

Kohei MARUMO* and Rodney C. WOLFF†

On optimal smoothing of density estimators obtained from orthogonal polynomial expansion methods‡

July 9, 2015

Abstract

We discuss the application of orthogonal polynomial to estimation of probability density functions, particularly for accessing features of a portfolio's profit/loss distribution. Such expansions are given by the sum of known orthogonal polynomials multiplied by an associated weight function.

However, naïve applications of expansion methods are flawed. The shape of the estimator's tail can undulate, under the influence of the constituent polynomials in the expansion, and can even exhibit regions of negative density.

This paper presents techniques to redeem these flaws and to improve quality of risk estimation. We show that by targeting a smooth density which is sufficiently close to the target density, we can obtain expansion-based estimators which do not have the shortcomings of equivalent naïve estimators. In particular, we apply optimisation and smoothing techniques which place greater weight on the tails than the body of the distribution.

Numerical examples using both real and simulated data illustrate our approach. We further outline how our techniques can apply to a wide class of expansion methods, and indicate opportunities to extend to the multivariate case, where distributions of individual component risk factors in a portfolio can be accessed for the purpose of risk management.

Keywords: Delta-Gamma-Vega-Normal model; Expected Shortfall; Hermite polynomials; Historical Simulation method; Value at Risk.

*WH Bryan Mining and Geology Research Centre, The University of Queensland, Brisbane QLD 4072, AUSTRALIA, and Faculty of Economics, Saitama University, JAPAN. Email: marumo@mail.saitama-u.ac.jp Corresponding author.

†WH Bryan Mining and Geology Research Centre, The University of Queensland, Brisbane QLD 4072, AUSTRALIA. Email: rodney.wolff@uq.edu.au.

‡ *Declarations of Interest:* KM was funded in part by Japan Society for the Promotion of Science KAKENHI (grant number 25380387) and by a Visiting Scholars grant from The University of Queensland. RCW was funded in part by the Australian Research Council under its Discovery Project scheme (file number DP1096326). The authors otherwise report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

1 Introduction

The measurement of risk in financial portfolios is critical in the current regulatory (Basel Committee, 2013) and operating (McAleer et al., 2013) environments. From a statistical point of view, methods of quantifying such risk amount to estimating quantiles, or similar, of the profit or loss (PL) distribution for the portfolio. Specifically, the two most commonly used risk measures, Value at Risk (VaR) and Expected Shortfall (ES), are defined, respectively, as a quantile of the PL distribution, and the expected loss given that the loss exceeds the VaR.

In practice, methods for risk measurement can be classified into two categories. One is the use of parametric models for risk factors, and the other is the non-parametric methods, typically the Historical Simulation (HS) method. See, for example, Jorion (2007), Mina and Xiao (2001) and Duffie and Pan (1997). A range of academic literature is devoted to discussing parametric models for expressing the dynamics of risk factors and their distributions, as well as methods for estimating the relevant parameters, while the articles discussing HS and other non-parametric methods are relatively limited. This is an interesting contrast to the strong popularity of the HS method among commercial banks pointed out by Pérignon and Smith (2010). They attribute this popularity of the HS method over the parametric methods to the following: Firstly, risk aggregation is much easier with the HS method than with parametric models; secondly, the HS method, which implicitly uses the empirical distribution of the risk factors, is less likely to be exposed to large estimation error and model risk; and third, since the HS method is concerned with the unconditional distribution of risk factors, the risk measures calculated using the HS method are less volatile than when using parametric models that consider the conditional distributions. The third point can be understood as a dislike of the practical difficulty in allocating the capital to a volatile risk measure, even when the underlying model is considered to be ‘good’ in some sense.

Despite its popularity, however, the HS method can suffer from insufficiency of historical observations, especially in the tails of the distribution, where the observations are sparse. This often causes striking discreteness in the empirical distribution function which is undesirable in risk measurement. Hence, we desire a smooth function that approximates the empirical distribution function. This is one of the motivations in this paper.

Having mentioned the popularity of the HS method, we are aware of the importance of the parametric methods. For instance, some financial products may not have enough historical data for the HS method, due to the illiquidity of the market or short history of the products. In such cases, we may need to turn to parametric models that allow us to fill the insufficiency in the information by model assumptions. Further, it is often pointed out that the HS method does not have a solution for the time aggregation problem: this problem can be addressed by using parametric models.

Among many parametric models, the use of the Normal distribution is still common in practice¹, despite its shortcomings such as the thin tails pointed

¹In fact, many recent text books mention the use of the Normal distribution. See, for

out by many authors. The reason can be its tractability; rough approximations for risk aggregations can be obtained by the Delta-Normal method.

In the present paper, we examine Hermite expansions as a method which not only can produce a smooth function that approximates the empirical distribution function but is also capable of approximating the distribution function in the case of parametric models. Specifically, they can be used for approximating empirical distributions as well as distributions from parametric models, such as the Delta-Gamma-Vega-Normal model. Thus, Hermite expansion can do most of the things that the Normal approximations can do, and moreover, it is capable of handling heavy tails. Details will be reviewed in the following sections.

There are other competitive methods besides Hermite expansions which deserve comment. If our purpose is limited to smoothing empirical distribution functions then the first candidate is the kernel methods (Silverman, 1986). Its unfavourable characteristics, however, include the difficulty in dealing with the bias-variance tradeoff and the heavy computational load in calculating quantiles. These seriously inhibit its successful application to risk measurement.

The method based on classical extreme value theory is often discussed in the context of risk measurement (McNeil et al., 2005, chap.7). Its premise is that the tails of most of the distributions can be approximated by the generalised Pareto distribution, and thus the tails can be approximated by fitting the distribution. It is not necessarily straightforward to apply this idea to the risk aggregation. The tail of the aggregated risk can be largely affected by the bodies of individual risk factors, while the extreme value theory is focused on dealing with the tails of distributions, ignoring the bodies.

If we are to concentrate on only the first four moments — the mean, variance, skewness and kurtosis — Simonato (2011) and Leccadito et al. (2014) propose the use of the Johnson distribution and a non-linearly transformed Normal variable, respectively. They demonstrate that their methods can reproduce risk measures close to those of the jump-diffusion model, which is chosen as the target. Their results encourage us to limit the input from the market to the first four moments and to use their methods, when we know that the market dynamics is reliably described by the jump-diffusion model. However, for other general situations where we do not know the market dynamics, their results suggest that we need more information in order to obtain more reliable estimators for the risk measures. In particular, Simonato (2011, Exhibit 10) points out that the risk measures calculated by the Gram-Charlier series are different from the target by more than 10%, given the same first four moments. Hence, risk measures can vary by more than 10% depending on the method employed. Further information such as higher order moments might reduce this variation. With respect to this point, methods using the Hermite expansions are capable of handling the higher order moments.

The Hermite expansion expresses a probability density function (pdf) as an infinite sum of Hermite polynomials multiplied by an associated weight function, assuming that the infinite sum converges. See Szegő (1975), Jackson

example, Baker and Filbeck (2014, p.267), Bessis (2015, chap.13). Barrieu and Scandolo (2015) also mention the use of the Delta-Normal method in practice.

(1963), Freud (1971) and Takahashi (2006) for the theoretical background of Hermite expansions. Loosely speaking, the Hermite expansion modifies a Normal approximation to the target distribution by including successive adjusting terms. Under expectation, the Hermite polynomials in those terms are linear combinations of moments of the target distribution, assuming that they exist. The validity of expansion methods depends only on the existence of moments of the target distribution up to some finite degree.

