No. 4

Kohei MARUMO* and Rodney WOLFF†

A Non-parametric Method for Approximating Joint Densities and Copula Functions for Financial Markets

February 2013

*marumo@mail.saitama-u.ac.jp. Faculty of Economics, Saitama University Japan

†rodney.wolff@uq.edu.au. WH Bryan Mining and Geology Research Centre, The University of Queensland Australia

Abstract

When we measure the market risk of a portfolio with multiple of risk factors, we, sometimes implicitly, deal with the risk factors' joint distribution. However, only a few methods are available to render tractable forms of multivariate distributions for risk aggregation.

This paper discusses approximation techniques using the Hermite expansion for marginal and joint density functions. These techniques (or expansion methods) approximate probability density functions by a sum of Hermite polynomials multiplied by the associated weight function. The advantage of the use of expansion methods is that they only require the moments of the target distributions up to some finite degree, assuming they exist.

The biggest shortcoming of the expansion methods is their poor approximation quality. This paper introduces techniques to redeem this problem, and considers application to risk aggregation. We also approximate joint density functions and show that expansion methods are applicable to approximating conditional expectations and copula density functions. Numerical examples for bivariate cases show that our approximations can capture characteristics of original observations. Such techniques may facilitate the further investigation of non-linear dependence structures among risk factors in the financial markets.

Keywords: Conditional expectations; Copulas; Hermite polynomials; Orthogonal expansion; Smoothing methods.

1 Introduction

Handling the joint distributions of risk factors is a key step in market risk measurement. Our portfolios are normally exposed to multiple of risk factors, and it is a common process to aggregate these risks and obtain the distribution of the profit or loss (PL) of the portfolio. This is because the PL distributions are often considered to be vital for market risk measurement. For instance, the two most commonly used risk measures, Value at Risk (VaR) and Expected Shortfall (ES) are defined as a quantile of the PL distribution, and the expected loss

given that the loss exceeds the VaR, respectively[1]. In reality, however, only a few methods are available to render tractable forms of multivariate distributions for risk aggregation. Use of the multivariate Normal distribution is one of the most common methods (Jorion, 2007; Malz, 2000). Among the many advantages of this method is that it only requires the mean vector and the covariance matrix of the risk factors. Further, the portfolio value is expressed as a Normal variable, as long as the PL of the portfolio is associated with the risk factors in a linear form. The shortcomings corresponding to these advantages are the ignorance of heavy tails and non-linear dependence structures, which are often observed in the markets, and the difficulty in dealing with non-linear risks.

As an alternative, the historical simulation (HS) method implicitly uses the empirical distribution of risk factors (Jorion, 2007). The empirical distribution is capable of capturing the characteristics of market observations such as heavy tails or non-linear dependence structures among risk factors. Furthermore, the measurement of non-linear risks is much easier with the HS method than with the Normal-based methods; we can easily obtain the 'virtual' empirical distribution of the PL from that of risk factors. However, discreteness, especially in the tails where the observations are sparse, is an unfavourable feature of the HS method. Also, the HS method does not render joint density functions, which give comprehensive visual images of dependence structures[2].

A less common method, which is becoming increasingly popular, is to use a copula to describe the dependence structure among risk factors (Junker and May, 2005; Natale, 2006). Although non-parametric copula estimation has been proposed[3], most copula-based methods require specification of the parametric copula function *a priori*.

Besides these methods, we discuss practical use of the Hermite expansion.

[1]See, for example, Jorion (2007), Mina and Xiao (2001) or Duffie and Pan (1997), for the calculation of VaR.

[2]It is also pointed out that the HS method does not have a solution to the *time aggregation problem*, which is concerned with measuring risks with time horizons longer than one day. See, for instance, Marumo and Wolff (2007).

[3]See Capéraà et al. (1997) for example.

The Hermite expansion expresses probability density functions (pdfs), by an infinite sum of Hermite polynomials multiplied by an associated weight function, assuming the infinite sum converges[4]. We consider approximating pdfs by a finite sum. We call such techniques expansion methods. One of the advantages of expansion methods over other methods to approximate distributions, such as those based on extreme value theory, is that the expansion methods only require the moments (and cross-moments for the multivariate case) of the target distribution up to some finite degree. Therefore, they have a potential to be applied to a wide range of situations, including market risk measurement.

One of the biggest drawbacks of expansion methods, however, is their poor approximation quality. Due to the polynomials involved in the approximations, they often exhibit negative density. Further, the infinite sum can be divergent, or the convergence is often very slow, and therefore the approximation quality can be sensitive to the choice of the order of expansion. Because of such difficulties, application of expansion methods has been rather limited. For instance, Marumo and Wolff (2007) apply univariate expansion methods to approximating the pdfs of asset return distributions. Mauleón and Perote (2000) consider a second-order multivariate approximation.

We introduce techniques to improve the approximation quality of expansion methods. Numerical examples show that these techniques can be effective, even with adverse cases.

We also discuss the application to market risk measurement. We show that expansion methods can redeem the shortcomings of the Normal-based methods.

Further, we extend the expansion methods more generally to multivariate distributions. We also show that, using our expansion methods, conditional moments and copula density functions can also be approximated. These also provide a comprehensive view on dependence structures among risk factors.

[4]See Szegö (1975), Jackson (1963), Freud (1971) or Takahashi (2006) for theoretical background of the Hermite expansion.

The structure of this paper is as follows. In Section 2, we review the Hermite expansion for univariate cases and introduce techniques to improve the approximation quality. Section 3 discusses the application to risk measurement. In Section 4, we extend expansion techniques to multivariate density functions and discuss further applications for dealing with conditional moments and copula density functions. Numerical examples are also shown. Section 5 contains the closing discussion.

2 Hermite polynomials and expansion methods for univariate distributions

In this section, we review Hermite polynomials and expansion methods for univariate cases, and introduce techniques to improve the approximation quality.

2.1 Hermite polynomials

The series of Hermite polynomials $\{\mathrm{He}_k(x)\}$ are defined as[5]

$$\mathrm{He}_k(x) = \frac{1}{\sqrt{k!}} \mathrm{e}^{\frac{x^2}{2}} \frac{\mathrm{d}^k}{\mathrm{d}x^k} \mathrm{e}^{-\frac{x^2}{2}}, \ k = 0, 1, 2, \ldots. \tag{1}$$

The most important property of this series is that it satisfies

$$\int_{-\infty}^{\infty} \phi(x) \mathrm{He}_k(x) \mathrm{He}_l(x) \mathrm{d}x = \begin{cases} 0 & (k \neq l) \\ 1 & (k = l) \end{cases}, \tag{2}$$

where $\phi(x) = \mathrm{e}^{-x^2/2}/\sqrt{2\pi}$, the pdf of the standard Normal distribution, is the associated weight function. This property is called orthonormality with respect to ϕ.

2.2 Hermite expansion and expansion methods

Consider a pdf f. Assume that f satisfies

$$\int_{-\infty}^{\infty} \frac{\left\{ f(u) - \phi(u) \sum_{k=0}^{\infty} C_k \mathrm{He}_k(u) \right\}^2}{\phi(u)} \mathrm{d}u = 0 \tag{3}$$

[5]There are some variations for the definition of the Hermite polynomials, including $\mathrm{He}_k(x) = \mathrm{e}^{x^2} \frac{\mathrm{d}^k}{\mathrm{d}x^k} \mathrm{e}^{-x^2}$, which are orthogonal but may not be necessarily orthonormal.

for some real coefficients $C_0, C_1, \ldots$. Condition (3) is satisfied iff f is bounded and

$$f(x) = \phi(x) \sum_{k=0}^{\infty} C_k \mathrm{He}_k(x), \tag{4}$$

for almost everywhere x. For such f that satisfies Equation (3), let us call the expression in Equation (4) the Hermite expansion of f.

The infinite sum on the right hand side of Equation (4) is known to be convergent if[6]

$$\int_{-\infty}^{\infty} \frac{\{f(u)\}^2}{\phi(u)} \mathrm{d}u < \infty, \tag{5}$$

that is, if $f(x)/\sqrt{\phi(x)}$ is square integrable. Roughly speaking, such f has to be bounded and have 'thin' tails. Using Equations (2), (3) and (4), it can be shown that, when the Condition (5) is satisfied, we have

$$\int_{-\infty}^{\infty} \frac{\{f(u)\}^2}{\phi(u)} \mathrm{d}u = \sum_{k=0}^{\infty} C_k^2 < \infty, \tag{6}$$

which corresponds to the Perceval identity for the Hermite system.

For f's whose Hermite expansion is convergent, the coefficients $\{C_k\}$ can be determined, using Equation (2), as:

$$C_k = \int_{-\infty}^{\infty} \mathrm{He}_k(u) f(u) \mathrm{d}u, \ k = 0, 1, \ldots. \tag{7}$$

Note that $\mathrm{He}_0(x) \equiv 1$ and therefore $C_0 = 1$ is required so that $\int_{-\infty}^{\infty} f(u) \mathrm{d}u = 1$ is satisfied.

Let X be a random variable with pdf f. Then, from Equation (7) we have $C_k = \mathrm{E}(\mathrm{He}_k(X))$, which is a linear combination of moments of X up to kth order. This implies that, given the moments $\mathrm{E}(X), \ldots, \mathrm{E}(X^n)$, Equation (4) can be approximated by a sum up to finite n:

$$f(x) \simeq f_n(x) = \phi(x) \sum_{k=0}^{n} \mathrm{E}(\mathrm{He}_k(X)) \mathrm{He}_k(x). \tag{8}$$

Let us call such approximation methods based on Equation (8) the expansion methods[7].

[6] See textbooks for Fourier Analysis such as Takahashi (2006).

[7] Besides using the Hermite expansion, the Laguerre expansion is sometimes applied to densities with non-negative support. See Marumo and Wolff (2007) or Marumo (2007).

2.3 Difficulties in applying expansion methods

One of the biggest drawbacks of the expansion methods is their poor approximation quality. It is pointed out by Gordy (2002) that naive use of the Hermite expansion — plugging the moments of a risk factor into Equation (8) directly — can result in a very poor approximation.

In fact, the expansion methods allow the approximations to take negative values, and such negative density is often too large to ignore, especially in the tails, where the pdf is close to 0.

For the cases where the Hermite expansion is convergent, and where the convergence is reasonably fast, we might deal with this problem by increasing the order of expansion n in Equation (8). The Hermite expansion, however, is often divergent or, its convergence can be very slow. In such cases, the approximation quality is sensitive to the order of the expansion. Marumo and Wolff (2007) and Jaschke (2002) studies the relationship between the order and the approximation quality; however, general criteria for determining an optimal order of expansions have not been proposed as far as we can determine.

In the following Subsections, we firstly take a close look at these difficulties using an example where we approximate the empirical distribution, and then introduce techniques to deal with them.

2.4 Naive application to approximating empirical distributions

Assume that we have i.i.d. samples $X^{(1)}, \ldots, X^{(N)}$ from a distribution with an unknown pdf f. A naive idea would be to use the sample moments to obtain the coefficients for the expansion. That is, we use the coefficients

$$\hat{C}_k = \frac{1}{N} \sum_{i=1}^{N} \mathrm{He}_k(X^{(i)}),\ k = 0, \ldots, n$$

to obtain the estimator of f,

$$\hat{f}_n(x) = \phi(x) \sum_{k=0}^{n} \hat{C}_k \mathrm{He}_k(x). \tag{9}$$

This approximation often results in a poor quality, as mentioned above. As shown later, using the sample moments implicitly assumes that our target distribution is the empirical distribution, whose pdf might be expressed using the Dirac delta function δ as

$$f^E(x) = \frac{1}{N}\sum_{i=1}^{N}\delta(x - X^{(i)}).$$

Since $f^E(x)/\sqrt{\phi(x)}$ is not square integrable, the expansion in Equation (9) is divergent as $n \to \infty$. This means that $\sum_{k=0}^{\infty}\hat{C}_k^2$ is also divergent.

