MAKING CORNISH–FISHER FIT FOR RISK MEASUREMENT

John D. Lamb[1], Maura E. Monville[2], Kai-Hong Tee[3]

[1]University of Aberdeen Business School, Aberdeen, UK
[2]University of Surrey, Guildford, UK
[3]Loughborough University Business School, Loughborough, UK

- We show how to fit a Cornish–Fisher distribution with exact skewness and kurtosis
- The method extends to multivariate distribution fitting
- The distributions fit financial data well
- They combine with estimation-error reduction and improve estimates of VaR and CVaR

ABSTRACT

The truncated Cornish–Fisher inverse expansion is well known and has been used to approximate value-at-risk and conditional value-at-risk. The following are also known. The expansion is available only for a limited range of skewness and kurtosis. The distribution approximation it gives is poor for larger values of skewness or kurtosis. We develop a computational method to find a unique corrected Cornish–Fisher distribution efficiently for a wide range of skewness and kurtosis. We show it has a unimodal density and a quantile function that is twice continuously differentiable as a function of mean, variance, skewness and kurtosis. We extend the univariate distribution to a multivariate Cornish–Fisher distribution and show it can be used together with estimation-error reduction methods to improve risk estimation. We show how to test goodness-of-fit. We apply the Cornish–Fisher distribution to fit hedge-fund returns and estimate conditional value-at-risk. We conclude that the Cornish–Fisher distribution is useful in estimating risk, especially in the multivariate case where we must deal with estimation error.

Keywords: conditional value-at-risk; estimation error; goodness-of-fit; kurtosis; skewness

1 INTRODUCTION

Consider how we might estimate the risk in investing in some combination of n assets. We may choose one or more risk measures such as variance, value-at-risk (VaR), deviation measures or conditional value-at-risk (CVaR). Whatever we choose, we should deal with estimation error (Herold and Maurer, 2006) whenever we estimate the risk measure. This is an error that arises from using $n > 1$ sample statistics to estimate n population statistics. Its effect is that the asset with lowest measured risk has underestimated risk while the asset with highest measured risk has overestimated risk. The effect increases rapidly with n. It is made worse when we consider not just n assets but combinations of them.

If variance were the only risk measure of interest, then the covariance-shrinkage method of Ledoit and Wolf (2004) might be enough. But usually we want risk measures like CVaR that change as skewness and kurtosis change. If we want to correct anything beyond mean and variance, and to compute estimates for combinations of assets we cannot simply adjust the data, for example, by shifting and scaling. Rather, we want a multivariate estimate of the distribution.

Our choice of Cornish–Fisher distribution may seem surprising. Although it has been used in risk estimation (Bali et al., 2007; Liang and Park, 2007), it has two obvious problems: it is valid only for a small range of skewness and kurtosis, and it is well-known to be inaccurate for estimating distributions not close to normal. Recently Maillard (2012) suggested a correction. We improve this correction and develop a method to find it quickly with guaranteed convergence to a unique set of parameters for skewness and kurtosis in a region ($\hat{R}$ of Figure 1) large enough to fit most asset returns. We then extend it to a multivariate distribution.

We show the corrected Cornish–Fisher distribution has many desirable properties. When fitting normal data we get a normal distribution. Otherwise, the density function f is smooth, unimodal (has only one maximum) and flat-tailed: that is, $f(x) > 0$ for $x \in \mathbb{R}$. These are properties we would usually expect in asset distributions. We also show Cornish–Fisher VaR and CVaR risk estimates are twice continuously differentiable functions of the coefficients of a linear combination of assets. Not only do we expect this of the true risk measures, but it also helps in portfolio selection.

Formally, suppose we have n random variables whose distribution function is unknown except that its shape must generalize the normal distribution. Suppose also we know the mean μ_i, variance σ_i^2, skewness κ_{3i} and (excess) kurtosis κ_{4i} of each variable. We seek a family $\mathscr{F}(\mathbf{p}_i)$ ($\mathbf{p}_i = (\mu_i, \sigma_i, \kappa_{3i}, \kappa_{4i})$, $i = 1, \ldots, n$) of distributions to model the random variables, ideally one that we can use to fit a multivariate distribution of all of them. Typically, as in our example, κ_{3i} and κ_{4i} will differ substantially from zero.