One of the biggest drawbacks of Hermite expansion methods, however, is poor approximation quality. The presence of polynomials in the approximations usually leads to oscillation of the density estimators in the tails, including exhibiting regions of negative density. This, of course, is a serious issue for extreme quantile estimation, and not least is statistically nonsensical. Further, the infinite sum can be divergent, or convergence can be very slow, and therefore approximation quality can be sensitive to the choice of the order of expansion. Due to these kinds of 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.

In the present paper, we introduce techniques to improve the approximation quality of expansion methods, in terms of addressing both tail behaviour and convergence. We present an automated procedure which jointly optimises (a) the fit of the distribution, where greater weight is put on the tails than the body of the distribution, and (b) smoothness of the estimator, in terms of curvature. We present numerical examples which show the effectiveness of these methods, even in adverse cases, using both simulated and real data. Our approach is a unified method for both smoothed approximations for empirical distributions (of particular relevance to the HS method) and for distributions from parametric models. This has application in such situations as where a bank wishes to use a parametric model for measuring the risk of a certain asset class — in cases where sufficient historical data are not available — and to incorporate it into HS-method-based risk measurement.

We acknowledge that greater insight would be obtained by focusing on multivariate estimation. In the case of PL distributions, obtaining the joint distribution of possibly dependent and non-linear risk factors would give far greater insight into how risk is aggregated across component risk factors to render PL risk. For simplicity of the present exposition, though, we focus on the univariate case, and note that the extension to multivariate estimation is described in an as yet unpublished manuscript by Marumo and Wolff (2015).

We further point out that our treatment can easily apply to other expansion methods, such as the Laguerre expansion, whose weight function is the gamma distribution pdf, and the Chebychev expansion, whose weight function is the uniform distribution pdf. Thus, respectively, our method can deal with distributions with semi-infinite and bounded support, as well as doubly-infinite support. We expect that these expansions can be applied to credit and operational risk measurement, and to firm-wide risk aggregation. Again, for simplicity of exposition, we focus on the Hermite expansion.

The structure of this paper is as follows. In Section 2, we review the Hermite expansion for univariate cases. We describe in Section 3 the methodology to

improve the approximation quality, in terms of targeting the fit in the tails and optimising smoothness of the estimator. Numerical examples are presented in Section 4 and an exposition in the context of risk management is in Section 5. Concluding remarks are given in Section 6.

2 Background

2.1 Hermite expansions

The series of Hermite polynomials $\{\text{He}_k(x)\}$ is defined as

$$\text{He}_k(x) = \frac{1}{\sqrt{k!}} e^{\frac{x^2}{2}} \frac{d^k}{dx^k} e^{-\frac{x^2}{2}}, \quad k = 0, 1, 2, \dots \quad (1)$$

The most important property of this series is that it satisfies

$$\int_{-\infty}^{\infty} \phi(x) \text{He}_k(x) \text{He}_\ell(x) dx = \begin{cases} 0 & (k \neq \ell) \\ 1 & (k = \ell) \end{cases}, \quad (2)$$

where $\phi(x) = e^{-x^2/2}/\sqrt{2\pi}$, the pdf of the standard Normal distribution, is the associated weight function. This property is called *orthonormality with respect to ϕ* . There are some variations for the definition of the Hermite polynomials, including $\text{He}_k(x) = e^{x^2} \frac{d^k}{dx^k} e^{-x^2}$, which are orthogonal but may not be necessarily orthonormal. We employ the one expressed in Equation (1) for simplicity, but parallel discussions can be made with other definitions.

Now consider a pdf f . Assume that f satisfies

$$\int_{-\infty}^{\infty} \frac{\{f(u) - \phi(u) \sum_{k=0}^{\infty} c_k \text{He}_k(u)\}^2}{\phi(u)} du = 0 \quad (3)$$

for some real coefficients $\mathbf{c} = (c_0, c_1, \dots)$. Condition (3) is satisfied if and only if f is bounded and

$$f(x) = \phi(x) \sum_{k=0}^{\infty} c_k \text{He}_k(x), \quad (4)$$

for almost everywhere x , using a standard Lebesgue-type argument. For such f that satisfies Equation (3), we call the right hand side of Equation (4) the *Hermite expansion of f* , and denote it by $f(\cdot|\mathbf{c})$.

For a real function f , the infinite sum in $f(x|\mathbf{c})$ is known to be convergent for every x if

$$\int_{-\infty}^{\infty} \frac{\{f(u)\}^2}{\phi(u)} du < \infty, \quad (5)$$

that is, if $f(x)/\sqrt{\phi(x)}$ is square integrable. See, for instance, Takahashi (2006, p.243). 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 the Hermite expansion $f = f(\cdot|\mathbf{c})$ and that

$$\int_{-\infty}^{\infty} \frac{\{f(u)\}^2}{\phi(u)} du = \sum_{k=0}^{\infty} c_k^2 < \infty, \quad (6)$$

which corresponds to the Parseval identity for the Hermite system. See again Takahashi (2006, p.244).

For pdfs f whose Hermite expansion is convergent, the coefficients $\{c_k\}$ can be determined using Equation (2):

$$c_k = \int_{-\infty}^{\infty} \text{He}_k(u) f(u) du, \quad k = 0, 1, \dots \quad (7)$$

Note that $\text{He}_0(x) \equiv 1$ and therefore $c_0 = 1$ is required so that $\int_{-\infty}^{\infty} f(u|\mathbf{c}) du = 1$ is satisfied.

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

$$f(x) \simeq f_n(x|\mathbf{c}_n) = \phi(x) \sum_{k=0}^n E(\text{He}_k(X)) \text{He}_k(x), \quad (8)$$

where $\mathbf{c}_n = (E(\text{He}_0(X)), \dots, E(\text{He}_n(X)))$. We call such approximation methods based on Equation (8) *expansion methods*. As noted in the Introduction, besides using the Hermite expansion, the Laguerre expansion is sometimes applied to densities with non-negative support. See Marumo and Wolff (2007) and Marumo (2007).

2.2 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 naïve 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, expansion methods allow the approximations to take negative values. Such negative density can become large, especially in the tails, where the pdf is close to 0.

For cases where the Hermite expansion is convergent, and where convergence is reasonably fast, we increase 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) study 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 remainder of this section, we firstly consider these difficulties carefully using an example where we approximate the empirical distribution, and then introduce techniques to deal with them.

2.3 Naïve application to approximating empirical distributions

We outline here the mathematical issues which determine convergence or otherwise of Hermite expansions. These will guide us in constructing our methodology in the next section.

Assume that we have i.i.d. samples X_1, \dots, X_N from a distribution with an unknown pdf f . A naïve idea would be to use the sample moments to obtain the coefficients for the expansion. That is, we use the coefficients

$$\hat{c}_k = \frac{1}{N} \sum_{i=1}^N \text{He}_k(X_i), \quad k = 0, \dots, n$$

to obtain the estimator of f ,

$$f_n(x|\hat{\mathbf{c}}_n) = \phi(x) \sum_{k=0}^n \hat{c}_k \text{He}_k(x), \quad (9)$$

where $\hat{\mathbf{c}}_n = (\hat{c}_0, \dots, \hat{c}_n)$. 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 can be expressed using the Dirac delta function δ as

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

Since $f^N(x)/\sqrt{\phi(x)}$ is not square integrable, the expansion in Equation (9) is divergent as $n \rightarrow \infty$. 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 convergent Hermite expansion. Using Equations (1) and (9), we have

$$F_n(x|\hat{\mathbf{c}}_n) = \int_{-\infty}^x f_n(u|\hat{\mathbf{c}}_n) du = \Phi(x) + \phi(x) \sum_{k=1}^n \frac{\hat{c}_k}{\sqrt{k}} \text{He}_{k-1}(x), \quad (10)$$

where Φ is the df of the standard Normal distribution. Note that $c_0 = 1$. This corresponds to the naïve approximation for the empirical df.