To get a clearer view on this point, we consider the distribution function (df), and verify that it has a converging Hermite expansion. Using Equations (1) and (9), we have

$$\hat{F}_n(x) = \int_{-\infty}^{x}\hat{f}_n(u)\mathrm{d}u = \Phi(x) + \phi(x)\sum_{k=1}^{n}\frac{\hat{C}_k}{\sqrt{k}}\mathrm{He}_{k-1}(x), \tag{10}$$

where Φ is the df of the standard Normal distribution[8]. This corresponds to the naive approximation for the empirical df.

For the empirical distribution, we have

$$F^E(x) = \int_{-\infty}^{x} f^E(u)\mathrm{d}u = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{\{X^{(i)}\le x\}},$$

where $\mathbf{1}_{\{\}}$ denotes the indicator function. It is easily shown[9] that

$$I_0 = \int_{-\infty}^{\infty}\{F^E(u) - \Phi(u)\}^2/\phi(u)\mathrm{d}u < \infty. \tag{11}$$

This implies that $F^E(x) - \Phi(x)$ has a converging Hermite expansion of the form:

$$F^E(x) - \Phi(x) = \phi(x)\sum_{k=1}^{\infty}D_k\mathrm{He}_{k-1}(x), \tag{12}$$

[8]Note that $C_0 = 1$ is used in Equation (10).

[9]Split the integral into three parts; $(\int_{-\infty}^{-a} + \int_{-a}^{a} + \int_{a}^{\infty})\{F^E(u) - \Phi(u)\}^2/\phi(u)\mathrm{d}u$ for some large a. It is trivial to show that the second integral is bounded. For large enough a, we have $F^E(x) = 0$ for $x \le -a$. Since $\Phi(x) - \phi(x)$ is decreasing for $x \le -1$ with $\lim_{x\to-\infty}(\Phi(x) - \phi(x)) = 0$, we have $\Phi(x) < \phi(x)$ for $x \le -1$. Therefore, $\int_{-\infty}^{-a}\{0 - \Phi(u)\}^2/\phi(u)\mathrm{d}u < \int_{-\infty}^{-a}\phi(u)\mathrm{d}u = \Phi(-a) < \infty$ for large a. This proves that the first integral is bounded. By symmetry, the third integral can be shown to be bounded.

where $\{D_k\}$ are real coefficients. We show that these coefficients are given by $D_k = \hat{C}_k/\sqrt{k}$ as follows:

Consider the weighted integrated square difference (WISD) between left and right hand sides of Equation (12),

$$\begin{aligned} I_1 &= \int_{-\infty}^{\infty} \frac{\left\{\phi(u)\sum_{k=1}^{\infty} D_k \mathrm{He}_{k-1}(u) - F^E(u) + \Phi(u)\right\}^2}{\phi(u)} \mathrm{d}u \\ &= \sum_{k=1}^{\infty} D_k^2 - \frac{2}{N}\sum_{k=1}^{\infty} D_k \sum_{i=1}^{N} \int_{-\infty}^{\infty} \mathrm{He}_{k-1}(u)(\mathbf{1}_{\{X^{(i)} \le u\}} - \Phi(u))\mathrm{d}u + I_0 \\ &= \sum_{k=1}^{\infty} D_k^2 - 2\sum_{k=1}^{\infty} \frac{D_k \hat{C}_k}{\sqrt{k}} + I_0, \end{aligned} \tag{13}$$

where we use the identity $\int \mathrm{He}_{k-1}(u)\mathrm{d}u = -\mathrm{He}_k(u)/\sqrt{k} + \text{constant}$, to obtain the third equality. We have the following three facts:

- From Equation (12), $\{D_k\}$ are the coefficients that let $I_1 = 0$, and existence of such $\{D_k\}$ is suggested by Condition (11),
- Since I_1 is a integrated square, $I_1 \ge 0$, and
- From Equation (13), I_1 is minimised when $D_k = \hat{C}_k/\sqrt{k}$ for $k = 1, 2, \ldots$.

These verify

$$F^E(x) = \Phi(x) + \phi(x)\sum_{k=1}^{\infty} \frac{\hat{C}_k}{\sqrt{k}} \mathrm{He}_{k-1}(x) = \lim_{n\to\infty} \hat{F}_n(x). \tag{14}$$

Therefore, the naive approximation for the df in Equation (10) is convergent, and we have

$$I_0 = \int_{-\infty}^{\infty} \left\{F^E(u) - \Phi(u)\right\}^2 / \phi(u)\mathrm{d}u = \sum_{k=1}^{\infty} \frac{\hat{C}_k^2}{k} < \infty,$$

whereas $\sum_{k=0}^{\infty} \hat{C}_k^2$ is divergent, as shown previously.

2.5 Techniques to improve the approximation quality

In the previous Subsection, we review naive approximations for empirical distributions and their convergence properties; the approximation for dfs is convergent, while the approximation for pdfs is divergent. As we will see in the

numerical examples, the convergence of the approximation for dfs, however, is very slow and it may not be suitable for use in practice[10].

The most popular way to redeem such fragility of the Hermite expansion is to use the saddlepoint approximation, as suggested by Gordy (2002). This is available for the cases where the cumulant generating function of the target distribution is of a tractable form. This condition can significantly narrow the scope of applicable cases.

Instead of using saddlepoint approximations, we introduce three techniques, standardisation, smoothing, and optimisation, which, in combination with expansion methods, result in a better practical approximation quality in a wider range of examples than investigated to date. Smoothing and optimisation have not been considered in the literature as far as we can determine[11]. Standardisation is a very common technique.

The advantage of our techniques over the saddlepoint approximations is that they only require the moments of the target distribution up to some finite degree.

Standardisation

We can view Equation (8) as approximating the target pdf by the pdf of the standard Normal distribution along with correction terms involving polynomials. If the target pdf is close to that of the standard Normal, then the correction can be small and therefore we expect that the expansion method provides a good approximation.

One obvious way to make the target pdf closer to that of standard Normal is to standardise the variable so that the first and second moments of the target pdf are equal to 0 and 1. Actually, this technique is used, sometimes implicitly, in most existing applications.

Let $\mu = \mathrm{E}(X)$, $\sigma^2 = \mathrm{E}(X^2) - \mu^2$, $X' = (X - \mu)/\sigma$ and $f^{X'}$ be the pdf of

[10] An intuitive account for this slow convergence can be the large order of the polynomial required to approximate functions with discontinuous points.

[11] A similar technique to optimisation for approximating a discrete distribution is found in Hall (1983). He considers a case with a different expansion formula from the present paper.

X'. We apply the Hermite expansion in Equation (8) to X' to obtain its approximation $f_n^{X'}$. The pdf of X can be approximated by $f_n(x) = f_n^{X'}\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}$.

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

We apply standardisation to all of the examples in this paper, and hereafter we call an expansion to which only standardisation is applied 'a naive expansion.'

Smoothing

As discussed above, a poor approximation quality can be due to the divergence or slow convergence of the Hermite expansion. Instead of approximating the target distribution itself, we aim a smooth function near the target, for which the convergence of the Hermite expansion is expected to be faster.

One way is to use the curvature of the target function; that is, we can seek for a function which is 'near' the target function and whose second derivative is 'small' in some sense.

As an example, we consider smoothing the approximation for the empirical distribution F^E in Equation (14). Assume that a function F^S is near F^E, in the sense that the WISD,

$$\int_{-\infty}^{\infty} \left\{F^S(u) - F^E(u)\right\}^2 / \phi(u) \mathrm{d}u$$

is small, and has the converging Hermite expansion

$$F^S(x) = \Phi(x) + \phi(x) \sum_{k=1}^{\infty} \frac{C_k}{\sqrt{k}} \mathrm{He}_{k-1}(x),$$

for some real coefficients $\{C_k\}$. Uing Equation (1), the second derivative of F^S is given as

$$F^{S''}(x) = \phi(x) \sum_{k=0}^{\infty} \sqrt{k+1} C_k \mathrm{He}_{k+1}(x).$$

We can use the weighted integrated square curvature (WISC),

$$\int_{-\infty}^{\infty} \left\{F^{S''}(u)\right\}^2 / \phi(u) \mathrm{d}u$$

as a measure of curvature.

In the two integrals, the WISD and WISC, the weight $1/\phi$ works in two ways. One is to put more importance on the difference and curvature in the tail than those of the centre of the distribution. The other is that it makes the integrals tractable. In fact, using the Perceval identity, we have

$$\mathrm{WISD} = \sum_{k=1}^{\infty} \frac{(C_k - \hat{C}_k)^2}{k}$$
$$\mathrm{WISC} = \sum_{k=0}^{\infty} (k+1) C_k^2.$$

Now, let us consider the average of two integrals, weighted by $0 \le q \le 1$,

$$\begin{aligned} I_q^S &= (1-q)\mathrm{WISD} + q\mathrm{WISC} \\ &= (1-q) \sum_{k=1}^{\infty} \frac{(C_k - \hat{C}_k)^2}{k} + q \sum_{k=0}^{\infty} (k+1) C_k^2 \\ &= \sum_{k=1}^{\infty} \left(\frac{1-q+qk(k+1)}{k} C_k^2 - 2\frac{1-q}{k} C_k \hat{C}_k \right) + (1-q) I_0 + q, \end{aligned}$$

where the weight q determines the relative importance of curvature over fidelity. This is minimised when

$$C_k = \hat{C}_k^S = \frac{1-q}{1-q+qk(k+1)} \hat{C}_k.$$

Thus, we might use the finite expansion

$$\hat{F}_n^S(x) = \Phi(x) + \phi(x) \sum_{k=1}^{n} \frac{\hat{C}_k^S}{\sqrt{k}} \mathrm{He}_{k-1}(x),$$

as the smoothed approximation for F^E.

Since $\hat{C}_k^S$ has $k(k+1)$ in the denominator, the convergence of the weighted square integral,

$$\int_{-\infty}^{\infty} \left\{ \hat{F}_n^S(u) - \Phi(u) \right\}^2 / \phi(u) \mathrm{d}u = \sum_{k=1}^{n} \frac{(\hat{C}_k^S)^2}{k}$$

is much faster than that of $\hat{F}_n$; that is, the convergence $\hat{F}_n^S \to F^S$ is much faster than that of $\hat{F}_n \to F^E$. Clearly, F^S is not the target itself, but we expect that it is a smooth function near F^E, especially when q is small. Other than its fast convergence, $\hat{F}_n^S$ has a favourable property. The derivative of $\hat{F}_n^S$ is given by

$$\frac{\mathrm{d}}{\mathrm{d}x} \hat{F}_n^S(x) = \hat{f}_n^S(x) = \phi(x) \sum_{k=0}^{n} \hat{C}_k^S \mathrm{He}_k(x),$$

and we have

$$\int_{-\infty}^{\infty} \frac{\{\hat{f}_n^S(u)\}^2}{\phi(u)} \mathrm{d}u = \sum_{k=0}^{n} (\hat{C}_k^S)^2. \tag{15}$$

Since $\sum_{k=1}^{\infty} \hat{C}_k^2/k$ is convergent, and from the fact that $(\hat{C}_k^S)^2 < \hat{C}_k^2/k$ for large enough k, $\sum_{k=0}^{\infty} (\hat{C}_k^S)^2$ is also convergent. This suggests that the approximation for the pdf is convergent.

In this example, we smooth the df, however, for the cases where the Hermite expansion for the pdf is available, we can smooth the pdf in a similar manner. An example is given later in *optimisation with smoothing*.

Optimisation[12]

Similarly to the previous example, consider a situation where we use i.i.d. samples $X^{(1)}, \ldots, X^{(N)}$ from the pdf f to obtain the natural estimators $\hat{C}_k = \sum_{i=1}^{N} \mathrm{He}_k(X^{(i)})/N$. Also assume that f can be expanded as in Equation (4).

