

for all $0 \leq u_1 \leq u_2 \leq 1$ and $0 \leq v_1 \leq v_2 \leq 1$.

Let $F(x,y)$ be the joint distribution of X and Y and $F_X(x)$ and $F_Y(y)$ be their marginal distributions. Then, Sklar's theorem[4] yields that there is a unique copula representation for F, given by

$$F(x,y) = C(F_X(x), F_Y(y)). \tag{6.16}$$

Assume that F is differentiable. Then, from Equation (6.16), the joint density function is given by

$$f(x,y) = f_X(x) f_Y(y) c(F_X(x), F_Y(y)), \tag{6.17}$$

where

$$c(u,v) = \frac{\partial^2}{\partial u \partial v} C(u,v)$$

is the copula density and f_X and f_Y are the marginal densities of X and Y, respectively.

[4]For details about Sklar's theorem and other properties of copulas, see for example Nelsen (1999).

6.3.2 Basic formula and related techniques

From Equations (6.3), (6.5) and (6.6), we have an expression for c in Equation (6.17) as a rational function:

$$\begin{aligned} c(F_X(x), F_Y(y)) &= \frac{f(x,y)}{f_X(x)f_Y(y)} \\ &= \frac{\sum_{k,l} C^{XY}_{kl} g_k(x) g_l(y)}{\sum_{k,l} C^X_k C^Y_l g_k(x) g_l(y)}. \end{aligned} \tag{6.18}$$

or

$$c(u,v) = \frac{\sum_{k,l} C^{XY}_{kl} g_k(F_X^{-1}(u)) g_l(F_Y^{-1}(v))}{\sum_{k,l} C^X_k C^Y_l g_k(F_X^{-1}(u)) g_l(F_Y^{-1}(v))}. \tag{6.19}$$

Note that it is also possible to remove the marginal functions F_X and F_Y from Equation (6.19) by applying an expansion with support $[0,1]$ to the variables $(U,V) = (F_X(X), F_Y(Y))$. However, the advantage of our method is that it only requires moments and cross-moments of the original variables. See Freedman (1981) for an expansion with closed support.

Standardisation and optimisation

From Equations (6.11) and (6.18), we find that the copula density can be expressed using the conditional expectation

$$c(F_X(x), F_Y(y)) = \frac{\sum_{l=0}^{\infty} \mathrm{E}(g_l(Y)|X=x) g_l(y)}{\sum_{l=0}^{\infty} C^Y_l g_l(y)}. \tag{6.20}$$

This suggests that standardisation and optimisation are available for the copula density as long as they are available for the calculation of conditional expectation.

6.3.3 Numerical examples

Figures 6.7 and 6.8 are the examples using the same data sets as in previous sections. Figure 6.9 is a copula density obtained by applying standardised and optimised Hermite expansion to 500 random samples from the Clayton copula

($\theta = 2.882$) with the Normal marginal distributions. The Clayton copula function with parameter θ is defined by $C(u, v) = (u^\theta + v^\theta - 1)^{\frac{1}{\theta}}$. By comparing Figure 6.9 with the theoretical shape of the Clayton copula density in Figure 6.10, we can get some idea how close the copula by the expansion can be to the 'true' copula density. In fact, except for the tails where the observations are sparse, the expansion produces a shape similar to that of the Clayton copula density.

From Figures 6.7 and 6.8, we can see that the dependence structures inherent in the Nikkei 225 data set is different from that in the S&P 500 data set.

We can see from all of the Figures 6.7, 6.8 and 6.9, that the approximations are erratic near the boundary. This is most likely to be because, as we can see from Equations (6.19) and (6.20), our method uses rational functions, which can be erratic at the extremes, for approximating the copula densities.

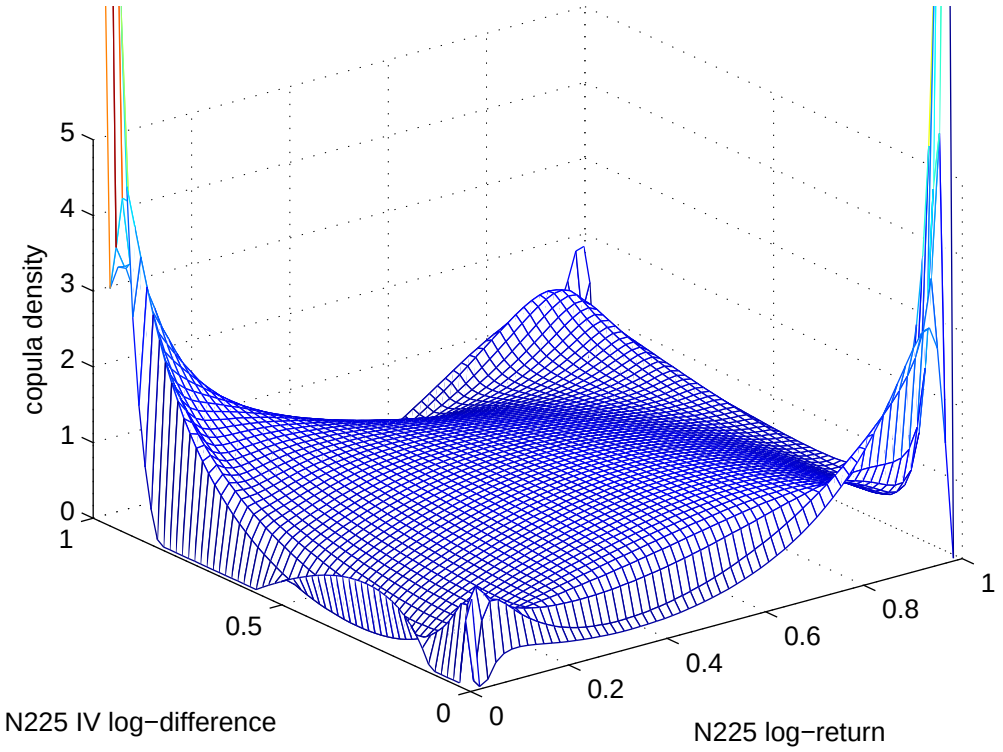

Figure 6.7: Copula density obtained by applying standardised and optimised Hermite expansion to Nikkei 225 log-return and Nikkei 225 IV log-difference. The data used are the same as in Figures 6.1 and 6.2. Sums in Equation (6.20) are taken up to $n = 8$.