For the empirical distribution, we have

$$F^N(x) = \int_{-\infty}^x f^N(u) du = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{X_i \leq x\}},$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function. Now

$$I_0 = \int_{-\infty}^{\infty} \{F^N(u) - \Phi(u)\}^2 / \phi(u) du < \infty. \quad (11)$$

To see this, split the integral into three parts: $(\int_{-\infty}^{-a} + \int_{-a}^a + \int_a^{\infty}) \{F^N(u) - \Phi(u)\}^2 / \phi(u) du$, for some large a . It is trivial to show that the second integral

is bounded. For large enough a , we have $F^N(x) = 0$ for $x \leq -a$. Since $\Phi(x) - \phi(x)$ is decreasing for $x \leq -1$ with $\lim_{x \rightarrow -\infty} (\Phi(x) - \phi(x)) = 0$, we have $\Phi(x) < \phi(x)$ for $x \leq -1$. Therefore, $\int_{-\infty}^{-a} \{0 - \Phi(u)\}^2 / \phi(u) du < \int_{-\infty}^{-a} \phi(u) du = \Phi(-a) < \infty$ for large a . This proves that the first integral is bounded. By symmetry, the third integral can be shown to be bounded. Consequently, $F^N(x) - \Phi(x)$ has a convergent Hermite expansion of the form

$$F^N(x) - \Phi(x) = \phi(x) \sum_{k=1}^{\infty} d_k \text{He}_{k-1}(x), \quad (12)$$

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

By applying Equation (7) to $F^N(x) - \Phi(x)$, for fixed N , we have that

$$\begin{aligned} d_k &= \int_{-\infty}^{\infty} \text{He}_{k-1}(u) \{F^N(x) - \Phi(x)\} du \\ &= \int_{-\infty}^{\infty} \text{He}_{k-1}(u) \frac{1}{N} \sum_{i=1}^N \{\mathbf{1}_{\{X_i \leq u\}} - \Phi(u)\} du \\ &= \frac{1}{N} \sum_{i=1}^N \int_{-\infty}^{\infty} \text{He}_{k-1}(u) \{\mathbf{1}_{\{X_i \leq u\}} - \Phi(u)\} du, \end{aligned}$$

where Fubini's condition

$$\int_{-\infty}^{\infty} \sum_{i=1}^N \left| \text{He}_{k-1}(u) \sum_{i=1}^N \{\mathbf{1}_{\{X_i \leq u\}} - \Phi(u)\} \right| du < \infty$$

is satisfied. Now

$$\begin{aligned} &\int_{-\infty}^{\infty} \text{He}_{k-1}(u) \{\mathbf{1}_{\{X_i \leq u\}} - \Phi(u)\} du \\ &= \int_{-\infty}^{X_i} \text{He}_{k-1}(u) \{-\Phi(u)\} du + \int_{X_i}^{\infty} \text{He}_{k-1}(u) \{1 - \Phi(u)\} du \\ &= \left[\frac{\text{He}_k(u)}{\sqrt{k}} \Phi(u) \right]_{-\infty}^{X_i} - \int_{-\infty}^{X_i} \frac{\text{He}_k(u)}{\sqrt{k}} \phi(u) du \\ &\quad + \left[-\frac{\text{He}_k(u)}{\sqrt{k}} \{1 - \Phi(u)\} \right]_{X_i}^{\infty} - \int_{X_i}^{\infty} \frac{\text{He}_k(u)}{\sqrt{k}} \phi(u) du \\ &= \frac{\text{He}_k(X_i)}{\sqrt{k}} - \int_{-\infty}^{\infty} \frac{\text{He}_k(u)}{\sqrt{k}} \phi(u) du \\ &= \frac{\text{He}_k(X_i)}{\sqrt{k}} - 0, \end{aligned}$$

where we use the identity $\text{He}_{k-1}(u) = -\text{He}'_k(u) / \sqrt{k}$. These verify that

$$F^N(x) = \Phi(x) + \phi(x) \sum_{k=1}^{\infty} \frac{\hat{c}_k}{\sqrt{k}} \text{He}_{k-1}(x) = \lim_{n \rightarrow \infty} F_n(x | \hat{\mathbf{c}}_n). \quad (13)$$

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

$$I_0 = \int_{-\infty}^{\infty} \{F^N(u) - \Phi(u)\}^2 / \phi(u) du = \sum_{k=1}^{\infty} \frac{\hat{c}_k^2}{k} < \infty,$$

whereas $\sum_{k=0}^{\infty} \hat{c}_k^2$ is divergent, as shown previously. Further, it can be easily verified that the choice of coefficients $d_k = \hat{c}_k / \sqrt{k}$, $k = 1, \dots, n < \infty$ minimises the WISD

$$\int_{-\infty}^{\infty} \left\{ \Phi(u) + \phi(u) \sum_{k=1}^n d_k \text{He}_{k-1}(u) - F^N(u) \right\}^2 / \phi(u) du.$$

Therefore, the naïve approximation $F_n(\cdot | \hat{\mathbf{c}}_n)$ is the best approximation of F^N within $\{F_n(\cdot | \mathbf{c}_n); \mathbf{c}_n \in \mathbb{R}^n\}$ in the sense that it has the smallest WISD.

3 Methodology

In the previous section, we recalled naïve approximations for empirical distributions and their convergence properties: the approximation for dfs is convergent, while the approximation for pdfs is divergent. As we shall 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. An intuitive reason for this slow convergence is the large order of the polynomial required to approximate functions with discontinuities.

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. Standardisation is a very common technique, though.

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.

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.

3.1 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 is 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, respectively. Actually, this technique is used, sometimes implicitly, in most existing applications.

Let $\mu = E(X)$, $\sigma^2 = E(X^2) - \mu^2$, $X' = (X - \mu)/\sigma$ and $f^{X'}$ be the pdf of X' . We apply the Hermite expansion in Equation (8) to X' to obtain its approximation $f_n^{X'}(\cdot|\mathbf{c}_n)$. The pdf of X can be approximated by $f_n(x|\mathbf{c}_n) = f_n^{X'}\left(\frac{x-\mu}{\sigma}|\mathbf{c}_n\right)/\sigma$.

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 *naïve expansion*.

3.2 Smoothing

As discussed above, poor approximation quality can be due to the divergence or slow convergence of the Hermite expansion. Instead of approximating the target distribution itself, we change our target to a smooth function which is close to the original target. We expect that the convergence is accelerated by this change.