We can use the normal family for $\mathscr{F}(\mathbf{p})$ if skewness and kurtosis are zero. If skewness is zero we can use the elliptical family of distributions. And if we ignore kurtosis and can limit skewness to $(-1, 1)$ we can use the skew-normal distribution family. But none of these conditions are likely in financial data. The Box–Cox power exponential (BCPE) distribution (Rigby and Stasinopoulos, 2004) appears a plausible alternative to $\mathscr{F}(\mathbf{p})$, because it is a four-parameter family of distributions. However, its quantile function does not in general have a continuous derivative and, if the distribution is not symmetric, it is truncated in one or other of its tails. Hence our approach.

We find a multivariate distribution that fits a mean vector, covariance matrix, co-skewness and co-kurtosis tensors. The use of higher co-cumulants is not uncommon in the literature. They are used to evaluate investment fund performance Moreno and Rodríguez (2009) and portfolio allocation (Hitaj et al., 2012). Jondeau et al. (2017) even develop a model to explain the drivers of co-skewness and co-kurtosis, giving rise to a better explanation of stock returns. It is well known that higher co-cumulants are difficult to estimate accurately and give rise to estimation error. Future research may explore how to deal well with these issues. The multivariate Cornish–Fisher distribution gives us a practical way to estimate risk measures when we do not estimate skewness and kurtosis directly from data.

Let X be a random variable. Writing

$$F_X(x) = \Phi(u), \tag{1}$$

we can derive a power series expansion (Cornish and Fisher, 1938; Hill and Davis, 1968):

$$x = \sum_{k=0}^{\infty} a_k u^k. \tag{2}$$

We call it the *Cornish–Fisher inverse expansion.* Here a_k are polynomials in the cumulants of X. The expansion truncated to four terms is commonly used to approximate F_X, for example to approximate conditional CVaR. We use it with a systematic correction. We write $C^2_{\mathbf{q}}$ for the set of functions of $\mathbf{q}$ that are twice continuously differentiable over some region of interest.

We use hedge-fund data and CVaR to illustrate the Cornish–Fisher distribution in risk estimation. For simplicity we ignore time-series effects in estimating cumulants in the data, though, see for example Gabrielsen et al. (2015), we can easily include them.

Sections 2 shows how to correct the Cornish–Fisher expansion to get a distribution. Section 3 shows it has desirable properties for risk measurement and discusses generalizations, the most important of which is a multivariate Cornish–Fisher distribution. Section 4 provides a practical example that shows how to fit the distribution and how to estimate CVaR. Supplement A – Supplement D of the online supplement give details of our derivation.

2 THE CORNISH–FISHER EXPANSION

Suppose X has mean 0 and variance 1. Then equation (2) gives

$$F_X^{-1}(u) = \sum_{k=0}^{\infty} a_k \left(\Phi^{-1}(u)\right)^k,$$

where each a_k is a polynomial in the cumulants of X and Φ is the standard normal distribution function. To use this in practice we need to truncate the series. For the approximation to be increasing the highest power of k must be odd. In practice the fourth order ($k = 3$) approximation is used. This gives (Cornish and Fisher, 1938) $a_2 = -a_0 = s$, $a_1 = 1 + 5s^2 - 3k$ and $a_3 = k - 2s^2$ with $s = \kappa_3/6$ and $k = \kappa_4/24$. So the fourth order expansion (see equation (2)) is

$$x = \xi(u) = -s + (1 + 5s^2 - 3k)u + su^2 + (k - 2s^2)u^3, \tag{3}$$

giving quantile function

$$\tilde{F}^{-1}(u) = -s + (1 + 5s^2 - 3k)\Phi^{-1}(u) + s\left(\Phi^{-1}(u)\right)^2 + (k - 2s^2)\left(\Phi^{-1}(u)\right)^3. \tag{4}$$

Following Maillard (2012), we treat s and k as parameters because equations (6) and (7) show $\tilde{F}$ does not have skewness $s/6$ or kurtosis $k/24$ unless $s = k = 0$. We write $q = s^2$ to simplify expressions containing only even powers of s. $\tilde{F}$ can only be a distribution function if ξ is a strictly increasing function of u. It is straightforward to show (Maillard, 2012) that ξ is strictly increasing in the region R given by $q < 3 - 2\sqrt{2}$ and

$$\frac{1 + 11q - \sqrt{q^2 - 6q + 1}}{6} < k < \frac{1 + 11q + \sqrt{q^2 - 6q + 1}}{6}. \tag{5}$$