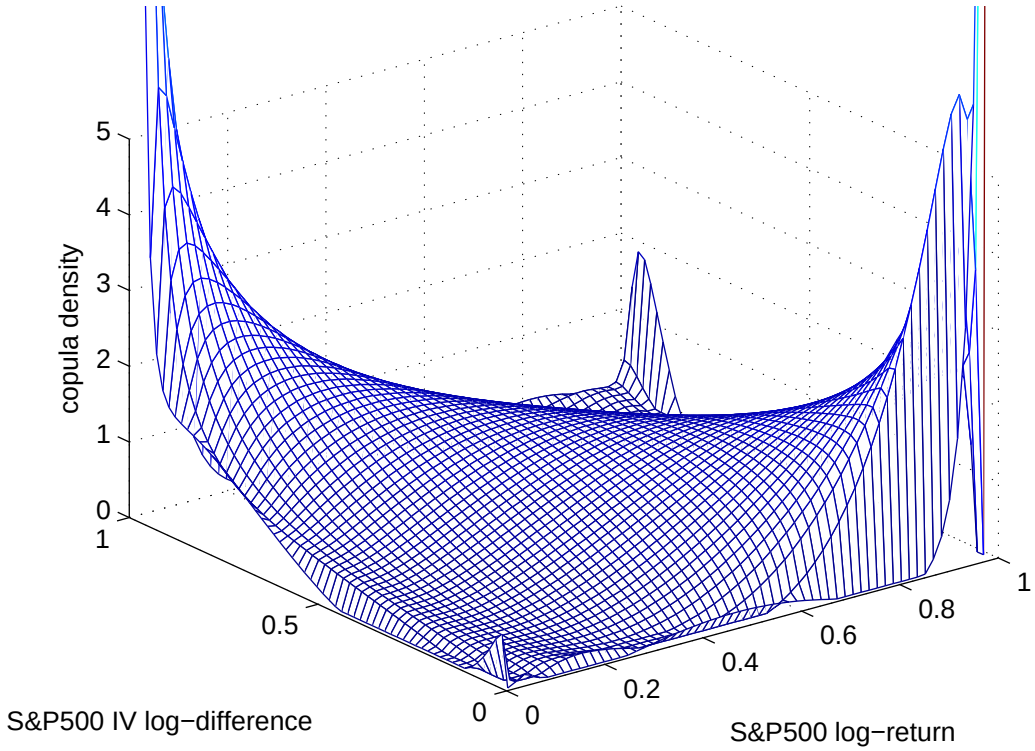

Figure 6.8: Copula density obtained by applying standardised and optimised Hermite expansion to S&P 500 log-return and S&P 500 IV log-difference. The data used are the same as in Figure 6.3. Standardised and optimised. Sums in Equation (6.20) are taken up to $n = 8$.

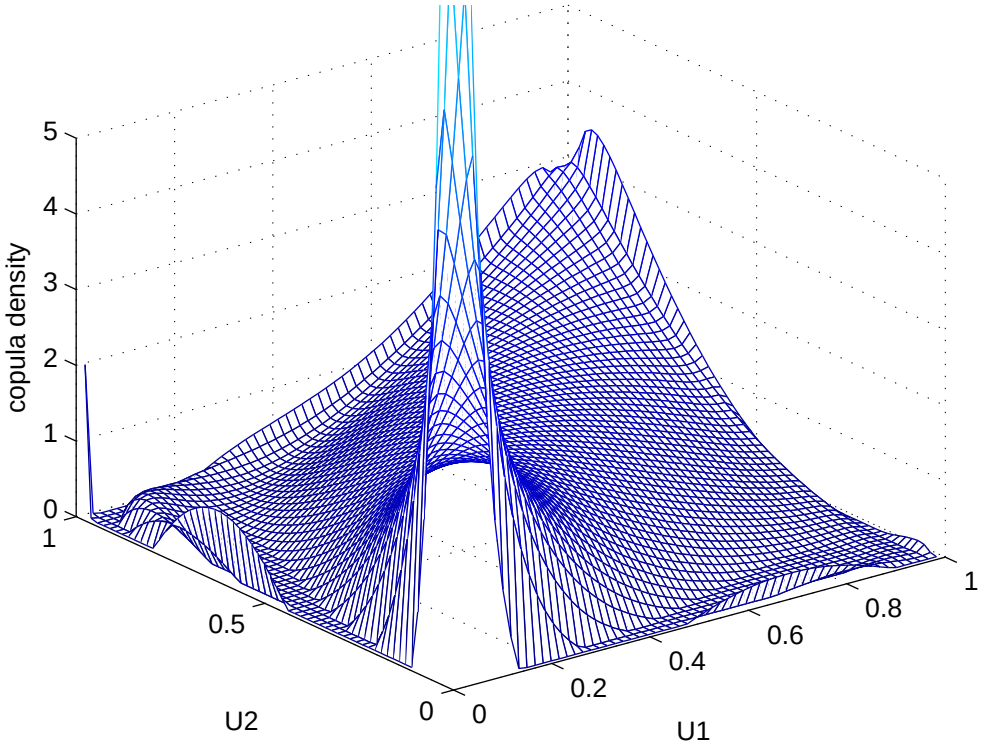

Figure 6.9: Copula density obtained by applying standardised and optimised Hermite expansion to 500 random samples from the Clayton copula ($\theta = 2.882$) with the Normal marginal distributions. Sums in Equation (6.20) are taken up to $n = 8$. By comparing this with theoretical value of the Clayton copula density in Figure 6.10, we can get some idea of how close the copula approximated by the expansion can be to the 'true' copula density. In fact, except for the tails where the observations are sparse, the expansion produces a shape similar to that of the Clayton copula density.