One way of doing this is to use the curvature of the target function; that is, we can seek 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^N in Equation (13). Assume that a function $F(\cdot|\mathbf{c}^S)$ has the convergent Hermite expansion

$$F(x|\mathbf{c}^S) = \Phi(x) + \phi(x) \sum_{k=1}^{\infty} \frac{c_k^S}{\sqrt{k}} \text{He}_{k-1}(x),$$

for some real coefficients $\mathbf{c}^S = (c_0^S, c_1^S, \dots)$, and is near F^N , in the sense that the weighted integrated squared difference (WISD),

$$\int_{-\infty}^{\infty} \{F(u|\mathbf{c}^S) - F^N(u)\}^2 / \phi(u) du,$$

is small. Using Equation (1), the second derivative of $F(\cdot|\mathbf{c}^S)$ is given by

$$F''(x|\mathbf{c}^S) = \phi(x) \sum_{k=0}^{\infty} \sqrt{k+1} c_k^S \text{He}_{k+1}(x).$$

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

$$\int_{-\infty}^{\infty} \{F''(u|\mathbf{c}^S)\}^2 / \phi(u) du,$$

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 both the difference and curvature in the tails than those of the centre of the distribution. The other is that it makes the integrals tractable. In fact, using the Parseval identity, we have

$$\text{WISD} = \sum_{k=1}^{\infty} \frac{(c_k^S - \hat{c}_k)^2}{k} \quad \text{and} \quad \text{WISC} = \sum_{k=0}^{\infty} (k+1)(c_k^S)^2.$$

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

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

where the weight q determines the relative importance of curvature over fidelity. Then I_q^S is minimised when

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

Thus, we use these values to obtain the finite expansion

$$F_n(x|\hat{\mathbf{c}}_n^S) = \Phi(x) + \phi(x) \sum_{k=1}^n \frac{\hat{c}_k^S}{\sqrt{k}} \text{He}_{k-1}(x).$$

Let us call this expansion the smoothed approximation for F^N .

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

$$\int_{-\infty}^{\infty} \{F_n(u|\hat{\mathbf{c}}_n^S) - \Phi(u)\}^2 / \phi(u) du = \sum_{k=1}^n \frac{(\hat{c}_k^S)^2}{k}$$

is much faster than that of $F_n(\cdot|\hat{\mathbf{c}}_n)$; that is, the convergence $F_n(\cdot|\hat{\mathbf{c}}_n^S) \rightarrow F(\cdot|\mathbf{c}^S)$ is much faster than that of $F_n(\cdot|\hat{\mathbf{c}}_n) \rightarrow F^N$. Clearly, $F(\cdot|\mathbf{c}^S)$ is not the target itself, but we expect that it is a smooth function near F^N , especially when q is small. The choice of q will be discussed briefly in Section 3.8.

Other than its fast convergence, $F_n(\cdot|\hat{\mathbf{c}}_n^S)$ has a favourable property. The derivative of $F_n(\cdot|\hat{\mathbf{c}}_n^S)$ is given by

$$\frac{d}{dx} F_n(x|\hat{\mathbf{c}}_n^S) = f_n(x|\hat{\mathbf{c}}_n^S) = \phi(x) \sum_{k=0}^n \hat{c}_k^S \text{He}_k(x),$$

and we have that

$$\int_{-\infty}^{\infty} \frac{\{f_n(u|\hat{\mathbf{c}}_n^S)\}^2}{\phi(u)} du = \sum_{k=0}^n (\hat{c}_k^S)^2. \quad (14)$$

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 Section 3.4.

3.3 Optimisation

Similarly to the previous example, consider a situation where we use a random sample X_1, \dots, X_N from the unknown pdf f to obtain the natural estimators $\hat{c}_k = \sum_{i=1}^N \text{He}_k(X_i)/N$. We further assume that f has the converging Hermite expansion $f = f(\cdot|\mathbf{c})$, where $\mathbf{c} = (c_0, c_1, \dots)$ are not known.

We consider a class of estimator

$$f_n(x|\boldsymbol{\alpha}_n, \hat{\mathbf{c}}_n) = \phi(x) \sum_{k=0}^n \alpha_k \hat{c}_k \text{He}_k(x), \quad (15)$$

and search for the coefficients $\boldsymbol{\alpha}_n = (\alpha_0, \dots, \alpha_n)$, so that the estimator of the weighted mean integrated square error (WMISE)

$$\mathbb{E} \left(\int_{-\infty}^{\infty} \frac{\{f_n(u|\boldsymbol{\alpha}_n, \hat{\mathbf{c}}_n) - f(u|\mathbf{c})\}^2}{\phi(u)} du \right) \quad (16)$$

is minimised. (Hall (1983) considers a different expansion formula, in which the MISE is not weighted.) The weight $1/\phi$ in Expression (16) works in two ways, similarly to the previous example: one is to put more importance on the error in the tail than at the centre of the distribution, and the other is that it makes Expression (16) tractable. It can be shown by using the Parseval identity that the WMISE in Expression (16) is equal to

$$\sum_{k=0}^n \alpha_k^2 \mathbb{E}(\hat{c}_k^2) - 2 \sum_{k=0}^n \alpha_k c_k^2 + \sum_{k=0}^{\infty} c_k^2. \quad (17)$$

Now we estimate the WMISE in Expression (17). Obviously, \hat{c}_k^2 is an unbiased estimator of $\mathbb{E}(\hat{c}_k^2)$. An unbiased estimator for c_k^2 is given by

$$\frac{N\hat{c}_k^2 - \hat{b}_k^2}{N-1},$$

where $\hat{b}_k^2 = N^{-1} \sum_{i=1}^N \{\text{He}_k(X_i)\}^2$, however, this estimator can take negative values, while the true value of c_k^2 is non-negative. Therefore, we use a biased estimator

$$\frac{(N\hat{c}_k^2 - \hat{b}_k^2)_+}{N-1},$$

where $(x)_+ = \max\{x, 0\}$, instead. We consider this estimator to be better, in that it is non-negative.

Hence, an estimator for the WMISE is given by

$$\sum_{k=0}^n \alpha_k^2 \hat{c}_k^2 - 2 \sum_{k=0}^n \alpha_k \frac{(N\hat{c}_k^2 - \hat{b}_k^2)_+}{N-1} + \sum_{k=0}^{\infty} \frac{(N\hat{c}_k^2 - \hat{b}_k^2)_+}{N-1}. \quad (18)$$

Now we consider $\boldsymbol{\alpha}_n$ which minimises Expression (18). Firstly, $\alpha_0 = 1$ is required so that $\int_{-\infty}^{\infty} f_n(u|\boldsymbol{\alpha}_n, \hat{\mathbf{c}}_n) du = 1$ 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 $k = 3, \dots, n$, Expression (18) is minimised when

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

Note that this method can be regarded as implicitly using shrinkage estimators.

3.4 Optimisation with smoothing

For the approximation in the form of Equation (15), the second derivative is derived using Equation (1) as

$$f_n''(x|\boldsymbol{\alpha}_n, \hat{\mathbf{c}}_n) = \frac{d^2}{dx^2} f_n(x|\boldsymbol{\alpha}_n, \hat{\mathbf{c}}_n) = \phi(x) \sum_{k=0}^n \sqrt{(k+1)(k+2)} \alpha_k \hat{c}_k \text{He}_{k+2}(x). \quad (19)$$

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

$$\mathbb{E} \left(\int_{-\infty}^{\infty} \frac{\{f_n''(u|\boldsymbol{\alpha}_n, \hat{\mathbf{c}}_n)\}^2}{\phi(u)} du \right),$$

which can be simplified to

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

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

$(1-q)\text{WMISE} + q\text{WMISC}$

$$= \sum_{k=0}^n \alpha_k^2 \{1 - q + q(k+1)(k+2)\} \hat{c}_k^2 - 2 \sum_{k=0}^n \alpha_k \frac{\left(N\hat{c}_k^2 - \hat{b}_k^2\right)_+}{N-1} + \sum_{k=0}^{\infty} \frac{\left(N\hat{c}_k^2 - \hat{b}_k^2\right)_+}{N-1}, \quad (20)$$