The moments of the fourth-order Cornish–Fisher inverse expansion are given by Maillard

(2012) (correcting a misprint and writing as functions of s and k):

$$\begin{aligned}
\mu_1(s,k) &= 0, \\
\mu_2(s,k) &= 1 + 6k^2 - 24s^2k + 25s^4, \\
\mu_3(s,k) &= 6s - 76s^3 + 510s^5 + 36sk - 468s^3k + 108sk^2, \\
\mu_4(s,k) &= 3 + 3348k^4 - 28080s^2k^3 + 1296k^3 - 6048s^2k^2 \\
&\qquad + 252k^2 - 123720s^6k + 8136s^4k - 504s^2k \\
&\qquad + 24k + 64995s^8 - 2400s^6 - 42s^4 + 88380k^2s^4.
\end{aligned} \tag{6}$$

The skewness and kurtosis are given by

$$\hat{s}(s,k) = \frac{\mu_3(s,k)}{(\mu_2(s,k))^{3/2}} \quad \text{and} \quad \hat{k}(s,k) = \frac{\mu_4(s,k)}{(\mu_2(s,k))^2} - 3. \tag{7}$$

Equations (6) and (7) implicitly define a function

$$G(s,k) = \big(\hat{s}(s,k), \hat{k}(s,k)\big)^\top \tag{8}$$

Maillard (2012) suggests G might be invertible and tabulates some values. Figure 1 shows an empirical plot of R (top) and of $\hat{R} = \{G(s,k) : (s,k) \in R\}$ (bottom). If a distribution has skewness κ_3 and kurtosis κ_4 then we must have $\kappa_4 \geq \kappa_3^2 - 2$ and the gray regions show the areas that are excluded by this inequality. The points and dashed lines are discussed in Section 4. Both plots show lines on which s or k is constant.

G is not globally invertible. We show in Supplement A that we can write its Jacobian determinant as $|J(s,k)| = \mu_2^{9/2} S(q,k)/144$ where $S(q,k)$ is a polynomial in q and k, which has a root at approximately $(0, -0.139) \notin R$. However, G remain useful if we can establish two things. First, that G has a unique inverse for $(\kappa_3, \kappa_4) \in \hat{R}$ so that $\tilde{F}^{-1}(u)$ is a twice continuously differentiable function of κ_3 and κ_4. Second, that there is an efficient method to obtain this inverse.

The method we use is essentially Newton's method, which is well-known and efficient. Define

$$J(s,k) = \begin{bmatrix} \frac{\partial \hat{s}}{\partial s} & \frac{\partial \hat{s}}{\partial k} \\ \frac{\partial \hat{k}}{\partial s} & \frac{\partial \hat{k}}{\partial k} \end{bmatrix},$$

the Jacobian matrix of G. Supplement A shows G has a unique inverse for $(\kappa_3, \kappa_4) \in \hat{R}$ by showing $|J(s,k)| > 0$ for $(s,k) \in R$.

We solve $G(s,k) = (\kappa_3, \kappa_4)$ starting from (s_0, k_0) by Newton's method. We compute (s_j, k_j) iteratively using

$$J(s_j,k_j)\begin{pmatrix} \tilde{s}_j \\ \tilde{k}_j \end{pmatrix} = -\begin{pmatrix} \hat{s}(s_j,k_j) - \kappa_3 \\ \hat{k}(s_j,k_j) - \kappa_4 \end{pmatrix}, \qquad \begin{pmatrix} s_{j+1} \\ k_{j+1} \end{pmatrix} = \begin{pmatrix} s_j \\ k_j \end{pmatrix} + \begin{pmatrix} \tilde{s}_j \\ \tilde{k}_j \end{pmatrix}. \tag{9}$$

For Newton's method to converge we require that $\hat{s}(s,k)$ and $\hat{k}(s,k)$ be continuously differentiable and that $J(s,k)$ be nonsingular wherever we evaluate it. We note that

$$\mu_2(s,k) = 1 + 6k^2 - 24qk + 25q^2 \geq 1 + 6(k^2 - 4qk + 4q^2) \geq 1 + 6(k - 2q)^2 \geq 1.$$