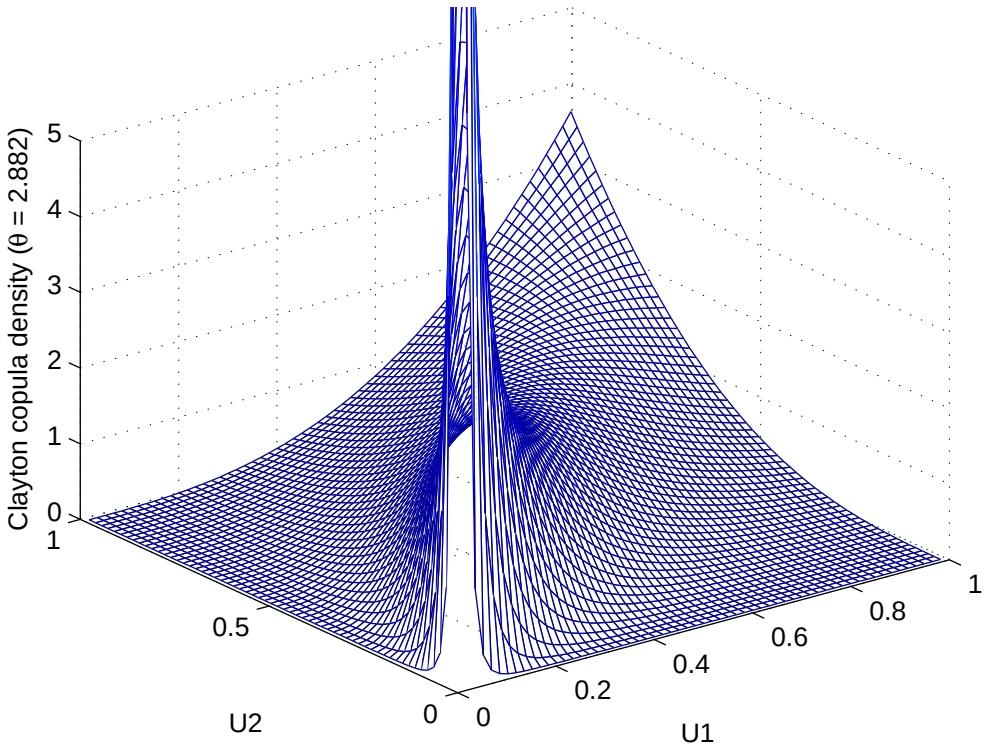

Figure 6.10: The Clayton copula density ($\theta = 2.882$). The Clayton copula function with parameter θ is given by $C(u, v) = (u^\theta + v^\theta - 1)^{\frac{1}{\theta}}$. (See also Figure 6.9.)

6.4 Extension to $p-$variate case

Let $f(x_1, \ldots, x_p)$ be the joint density function of $(X_1, \ldots, X_p)$, $p \geq 3$. Similarly to Section 6.1, we have

$$f(x_1, \ldots, x_p) = w(x_1) \cdots w(x_p) \times \sum_{k_1, \ldots, k_p} \mathrm{E}(g_{k_1}(X_1) \cdots g_{k_p}(X_p)) g_{k_1}(x_1) \cdots g_{k_p}(x_p). \tag{6.21}$$

Standardisation is done by a linear transformation of $(X_1, \ldots, X_p)$ into $(Z_1, \ldots, Z_p)$ so that the covariances between Z_i and Z_j $(i \neq j)$ are 0. Such a linear transformation can be found, for example, by inverting the Cholesky decomposition of the covariance matrix of the original variables. Note that Z_i obtained in this way is a linear combination of $X_1, \ldots, X_i$.

For the case where we have N i.i.d. samples, we can make a parallel discussion on optimisation to that of the bivariate case.

Conditional expectations

Consider the expectation of $g_{k_{i+1}}(X_{i+1}) \cdots g_{k_p}(X_p)$ under the condition $X_1 = x_1, \ldots, X_i = x_i$. From Equation (6.21), we have

$$\mathrm{E}(g_{k_{i+1}}(X_{i+1}) \cdots g_{k_p}(X_p) | X_1 = x_1, \ldots, X_i = x_i) = \frac{\sum_{k_1, \ldots, k_i} \mathrm{E}(g_{k_1}(X_1) \cdots g_{k_p}(X_p)) g_{k_1}(x_1) \cdots g_{k_i}(x_i)}{\sum_{k_1, \ldots, k_i} \mathrm{E}(g_{k_1}(X_1) \cdots g_{k_i}(X_i)) g_{k_1}(x_1) \cdots g_{k_i}(x_i)}. \tag{6.22}$$

Conditional moments such as $\mathrm{E}(X_{i+1}^{k_{i+1}} \cdots X_p^{k_p} | X_1, \ldots, X_i)$ are obtained by solving a system of linear equations of $\mathrm{E}(g_{r_{i+1}}(X_{i+1})) \cdots g_{r_p}(X_p) | X_1, \ldots, X_i)$, $r_{i+1} \in \{0, \ldots, k_{i+1}\}$, $\ldots$, $r_p \in \{0, \ldots, k_p\}$.

Standardisation is done by applying Equation (6.22) to $(Z_1, \ldots, Z_p)$ and solving a system of linear equations, to obtain the conditional moments of $(X_1, \ldots, X_p)$. We can make a parallel discussion on optimisation to that of bivariate case.

Copula density

Using Equation (6.21), a p-variate copula density function can be expressed as

$$\begin{aligned} & c(F_1(x_1), \ldots, F_p(x_p)) \\ & \quad = \frac{f(x_1, \ldots, x_p)}{f_1(x_1) \cdots f_p(x_p)} \\ & \quad = \frac{\sum_{k_1,\ldots,k_p} \mathrm{E}(g_{k_1}(X_1) \cdots g_{k_p}(X_p)) g_{k_1}(x_1) \cdots g_{k_p}(x_p)}{\sum_{k_1,\ldots,k_p} \mathrm{E}(g_{k_1}(X_1)) \cdots \mathrm{E}(g_{k_p}(X_p)) g_{k_1}(x_1) \cdots g_{k_p}(x_p)}, \end{aligned} \tag{6.23}$$

where F_i is the marginal distribution function of X_i and f_i is its density function. From Equations (6.22) and (6.23), we have

$$\begin{aligned} & c(F_1(x_1), \ldots, F_p(x_p)) = \\ & \quad \frac{\sum_{k_2} \mathrm{E}(g_{k_2}(X_2)|X_1 = x_1) g_{k_2}(x_2)}{\sum_{k_2} \mathrm{E}(g_{k_2}(X_2)) g_{k_2}(x_2)} \times \cdots \\ & \qquad \times \frac{\sum_{k_p} \mathrm{E}(g_{k_p}(X_p)|X_1 = x_1, \ldots, X_{p-1} = x_{p-1}) g_{k_p}(x_p)}{\sum_{k_p} \mathrm{E}(g_{k_p}(X_p)) g_{k_p}(x_p)}, \end{aligned}$$

and we can apply standardisation and optimisation to this whenever they are available for the conditional moments.

6.5 Summary of this Chapter

We showed that simple extensions of univariate expansion methods are applicable to joint density functions, and demonstrated using bivariate case. It is also notable that, using our multivariate expansions, the conditional moments and the copula densities can be approximated by rational functions.

Related techniques such as standardisation, use of the Laguerre expansions and optimisation can make the approximation quality sufficient for many purposes such as visualising the non-linear dependence structure, approximating the conditional mean, and approximating the copula density; however, even with these techniques, the expansions showed negative density. Dealing with this problem can be our future work.

Chapter 7

Conclusion

In this thesis, we firstly posed three problems related to market risk measurement: description, time aggregation and risk aggregation problems. We reviewed that existing methods including those using the Normal distribution and HS methods, can provide solutions to one or two of these problems; however, they have difficulty in solving at least one of these problems.

Here, we focused on the potential ability of expansion methods to solve all three problems. Expansion methods can approximate a target distribution using its moments only, and therefore have a potential to be applied to a wide range of situations, although they can lack robustness.