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

From a similar discussion to that in Section 3.3, 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 a factor of order k^2 in the denominator of this smoothed α_k , the sum

$$\sum_{k=0}^n (\alpha_k \hat{c}_k)^2 = \int_{-\infty}^{\infty} \frac{\{f_n(u|\boldsymbol{\alpha}_n, \hat{\mathbf{c}}_n)\}^2}{\phi(u)} du \quad (21)$$

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

3.5 Asymptotics in N

We make the following comments in relation to asymptotics in N . For

$$F^N(x) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{X_i \leq x\}}$$

we have that $F^N(x) \rightarrow F(x)$ in probability as $N \rightarrow \infty$, from the Law of Large Numbers, where F is the true df of X . Further, for fixed N ,

$$F_n(x|\hat{\mathbf{c}}_n) = \Phi(x) + \phi(x) \sum_{k=1}^n \frac{\hat{c}_k}{\sqrt{k}} \text{He}_{k-1}(x) \rightarrow F^N(x)$$

as $n \rightarrow \infty$, almost everywhere x . However, $F_n(x|\hat{\mathbf{c}}_n) \rightarrow F(x)$ is not true when

$$\sum_{k=1}^{\infty} \frac{\hat{c}_k^2}{k}$$

diverges or, equivalently, when

$$\int_{-\infty}^{\infty} \frac{\{F(x) - \Phi(x)\}^2}{\phi(x)} dx = \infty.$$

To summarise: $F_n^N \rightarrow F^N$ as $n \rightarrow \infty$, and $F^N \rightarrow F$ as $N \rightarrow \infty$, but $F_n^N \rightarrow F$ as $N \rightarrow \infty$ and $n \rightarrow \infty$ only when

$$\int_{-\infty}^{\infty} \frac{\{F(x) - \Phi(x)\}^2}{\phi(x)} dx < \infty.$$

3.6 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 a log Normal distribution, can be problematic, whereupon we might turn to Monte Carlo methods or a first order approximation. Expansion methods have the potential to be applied to such cases as long as the moments of the target distribution are available.

Indeed, we treat an example of the distribution of a call option premium in Section 5.

However, naïve applications of expansion methods are unlikely to work when either the target function is not bounded, or the target function has heavy tails. For the former, we can smooth the target, as considered for empirical distributions. For the latter, for instance, we can truncate the tails at some large but finite points, so that $f(x)/\sqrt{\phi(x)}$ is square integrable. One obvious drawback here is the challenge of deriving the truncated moments.

Developing techniques for applying expansion methods in such situations is our next task.

3.7 Choice of order of expansion n

Suppose that

$$f_n(x|\mathbf{c}_n) = \phi(x) \sum_{k=0}^n c_k \text{He}_k(x)$$

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

$$\int_{-\infty}^{\infty} \frac{\{f_n(u|\mathbf{c}_n) - f_{n-1}(u|\mathbf{c}_{n-1})\}^2}{\phi(u)} dx = c_n^2,$$

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 $f_n(\cdot|\mathbf{c}_n)$ 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, this idea is not practical since it requires evaluating an infinite sum. As an alternative, we observe the series c_0^2, \dots, c_N^2 for some large N , and see if c_{n+1}^2, \dots, c_N^2 are ‘small enough’ compared to c_0^2, \dots, c_n^2 . We expect such n can be found easily if convergence is fast.

3.8 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, we can obtain fair approximation quality for the range $q = 0.005 \sim 0.01$. The examples will further show that this range for q has a strong impact on our method, despite it being located apparently near to the boundary of the theoretical range for q .

4 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, in contradistinction, we consider three adverse cases and examine the success or otherwise of expansion methods. The first and second cases apply expansion methods to empirical distributions. The third case applies expansion methods to approximating the Gamma density as an example of application to a parametric model.

By way of metrics in our numerical studies, we use the popular risk measures used in the finance industry, VaR and ES (Basel Committee, 2013). As already mentioned, VaR is defined as a quantile of a financial assets distribution, and ES is defined as the expected loss given that the loss exceeds VaR, respectively. If the random variable Z denotes the financial asset in question, then the VaR with a confidence level of α is defined as

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

and the corresponding ES is defined as

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

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

$$-E(Z | Z \leq -\text{VaR}_{0.99}) = -\int_{-\infty}^{-\text{VaR}_{0.99}} u f_Z(u) du / 0.01. \quad (22)$$

For the cases where Z denotes loss, the VaR and ES can be defined as $\text{VaR}_\alpha = \inf\{z | P(Z \leq z) \geq \alpha\}$ and $\text{ES}_\alpha = E(Z | Z \geq \text{VaR}_\alpha)$, respectively. In the numerical examples, we use these definitions for the risk measures on the right tails.

4.1 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 long right tail and non-negative support. The shape and scale parameters are set to be 7 and 1, respectively, with a skewness of 0.756 and an excess kurtosis of 0.857. 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 (sample skewness of 0.915), and heavy-tailed (sample excess kurtosis of 1.447), compared to the Normal distribution. The order of expansion is $n = 10$, and the weight for smoothness is $q = 0.01$, where used. Figure 1 shows the shapes of the target pdf and df, and Table 1 compares the risk measures of VaR and ES, as discussed in Section 5, for the target distribution and those of approximation.

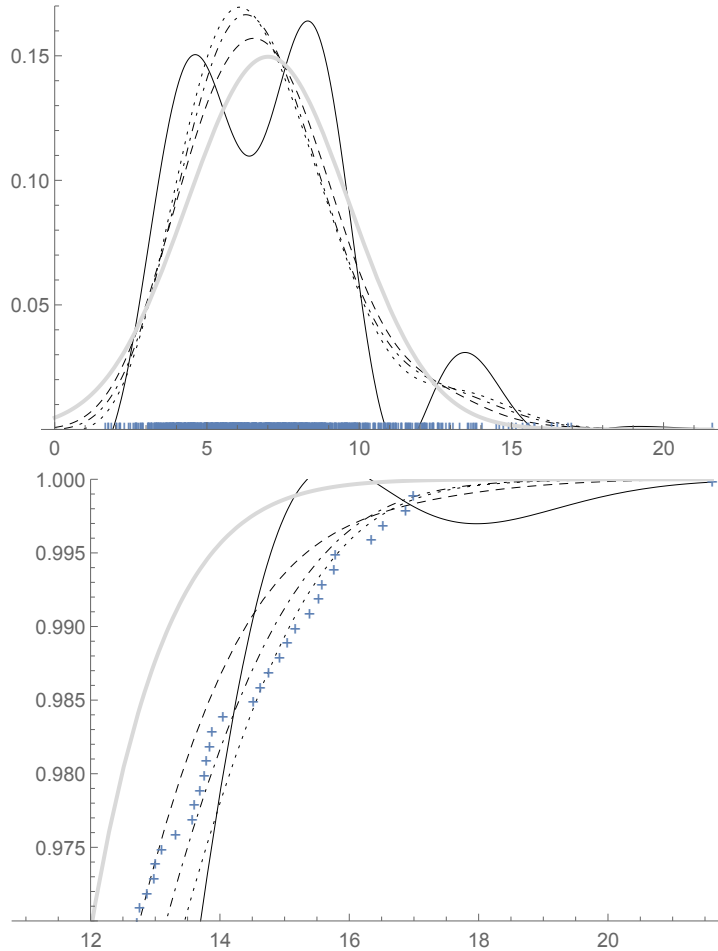


Figure 1: The pdfs (upper) and the magnified right tails of dfs (lower): the naïve expansion (solid curves), the smoothed expansion (dashed curves), the optimised expansion (dotted curves), and the optimised and smoothed expansion (dot-dashed curves), respectively. The order of expansions is $n = 10$ and the weight for smoothness is $q = 0.01$. The ‘+’ symbols show the 1,000 pseudo-random samples from the gamma distribution with shape parameter 7 and scale parameter 1. The thick grey curves show the pdf and df of the Normal approximation.