We consider a class of estimator

$$\hat{f}_n^O(x) = \phi(x) \sum_{k=0}^{n} \alpha_k \hat{C}_k \mathrm{He}_k(x), \tag{16}$$

and search for the coefficients α_k, $k = 0, 1, \ldots, n$, so that the estimator of the weighted mean integrated square error (WMISE)

$$\mathrm{E}\left(\int_{-\infty}^{\infty} \frac{\left\{ \hat{f}_n^O(u) - f(u) \right\}^2}{\phi(u)} \mathrm{d}u \right) \tag{17}$$

is minimised[13]. The weight $1/\phi$ in Expression (17) works in two ways, similarly to the previous example; one is to put more importance on the error in the tail than that of the centre of the distribution, and the other is that it makes Expression (17) tractable. It can be shown by using the Perceval identity that the WMISE in Expression (17) is equal to

$$\sum_{k=0}^{n} \alpha_k^2 \mathrm{E}\left(\hat{C}_k^2\right) - 2 \sum_{k=0}^{n} \alpha_k C_k^2 + \sum_{k=0}^{\infty} C_k^2. \tag{18}$$

[12]This method can be regarded as implicitly using the shrinkage estimators.

[13]Hall (1983) considers a different expansion formula, in which the MISE is not weighted.

Now we estimate the WMISE in Expression (18). Obviously, $\hat{C}_k^2$ is an unbiased estimator of $\mathrm{E}\left(\hat{C}_k^2\right)$. An unbiased estimator for C_k^2 is given by

$$\frac{N\hat{C}_k^2 - \hat{B}_k^2}{N-1},$$

where $\hat{B}_k^2 = N^{-1}\sum_{i=1}^{N}\left\{\mathrm{He}_k(X^{(i)})\right\}^2$, however, this estimator can take a negative value, while the true value C_k^2 is non-negative. Therefore, we use a biased (but considered to be better) estimator

$$\frac{\left(N\hat{C}_k^2 - \hat{B}_k^2\right)_+}{N-1},$$

where $(x)_+ = \max\{x, 0\}$, instead.

Hence, an estimator for WMISE is given by

$$\sum_{k=0}^{n}\alpha_k^2\hat{C}_k^2 - 2\sum_{k=0}^{n}\alpha_k\frac{\left(N\hat{C}_k^2 - \hat{B}_k^2\right)_+}{N-1} + \sum_{k=0}^{\infty}\frac{\left(N\hat{C}_k^2 - \hat{B}_k^2\right)_+}{N-1}. \tag{19}$$

Now we consider $\{\alpha_k\}$ which minimises Expression (19). Firstly, $\alpha_0 = 1$ is required so that $\int_{-\infty}^{\infty}\hat{f}_n^O(u)\mathrm{d}u = 1$ is satisfied. If the variable is already standardised so that the first and second moments are identical to those of ϕ, we have $C_1 = C_2 = 0$, and therefore we set $\alpha_1 = \alpha_2 = 0$. For $k = 3, \ldots, n$, Expression (19) is minimised when

$$\alpha_k = \frac{\left(N\hat{C}_k^2 - \hat{B}_k^2\right)_+}{(N-1)\hat{C}_k^2}.$$

Optimisation with smoothing

For the approximation in the form of Equation (16), the second derivative is derived using Equation (1) as

$$\hat{f}_n''(x) = \frac{\mathrm{d}^2}{\mathrm{d}x^2}\hat{f}_n(x) = \phi(x)\sum_{k=0}^{n}\sqrt{(k+1)(k+2)}\alpha_k\hat{C}_k\mathrm{He}_{k+2}(x). \tag{20}$$

We might define a weighted mean integrated square curvature (WMISC) by

$$\mathrm{E}\left(\int_{-\infty}^{\infty}\frac{\left\{\hat{f}_n''(u)\right\}^2}{\phi(u)}\mathrm{d}u\right),$$

which can be simplified to

$$\sum_{i=0}^{n}(k+1)(k+2)\alpha_k^2 \mathrm{E}(\hat{C}_k^2).$$

An unbiased estimator for this quantity is

$$\sum_{i=0}^{n}(k+1)(k+2)\alpha_k^2 \hat{C}_k^2.$$

Then we search for $\{\alpha_k\}$ which minimises

$$\begin{aligned}(1-q)\mathrm{WMISE} + q\mathrm{WMISC} = &\sum_{k=0}^{n}\alpha_k^2\{1-q+q(k+1)(k+2)\}\hat{C}_k^2 \\ &- 2\sum_{k=0}^{n}\alpha_k\frac{\left(N\hat{C}_k^2 - \hat{B}_k^2\right)_+}{N-1} \\ &+ \sum_{k=0}^{\infty}\frac{\left(N\hat{C}_k^2 - \hat{B}_k^2\right)_+}{N-1}, \end{aligned} \tag{21}$$

for some $0 \leq q \leq 1$, where q determines the relative importance we put on smoothness.

From a similar discussion to that in optimisation, we can set $\alpha_0 = 1, \alpha_1 = \alpha_2 = 0$, and

$$\alpha_k = \frac{\left(N\hat{C}_k^2 - \hat{B}_k^2\right)_+}{\{1-q+q(k+1)(k+2)\}(N-1)\hat{C}_k^2},$$

for $k \geq 3$.

Since we have k^2 term in the denominator of this smoothed α_k, the sum

$$\sum_{k=0}^{n}(\alpha_k\hat{C}_k)^2 = \int_{-\infty}^{\infty}\frac{\left\{\hat{f}_n(u)\right\}^2}{\phi(u)}\mathrm{d}u \tag{22}$$

converges much faster than that without smoothing, as $n \to \infty$.

Approximating distributions from parametric models

We often wish to obtain distributions from parametric models. However, this is sometimes a tricky problem. For instance, a situation as common as deriving the distribution of a call option premium using the plain Black and Scholes

formula, and assuming that the risk factors have the log Normal distribution, can be a trouble for which we might turn to Monte Carlo method or the first order approximation[14]. Expansion methods have a potential to be applied to such cases as long as the moments of the target distribution are available.

There can be, however, two types of situations where naive applications of expansion methods do not work;

- the target function is not bounded, and
- the target function has thick tails.

For the first type, we might consider smoothing the target, as considered for empirical distributions.

For the second type, for instance, we might truncate the tails at some large but finite points, so that $f(x)/\sqrt{\phi(x)}$ is square integrable. One obvious drawback of this method is the possible difficulty in deriving the truncated moments.

Developing techniques for applying expansion methods in such situations is our future issue.

Choice of order of expansion n

Suppose that

$$f_n(x) = \phi(x) \sum_{k=0}^{n} C_k \mathrm{He}_k(x)$$

converges to f as $n \to \infty$. Since we have

$$\int_{-\infty}^{\infty} \frac{\{f_n(u) - f_{n-1}(u)\}^2}{\phi(u)} \mathrm{d}x = C_n^2,$$

we can view C_n^2 as 'the amount of change in the shape of the function' when we increase the order from $n-1$ to n. Therefore, we expect that C_n^2 converges to 0 as f_n converges to f.

Based on these facts, an obvious idea would be that we stop the expansion at some n^* for which $\sum_{k=n^*+1}^{\infty} C_k^2$ is 'much smaller' than $\sum_{k=0}^{n^*} C_k^2$. However,

[14] See Numerical examples in Section 2.6.

this idea is not practical since it requires evaluating an infinite sum. As an alternative, we observe the series $C_0^2, \ldots, C_N^2$ for some large N, and see if $C_{n+1}^2, \ldots, C_N^2$ are 'small enough' compared to $C_0^2, \ldots, C_n^2$. We expect such n can be easily found if the convergence is fast.

Choice of smoothness weight q

In the smoothing techniques introduced above, the weight for smoothness q is an arbitrary parameter. Obviously, $q = 0$ is equivalent to not smoothing, and $q = 1$ is equivalent to approximating the target by the Normal distribution. For $0 < q < 1$, the approximation lies in-between these two cases. Considering the fact that q determines the difference between the original target function and the smoothed target, a smaller q might be more desirable, however, a general criterion for choosing an optimum in some sense has not been developed.

As we see in the Numerical examples, $q = 0.005 \sim 0.01$ can provide fair approximation quality.

2.6 Numerical examples

From the discussion so far, it is obvious that approximations using the Hermite expansions are likely to perform well when the target distribution is close to the Normal distribution. Here, we consider three adverse cases and see how expansion methods work. The first and second cases apply the expansion methods to empirical distributions. The third case applies the expansion methods to approximating the Gamma density as an example of application to a parametric model.

Random samples from the Gamma distribution

We take our first example from an empirical distribution generated from the Gamma distribution, which is skewed and has heavy right tail and non-negative support. The shape and scale parameters are set to be 7 and 1, respectively, with a skewness of .756 and an excess kurtosis of .857. We generate a set of 1,000 pseudo-random samples from this Gamma distribution and apply the

expansion methods to its empirical distribution using the techniques described above . The empirical distribution, which is the target of the approximation, is fairly skewed (the sample skewness of .915), and heavily tailed (the sample excess kurtosis of 1.447), compared to the Normal distribution. The order of expansion is $n = 10$, and the weight for smoothness is $q = .01$, where used. Figure 1 shows the shapes of the target pdf and df, and Table 1 compares the risk measures of the target distribution and those of approximations[15].

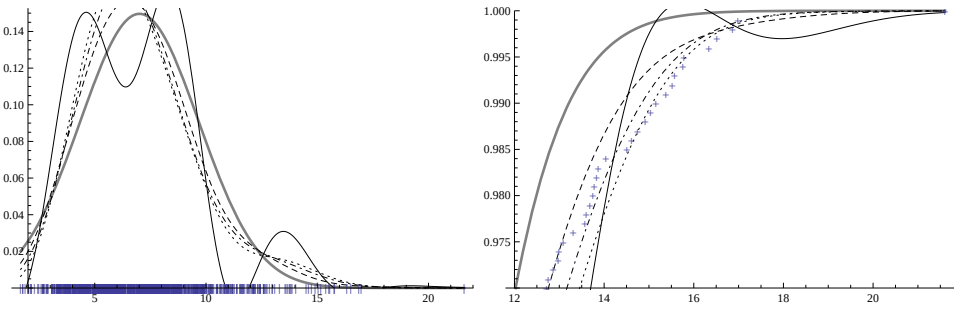

Figure 1: The pdfs (left) and the magnified right tails of dfs (right): the naive expansion (solid curves), the smoothed expansion (dashed curves), the optimised expansion (dotted curves), and the optimised and smoothed expansion (dot-dashed curves), respectively. The order of expansions is $n = 10$ and the weight for smoothness is $q = .01$. The '+' symbols show the 1,000 pseudo-random samples from the gamma distribution with shape parameter 7 and scale parameter 1. The thick gray curves show the pdf and df of the Normal distribution.

We find from Figure 1 that the naive expansion is erratic and shows negative density, while other expansions shows fair approximation quality even in the tail part. This confirms that smoothing and optimising can redeem the fragility of the naive expansion in an adverse example.

Table 1 shows that approximations underestimate the risk in most cases, however, the relative errors are much smaller with expansion methods than with the Normal approximations.

[15]See Section 3 for risk measures.

Level	99%		99.5%	
Measure	VaR	ES	VaR	ES
Empirical	15.16	16.10	15.78	17.35
Normal	-12.84%	-14.43%	-12.04%	-15.14%
Naive	–	–	–	–
Smoothed	-5.03%	-4.29%	-2.82%	-3.33%
Optimised	-0.55%	-2.59%	0.28%	-3.58%
Opt+Smth	-2.13%	-3.73%	-0.94%	-4.50%

Table 1: The risk measures of the empirical distribution and relative errors of the approximations. The order of expansion is $n = 10$ and the weight for smoothness is $q = .01$. The minus signs indicate that the approximations underestimate the risk. Fig 1 shows that the naive expansion exhibits negative density at around $x = 11$ and is not reliable at 99% and 99.5%tile.

Empirical distribution from market observations

From the real world example, we approximate the empirical distribution of 500 daily log-differences of the European call option implied volatility (three-month, at-the-money, hereafter denoted IV) on the S&P 500. It has the sample skewness of $-.251$ and the sample excess kurtosis of 1.782. Other summary statistics are shown in Table 5. We apply the expansion methods to this empirical distribution (Figure 2). Although, we are aware that the risk measures of the IV itself do not so much sense, we calculate them to demonstrate the performance of the approximations (Table 2) The order of expansion is $n = 10$ for all approximations, and the weight for smoothness is $q = .01$.