Then we proposed three techniques, combinations of which can provide robustness for the conventional use of expansion methods: standardisation, use of Laguerre expansions and optimisation. It is notable that use of Laguerre expansions and optimisation have not been introduced in the literature to the best of our knowledge.

Using expansion methods together with these techniques, the three problems were dealt with as follows.

The description problem — how we describe the distributions and dependence structure of risk factors — was shown solved: the marginal distributions of risk

factors can be described in terms of their moments (Chapter 3), and the dependence structure among them can be described in terms of their cross-moments (Chapters 5 and 6).

In Chapter 4, it was shown that, given autocross-moments, expansion methods can be applied to the time aggregation problem: how we deal with risk factors with a long time horizon. Since the HS methods do not have a reasonable solution to this problem, we discussed the idea that approximations by expansion methods can be used as proxies.

In Chapter 5, we showed that expansion methods can provide a reasonable solution to the risk aggregation problem: how we derive the PL distribution given a description of the risk factors. Here, expansion methods do not assume a certain class of models for the risk factors or the portfolio. They only require two conditions: that the moments and cross-moments of risk factors are available, and that the function d, which links risk factors to the PL, can be approximated by its Taylor expansion.

Numerical examples using observations on real markets and GARCH models showed that the solutions to these problems by expansion methods can be fairly reasonable. They provided good approximations for empirical distributions, where empirical distributions were available. We might deduce from this that the approximation quality of expansion methods can also be good for the cases where empirical or 'true' distributions are not available, such as applications to the time aggregation or GARCH models.

Based on the findings in these Chapters, we might conclude that expansion methods provide reasonable solutions to all three problems. As discussed in Section 2.2.2, HS methods can also provide reasonable solutions to description and risk aggregation problems, when a sufficient number of observations is available. However, the expansion methods can offer solutions even to the cases where HS methods cannot be applied, such as the time aggregation and cases with

parametric models including the GARCH models.

In Chapter 6, we proposed methods to deal with joint distributions using expansion methods. These new methods have a potential for further applications. In fact, applications to approximating conditional expectations and copula densities have not been introduced in the literature as far as we can determine. Numerical examples for two dimensional distributions showed that expansion methods can successfully produce a visual representation of the dependence structure between two random variables from their moments and cross-moments. At the same time, one example showed visually recognisable negative density, even with our techniques to provide robustness. This may be suggesting that another technique to provide further robustness can be useful for multivariate cases. This can be our future work.

Also, our future work may involve comparison of expansion methods with other competing methods. For instance, kernel methods can do many of the things that expansion methods can, as long as a sufficient number of observations are available. However, unlike for our expansion methods, it may not be straightforward to apply kernel methods to the cases where variables are given as models, rather than as a set of observations. Besides kernel methods, it can be of interest to compare our methods with expansion methods which use other orthogonal systems, such as wavelets[1]. We notice that, however, our expansion methods only require the moments and cross-moments of the target distributions and that this may not be the case with other expansion methods in general. We expect that this point is an advantage for our methods over expansion methods based on other orthogonal systems.

Since our expansion methods do not require intensive computations — once the moments of the target distribution are obtained, the approximation only requires evaluation of polynomials and of an associated weight function — and

[1]See (Hall & Neumeyer 2006), for instance.

the logic behind the methods can be easily understood, we might expect that they can be adapted to further use in industry and the academy.

Appendix A

Appendix

A.1 Schmidt's process

Here, we review the Schmidt's process, which can be used for obtaining an orthonormal polynomial basis $\{g_k\}$ for a given weight function w with support $\mathcal{S}$, which satisfies

$$\int_{\mathcal{S}} w(u)\mathrm{d}u = 1$$

and $w(x) \geq 0$ for $x \in \mathcal{S}$. Jackson (1963) explains a more general method where the basis is not limited to polynomials.

Firstly,

$$g_0(x) \equiv 1$$

is given by a similar discussion to those in Section 2.2.3.

For $k = 1$, let

$$G_1(x) = x - c_{1,0}g_0(x),$$

where

$$c_{1,0} = \int_{\mathcal{S}} w(u)ug_0(u)\mathrm{d}u,$$

and

$$e_1 = \int_{\mathcal{S}} w(u)\{G_1(u)\}^2\mathrm{d}u,$$

then g_1 is given by

$$g_1(x) = \frac{G_1(x)}{\sqrt{e_1}}.$$

For $k = 2, 3, \ldots$, similarly define

$$G_k(x) = x^k - \sum_{i=0}^{k-1} c_{k,i} g_i(x),$$

where

$$c_{k,i} = \int_{\mathcal{S}} w(u) u^k g_i(u) \mathrm{d}u,$$

and

$$e_k = \int_{\mathcal{S}} w(u) \{G_k(u)\}^2 \mathrm{d}u,$$

then g_k is given by

$$g_k(x) = \frac{G_k(x)}{\sqrt{e_k}}.$$

We find from this procedure, that w needs to have moments up to the $2n$th order to allow construction of the orthonormal polynomials up to nth order.

A.2 Tractable formula for MISE

We derive Equation (3.2) in Section 3.1. We have

$$\begin{aligned}
\frac{\left\{\hat{f}(u) - f(u)\right\}^2}{w(u)} &= w(u) \left\{ \sum_{k=0}^{n} \alpha_k \hat{C}_k g_k(u) - \sum_{k=0}^{\infty} C_k g_k(u) \right\}^2 \\
&= w(u) \Bigg\{ \sum_{k=0}^{n} \sum_{l=0}^{n} \alpha_k \alpha_l \hat{C}_k \hat{C}_l g_k(x) g_l(x) \\
&\qquad -2 \sum_{k=0}^{n} \sum_{l=0}^{\infty} \alpha_k \hat{C}_k C_l g_k(u) g_l(u) \\
&\qquad + \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} C_k C_l g_k(u) g_l(u) \Bigg\}.
\end{aligned}$$

By orthonormality, we have

$$\int_{\mathcal{S}} \frac{\left\{\hat{f}(u) - f(u)\right\}^2}{w(u)} \mathrm{d}u = \sum_{k=0}^{n} \alpha_k^2 \hat{C}_k^2 - 2 \sum_{k=0}^{n} \alpha_k \hat{C}_k C_k + \sum_{k=0}^{\infty} C_k^2.$$

By taking expectation of both sides, we have Equation (3.2). Since

$$\mathrm{E}\left(\frac{N\hat{C}_k^2 - \hat{B}_k^2}{N-1}\right) = C_k^2,$$

Equation (3.3) is an unbiased estimator for Equation (3.2).

A.3 Definitions of terms associated with stochastic processes

Here we give some definitions of terms associated with stochastic processes, which can help with understanding the definition of GARCH models in Section 3.2.2. These ideas are given rather axiomatically, and therefore may not have a clear relationship with intuitive interpretation.