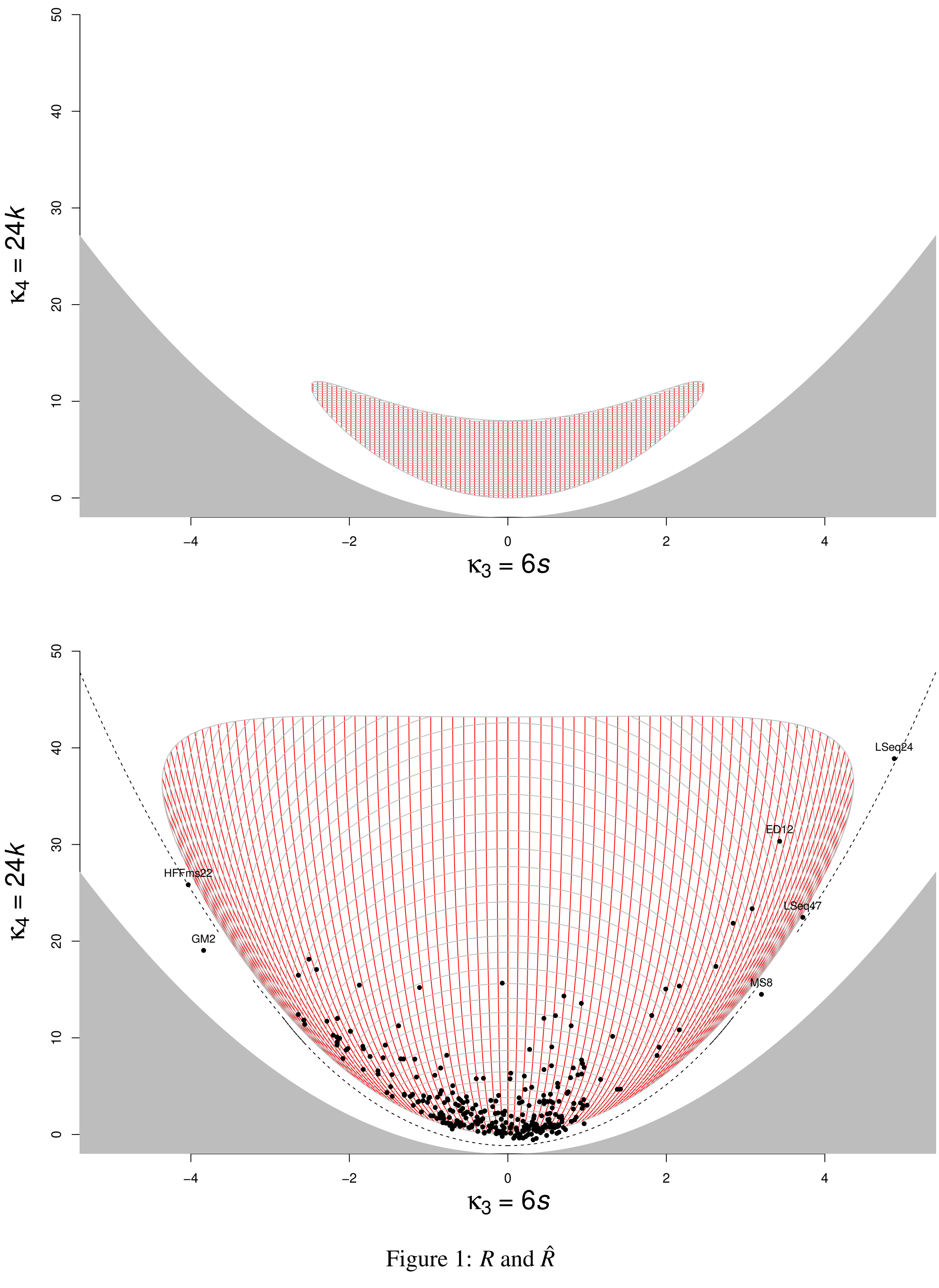

Figure 1: R and $\hat{R}$

Hence $\hat{s}$ and $\hat{k}$ are twice continuously differentiable. Since $G(0,0) = (0,0)$ we set $(s_0, k_0) = (0,0)$. By symmetry we can assume $\kappa_3 \geq 0$ and $s \geq 0$. Then to ensure $J(s,k) > 0$ we restrict Newton's method, replacing equation (9) (right) with

$$\begin{pmatrix} s_{j+1} \\ k_{j+1} \end{pmatrix} = \begin{pmatrix} s_j \\ k_j \end{pmatrix} + \alpha \begin{pmatrix} \tilde{s}_j \\ \tilde{k}_j \end{pmatrix} \tag{10}$$

where we choose $\alpha \in (0,1]$ so that (i) (s_{j+1}, k_{j+1}) is in a convex subset containing the right side of R (see Figure 1) and (ii) the solution is improving. Supplement A explains the detail.