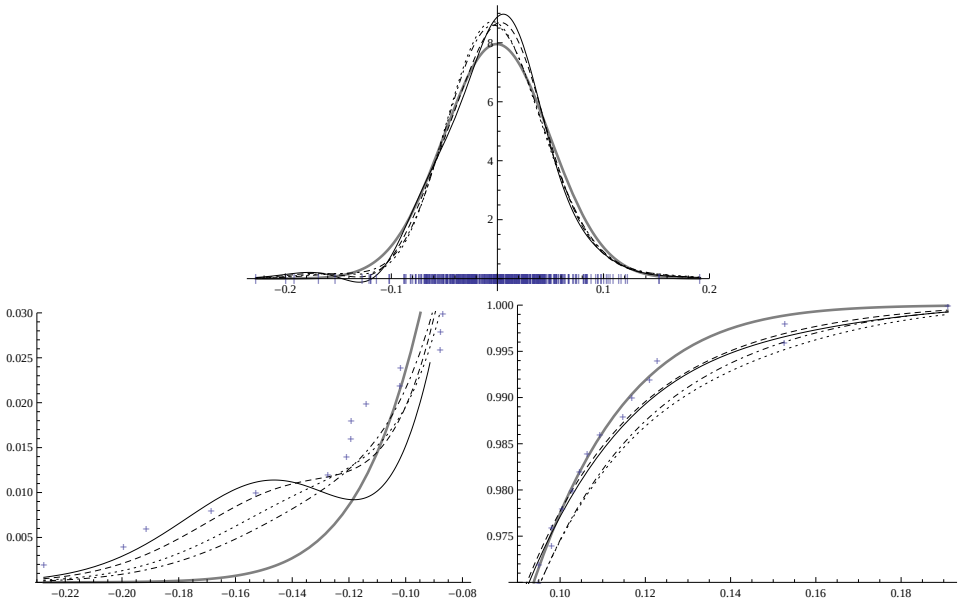

Figure 2: The pdfs (upper) and the magnified left and right tails of dfs (lower): the naive expansion (solid curves), the smoothed expansion (dashed curves), the optimised expansion (dotted curves), and the optimised and smoothed expansion (dot-dashed curves), respectively. The order of expansions is $n = 10$ and the weight for smoothness is $q = .01$. The '+' symbols show the 500 samples of the daily log-difference of S&P 500 IV. The thick gray curves show the pdf and df of the Normal distribution.

This example shows similar features to the previous one; the naive expansion shows negative density in the left tail, while others approximate the

Level	0.5%		1%	
Measure	VaR	ES	VaR	ES
Empirical ($\times -10^{-1}$)	1.914	2.061	1.528	1.879
Normal	-32.29%	-29.42%	-23.32%	-28.64%
Naive	–	–	–	–
Smoothed	-6.39%	-2.78%	-2.13%	-2.77%
Optimised	-12.44%	-7.55%	-10.15%	-8.71%
Opt+Smth	-17.04%	-11.21%	-14.40%	-12.97%

Level	99%		99.5%	
Measure	VaR	ES	VaR	ES
Empirical ($\times 10^{-1}$)	1.169	1.428	1.526	1.654
Normal	-0.64%	-6.81%	-15.69%	-12.67%
Naive	4.20%	4.00%	-7.47%	0.88%
Smoothed	3.63%	2.41%	-8.43%	-1.33%
Optimised	10.19%	10.12%	-0.72%	6.15%
Opt+Smth	8.25%	6.84%	-4.00%	2.62%

Table 2: The risk measures of the empirical distribution and relative errors of the approximations. The risk is measured on the left (Levels 0.5% and 1%) and right (Levels 99% and 99.5%) tails. The order of expansion is $n = 10$ and the weight for smoothness is $q = .01$. The minus signs indicate that the approximations underestimate the risk. Fig 2 shows that the naive expansion exhibits negative density at around $x = -0.14$ and is not reliable at 0.5% and 1%tile.

empirical distribution better than the Normal distribution.

In order to investigate the convergence property, we observe the series of the squared coefficients of the expansions, $\{C_n^2/n\}$ for this example (Figure 3).

The left plots in the Figure 3 shows that the squared coefficients of naive and optimised expansion does not seem to be converging toward 0, at least up

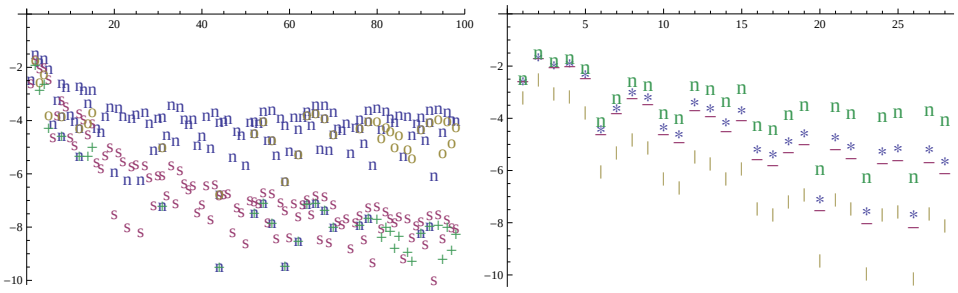

Figure 3: Squared coefficients of the expansions up to order 100 (left) and 30 (right). The vertical axes are in a log scale with base 10. The symbols in the left plots denote naive (n), smoothed (s), optimised (o), and optimised and smoothed (+) expansions, respectively. The weight for smoothness is $q = .01$. The symbols in the right plots denote $q = .1$ (|), .01 ($-$), .005 ($*$), and 0 (n), in the smoothed expansions, respectively.

to order 100, while those of smoothed expansions attenuate rather quickly. The right plots shows that squared coefficients of expansions with larger q converges faster. We might use such plots for determining the order of expansion. For instance, we can stop increasing the order before the squared coefficients are constantly below, say, 10^{-4}. Since the first few large values are $\sim 10^{-2}$, the relative importance of squared coefficients that are below 10^{-4} can be less than 1% of the large values. The order determined in such way is around 10 to 15 for smoothed expansions, however, the required order can be much larger for naive and optimised expansions.

The Gamma pdf

As an example of an application to a parametric model, we consider approximating the pdf of the Gamma distribution. We set the shape parameter 7 and the scale parameter 1, the same as the previous example where we approximate the empirical distribution. Due to the heavy right tail, this Gamma pdf weighted by $1/\sqrt{\phi(x)}$ is not square integrable, and therefore does not have a converging Hermite expansion. As mentioned previously, one possible solution to this problem is to truncate the tail so that the Hermite expansion is convergent. In this example, we truncate the Gamma distribution at $x = 30$, which

corresponds to $1 - 1.173 \times 10^{-7}$ quantile. We consider this truncation point to be reasonably far right. The truncated Gamma distribution still has adverse features; It has a skewness of 0.75585 and an excess kurtosis of 0.85638, which are close to those before truncation, a skewness of 0.75592 and an excess kurtosis of 0.85714.

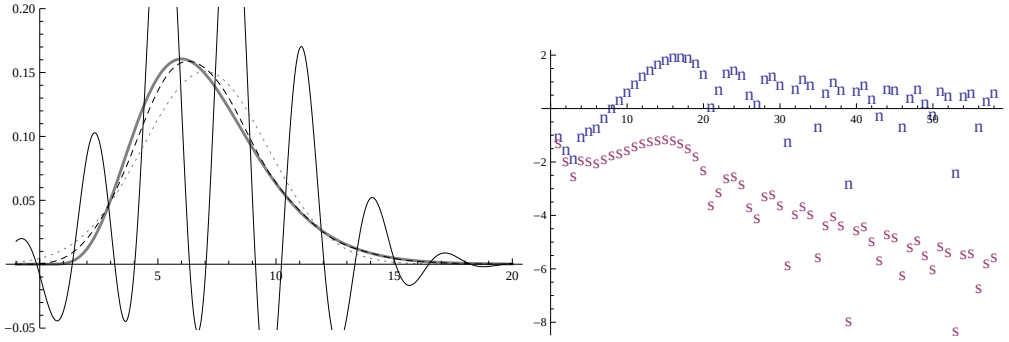

Figure 4: The pdfs (left) and the squared coefficients $\{C_k^2\}$ up to order 60 (right). In the left plots, the truncated Gamma distribution (thick gray), naive expansion (solid), the smoothed expansion (dashed), the Normal distribution (dotted), are shown. The order of expansions is $n = 30$ and the weight for smoothness is $q = .005$. In the right plots, the symbols 'n' denote the naive expansion and the symbols 's' denote the smoothed expansion, respectively. The vertical axis in the right plots is in a log scale with base 10.

The left plots in Figure 4 shows that the approximation quality of the naive expansion is very poor while smoothed expansion has much better quality, at least for the order of expansion $n = 30$. From Table 3 we find that the relative error of the risk measures with the smoothed expansion is at most 3.42%, which is much smaller than those with the Normal approximations.

From the right plots in Figure 4, we find that the convergence of the naive expansion is very slow, while the smoothed expansion converges much faster with smoothness weight $q = .005$. This example suggests that truncating and smoothing can make the expansion methods applicable to some adverse cases. One obvious drawback is, as mentioned previously, that the truncated moments may not be available for the cases where the pdf is not known.

Level	99%		99.5%	
Measure	VaR	ES	VaR	ES
Gamma	14.57	16.10	15.66	17.15
Normal	-9.72%	-12.27%	-11.78%	-14.55%
Naive	–	–	–	–
Smoothed	-1.63%	-2.75%	-2.37%	-3.42%

Table 3: The risk measures of the empirical distribution and relative errors of the approximations. The order of expansion is $n = 30$ and the weight for smoothness is $q = .005$. The minus signs indicate that the approximations underestimate the risk. Fig 4 shows that the naive expansion exhibits erratic shape and is not reliable.

3 Application to risk measurement

3.1 Risk measurement and risk aggregation

VaR and ES are very popular risk measures in practice[16]. As already mentioned, VaR is defined as a quantile of a PL distribution, and ES, as the expected loss given that the loss exceeds VaR, respectively. Assume that a random variable Z denote the PL of our portfolio, where the loss is denoted as a negative profit. Then, the VaR with a confidence level of α is defined as

$$\mathrm{VaR}_\alpha = -\sup\{z|P(Z \leq z) \leq 1-\alpha\},$$

and the corresponding ES is defined as

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

For instance, if Z has a continuous df F_Z and a pdf f_Z, then the 99% VaR is derived as $-z$ that satisfies $F_Z(z) = 0.01$, and the 99% ES is given by

$$-\mathrm{E}(Z|Z \leq -\mathrm{VaR}_{0.99}) = -\int_{-\infty}^{-\mathrm{VaR}_{0.99}} u f_Z(u)\mathrm{d}u/0.01. \tag{23}$$

[16]See Basel Committee (2012).

For the cases where Z denote the loss of the portfolio, the VaR and ES can be defined as $\mathrm{VaR}_\alpha = \inf\{z|P(Z \leq z) \geq \alpha\}$ and $\mathrm{ES}_\alpha = \mathrm{E}(Z|Z \geq \mathrm{VaR}_\alpha)$, respectively[17].

Thus, obtaining a PL distribution is vital for risk measurement. Theoretically, the process of aggregating multiple of risk factors and deriving a PL distribution can be described as follows: Assume that our portfolio is exposed to K risk factors, $X_1, \ldots, X_K$, in such a way that Z, the PL of our portfolio, is expressed using a K variate function ψ as

$$Z = \psi(X_1, \ldots, X_K).$$

Then, the PL distribution function is given by

$$F_Z(z) = P(Z \leq z) = \int_{-\infty}^{z} \mathrm{d}P(\psi(X_1, \ldots, X_K) \leq u).$$

Even in those cases where we have complete information on the risk factors' joint distribution, however, calculating this integration can be a tricky task when ψ is not a linear function, or when $X_1, \ldots, X_K$ do not have the K-variate Normal distribution[18]. Further, it may be rather exceptional that the risk factors' joint distribution is completely known.