Standard references for discrete time stochastic processes include Williams (1991).

A set called the sample space is denoted by Ω. Each element $\omega \in \Omega$ might be intuitively interpreted as being a realisation of one possible occurrence, or a possible sample path, and Ω as being the set of all possible occurrences.

A $\sigma-$algebra $\mathcal{F}$ on Ω is defined as a family of subsets of Ω which satisfies the following.

- $\Omega \in \mathcal{F}$.
- $A \in \mathcal{F} \Rightarrow A^C \in \mathcal{F}$, where $A \cup A^C = \Omega$ and $A \cap A^C = \emptyset$.
- $A_1 \in \mathcal{F}, A_2 \in \mathcal{F}, \ldots \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

A probability measure P on $\mathcal{F}$ can be defined as a map $P : \mathcal{F} \mapsto [0, 1]$ which satisfies the following.

- $A \in \mathcal{F} \Rightarrow 0 \leq P(A) \leq 1$.

- $P(\Omega) = 1$.
- $A_i \cap A_j = \emptyset, i \neq j \Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

A combination of these three, $(\Omega, \mathcal{F}, P)$ is called the probability space. For a given probability space $(\Omega, \mathcal{F}, P)$, a filtration is an increasing sequence of $\sigma-$algebra which satisfies $\{\emptyset, \Omega\} \subseteq \mathcal{F}(0) \subseteq \mathcal{F}(1) \subseteq \cdots \subseteq \mathcal{F}(\infty) \subseteq \mathcal{F}$. Intuitively, it can be interpreted as a sequence of sets of information available at corresponding times.

An $\mathcal{F}-$measurable random variable X is defined as a map $X : \Omega \mapsto \mathbf{R}$ which satisfies $\{\omega : X(\omega) \leq x\} \in \mathcal{F}$ for all $x \in \mathbf{R}$. A stochastic process $\{X(\omega)(t) : \omega \in \Omega, t = 1, 2, \ldots\}$ is a sequence of random variables, and it is adapted to a filtration $\{\mathcal{F}(t)\}$ if $X(t)$ is $\mathcal{F}(t)-$measurable for all $t = 1, 2, \ldots$.

Now let X be an $\mathcal{F}-$measurable random variable satisfying $\mathrm{E}(|X|) < \infty$ and $\mathcal{G} \subseteq \mathcal{F}$ be a $\sigma-$algebra. The conditional expectation of X given $\mathcal{G}$ is defined as a $\mathcal{G}-$measurable random variable Y which satisfies $\mathrm{E}(|Y|) < \infty$ and

$$\int_g Y \mathrm{d}P = \int_g X \mathrm{d}P, \tag{A.1}$$

for all $g \in \mathcal{G}$, and is denoted as $\mathrm{E}(X|\mathcal{G}) = Y$. It is known that such a random variable Y is unique, in an 'almost sure' sense. That is, let $\tilde{Y}$ be another $\mathcal{G}-$measurable random variable satisfying $\mathrm{E}(|\tilde{Y}|) < \infty$ and Equation (A.1), then $P(\tilde{Y} = Y) = 1$ is satisfied[1].

A.4 Incomplete gamma function and approximations

Here we show that

$$I_\alpha = \int_{-\infty}^{x_\alpha} u f(u) \mathrm{d}u, \tag{A.2}$$

[1]For proof, see, for example, Chapter 9 of Williams (1991).

where x_α is the $\alpha-$quantile of a random variable X and f is its pdf, can be approximated in term of incomplete gamma functions, when f is approximated by the Laguerre expansion with squaring. That is, f is approximated by

$$\hat{f}_X(x) = \frac{2\beta}{\mu}(x+M)\hat{f}_{L^*}\left(\frac{\beta}{\mu}(x+M)^2\right),$$

where

$$\hat{f}_{L^*}(z) = \frac{z^{\beta-1}\mathrm{e}^{-z}}{\Gamma(\beta)}\sum_{k=0}^{n} C_k L_k^{(\beta-1)}(z),$$

$\mu = \mathrm{E}((X+M)^2)$, $\beta = \mu^2/(E((X+M)^4) - \mu^2$ and M is some constant. Now,

$$x\hat{f}_X(x) = \frac{2\beta}{\mu}(x+M)^2\hat{f}_{L^*}\left(\frac{\beta}{\mu}(x+M)^2\right) - M\frac{2\beta}{\mu}(x+M)\hat{f}_{L^*}\left(\frac{\beta}{\mu}(x+M)^2\right),$$

and therefore I_α in Equation (A.2) can be approximated by

$$\begin{aligned}
\hat{I}_\alpha &= \int_{-M}^{x_\alpha} u\hat{f}(u)\mathrm{d}u \\
&= \sqrt{\frac{\mu}{\beta}}\int_0^{z_\alpha} z^{\frac{1}{2}}\hat{f}_{L^*}(z)\mathrm{d}z - M\int_0^{z_\alpha}\hat{f}_{L^*}(z)\mathrm{d}z \\
&= \sqrt{\frac{\mu}{\beta}}\int_0^{z_\alpha} z^{\frac{1}{2}}\hat{f}_{L^*}(z)\mathrm{d}z - M\alpha, \qquad \text{(A.3)}
\end{aligned}$$

where z_α is the $\alpha-$quantile of $\beta(X+M)^2/\mu$ given by $z_\alpha = \beta(x_\alpha+M)^2/\mu$. Since

$$z^{\frac{1}{2}}\hat{f}_{L^*}(z) = \frac{1}{\Gamma(\beta)}\sum_{k=0}^{n}\sum_{i=0}^{k} C_k l_{k,i}^{(\beta-1)} z^{\beta+i+\frac{1}{2}-1}\mathrm{e}^{-z},$$

where $l_{k,i}^{(\beta-1)}$ are constants determined by the Laguerre polynomials

$$L_k^{(\beta-1)}(z) = \sum_{i=0}^{k} l_{k,i}^{(\beta-1)} z^i,$$

Equation (A.2) can be expressed by

$$\hat{I}_\alpha = \sqrt{\frac{\mu}{\beta}}\frac{1}{\Gamma(\beta)}\sum_{k=0}^{n}\sum_{i=0}^{k} C_k l_{k,i}^{(\beta-1)}\Gamma\left(\beta+i+\frac{1}{2}, z_\alpha\right) - M\alpha,$$

where incomplete gamma functions for $a \geq 0$ and $b \geq 0$ are defined by

$$\Gamma(a,b) = \int_0^b u^{a-1}\mathrm{e}^{-u}\mathrm{d}u.$$

A.5 Effect of shift parameter on Laguerre expansion with squaring

Here we investigate the effect of the shift parameter M on the Laguerre expansion with squaring. Figure A.1 shows the RMSE defined by Equation (3.5) as a measure of difference plotted against the shift parameter M introduced in Section 3.1.