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

The minus signs in 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. We are aware that underestimations are unfavourable; however, considering the fact that the Normal distribution is still in use in practice,² the reduction in underestimation can be considered as a motivation for using expansion methods instead of Nor-

² See Barrieu and Scandolo (2015), Bessis (2015, chap.13) and Baker and Filbeck (2014, p.267).

Level	99%		99.5%	
	VaR	ES	VaR	ES
Empirical	15.16	16.10	15.78	17.35
Normal	-12.84%	-14.43%	-12.04%	-15.14%
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 absolute values of 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 $q = 0.01$. The minus signs in the body of the table indicate that the approximations underestimate the risk. Figure 1 shows that the naïve expansion exhibits negative density at around $x = 11$ and is not reliable at 99% and 99.5% levels, and hence there is no merit in reporting those results.

mal based methods. Further, we can comment that the approximations by expansions can approach the target as we increase n .

4.2 Empirical distribution from market observations

As a 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. The reason why we employ this example is that this data set is not close to the Normal distribution; it has sample skewness of -0.251 and sample excess kurtosis of 1.782 . We only comment here that the return series of S&P 500 itself is much closer to the Normal distribution and naturally all of its approximations performed well. 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 entirely make 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 $q = 0.01$.

This example shows similar features to the previous one: the naïve expansion shows negative density in the left tail, while others approximate the 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 Figure 3 show that the squared coefficients of naïve and optimised expansions do not seem to converge toward 0, at least up 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 can use such plots for determining the order of expansion. For instance, we can stop increasing the order before the squared coefficients are consistently below, say, 10^{-4} . Since the first few large values are $\sim 10^{-2}$, the

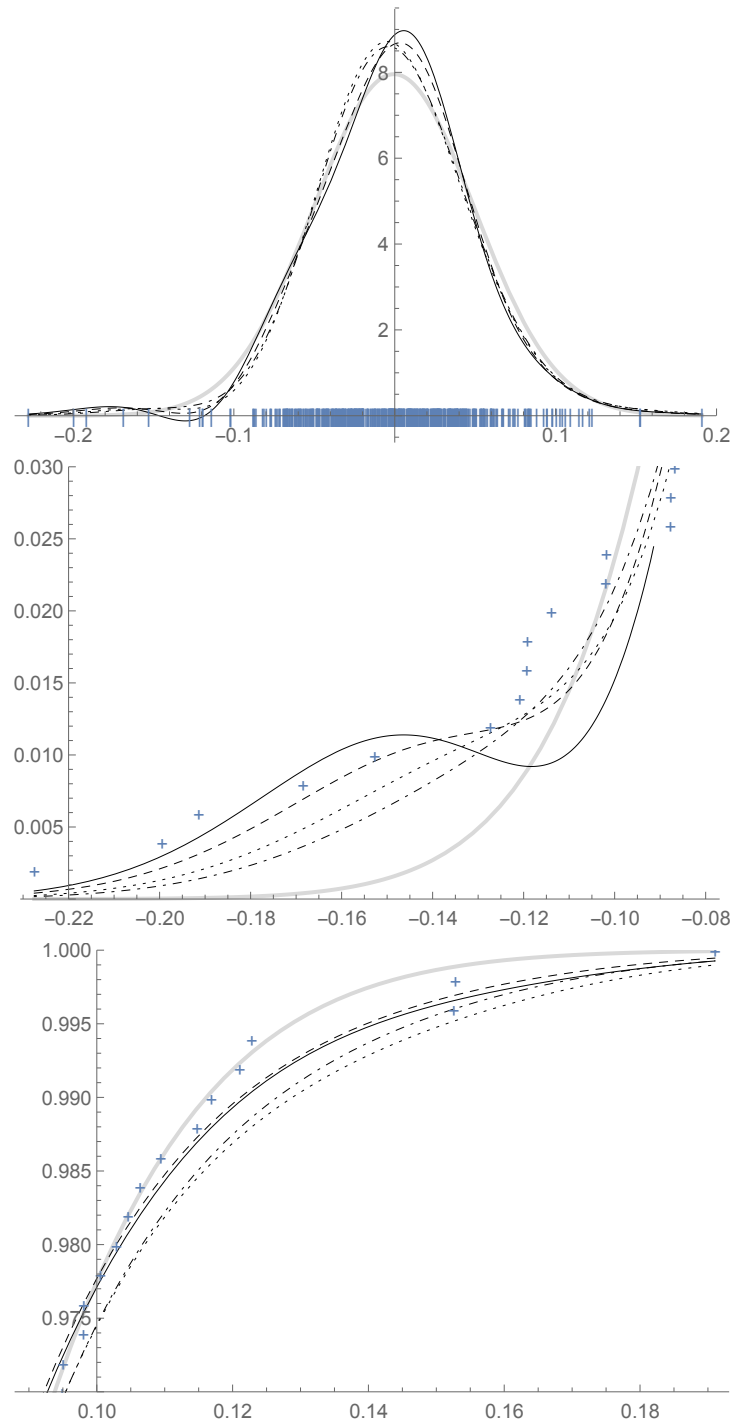


Figure 2: The pdfs (upper) and the magnified left and right tails of dfs (middle and lower, respectively): the naïve expansion (solid curves), the smoothed expansion (dashed curves), the optimised expansion (dotted curves), and the optimised and smoothed expansion (dot-dashed curves), respectively. The order of expansions is $n = 10$ and the weight for smoothness is $q = 0.01$. The ‘+’ symbols show the 500 samples of the daily log-difference of S&P 500 IV. The thick grey curves show the pdf and df of the Normal distribution.

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%
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%
naïve	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 absolute values of 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 $q = 0.01$. The minus signs in the body of the table indicate that the approximations underestimate the risk. Figure 2 shows that the naïve expansion exhibits negative density at around $x = -0.14$ and is not reliable at 0.5% and 1% levels, and hence there is no merit in reporting those results.

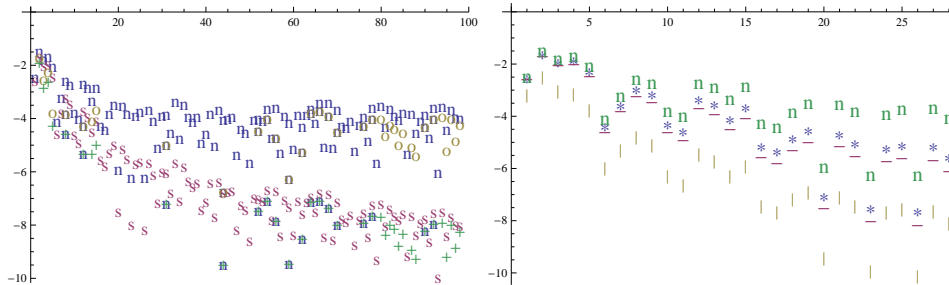


Figure 3: Squared coefficients of the expansions up to order 100 (left) and 30 (right). The vertical axes are in a log scale with base 10. The symbols in the left plots denote naïve (n), smoothed (s), optimised (o), and optimised and smoothed (+) expansions, respectively. The weight for smoothness is $q = 0.01$. The symbols in the right plots denote $q = 0.1$ (|), 0.01 (-), 0.005 (*), and 0 (n), in the smoothed expansions, respectively.

relative importance of squared coefficients that are below 10^{-4} is 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 naïve and optimised expansions.