In practice, most nonlinear optimizers should be able to invert G, even without explicit derivatives, provided s and k are constrained to lie in a convex subset of R. Our method is merely efficient and allows us to invert G for points very close to the dashed line in Figure 1 (bottom), where the Jacobian is singular. The final paragraph of Section 4.2 explains this line in more detail.

Suppose we want a distribution with mean μ, variance σ^2, skewness κ_3 and kurtosis κ_4. Suppose also μ and σ are finite and $(\kappa_3, \kappa_4) \in \hat{R}$. Then we can use Newton's method to evaluate $G^{-1}(\kappa_3, \kappa_4)$, giving us values for s and k and hence a_0, a_1, a_2 and a_3. Define

$$F^{-1}(u; \mathbf{p}) = \mu + \sigma \mu_2^{-1/2} \sum_{j=0}^{3} a_j \left(\Phi^{-1}(u)\right)^j, \tag{11}$$

where μ_2 is given by equation (6). $\tilde{F}$, defined by equation (4), has mean 0, variance μ_2 skewness κ_3 and kurtosis κ_4. Skewness and kurtosis are invariant under scaling and shifting. So F has mean μ, variance σ^2, skewness κ_3 and kurtosis κ_4. We call it the *Cornish–Fisher distribution with parameters μ, σ, κ_3 and κ_4*. Thus equation (11) defines a family of distributions for finite μ, finite $\sigma > 0$ and $(\kappa_3, \kappa_4) \in \hat{R}$. We write $Y \sim \mathscr{F}(\mathbf{p})$ to indicate Y is a distribution in this family. For reasons that become clearer at the end of Section 3 and in Section 4 we occasionally abuse notation by referring to F as a Cornish–Fisher distribution when $(\kappa_3, \kappa_4) \notin \hat{R}$.

The Cornish–Fisher distribution that approximates a distribution by putting $s = \kappa_3/6$ and $k = \kappa_4/24$ is widely used (Bali et al., 2007; Liang and Park, 2007). We call it the *uncorrected Cornish–Fisher distribution* and show in Section 4 that it is sometimes a poor approximation.

3 THE CORNISH–FISHER DISTRIBUTION AND A MULTIVARIATE GENERALIZATION

We show some properties of the Cornish–Fisher distribution and a multivariate expansion.

If we wish to estimate risk by fitting a distribution to data, then we should be concerned about some of the properties, particularly those of the left tail, of the distribution. We want VaR_α and CVaR_α to be well-behaved. Usually this means that we want them to be strictly increasing as functions of α. That is, as $\alpha \to 0$, there is no point at which risk becomes zero and no range over which it is not decreasing. Consequently we want a smooth distribution whose density does not become zero at some point in the left tail, as for example, the lognormal and gamma densities do.

Proposition 3.1 shows that the Cornish–Fisher distribution has these properties and consequently makes sense when estimating tail risk. We end Section 3.1 by showing that, even if we relax the region on which the Cornish–Fisher distribution is defined, CVaR_α often remains well-behaved for α not too small.

In the multivariate case we may wish to estimate the risk of some *convex combination*

$$\sum_{i=1}^{n} \lambda_i X_i$$

of assets $X_1,\dots,X_n$ satisfying $\sum_{i=1}^n \lambda_i = 1$ and $\lambda_i \geq 0$ for $i = 1,\dots,n$. Again we wish the risk measure to change smoothly as we change $\lambda = (\lambda_1,\dots,\lambda_n)^\top$. Proposition 3.2 shows that this will happen provided we choose a smooth risk measure like VaR or CVaR and provided skewness and kurtosis also change smoothly with λ. They do so for the multivariate distribution of Section 3.2, which preserve the covariance matrix and co-cumulant tensors.