From these plots, we can see that each curve has some oscillation near the origin and then converges to a value as M increases.

We might use the M which minimises the RMSE found within the oscillation near the origin for our Laguerre expansion with squaring. However, such optimisation may have the following problems. Firstly, such optimisation is only possible when an empirical distribution is available, while we also deal with the cases where empirical distributions are not available. Secondly, the optimisation has to be done numerically and therefore may be unfavourable in terms of computational load.

Instead, we find that the RMSE is almost constant for a large M, and its level is not largely different from the minimum. Therefore, it can be considered to be more practically convenient to use some large enough value which is within the range that the RMSE is almost constant for M in expansions than to search for an optimal M. In fact, we use five times the standard deviation as the value of M in our numerical examples, such as in Section 3.2.1, which show such choice of M can provide approximation quality comparable with or better than Hermite expansions and optimised Hermite expansions.

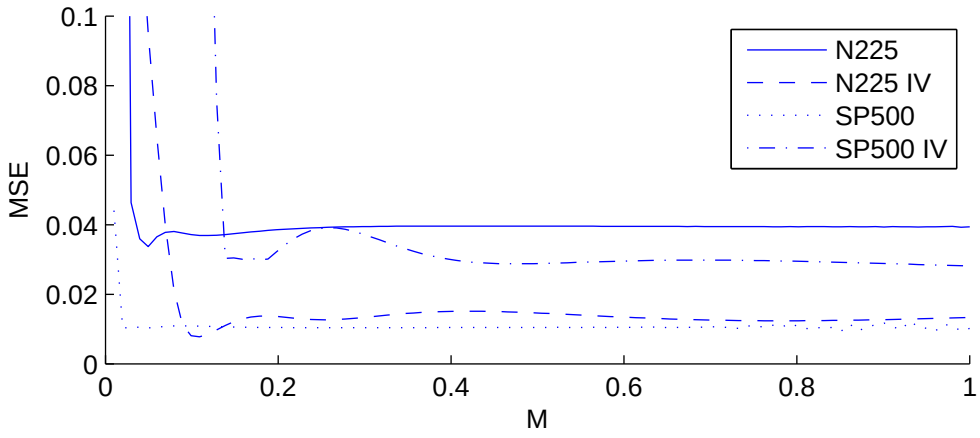

Figure A.1: RMSE defined by Equation (3.5) for Laguerre expansions with squaring, plotted against the shift parameter M (see Section 3.1) applied to the distributions of four risk factors (see Section 3.2.1): $n = 8$ in Equation (2.2).

A.6 Applications to a non-central t distribution

In Sections 3.2.2 and 4.2.2, we dealt with distributions for which higher order moments do not exist. In these examples, our approach was to include existing moments only into the expansions.

Here, we consider Monte Carlo experiments based on a different approach. Assume that a random variable X does not have moments of order higher than n_0. If the distribution function is known, we can generate a finite set of pseudo-independent random samples $\{X^{(i)}\}$, $i = 1, \ldots, N$ from X. Then, the sample moments $N^{-1} \sum_{i=1}^{N} (X^{(i)})^k$ exist even for $k > n_0$. Therefore, we can apply expansion methods to this empirical distribution. Such empirical moments converge to some finite value for $k \leq n_0$ as $N \to \infty$, however, they do not for $k > n_0$.

A non-central t distribution is one distribution for which moments of some high order do not exist. When Y has the Normal distribution with mean δ and variance 1 and V has the χ^2 distribution with degrees of freedom p, and Y and V are independent, the ratio of two random variables $X = Y/\sqrt{V/p}$ is defined

to have a non-central t distribution with degrees of freedom p and non-centrality parameter δ. It is known that moments exist only up to order $p - 1$[2].

Table A.1 shows summary statistics of the non-central t distribution with $p = 4$ and $\delta = 1$, for which moments only up to third order exist. Columns (a) and (b) show sample statistics of two sets of 500 pseudo random samples[3]. We can see that the sample kurtosis values of the two sets, for which the theoretical value does not exist, are very different from each other. It is notable that the sample skewness values, for which the theoretical value does exist, are also quite different from each other.

As such, we can see that the estimates of the moments can depend heavily on which sample set we use. In fact, by comparing expansions applied to sample set (a) in Figure A.2 with those applied to sample set (b) in Figure A.3, we can see that the approximation quality can be very different according to the choice of the sample set. Figure A.2 shows that, except for the Hermite expansion, the approximations by expansions are successful. However, in Figure A.3, all the expansions perform poorly, even with the techniques introduced in Section 3.1. Note that increasing the number of observations N may not improve the situation, since the sample moments of order higher than three do not converge as $N \to \infty$, as discussed above.

Analytically, a sufficient condition for convergence of a Laguerre expansion is given as $\int_0^\infty f(u)u^{-(\beta-1)/2}\mathrm{e}^{u/2}\mathrm{d}u < \infty$ (Freedman 1981). However, it can be easily shown, using the fact that non-central t distributions do not have a moment generating function, that non-central distributions do not satisfy this condition. Hence, expansions may not be valid.

We might conclude that expansions which use pseudo-sample moments when the theoretical moments do not exist are not very reliable. They may work fairly

[2]See standard text books in statistics including Minotani (1998).

[3]Here we use the Matlab command "nctrnd" to generate the pseudo-random samples.

well as in Figure A.2; however, whether it works or not largely depends on the choice of the sample set.

	theoretical	(a)	(b)
Number of observations	—	500	500
Mean	1.253	1.247	1.184
Std. dev.	1.559	1.460	1.684
Skewness	2.364	1.083	2.361
Kurtosis	—	6.273	10.468
Min	—	−4.040	−2.761
Max	—	9.000	12.838

Table A.1: Summary statistics of a non-central t distribution (with degrees of freedom 4 and non-centrality parameter 1) and of two sets (a) and (b) of 500 pseudo random samples.