4.3 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 approximated the empirical distribution. Due to the long right tail, this Gamma pdf weighted by $1/\sqrt{\phi(x)}$ is not square integrable, and therefore does not have a convergent 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 sufficiently far to the 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. Since the optimisation can be applied when we work on data sets, not on models, we only exhibit the results for naïve and smoothed expansions.

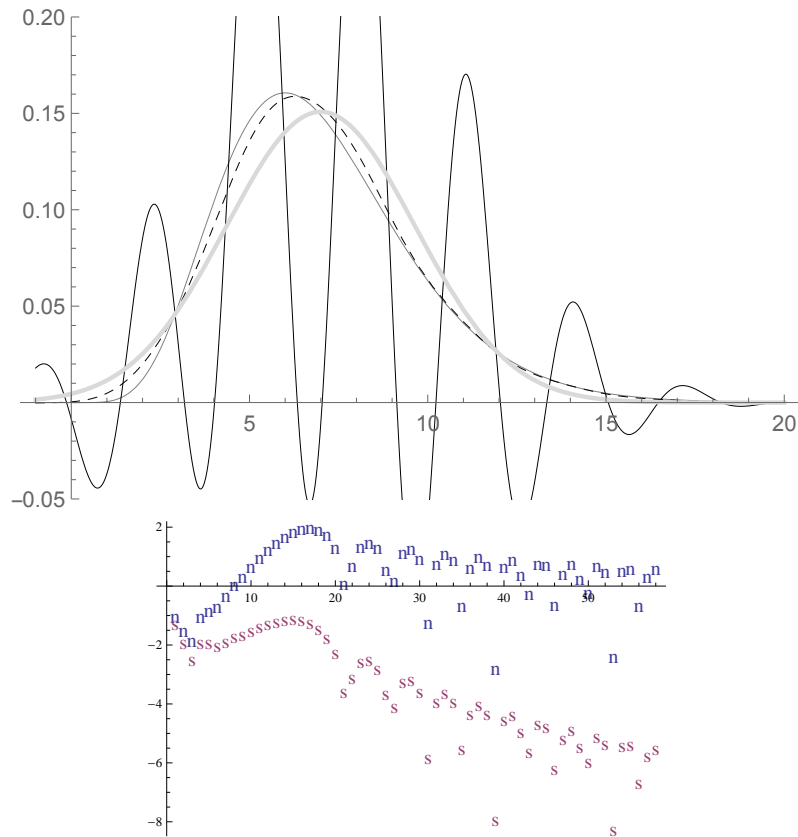


Figure 4: The pdfs (upper) and the squared coefficients $\{c_k^2\}$ up to order 60 (lower). In the upper plots, the truncated Gamma distribution (thin gray), naïve expansion (solid), the smoothed expansion (dashed), the Normal distribution (thick gray), are shown. The order of expansions is $n = 30$ and the weight for smoothness is $q = 0.005$. In the lower plots, the symbols ‘n’ denote the naïve 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.

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

Table 3: The absolute values of 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 $q = 0.005$. The minus signs indicate that the approximations underestimate the risk. Figure 4 shows that the naïve expansion exhibits erratic shape and is not reliable, and hence there is no merit in reporting those results.

The upper plots in Figure 4 show that the approximation quality of the naïve 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 lower plots in Figure 4, we find that the convergence of the naïve expansion is very slow, while the smoothed expansion converges much faster with smoothness weight $q = 0.005$. 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 obtaining the truncated moments may be a tricky task in some cases.

5 Application to risk measurement

5.1 Risk measurement and risk aggregation

We consider now VaR and ES in respect of obtaining risk measures for a PL distribution. Let Z denote the loss of the portfolio. 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, \dots, X_K , in such a way that Z , the PL of our portfolio, is expressed using a K variate function ψ as

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

Then the PL distribution function is given by

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

Even in those cases where we have complete information on the joint distribution of the risk factors, performing this integration can be a tricky task when ψ is not a linear function, or when X_1, \dots, X_K do not have the K -variate Normal

distribution. (Jorion (2007) describes current practice for coping with these problems.) Further, it is rather exceptional that the joint distribution of the risk factors is completely known.

5.2 Applying expansion methods

We demonstrate 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

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

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

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

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

exist and are available. We only require the availability of cross-moments, regardless of whether or not we know the joint distribution.

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

$$F_Z(z) \simeq F_n(z|\mathbf{c}_n) = \Phi(z) + \phi(z) \sum_{k=1}^n \frac{c_k}{\sqrt{k}} \text{He}_{k-1}(z)$$

for calculating the VaR. The pdf can be approximated by

$$f_Z(z) \simeq f_n(z|\mathbf{c}_n) = \phi(z) \sum_{k=0}^n c_k \text{He}_k(x),$$

and further, using Equation (1), we have

$$\begin{aligned} \int_{-\infty}^z u f_n(u|\mathbf{c}_n) du &= -\phi(z) + c_1 \{\phi(z)z - \Phi(z)\} \\ &+ \phi(z) \sum_{k=2}^n \frac{c_k}{\sqrt{k}} \left\{ z \text{He}_{k-1}(z) - \frac{1}{\sqrt{k-1}} \text{He}_{k-2}(z) \right\}, \end{aligned}$$

which can be used for calculating the ES: see Equation (22).

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

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

$$\begin{aligned} C(S, \sigma_I) &= S\Phi(d) - Ke^T\Phi(d - \sigma_I\sqrt{T}), \\ d &= \frac{\log(S/K) + (r + \sigma_I^2/2)T}{\sigma_I\sqrt{T}}, \end{aligned} \quad (23)$$

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 $0 < t \ll T$ ahead can be expressed as S_0e^X , and $\sigma_{I0}e^Y$, where (X, Y) has a bivariate Normal distribution, whose parameters are estimated from the historical observations. Then the PL of the option can be expressed as

$$\Delta C = C(S_0e^X, \sigma_{I0}e^Y) - C(S_0, \sigma_{I0}). \quad (24)$$

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

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

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

$$\Delta C \simeq \Delta C_{\delta\gamma\kappa} = \delta S_0 X + \frac{1}{2}(\delta S_0 + \gamma S_0^2) X^2 + \kappa \sigma_{I0} Y. \quad (25)$$

Even with the aid of this Taylor approximation, deriving the distribution of $\Delta C_{\delta\gamma\kappa}$ is not straightforward. However, the moments $\{E(\Delta C_{\delta\gamma\kappa}^k)\}$, $k = 0, 1, \dots$, can be obtained easily 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 a European call option (three months, at-the-money) on the S&P 500 for $t =$ one day. (See Table 5 for summary statistics of the data used. In this example, $t = 0$ is set to 18 Oct. 2006.).