3.2 Applying expansion methods

We discuss how F_Z can be approximated by the expansion methods. We impose two assumptions:

- The function ψ can be approximated by a finite Taylor expansion; that is, ψ has an approximation of a form:

$$\psi(x_1, \ldots, x_K) \simeq \sum_{i_1 + \cdots + i_K \leq n} a_{i_1, \ldots, i_K} x_1^{i_1} \cdots x_K^{i_K},$$

  where $\{a_{i_1, \ldots, i_K}\}$ are known constants.

[17] In the numerical examples, we used these definitions for the risk measures on the right tails.

[18] See Jorion (2007) for how the current practice cope with these problems.

- The cross-moments of the risk factors up to some finite order,

$$m_{i_1,\ldots,i_K} = \mathrm{E}(X_1^{i_1} \cdots X_K^{i_K})$$

  exist and are available. We may or may not know the joint distribution as long as the cross-moments are available.

Then, it is straightforward that the moments of Z can be approximated by linear combinations of risk factors' cross-moments, and therefore $C_k = \mathrm{E}(\mathrm{He}_k(Z))$ can be approximated. Thus, we can use the formula

$$F_Z(z) \simeq F_n(z) = \Phi(z) + \phi(z) \sum_{k=1}^{n} \frac{C_k}{\sqrt{k}} \mathrm{He}_{k-1}(z)$$

for calculating the VaR. The pdf can be approximated by

$$f_Z(z) \simeq f_n(z) = \phi(z) \sum_{k=0}^{n} C_k \mathrm{He}_k(x),$$

and further, using Equation (1), we have

$$\begin{aligned}\int_{-\infty}^{z} u f_n(u) \mathrm{d}u = &-\phi(z) + C_1\{\phi(z)z - \Phi(z)\} \\ &+ \phi(z) \sum_{k=2}^{n} \frac{C_k}{\sqrt{k}} \left\{ z\mathrm{He}_{k-1}(z) - \frac{1}{\sqrt{k-1}} \mathrm{He}_{k-2}(z) \right\},\end{aligned}$$

which can be used for calculating the ES (See Equation (23)).

3.3 The delta-gamma-vega-Normal model for an European call options

In order to see how the expansion methods can be applied to risk measurement, we apply the expansion methods to a very common problem; approximating the distribution of a change in the value (or 'profit or loss') of an European call option, where we assume that the underlying asset value and the IV have the log Normal distribution. The value of an European call option C can be expressed using the plain Black and Scholes formula,

$$\begin{aligned} C(S, \sigma_I) &= S\Phi(d) - K\mathrm{e}^T \Phi(d - \sigma_I\sqrt{T}), \\ d &= \frac{\log(S/K) + (r + \sigma_I^2/2)T}{\sigma_I\sqrt{T}}, \end{aligned} \tag{24}$$

where S, K, r, σ_I, T are the value of the underlying asset, strike price, risk free interest rate, IV, and time to maturity, respectively. Suppose that current asset value and the IV are S_0 and σ_{I0}, respectively. Assume that the asset value and the IV in the time $0 < t(<< T)$ ahead can be expressed as $S_0 \mathrm{e}^X$, and $\sigma_{I0}\mathrm{e}^Y$, where (X, Y) has the bivariate Normal distribution, whose parameters are estimated from the historical observations. Then, the PL of the option can be expressed as

$$\Delta C = C(S_0\mathrm{e}^X, \sigma_{I0}\mathrm{e}^Y) - C(S_0, \sigma_{I0}). \tag{25}$$

Deriving the distribution of ΔC is a key problem in measuring the risk of the option, however; this is a tricky task. Instead of deriving the distribution, in practice, we often approximate it using the Taylor expansion. We define the derivatives δ, γ and κ as

$$\begin{aligned}
\delta &= \frac{\partial}{\partial S} C(S, \sigma_{I0})\bigg|_{S=S_0} = \Phi(d), \\
\gamma &= \frac{\partial^2}{\partial S^2} C(S, \sigma_{I0})\bigg|_{S=S_0} = \frac{\phi(d)}{S_0\sigma_I\sqrt{T}}, \\
\kappa &= \frac{\partial}{\partial \sigma_I} C(S_0, \sigma_I)\bigg|_{\sigma_I=\sigma_{I0}} = S_0\phi(d)\sqrt{T},
\end{aligned}$$

then the approximation of ΔC, or the delta-gamma-vega model is given by

$$\Delta C \simeq \Delta C_{\delta\gamma\kappa} = \delta S_0 X + \frac{1}{2}(\delta S_0 + \gamma S_0^2)X^2 + \kappa\sigma_{I0}Y. \tag{26}$$

Even with the aid from this Taylor approximation, deriving the distribution of $\Delta C_{\delta\gamma\kappa}$ is not straightforward. However, the moments $\{\mathrm{E}(\Delta C_{\delta\gamma\kappa}^k)\}$, $k = 0, 1, \ldots$, can be easily obtained using the assumption that X and Y has bivariate Normal distribution. Therefore, we can apply the expansion methods to the distribution of $\Delta C_{\delta\gamma\kappa}$ using its moments. Figure 5 and Table 4 show approximations for the PL distribution of an European call option (three months, at-the-money) on S&P 500 (See Table 5 for summary statistics of the data used). They show that the approximation by the naive expansion is very close to the empirical distribution of the 20,000 pseudo-random samples from ΔC, and we find that the naive expansion is sufficient in this particular case.

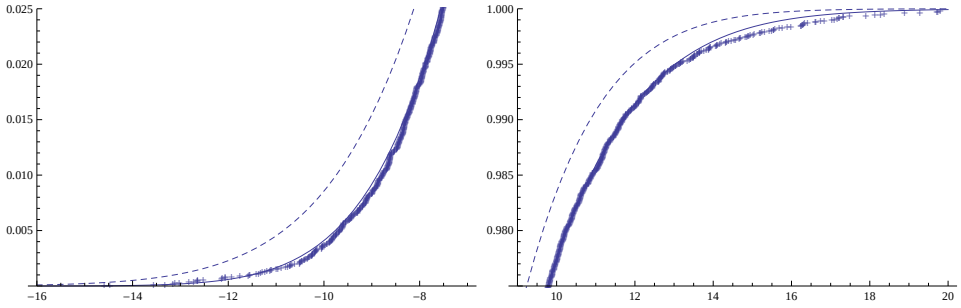

Figure 5: Magnified left and right tails of approximated dfs of $\Delta C_{\delta\gamma\kappa}$. The solid curve shows the naive expansion with order $n = 10$, and the dashed curve shows the Normal distribution. The '+' symbols exhibit the distribution of ΔC obtained by applying the Monte Carlo method with 20,000 trials to Equation (25).

Level	0.5%		1%	
Measure	VaR	ES	VaR	ES
Empirical ($\times -1$)	9.676	10.806	8.777	9.998
Normal	12.11%	13.33%	11.00%	12.45%
Naive	0.91%	0.19%	1.24%	0.56%
Smoothed	1.92%	1.18%	2.15%	1.53%

Level	99%		99.5%	
Measure	VaR	ES	VaR	ES
Empirical	11.679	13.838	13.077	15.355
Normal	-7.03%	-10.69%	-8.53%	-12.98%
Naive	0.18%	-2.21%	-0.36%	-3.83%
Smoothed	-0.35%	-2.83%	-0.94%	-4.50%

Table 4: The risk measures of the empirical distribution from the Monte Carlo method and relative errors of the approximations. The target of the approximations is the distribution of ΔC, which is not available. We compare the approximations with the empirical distribution obtained from the Monte Carlo method in this Table. The risk is measured on the left (Levels 0.5% and 1%) and right (Levels 99% and 99.5%) tails. The order of expansion is $n = 10$ and the weight for smoothness is $q = .01$. The minus signs indicate that the approximations underestimate the risk.

4 Expansion methods for multivariate distributions

In this Section we discuss expansion methods applied to multivariate density functions and their relationship with conditional expectations and copula functions. We firstly deal with bivariate distributions in some numerical examples, and then discuss the extension to general multivariate cases. The key finding is that the product of orthogonal polynomials can be used to expand a joint density function[19].

4.1 Expanding a joint density function

Let X and Y have a joint density function $f(x, y)$. Assume that the conditional density can be expanded using Equation (4) as

$$\begin{aligned} f(x|Y=y) &= \frac{f(x,y)}{\int_{-\infty}^{\infty} f(u,y)\mathrm{d}u} \\ &= \phi(x)\sum_{k=0}^{\infty}\int_{-\infty}^{\infty} f(u|Y=y)\mathrm{He}_k(u)\mathrm{d}u\ \mathrm{He}_k(x). \end{aligned}$$

By multiplying throughout by $\int_{-\infty}^{\infty} f(u, y)\mathrm{d}u$, we have

$$f(x,y) = \phi(x)\sum_{k=0}^{\infty}\int_{-\infty}^{\infty} f(u,y)\mathrm{He}_k(u)\mathrm{d}u\ \mathrm{He}_k(x). \tag{27}$$

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

$$f(x,y) = \phi(y)\sum_{l=0}^{\infty}\int_{-\infty}^{\infty} f(x,v)\mathrm{He}_l(v)\mathrm{d}v\ \mathrm{He}_l(y). \tag{28}$$

[19] Use of the multivariate orthogonal polynomials can also be a natural extension of the univariate methods in Section 2 to multivariate ones. However, in this paper, we show that the product of univariate orthogonal polynomials, which is much simpler than general multivariate orthogonal polynomials, can still be of a great use. Multivariate orthogonal polynomials are studied by, for example, Xu (1994) and Xu (1995). The fact that a product of univariate orthogonal systems can be used as a multivariate orthogonal system is pointed out by Hall (1983), for example.

By plugging Equation (28) into the right hand side of Equation (27), we have the expansion for a joint density function,

$$\begin{aligned} f(x,y) &= \phi(x)\sum_{k=0}^{\infty}\int_{u=-\infty}^{\infty}\phi(y)\sum_{l=0}^{\infty}\int_{v=-\infty}^{\infty} f(u,v)\mathrm{He}_l(v)\mathrm{d}v\ \mathrm{He}_l(y) \\ &\qquad \times \mathrm{He}_k(u)\mathrm{d}u\ \mathrm{He}_k(x) \\ &= \phi(x)\phi(y)\sum_{k,l=0}^{\infty}\mathrm{E}(\mathrm{He}_k(X)\mathrm{He}_l(Y))\mathrm{He}_k(x)\mathrm{He}_l(y). \end{aligned} \tag{29}$$

Since He_k are polynomials, $\mathrm{E}(\mathrm{He}_k(X)\mathrm{He}_l(Y))$ are linear combinations of moments and cross-moments of X and Y, and therefore all the information from X and Y which is used in the expansion is summarised in terms of their moments and cross-moments.

Based on Equation (29), given the moments and cross-moments of X and Y, $\mathrm{E}(X^sY^t)$, $s+t \leq n$, the joint density function of X and Y can be approximated by

$$f_n(x,y) = \phi(x)\phi(y)\sum_{k+l\leq n}\mathrm{E}(\mathrm{He}_k(X)\mathrm{He}_l(Y))\mathrm{He}_k(x)\mathrm{He}_l(y). \tag{30}$$

The usual factorisation holds in the case of independence.

Properties of bivariate expansions

From the discussion in the previous Section and by the construction of the bivariate expansion, it can be shown that a joint pdf f has a converging Hermite expansion of the form in Equation (29) if $f(x,y)/\sqrt{\phi(x)\phi(y)}$ is square integrable. In such cases, we have

$$\iint_{u,v=-\infty}^{\infty}\frac{\{f(u,v)\}^2}{\phi(u)\phi(v)}\mathrm{d}u\mathrm{d}v = \sum_{k,l=0}^{\infty}C_{k,l}^2 < \infty,$$

where $C_{k,l} = \mathrm{E}(\mathrm{He}_k(X)\mathrm{He}_l(Y))$.