3.1 Properties of the Cornish–Fisher distribution

Suppose $Y \sim \mathscr{F}(\mathbf{p})$. Equations (3) and (11) give

$$F^{-1}(u;\mathbf{p}) = \mu + \sigma \mu_2^{-1/2} \xi\left(\Phi^{-1}(u)\right). \tag{12}$$

We can rearrange equation (12) to get the distribution function

$$F(x) = u = \Phi\left(\xi^{-1}\left(\mu_2^{1/2}\frac{x-\mu}{\sigma}\right)\right). \tag{13}$$

In the R-package in the supplementary material we use Newton's method to evaluate ξ^{-1} rather than the formula for the roots of a cubic, because $\xi' > 0$ guarantees the root is unique and Newton's method converges.

The following two propositions are proved in Supplement B of the supplement. Note that the density function without the $\mu_2^{1/2}$ correction is given in Maillard (2013).

Proposition 3.1 *Let $Y \sim \mathscr{F}(\mathbf{p})$. The density function of Y is*

$$f(x) = \frac{\mu_2^{1/2}\phi(v)}{\sigma\xi'(v)},$$

with $v = \xi^{-1}\left(\mu_2^{1/2}(x-\mu)/\sigma\right)$. The distribution function is smooth, and $f(x)$ is unimodal and satisfies $f(x) > 0$ for $x \in \mathbb{R}$.

Note that smoothness and $f(x) > 0$ are more important for risk estimation. However, unimodality increases our confidence: when we estimate risk through a distribution function, we want that function to match as closely as possible the properties of the true distribution.

Proposition 3.2 *F^{-1} is twice continuously differentiable with respect to $\mathbf{p} = (\mu, \sigma, \kappa_3, \kappa_4)$.*

We can use the Cornish–Fisher distribution to estimate the α quantile, $F^{-1}(\alpha)$, and the (lower) α tail mean,

$$\mathrm{TM}(X;\alpha) = \mathbb{E}\left[X : X \leq F^{-1}(\alpha)\right]$$

of X. These are often seen in finance as value-at-risk (Tee, 2009), $\mathrm{VaR}(X;\alpha) = -F^{-1}(\alpha)$, and conditional-value-at-risk, $\mathrm{CVaR}(X;\alpha) = -\mathrm{TM}(X;\alpha)$. Acerbi (2002) shows

$$\mathrm{TM}(X;\alpha) = \frac{1}{\alpha}\int_0^\alpha F^{-1}(u)\,\mathrm{d}u.$$

Put

$$\tau_r(\alpha) = \frac{1}{\alpha\sqrt{2\pi}}\int_{-\infty}^{\Phi^{-1}(\alpha)} z^r \exp\left(-\frac{1}{2}z^2\right)\mathrm{d}z.$$

Then, if Y has Cornish–Fisher distribution approximating X,

$$\begin{aligned}
\text{TM}(Y;\alpha) &= \frac{1}{\alpha}\int_0^{\alpha}\left(\mu + \sigma\mu_2^{-1/2}\left(\sum_{r=0}^{3} a_r(\Phi^{-1}(u))^r\right)\right)\mathrm{d}u \\
&= \mu + \sigma\mu_2^{-1/2}\sum_{r=0}^{3} a_r \frac{1}{\alpha}\int_0^{\alpha}(\Phi^{-1}(u))^r\,\mathrm{d}u \\
&= \mu + \sigma\mu_2^{-1/2}\sum_{r=0}^{3} a_r \frac{1}{\alpha\sqrt{2\pi}}\int_{-\infty}^{\Phi^{-1}(\alpha)} z^r \exp\left(-\frac{1}{2}z^2\right)\mathrm{d}z \\
&= \mu + \sigma\mu_2^{-1/2}\sum_{r=0}^{3} a_r \tau_r(\alpha). \qquad (14)
\end{aligned}$$

This last expression is a twice continuously differentiable function of $\mathbf{p}$ and so, by Proposition 3.2, VAR, TM and CVaR are also in $C^2_{\mathbf{p}}$. To use this result in practice we need expressions to compute the partial derivatives of F^{-1}. Supplement C derives such expressions.