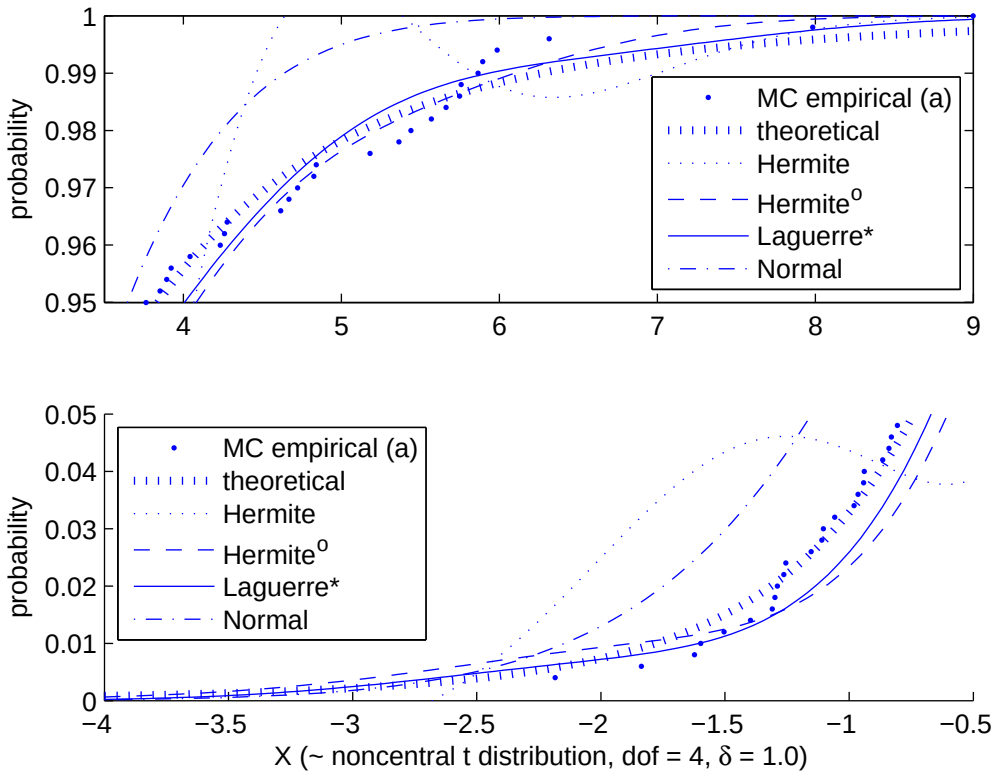

Figure A.2: Three expansions and the Normal approximation applied to sample set (a): $n = 8$ in Equation (2.2). See Table A.1 for sample set (a). The empirical distribution and theoretical value of the distribution function are also shown. Upper plots show the right tails and lower plots show the left tails.

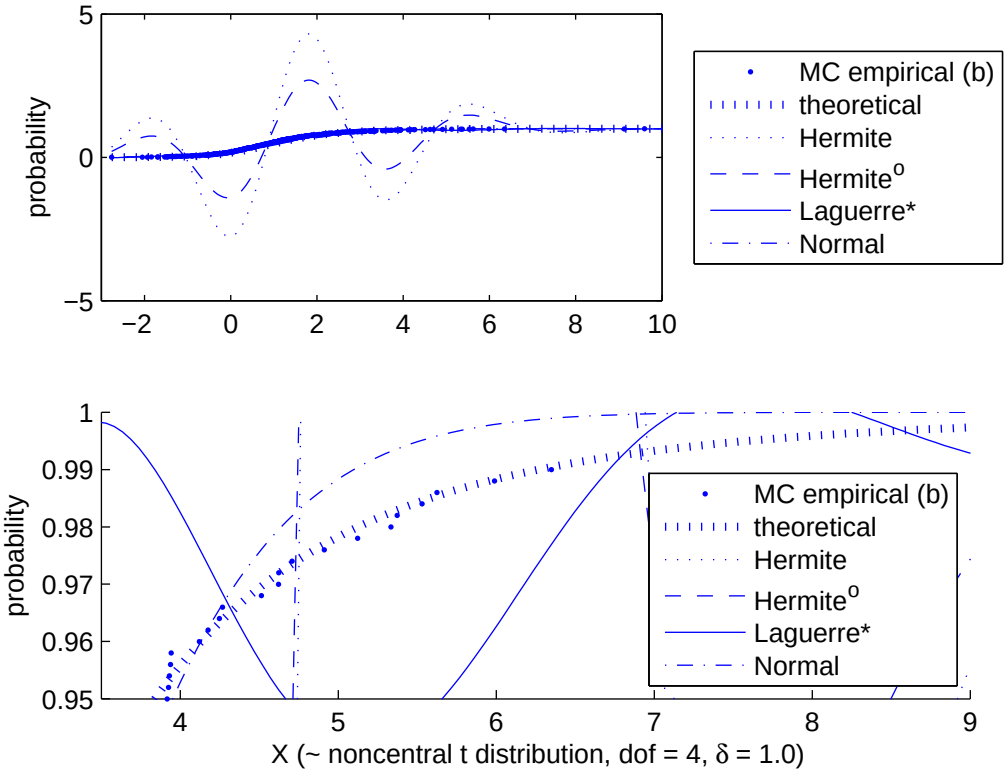

Figure A.3: Three expansions and the Normal approximation applied to sample set (b): $n = 8$ in Equation (2.2). See Table A.1 for sample set (b). The empirical distribution and theoretical value of the distribution function are also shown. Upper plots show the overall shape of the approximations and lower plots magnify the right tails.

Bibliography

Acerbi, C. & Tasche, D. (2002), 'On the coherence of expected shortfall', *Journal of Banking and Finance* **26**.

Aït-Sahalia, Y. (2002), 'Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach', *Econometrica* **70**(1), 223–262.

Ari Hauksson, A., Dacorogna, M., Domenig, T., Müller, U. & Samorodnitsky, G. (2001), 'Multivariate extremes, aggregation and risk estimation', *Quantitative finance* **1**, 79–95.

Artzner, P., Delbaen, F., Eber, J. M. & Heath, D. (1999), 'Coherent measures of risk', *Mathematical Finance* **9**(3), 203–228.

Barakat, R. (1996), 'Second-order statistics of integrated intensities and detected photons, the exact analysis', *Journal of Modern Optics* **43**(6), 1237–1252.

Barakat, R. (1999), 'First-order intensity and log-intensity probability density functions of light scattered by the turbulent atmosphere in terms of lower-order moments', *Journal of Optical Society of America* **16**(9), 2269–2274.

Basel Committee (1996), 'Amendment to the capital accord to incorporate market risks'.

Black, F. & Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**, 637–659.

Buckland, S. T. (1992), 'Fitting density functions with polynomials', *Applied Statistics* **41**(1), 63–76.

Capéraà, P., Fougères, A. L. & Genest, C. (1997), 'A nonparametric estimation procedure for bivariate extreme value copulas', *Biometrika* **84**(3), 567–577.

Carnero, M. A., Peña, D. & Ruiz, E. (2004), 'Persistence and kurtosis in GARCH and stochastic volatility models', *Journal of Financial Econometrics* **2**(2), 319–342.

Cont, R. & Fonseca, J. (2002), 'Dynamics of implied volatility surfaces', *Journal of Quantitative Finance* **2**, 45–60.