The joint df is given by

$$\begin{aligned} F(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\mathrm{d}v\mathrm{d}u \\ &= -\Phi(x)\Phi(y) + \Phi(x)F_Y(y) + F_X(x)\Phi(y) \\ &\qquad + \phi(x)\phi(y)\sum_{k,l=1}^{\infty}\frac{C_{k,l}}{\sqrt{kl}}\mathrm{He}_{k-1}(x)\mathrm{He}_{l-1}(y), \end{aligned}$$

where F_X and F_Y are the marginal dfs;

$$F_X(x) = \Phi(x) + \phi(x) \sum_{k=1}^{\infty} \frac{C_{k,0}}{\sqrt{k}} \mathrm{He}_{k-1}(x),$$
$$F_Y(y) = \Phi(y) + \phi(y) \sum_{l=1}^{\infty} \frac{C_{0,l}}{\sqrt{l}} \mathrm{He}_{l-1}(y).$$

Thus, F has a converging Hermite expansion if F_X and F_Y have convergent Hermite expansions and if

$$\iint_{u,v=-\infty}^{\infty} \frac{\{F(u,v) + \Phi(u)\Phi(v) - F_X(u)\Phi(v) - \Phi(u)F_Y(v)\}^2}{\phi(u)\phi(v)} \mathrm{d}u\mathrm{d}v$$
$$= \sum_{k,l=1}^{\infty} \frac{C_{k,l}^2}{kl} < \infty$$

In order to measure the difference between two dfs, F_1 and F_2, we use the WISD, $\int_{-\infty}^{\infty} \{F_1(u) - F_2(u)\}^2/\phi(u)\mathrm{d}u$, for univariate cases. In fact, we use the Perceval identity in Equation (6) to relate the WISD with the coefficients in the Hermite expansions. For the bivariate cases, however, the natural extension of WISD, $\iint_{u,v=-\infty}^{\infty} \{F_1(u,v) - F_2(u,v)\}^2/\{\phi(u)\phi(v)\}\mathrm{d}u\mathrm{d}v$, is not useful since this quantity diverges unless they have common marginals. Instead, we might consider a WISD with an adjustment for the marginals to obtain a tractable form:

$$\iint_{u,v=-\infty}^{\infty} \{F_1(u,v) - F_2(u,v) - (F_{1X}(u) - F_{2X}(u))\Phi(v)$$
$$- \Phi(u)(F_{1Y}(v) - F_{2Y}(v))\}^2/\{\phi(u)\phi(v)\}\mathrm{d}u\mathrm{d}v$$
$$= \sum_{k,l=1}^{\infty} \frac{(C_{1\ k,l} - C_{2\ k,l})^2}{kl},$$

where F_{iX} and F_{iY} are the marginal dfs of F_i, $i = 1, 2$, and $C_{i\ k,l}/\sqrt{kl}$ are the coefficients of their Hermite expansions. Then we consider the WISDs for the two marginals as univariate distributions.

Similarly to univariate cases, it can be shown that the Hermite expansion of a pdf of a empirical distribution is divergent, while that of a df is convergent;

that is,

$$\sum_{k,l=0}^{\infty} \hat{C}_{k,l}^2 = \infty,$$
$$\sum_{k,l=1}^{\infty} \frac{\hat{C}_{k,l}^2}{kl} + \sum_{k=1}^{\infty} \frac{\hat{C}_{k,0}^2}{k} + \sum_{l=1}^{\infty} \frac{\hat{C}_{0,l}^2}{l} < \infty,$$

where $\hat{C}_{k,l} = N^{-1} \sum_{i=1}^{N} \mathrm{He}_k(X^{(i)})\mathrm{He}_l(Y^{(i)})$ are the natural estimators given N i.i.d. samples $(X^{(i)}, Y^{(i)})$.

4.2 Application and related techniques

As discussed in Section 2.5, we expect that a naive application of the Hermite expansion can result in a poor approximation, and that the same techniques — standardisation, smoothing and optimisation — may provide expansion methods with greater robustness. We discuss the application of these techniques in the bivariate case and show numerical examples.

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 now have the cross-moment term $\mathrm{E}(XY)$. Our bivariate weight function is $\phi(x)\phi(y)$, which is the joint density function of two independent Normal variables. This suggests that a transformation should be done so that the covariance of the two variables 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. This is merely a two-dimensional Cholesky decomposition.

Other moments can be matched to those of the weight function in the same way as for the univariate case.

That is, 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

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

Smoothing

We firstly consider smoothing a joint pdf. Assume that a pdf f has a converging Hermite expansion

$$f(x,y) = \phi(x)\phi(y)\sum_{k,l=0}^{\infty} C_{k,l}\mathrm{He}_k(x)\mathrm{He}_l(y),$$

and suppose that we wish to smooth f in order to improve its convergence speed. Similarly to the univariate cases, we search for a smooth function

$$f^S(x,y) = \phi(x)\phi(y)\sum_{k,l=0}^{\infty} C^S_{k,l}\mathrm{He}_k(x)\mathrm{He}_l(y)$$

near f. The second derivatives are given by

$$\frac{\partial^2}{\partial x^2}f^S(x,y) = \phi(x)\phi(y)\sum_{k,l=0}^{\infty}\sqrt{(k+1)(k+2)}C^S_{k,l}\mathrm{He}_{k+2}(x)\mathrm{He}_l(y),$$
$$\frac{\partial^2}{\partial y^2}f^S(x,y) = \phi(x)\phi(y)\sum_{k,l=0}^{\infty}\sqrt{(l+1)(l+2)}C^S_{k,l}\mathrm{He}_k(x)\mathrm{He}_{l+2}(y),$$

and the WISC is given by[20]

$$\iint_{u,v=-\infty}^{\infty}\frac{\left\{\frac{\partial^2}{\partial u^2}f^S(u,v)\right\}^2 + \left\{\frac{\partial^2}{\partial v^2}f^S(u,v)\right\}^2}{\phi(u)\phi(v)}\mathrm{d}u\mathrm{d}v$$
$$= \sum_{k,l=0}^{\infty}\{(k+1)(k+2)+(l+1)(l+2)\}\left(C^S_{k,l}\right)^2.$$

The WISD is given by

$$\iint_{u,v=-\infty}^{\infty}\frac{\{f^S(u,v)-f(u,v)\}^2}{\phi(u)\phi(v)}\mathrm{d}u\mathrm{d}v = \sum_{k,l=0}^{\infty}(C^S_{k,l}-C_{k,l})^2.$$

For $0 < q < 1$, the coefficients that minimises $(1-q)\mathrm{WISD} + q\mathrm{WISC}$ is given by

$$C^S_{k,l} = \frac{1-q}{1-q+q(k+1)(k+2)+q(l+1)(l+2)}C_{k,l}.$$

[20] We can define WISC using other measures such as the Hessian determinants in a similar manner, however, we demonstrate here the use of the second derivatives for simplicity.

In order to smooth empirical distributions, we need to start with the joint dfs, not the joint pdfs, since the joint pdfs of empirical distributions do not have converging Hermite expansions. Smoothing joint dfs can be done similarly to smoothing joint pdfs except that the WISD and WISC need to be adjusted for the marginals, so that the integrations have finite values. That is, the coefficients for a smoothed empirical distribution function is derived to be

$$\begin{aligned}
\hat{C}^S_{k,0} &= \frac{1-q}{1-q+qk(k+1)}\hat{C}_{k,0}, && k \geq 1,\ l = 0,\\
\hat{C}^S_{0,l} &= \frac{1-q}{1-q+ql(l+1)}\hat{C}_{0,l}, && k = 0,\ l \geq 1,\\
\hat{C}^S_{k,l} &= \frac{1-q}{1-q+qk(k+1)+ql(l+1)}\hat{C}_{k,l}, && k,l \geq 1,
\end{aligned}$$

and $\hat{C}^S_{0,0} = 1$.

Optimisation

We can make a parallel discussion to the univariate case in Section 2.5 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

$$f(x,y) = \phi(x)\phi(y)\sum_{k,l=0}^{\infty} C_{k,l}\mathrm{He}_k(x)\mathrm{He}_l(y).$$

Let $\hat{C}_{k,l} = N^{-1}\sum_{i=1}^{N}\mathrm{He}_k(X^{(i)})\mathrm{He}_l(Y^{(i)})$ then $\mathrm{E}(\hat{C}_{k,l}) = C_{k,l}$. We consider a class of estimator

$$\hat{f}^O_n(x,y) = \phi(x)\phi(y)\sum_{k+l\leq n}\alpha_{k,l}\hat{C}_{k,l}\mathrm{He}_k(x)\mathrm{He}_l(y)$$

and optimise the coefficients $\alpha_{k,l}$, $k + l \leq n$, so that the WMISE

$$\mathrm{E}\left(\iint_{u,v=-\infty}^{\infty}\frac{\left\{\hat{f}^O_n(u,v) - f(u,v)\right\}^2}{\phi(u)\phi(v)}\mathrm{d}u\mathrm{d}v\right) \tag{31}$$

is minimised. Thus Expression (31) is equal to

$$\sum_{k+l\leq n}\alpha^2_{k,l}\mathrm{E}\left(\hat{C}^2_{k,l}\right) - 2\sum_{k+l\leq n}\alpha_{k,l}C^2_{k,l} + \sum_{k,l=0}^{\infty}C^2_{k,l}. \tag{32}$$

An unbiased estimator for Expression (32) is given by

$$\sum_{k+l\leq n} \alpha_{k,l}^2 \hat{C}_{k,l}^2 - 2 \sum_{k+l\leq n} \alpha_{k,l} \frac{N\hat{C}_{k,l}^2 - \hat{B}_{k,l}^2}{N-1} + \sum_{k,l=0}^{\infty} \frac{N\hat{C}_{k,l}^2 - \hat{B}_{k,l}^2}{N-1},$$

where $\hat{B}_{k,l}^2 = N^{-1}\sum_{i=1}^{N} \left\{ \mathrm{He}_k(X^{(i)}) \mathrm{He}_l(Y^{(i)}) \right\}^2$. From a similar discussion to that in Section 2.5, we use a biased but considered to be better estimator

$$\sum_{k+l\leq n} \alpha_{k,l}^2 \hat{C}_{k,l}^2 - 2 \sum_{k+l\leq n} \alpha_{k,l} \frac{\left(N\hat{C}_{k,l}^2 - \hat{B}_{k,l}^2\right)_+}{N-1} + \sum_{k,l=0}^{\infty} \frac{\left(N\hat{C}_{k,l}^2 - \hat{B}_{k,l}^2\right)_+}{N-1}. \tag{33}$$

Again, $\alpha_{0,0} = 1$ is required so that $\iint_{u,v=-\infty}^{\infty} \hat{f}_n^O(u,v)\mathrm{d}u\mathrm{d}v = 1$ is satisfied. If the variables are standardised, we can set $\alpha_{k,l} = 0$ for $k+l = 1, 2$. For $k+l \geq 3$, Equation (33) is minimised when

$$\alpha_{k,l} = \frac{\left(N\hat{C}_{k,l}^2 - \hat{B}_{k,l}^2\right)_+}{(N-1)\hat{C}_{k,l}^2}.$$

Optimisation with smoothing

The derivation is almost parallel to that for the univariate case and $\{\alpha_{k,l}\}$ are given by $\alpha_{0,0} = 1$, $\alpha_{k,l} = 0$ for $k + l = 1, 2$, and

$$\alpha_{k,l} = \frac{\left(N\hat{C}_{k,l}^2 - \hat{B}_{k,l}^2\right)_+}{\{1 - q + q(k+1)(k+2) + q(l+1)(l+2)\}(N-1)\hat{C}_{k,l}^2}$$

for $k + l \geq 3$.

Numerical examples

We show numerical examples using two data sets: 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)[21] on the Nikkei 225;

[21] The implied volatility was obtained using the standard Black and Scholes formula.

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. The reason why we consider the combination of a stock index and an implied volatility (or IV) of an option on the index is that both of these variables are important for measuring market risk of an option premium[22]. Summary statistics are given in Table 5.