The tail means and quantiles are only defined if $\xi(u)$ is strictly increasing on $\mathbb{R}$. Suppose, however, that $\xi(u)$ is increasing on (u_1, u_2) and that $\alpha \in (\alpha_1, \alpha_2)$ with $\alpha_1 = \Phi(u_1)$ and $\alpha_2 = \Phi(u_2)$. Then both $F^{-1}(Y;\alpha)$ and $\text{TM}(Y;\alpha)$ are sensibly defined and are the quantile and tail mean of a distribution given by

$$F'(x) = \begin{cases} F_1(x), & x \le x_1, \\ F(x), & x_1 < x < x_2, \\ F_2(x), & x \ge x_2, \end{cases}$$

for $x_1 = F^{-1}(u_1)$ and $x_2 = F^{-1}(u_2)$ if we can find strictly increasing continuous functions F_1 and F_2 satisfying $F_1(x_1) = \alpha_1$, $F_2(x_2) = \alpha_2$,

$$\begin{aligned}
\frac{1}{\alpha_1}\int_0^{\alpha_1} F_1^{-1}(u)\,\mathrm{d}u &= \mu + \sigma\mu_2^{-1/2}\sum_{r=0}^{3} a_r\tau_r(\alpha_1), \qquad (15) \\
\frac{1}{1-\alpha_2}\int_{\alpha_2}^{1} F_2^{-1}(u)\,\mathrm{d}u &= \mu + \sigma\mu_2^{-1/2}\sum_{r=0}^{3} a_r\tau_r(1-\alpha_2).
\end{aligned}$$

Such functions are easy to find. For example, $F_1^{-1}(u) = mu + x_1 - m\alpha_1$ satisfies $F_1(x_1) = \alpha$ and

$$\frac{1}{\alpha_1}\int_0^{\alpha_1} F^{-1}(u)\,\mathrm{d}u = x_1 - \frac{1}{2}m\alpha_1,$$

and it is straightforward to show

$$\mu + \sigma\mu_2^{-1/2}\sum_{r=0}^{3} a_r\tau_r(\alpha) < x_1.$$

So if we choose m satisfying equation (15) holds then $m > 0$ and so F_1 is, as required, strictly increasing. It follows that if $(\kappa_3, \kappa_4) \notin \hat{R}$ but $\alpha \in (\alpha_1, \alpha_2)$ we can still use the Cornish–Fisher distribution to estimate quantiles and tail means. We return to this issue in Section 4.

3.2 Generalizations and the multivariate Cornish–Fisher distribution

We have established that the Cornish–Fisher distribution has very desirable properties and region of validity much larger than we might naively expect. We now consider generalizations.

Hill and Davis (1968) generalize Cornish–Fisher expansions to use non-normal distributions. So we might ask what happens if we replace Φ^{-1} in equation (11) with a non-normal quantile function. We have tried this and find that if we replace Φ^{-1} with the quantile function of a beta, gamma or lognormal distribution then F fails to be a distribution function except in degenerate cases. The log of a Cornish–Fisher distribution is usually well defined. But we find, in contrast with Proposition 3.1, it is only unimodal in the most degenerate case. So it has few practical uses.

In contrast, a multivariate Cornish–Fisher distribution is practically useful. We want to estimate the distribution of a vector $(X_1,\ldots,X_n)$ so that we may estimate risk measures not just of individual assets but of convex combinations of them: that is, of portfolios. We can write each portfolio as

$$X=\sum_{j=1}^{n}\lambda_j X_j$$

for $\lambda_1,\ldots,\lambda_n\geq 0$ satisfying $\lambda_1+\cdots+\lambda_n=1$.

One way to estimate a multivariate Cornish–Fisher distribution is to use data. If r_{jt} is the tth of T observations from X_j then

$$r_t=\sum_{j=1}^{n}\lambda_j r_{jt}\qquad(t=1,\ldots,T)$$

is a vector of observations of the portfolio, whose cumulants we can use to estimate a Cornish–Fisher distribution, as before.

We generalize this. For various reasons such as reducing estimation risk or accounting for time-series effects, we may not wish to estimate the cumulants directly from data but from tensors of means, covariance, co-skewness and co-kurtoses (Moreno and Rodríguez, 2009; Hitaj et al., 2012).