Daníelsson, J. (1997), 'Value-at-Risk and extreme returns', *LSE Financial Markets Group Discussion Paper* **273**.

Daníelsson, J. (2002), 'The emperor has no clothes: Limits to risk modelling', *Journal of Banking and Finance* **26**, 1273–1296.

Duffie, D. & Pan, J. (1997), 'An overview of value at risk', *Journal of Derivatives* **Spring**, 7–48.

Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997), *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag.

Engle, R. (1982), 'Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation', *Econometrica* **50**, 987–1008.

Fabozzi, F. J. (2007), *Bond Markets, Analysis and Strategies*, 6 edn, Prentice Hall.

Feuerverger, A. & Wong, A. C. M. (2000), 'Computation of value-at-risk for nonlinear portfolios', *Journal of Risk* **3**(1), 37–55.

Finger, C. C. (2006), 'How historical simulation made me lazy', *Risk Metrics Group Research Monthly* .

Freedman, R. S. (1981), 'On Gram-Charlier approximations', *IEEE Transactions on Communications* **COM-29**(2), 122–125.

Freud, G. (1971), *Orhtogonal Polynomials*, Pergamon Press.

Gaztañaga, E., Fosalba, P. & Elizalde, E. (2000), 'Gravitational evolution of the large-scale probability density distribution: The Edgeworth and Gamma expansions', *The Astrophysical Journal* **539**(2), 522–531.

Gordy, M. B. (2002), 'Saddlepoint approximation of CreditRisk+', *Journal of Banking & Finance* **26**, 1335–1353.

Gouriéroux, C. (1997), *ARCH models and Financial Applications*, Springer.

Group of Thirty (1993), 'Derivatives: Practices and principles'.

Hall, P. (1983*a*), 'Large sample optimality of least squares cross-validation in density estimation', *The Annals of Statistics* **11**(4), 1156–1174.

Hall, P. (1983*b*), 'Orthogonal series methods for both qualitative and quantitative data', *The Annals of Statistics* **11**(3), 1004–1007.

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, Springer.

Hall, P. & Neumeyer, N. (2006), 'Estimating a bivariate density when there are extra data on one or both components', *Biometrika* **93**(2), 439–450.

Heston, S. & Nandi, S. (2000), 'A closed form GARCH option pricing model', *The Review of Financial Studies* **13**, 585–625.

Jackson, D. (1963), *Fourier Series and Orthogonal Polynomials*, Vol. 6, The Mathematical Association of America.

Jaschke, S. R. (2002), 'The Cornish-Fisher-expansion in the context of delta-gamma-normal approximations', *Journal of Risk* **4**(4), 33–52.

Jensen, J. L. (1995), *Saddlepoint Approximations*, Clarendon Press, Oxford.

Jorion, P. (2001), *Value at Risk, Second Edition*, McGraw-Hill.

Junker, M. & May, A. (2005), 'Measurement of aggregate risk with copulas', *Econometric journal* **8**, 428–454.

Kim, S., Shephard, N. & Chib, S. (1998), 'Stochastic volatility: Likelihood inference and comparison with arch models', *Review of Economic Studies* **65**, 361–393.

Kolassa, J. E. (1997), *Series Approximation Methods in Statistics, 2nd edition*, Springer.

Lillo, F. & Mantegna, R. N. (2002), Dynamics of a financial market index after a crash, *in* 'Proceedings of the 8th conference "Economics and Finance"'.

Linsmeier, T. J. & Pearson, N. D. (1996), 'Risk measurment: An introduction to Value at Risk', Office for Futures and Options Research Working Paper.

Longin, F. M. (2000), 'From value at risk to stress testing: The extreme value approach', *Journal of Banking & Finance* **24**, 1097–1130.

Malz, A. M. (2000), Vaga risk and the smile, Technical Report 99-06, The RiskMetrics Group.

Markowitz, H. (1959), *Portfolio Selection: Efficient Diversification of Investments*, John Wiley.

Marumo, K. & Wolff, R. (2007), 'Expansion methods applied to asset return distributions', *Journal of Risk* **10**.

Massey, F. (1951), 'The Kolmogorov-Smirnov test for goodness of fit', *Journal of the American Statistical Association* **46**(253), 68–78.

Mauleón, I. & Perote, J. (2000*a*), 'The ability of multivariate Edgeworth-Sargan density capturing financial data behaviour', SSRN.

Mauleón, I. & Perote, J. (2000*b*), 'Testing densities with financial data: an empirical comparison of the Edgeworth-Sargan density to the student's t', *The European Journal of Finance* **6**, 225–239.

Mina, J. & Xiao, J. Y. (2001), Return to RiskMetrics: The evolution of a standard, Technical report, RiskMetrics.

Minotani, C. (1998), *Sugu-ni yakudatsu toukei bunpu (In Japanese)*, Tokyo-Tosho.

Natale, F. P. (2006), 'Optimization with tail-dependence and tail risk: A copula based approach for strategic asset allocation', *SSRN Working Paper* .

Nelsen, R. B. (1999), *An Introduction to Copulas*, Springer.

Perote, J. & Del Brío, E. (2001), 'Measuring value at risk of portfolios under the Edgeworth-Sargan distribution', Working Paper, Universidad de Salamanca.

Pickands III, J. (1975), 'Statistical inference using extreme order statistics', *The Annals of Statistics* **3**(1), 119–131.

Rachev, S. T., ed. (2003), *Handbook of heavy tailed distributions in finance*, Elsevier.

Resnick, S. I. (1987), *Extreme Values, Regular Variation, and Point Processes*, Springer-Verlag.

RiskMetrics (1994), *RiskMetrics Technical Document*, 1 edn.

RiskMetrics (1996), *RiskMetrics Technical Document*, 4 edn.

Rockafellar, R. T. & Uryasev, S. (2000), 'Optimisation of conditional value-at-risk', *Journal of Risk* **2**(3), 21–41.

Rogachev, A. (2007), 'Value-at-risk concept by Swiss private banks', *The Journal of Risk Finance* **8**(1), 72–78.

Schachter, B. & Butler, J. S. (1996), 'Improving value-at-risk estimates by combining kernel estimation with historical simulation', Working paper series, SSRN.

Szegö, G. (1975), *Orthogonal Polynomials*, 4 edn, American Mathematical Society.

Szegö, G. (2002), 'Measures of risk', *Journal of Banking and Finance* **26**(7), 1253–1272.

Vlaar, P. J. G. (2000), 'Value at risk models for Dutch bond portfolios', *Journal of Banking and Finance* **24**, 1131–1154.

Weron, R. (2004), *Handbook of Computational Statistics*, Springer, chapter IV.1, pp. 911–950.

Williams, D. (1991), *Probability with martingale*, Cambridge University Press.