	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	.654	5.018
Skewness ($\times 10^{-1}$)	−2.269	3.125	0.715	−2.510
Excess Kurtosis	1.013	1.957	.376	1.782
Min ($\times 10^{-1}$)	−0.423	−1.375	−0.185	−2.275
Max ($\times 10^{-1}$)	.352	1.996	.213	1.910
Correl. coef.	−0.3612		−0.7443	

Table 5: Summary statistics of the data sets.

Comparison of the expansion methods. Figures 6 and 7 shows numerical examples of bivariate expansions. Each example approximates the joint density of an index return and the log-difference of the implied volatility (IV). We show examples with the order $n = 20$, and weight for smoothness $q = 0.01$ where used.

By comparing the plots in Figures 6 and 7, we can see that smoothing and optimisation mitigates the fluctuations and negative density around the tail area.

The Figures also show that the approximations capture the characteristics of the original observations: Figure 6 shows a high concentration of the ob-

[22] See, for instance, Marumo (2007).

servations around the origin with only a few observations around them while Figure 7 shows a high concentration of the observations along the line from front-left to rear-right, which also can be seen as a large correlation coefficient in Table 5.

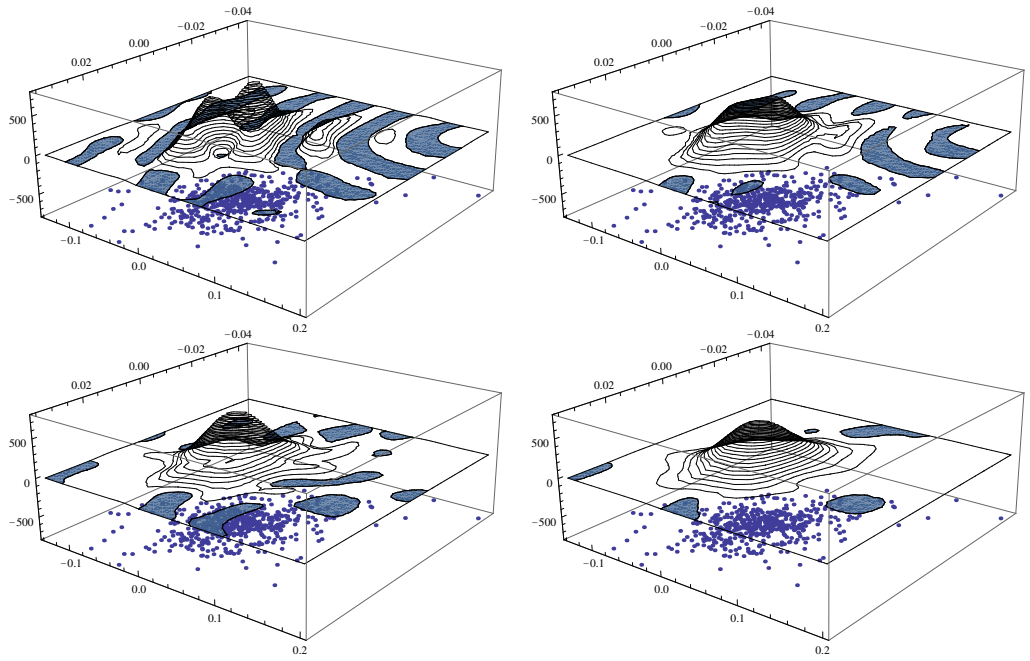

Figure 6: Approximated joint densities of the N225 log-return and the IV log-difference (3 month, at-the-money European call option). The front-left axes in the plots corresponds those of IV. The naive (upper left), smoothed (upper right), optimised (lower left), and optimised and smoothed (lower right) expansions are shown. The gray areas indicate those with negative density. The order of expansion is $n = 20$ and the weight for smoothness is $q = 0.01$ where used. The observations are scattered on the plane $z = -800$ to show the relationship. See Table 5 for a summary of the data set.

Figure 8 shows squared coefficients $\left\{\sum_{k+l=n} C_{k,l}^2/(\max\{k,1\}\max\{l,1\})\right\}$ up to order $n = 40$. This suggests that the convergence of naive and optimised expansions is slow and the convergence of optimised and smoothed expansions is the fastest among the four, as expected.

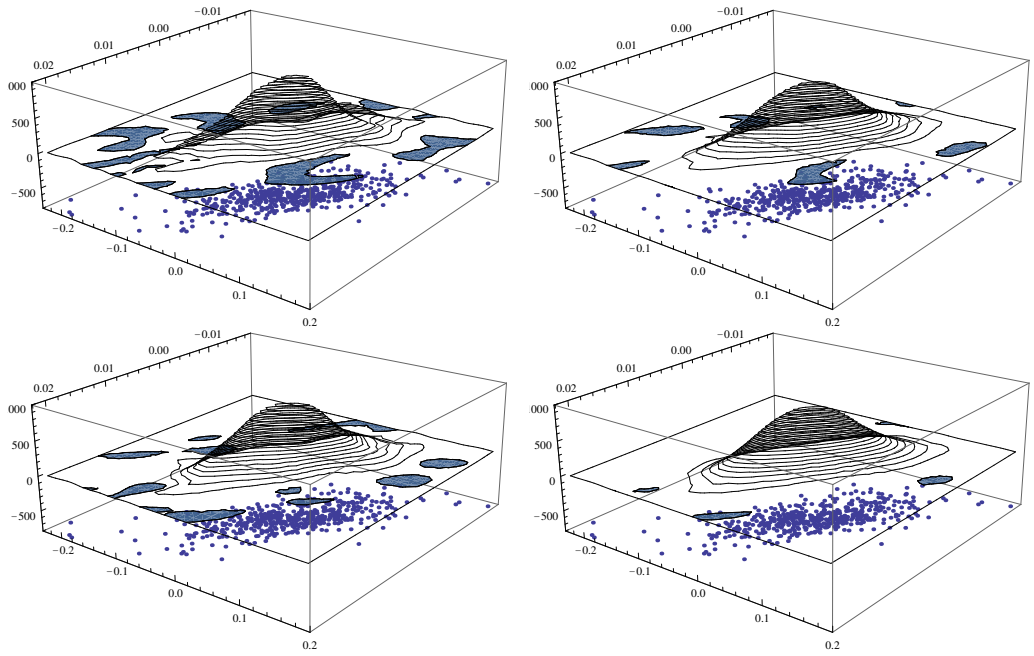

Figure 7: Approximated joint densities of the S&P 500 log-return and the IV log-difference (3 month, at-the-money European call option). The front-left axes in the plots corresponds those of IV. The naive (upper left), smoothed (upper right), optimised (lower left), and optimised and smoothed (lower right) expansions are shown. The gray areas indicate those with negative density. The order of expansion is $n = 20$ and the weight for smoothness is $q = 0.01$ where used. The observations are scattered on the plane $z = -800$ to show the relationship. See Table 5 for a summary of the data set.

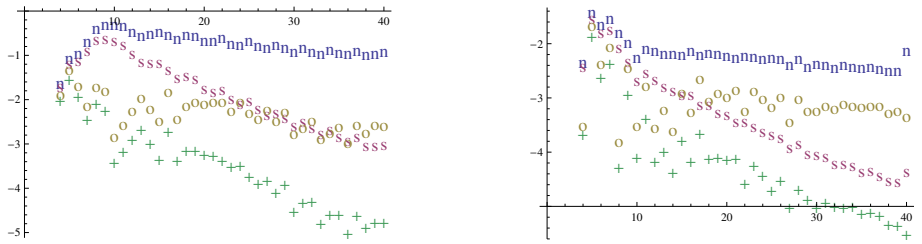

Figure 8: Squared coefficients of the expansions up to order 40. N255 (left) and S&P 500 (right). Plots are in log scale with base 10. The symbols in the left plots denote naive (n), smoothed (s), optimised (o), and optimised and smoothed (+) expansions, respectively.

4.3 Conditional expectations and expansion

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

$$\mathrm{E}(\mathrm{He}_r(Y)|X = x) = \frac{\int_{-\infty}^{\infty} \mathrm{He}_r(y) f(x, y) \mathrm{d}y}{\int_{-\infty}^{\infty} f(x, y) \mathrm{d}y},$$

where f is the joint pdf of (X, Y). By applying expansions to f, we have

$$\begin{aligned} \mathrm{E}(\mathrm{He}_r(Y)|X = x) &= \frac{\int_{-\infty}^{\infty} \mathrm{He}_r(y)\phi(x)\phi(y) \sum_{k,l=0}^{\infty} C_{k,l} \mathrm{He}_k(x)\mathrm{He}_l(y)\mathrm{d}y}{\phi(x) \sum_{k=0}^{\infty} C_{k,0}\mathrm{He}_k(x)} \\ &= \frac{\sum_{k=0}^{\infty} C_{k,r}\mathrm{He}_k(x)}{\sum_{k=0}^{\infty} C_{k,0}\mathrm{He}_k(x)}, \end{aligned} \tag{34}$$

where $C_{k,l} = \mathrm{E}(\mathrm{He}_k(X)\mathrm{He}_l(Y))$. We can approximate the conditional expectation by taking sums in Equation (34) up to some finite n.

The conditional moments such as $\mathrm{E}(Y^r|X = x)$ can be obtained by solving the system of linear equations in $\mathrm{E}(\mathrm{He}_0(Y)|X = x), \ldots, \mathrm{E}(\mathrm{He}_r(Y)|X = x)$.

The same techniques, standardising, smoothing, optimisation, and optimisation with smoothing, can be applied similarly to the approximations for the joint pdfs.

Numerical examples

Figure 9 is numerical examples using the same data sets as in Figures 6 and 7. The plots show the expected value of IV log-difference conditioned on stock indices log-return, or $\mathrm{E}(Y|X = x)$ (solid line) and $\pm$conditional standard deviation, or $\mathrm{E}(Y|X = x) \pm \sqrt{\mathrm{V}(Y|X = x)}$ (dotted lines). Both are obtained by applying the optimised and smoothed expansion with order $n = 20$ and weight for smoothness $q = 0.01$.

These plots shows that expansions captures non-linear dependence structure of the original observations.

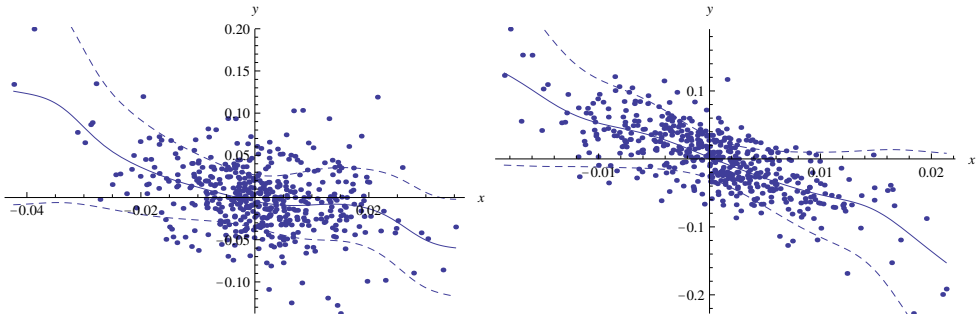

Figure 9: Expected value of IV log-difference conditioned on stock indices log-return, or $\mathrm{E}(Y|X=x)$ (solid line) and $\pm$conditional standard deviation, or $\mathrm{E}(Y|X=x) \pm \sqrt{\mathrm{V}(Y|X=x)}$ (dotted lines). Nikkei 225 (left) and S&P 500 (right). Both are obtained by applying the optimised and smoothed expansion with order $n = 20$ and weight for smoothness $q = .01$. The data used are the same as in Figures 6 and 7. Scattered dots represent observations.