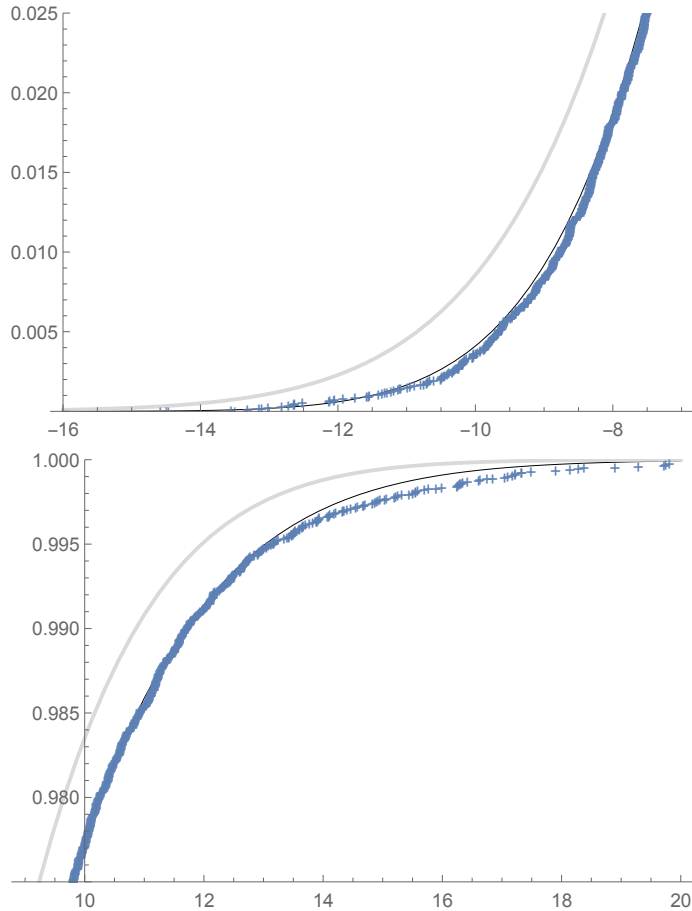


Figure 5: Magnified left and right tails of approximated dfs of $\Delta C_{\delta\gamma\kappa}$. The solid curve shows the naïve expansion, and the dotted line shows the smoothed expansion with the weight for smoothness $q = 0.005$. The order of the expansions is $n = 10$. The thick grey curve shows the approximation by the Normal distribution. The '+' symbols exhibit the distribution of ΔC obtained by applying the Monte Carlo method with 20,000 trials to Equation (24).

They show that the approximation by the naïve expansion is very close to the empirical distribution of the 20,000 pseudo-random samples from ΔC , and we find that the naïve expansion is sufficient in this particular case.

We comment that distributions other than bivariate normal distribution can be used as long as the moments and cross-moments are available.

6 Discussion

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

Smoothing and optimisation can make the approximation quality sufficient for many purposes in our particular examples. We proposed truncating the target distribution for the cases where the Hermite expansion is not conver-

Level	0.5%		1%	
	VaR	ES	VaR	ES
Empirical ($\times -1$)	9.676	10.806	8.777	9.998
Normal	12.11%	13.33%	11.00%	12.45%
naïve	0.91%	0.19%	1.24%	0.56%
Smoothed	1.65%	0.90%	1.91%	1.27%

Level	99%		99.5%	
	VaR	ES	VaR	ES
Empirical	11.679	13.838	13.077	15.355
Normal	-7.03%	-10.69%	-8.53%	-12.98%
naïve	0.18%	-2.21%	-0.36%	-3.83%
Smoothed	-0.21%	-2.66%	-0.78%	-4.32%

Table 4: The absolute values of 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 expansions is $n = 10$ and the weight for smoothness is $q = 0.005$. The minus signs in the body of the table indicate that the approximations underestimate the risk.

	SP500	SP500 IV
Observation period	From 25/10/2004 to 18/10/2006	
Number of observations	500	
Mean ($\times 10^{-4}$)	-4.425	-5.006
Std. dev. ($\times 10^{-2}$)	0.654	5.018
Skewness ($\times 10^{-1}$)	0.715	-2.510
Excess Kurtosis	0.376	1.782
Min ($\times 10^{-1}$)	-0.185	-2.275
Max ($\times 10^{-1}$)	0.213	1.910
Correl. coef.	-0.7443	

Table 5: Summary statistics of the data sets.

gent due to the heavy-tailedness; however, this may not be applicable when the truncated moments are not available. One possible solution to this is to optimise MISE subject to a constraint that the density is non-negative: see Hall and Presnell (1999).

By comparing the techniques, we found the following. For the cases where random samples are available, the optimised approximations performed better than the smoothed ones, in terms of their small errors from the target risk measures. This is due to the fact that the smoothing technique sacrifices

fidelity to the target in order to accelerate convergence. For some extreme cases where optimised expansions suffer from the negative densities within realistic order of expansion, such as the ones we deal with in an as yet unpublished manuscript by Marumo and Wolff (2015), the smoothing technique is expected to be a powerful means to deal with this problem.

For approximating the distributions from the parametric models, we found that the naïve expansion can still be of great use, as in Section 5. The example in Section 4.3 showed that smoothing may be applied to the cases where naïve expansions are divergent, once the problem about the truncated moments have overcome.

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 (kde), and we anticipate that an optimality criterion in the present case will have a similar character to the kde case.

In the example we dealt with in Section 5, the dependence structure between the risk factors was captured via their cross-moments. This implies the possibility of reconstructing the joint distribution from the cross- and marginal moments. We further extend this idea in the as yet unpublished manuscript by Marumo and Wolff (2015) and show that multivariate expansion methods can be used for capturing non-linear dependence structure between two variables and approximating copula densities.

References

- H. K. Baker and G. Filbeck. *Investment Risk Management*. Oxford University Press, 2014.
- P. Barrieu and G. Scandolo. Assessing financial model risk. *European Journal of Operational Research*, 242:546–556, 2015.
- Basel Committee. Fundamental review of the trading book: A revised market risk framework. Technical report, Basel Committee, 2013.
- J. Bessis. *Risk Management in Banking*. Wiley, 4th edition, 2015.
- D. Duffie and J. Pan. An overview of value at risk. *Journal of Derivatives*, Spring:7–48, 1997.
- 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.
- A. Leccadito, P. Toscano, and R. S. Tunaru. Value at risk and expected shortfall improved calculation based on the power transformation method. *Journal of Derivatives*, 22(2):67–81, 2014.
- 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.
- K. Marumo and R. C. Wolff. Orthogonal polynomial expansion estimators of joint distributions and copulas, with application to multivariate financial risk. *Manuscript submitted to Journal of Multivariate Analysis*, 2015.
- M. McAleer, J. Jimenez-Martin, and T. Perez-Amaral. Gfc-robust risk management strategies under the basel accord. *International Review of Economics and Finance*, 27:97–111, 2013.
- A. J. McNeil, R. Frey, and Embrechts. *Quantitative Risk Management*. Princeton, 2005.
- J. Mina and J. Y. Xiao. Return to RiskMetrics: The evolution of a standard. Technical report, RiskMetrics, 2001.
- C. Pérignon and D. R. Smith. The level and quality of value-at-risk disclosure by commercial banks. *Journal of Banking and Finance*, 34:362–377, 2010.
- B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall London, 1986.
- J.-G. Simonato. The performance of johnson distributions for computing value at risk and expected shortfall. *Journal of Derivatives*, 19(1):7–24, 2011.
- G. Szegő. *Orthogonal Polynomials*. American Mathematical Society, 4 edition, 1975.
- Y. Takahashi. *Jitsu Kansuu to Furie Kaiseki, (Real Functions and Fourier Analysis)*. Iwanami, Japan, 2006.