4.4 Copula functions and expansion

The two dimensional copula is a function $C : [0,1]^2 \mapsto [0,1]$ which satisfies[23] $C(t,0) = C(0,t) = 0$, $C(1,t) = C(t,1) = t$, for all $0 \leq t \leq 1$ and $C(u_1,v_1) - C(u_2,v_1) - C(u_1,v_2) + C(u_2,v_2) \geq 0$, 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[24] yields that there is a unique copula representation for F, given by

$$F(x,y) = C(F_X(x), F_Y(y)). \tag{35}$$

Assume that F is differentiable. Then, from Equation (35), the joint density function is given by $f(x,y) = f_X(x)f_Y(y)c(F_X(x), F_Y(y))$, 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. From Equations (29) and (4), we have an expression of c by a rational function:

$$\begin{aligned} c(F_X(x), F_Y(y)) &= \frac{f(x,y)}{f_X(x)f_Y(y)} \\ &= \frac{\sum_{k,l=0}^{\infty} C_{k,l} \mathrm{He}_k(x) He_l(y)}{\sum_{k,l=0}^{\infty} C_{k,0}C_{0,l} \mathrm{He}_k(x) \mathrm{He}_l(y)}. \end{aligned} \tag{36}$$

or

$$c(u,v) = \frac{\sum_{k,l=0}^{\infty} C_{k,l} \mathrm{He}_k(F_X^{-1}(u)) \mathrm{He}_l(F_Y^{-1}(v))}{\sum_{k,l=0}^{\infty} C_{k,0}C_{0,l} \mathrm{He}_k(F_X^{-1}(u)) \mathrm{He}_l(F_Y^{-1}(v))}. \tag{37}$$

Note that it is also possible to remove the marginal functions F_X and F_Y from Equation (37) 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, for instance, Freedman (1981) for an expansion with closed support.

[23] Intuitively, a two dimensional copula function can be understood as a df of two random variables U, V, where both U and V have uniform marginal distributions on $[0,1]$, and are possibly dependent on each other.

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

Standardisaton, smoothing and optimisation

From Equations (34) and (36), we find that the copula density can be expressed using the conditional expectation

$$c(F_X(x), F_Y(y)) = \frac{\sum_{l=0}^{\infty} \mathrm{E}(\mathrm{He}_l(Y)|X = x)\mathrm{He}_l(y)}{\sum_{l=0}^{\infty} C_{0,l}\mathrm{He}_l(y)}. \tag{38}$$

This suggests that the techniques such as standardisation, smoothing and optimisation are available as long as they are available for the calculation of the conditional expectation.

Numerical examples

Figure 10 is the examples using the same data sets as in previous sections. The left plot in Figure 11 is a copula density obtained by applying standardised and optimised Hermite expansion to 500 random samples from the Clayton copula ($\theta = 2$) with the Normal marginal distributions. The Clayton copula function with parameter θ is given by $C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}$. By comparing this with the theoretical shape of the Clayton copula density in the right plot, 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 reproduces the shape similar to the one of the Clayton copula density.

From the plots in Figure 10, 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 the plots in Figures 10 and 11, that the approximations are erratic near the boundary. This is most likely to be because our method uses rational functions, which can be erratic at the extremes, for approximationg the copula density[25].

Apart from that, they exhibit characteristics of dependence structure; for instance, the S&P 500 data set shows heavier concentration along the diagonal line from $(0, 1)$ to $(1, 0)$ than Nikkei 225 data set.

[25]See Equations (36) and (37).

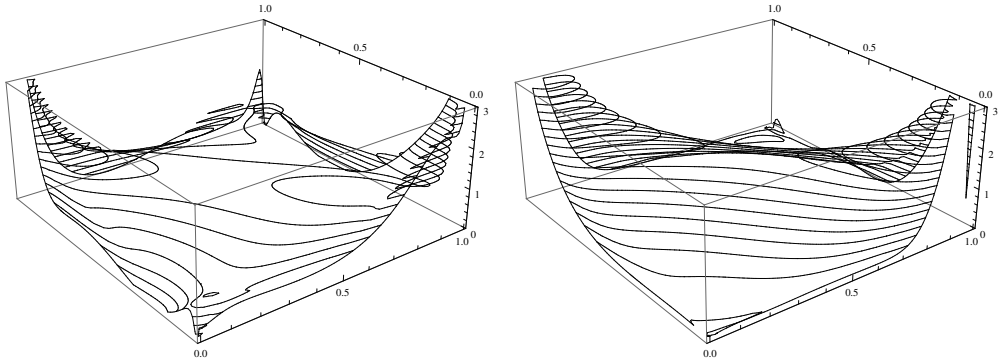

Figure 10: Copula density obtained by applying the optimised and smoothed expansion to 500 observations of the stock indices log-returns and the IV log-differences. Nikkei 225 (left) and S&P 500 (right). The front right axis corresponds to the indices log-return. The data used are the same as in Figures 6 and 7. The order of expansion is $n = 20$ and weight for smoothness is $q = .01$.

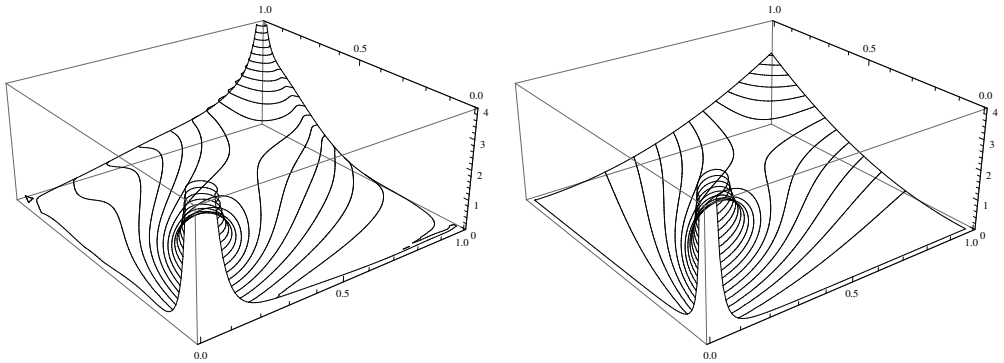

Figure 11: Copula density obtained by applying the optimised and smoothed expansion to 500 random samples from the Clayton copula ($\theta = 2$) with the Normal marginal distributions (left). The order of the expansion is $n = 20$ and the weight for smoothness is $q = .01$. The right plot shows the Clayton copula density. By comparing them with each other, 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 reproduce the shape similar to the one of the Clayton copula density.

4.5 Extension to the 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 4.1, we have

$$f(x_1, \ldots, x_p) = \\ \phi(x_1) \cdots \phi(x_p) \sum_{k_1, \ldots, k_p} \mathrm{E}(\mathrm{He}_{k_1}(X_1) \cdots \mathrm{He}_{k_p}(X_p)) \mathrm{He}_{k_1}(x_1) \cdots \mathrm{He}_{k_p}(x_p). \quad (39)$$

The condition for convergence is expressed as

$$\int \cdots \int_{u_1, \ldots, u_p} \frac{\{f(u_1, \ldots, u_p)\}^2}{\phi(u_1) \cdots \phi(u_p)} \mathrm{d}u_1 \cdots \mathrm{d}u_p < \infty.$$

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 linear transformation can be found, for example, by inverting the Cholesky decomposition of the covariance matrix of the original variables.

We can make a parallel discussion on smoothing and optimisation to the bivariate case.

Conditional expectations

Consider the expectation of $\mathrm{He}_{k_{i+1}}(X_{i+1}) \cdots \mathrm{He}_{k_p}(X_p)$ under the condition $X_1 = x_1, \ldots, X_i = x_i$. From Equation (39), we have

$$\mathrm{E}(\mathrm{He}_{k_{i+1}}(X_{i+1}) \cdots \mathrm{He}_{k_p}(X_p) | X_1 = x_1, \ldots, X_i = x_i) \\ = \frac{\sum_{k_1, \ldots, k_i} \mathrm{E}(\mathrm{He}_{k_1}(X_1) \cdots \mathrm{He}_{k_p}(X_p)) \mathrm{He}_{k_1}(x_1) \cdots \mathrm{He}_{k_i}(x_i)}{\sum_{k_1, \ldots, k_i} \mathrm{E}(\mathrm{He}_{k_1}(X_1) \cdots \mathrm{He}_{k_i}(X_i)) \mathrm{He}_{k_1}(x_1) \cdots \mathrm{He}_{k_i}(x_i)}. \quad (40)$$

Standardisation is done by applying Equation (40) 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 smoothing and optimisation to the bivariate case.

Copula density

Using Equation (39), a p-variate copula density function can be expressed as

$$\begin{aligned} c(F_1(x_1), \ldots, F_p(x_p)) &= \frac{f(x_1, \ldots, x_p)}{f_1(x_1) \cdots f_p(x_p)} \\ &= \frac{\sum_{k_1,\ldots,k_p} \mathrm{E}(\mathrm{He}_{k_1}(X_1) \cdots \mathrm{He}_{k_p}(X_p)) \mathrm{He}_{k_1}(x_1) \cdots \mathrm{He}_{k_p}(x_p)}{\sum_{k_1,\ldots,k_p} \mathrm{E}(\mathrm{He}_{k_1}(X_1)) \cdots \mathrm{E}(\mathrm{He}_{k_p}(X_p)) \mathrm{He}_{k_1}(x_1) \cdots \mathrm{He}_{k_p}(x_p)}, \end{aligned} \tag{41}$$

where F_i is the marginal df of X_i and f_i is its marginal pdf. From Equations (40) and (41), we have

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

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

5 Discussion

We reviewed the convergence properties of the Hermite expansion and proposed techniques, smoothing and optimising, to mitigate the fragility of naive applications of the Hermite expansion.

We also showed that the simple extension of univariate expansion methods is applicable to the joint density function, and demonstrated how it works using the bivariate case.

The techniques, smoothing and optimisation, can make the approximation quality sufficient for many purposes such as visualising non-linear dependence structure, approximating the conditional mean, and approximating the copula density; however, even with these techniques, the expansions showed negative density. One possible solution to this might be to optimise MISE subject to a constraint that the joint density is non-negative: see Hall and Presnell (1999).

The remaining issues include the problem of choosing an 'optimal' weight for smoothness q. This can be similar to the problem of choosing the optimal bandwidth in kernel density estimation, and we expect we find some criterion similarly.

Dealing with problems including these can be our future work.

References

Basel Committee. Consultative document — fundamental review of the trading book. Technical report, Basel Committee on Banking Supervision, 2012.

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

D. Duffie and J. Pan. An overview of value at risk. *Journal of Derivatives*, Spring:7–48, 1997.

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

G. Freud. *Orthogonal Polynomials*. Pergamon Press, 1971.

M. B. Gordy. Saddlepoint approximation of CreditRisk+. *Journal of Banking and Finance*, 26:1335–1353, 2002.

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

P. Hall and B. Presnell. Density estimation under constraints. *Journal of Computational and Graphical Statistics*, 8(2):259–277, 1999.

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

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

P. Jorion. *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, 3rd edition, 2007.

M. Junker and A. May. Measurement of aggregate risk with copulas. *Econometric journal*, 8:428–454, 2005.

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

K. Marumo. *Expansion Methods Applied to Distributions and Risk Measurement in Financial Markets*. PhD thesis, School of Economics and Finance, Queensland University of Technology, Australia, 2007.

K. Marumo and R. Wolff. Expansion methods applied to asset return distributions. *Journal of Risk*, 10(2):3–24, 2007.

I. Mauleón and J. Perote. The ability of multivariate Edgeworth-Sargan density capturing financial data behaviour. SSRN, 2000.

J. Mina and J. Y. Xiao. Return to RiskMetrics: The evolution of a standard. Technical report, RiskMetrics, 2001.

F. P. Natale. Optimization with tail-dependence and tail risk: A copula based approach for strategic asset allocation. *SSRN Working Paper*, November 2006.

R. B. Nelsen. *An Introduction to Copulas*. Springer, 1999.

G. Szegö. *Orthogonal Polynomials*. American Mathematical Society, 4 edition, 1975.

Y. Takahashi. *Jitsu Kansuu to Furie Kaiseki, (Real Functions and Fourier Analysis)*. Iwanami, Japan, 2006.

Y. Xu. Multivariate orthogonal polynomials and operator theory. *Transactions of the American Mathematical Society*, 343(1):193–202, May 1994.

Y. Xu. Christoffel functions and Fourier series for multivariate orthogonal polynomials. *Journal of Approximation Theory*, 82:205–239, 1995.