Write $\mu_1,\ldots,\mu_n$ and $\sigma_1^2,\ldots,\sigma_n^2$ for the means and variances of $X_1,\ldots,X_n$. Then co-skewness is defined as

$$S_{ijk}=\frac{\mathbb{E}\left[(X_i-\mu_i)(X_j-\mu_j)(X_k-\mu_k)\right]}{\sigma_i\sigma_j\sigma_k},$$

where i, j and k need not be distinct. S_{iii} is the skewness of X_i. Similarly, the co-kurtosis is

$$K_{ijkl}=\frac{\mathbb{E}\left[(X_i-\mu_i)(X_j-\mu_j)(X_k-\mu_k)(X_l-\mu_l)\right]}{\sigma_i\sigma_j\sigma_k\sigma_l}$$

and K_{iiii} is the kurtosis of X_i.

Suppose we have an estimates s_{ij} of $\operatorname{cov}(X_i,X_j)$ (variance if $i=j$) for $i,j=1,\ldots,n$. Then we estimate the variance of the portfolio X as

$$s_X=\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j s_{ij}.$$

Similarly we estimate the skewness as

$$S_X=\frac{1}{s_X^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\lambda_i\lambda_j\lambda_k S_{ijk}$$

and the kurtosis as

$$K_X = \frac{1}{s_X^4} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} \lambda_i \lambda_j \lambda_k \lambda_l K_{ijkl}.$$

Thus, any estimate of the four co-cumulant tensors allows us to construct a multivariate Cornish–Fisher distribution.

Section 1 summarises the univariate alternatives to the Cornish–Fisher distribution. We know of no general distribution defined by mean, covariance, co-skewness and co-kurtosis parameters that does not generalise one of these. The multivariate Cornish–Fisher distribution has the same advantages over multivariate alternatives as the univariate Cornish–Fisher distribution has. Its weakness is that it is not well defined for the rare (see Section 4) cases where it does not fit.

Note that the multivariate Cornish–Fisher distribution can be used whether we estimate the parameters directly from the data or using, for example, some time-series model. It is especially useful when the parameter estimates are not the same as those of the data so that we have no empirical distribution from which to estimate risk measures.

4 AN APPLICATION TO INVESTMENT FUND DISTRIBUTIONS

We now investigate the practical applicability of the Cornish–Fisher distribution. We do this in four ways. First, we compare the Cornish–Fisher distribution with the uncorrected version to get a broad picture of how much the difference matters. We do this by simulating both distributions using intended parameters and comparing with observed parameters. Second, we investigate how well a large data set fits the Cornish–Fisher distribution. We do this using goodness-of-fit tests and comparing with the uncorrected distribution. We also test the fit of the normal distribution to show that the data do not plausibly have zero skewness and kurtosis. Third, we investigate the value of using the Cornish–Fisher distribution with a risk estimate (CVaR). Estimating risk in this manner is most valuable when we are using a model that does not allow us to estimate risk directly from the data. But here we mainly use cases where an empirical CVaR estimate is also possible, so that we can get direct comparison of two methods. Finally, we investigate the extent to which we can fit the multivariate Cornish–Fisher distribution to the entire data set.

To illustrate the Cornish–Fisher distribution we consider the monthly returns of 339 investment funds from January 2000 to December 2012. We choose these data because investment returns often plausibly have a unimodal density that is too skewed, platykurtic or leptokurtic to be normal. These data originate from hedge funds. We collected monthly returns of the various strategies listed in Table 1. The data come from Morningstar (2014), which describes the strategies in detail.

The points in the lower chart of Figure 1 show the skewness and kurtosis (estimated by k-statistics) for each investment except GM15, HFFms9 and SF49, which have skewness 5.672, -8.099 and 6.26 out of the range of the chart. The prefix on the labeled funds shows its type. We choose a range of types to ensure a range of values of skewness and kurtosis. Table 1 shows the prefixes and fund types.

4.1 Comparison with the uncorrected Cornish–Fisher distribution

We first compare the Cornish–Fisher distribution with the uncorrected Cornish–Fisher distribution. Figure 2 shows the skewness and kurtosis of the funds of Figure 1 as black points. The gray points show the skewness and kurtosis of the uncorrected Cornish–Fisher distributions and we have joined

prefix	type	prefix	type
HFFeq	Hedge funds of funds for equity	MS	Multi-strategy
HFFms	Hedge funds of funds for multi-strategy	SF	Systematic futures
HFFmasy	Hedge funds of funds for macro systematic	GM	Global Macro
LSeq	Long–short equity strategy	ED	Event driven strategy

Table 1: Investment fund types

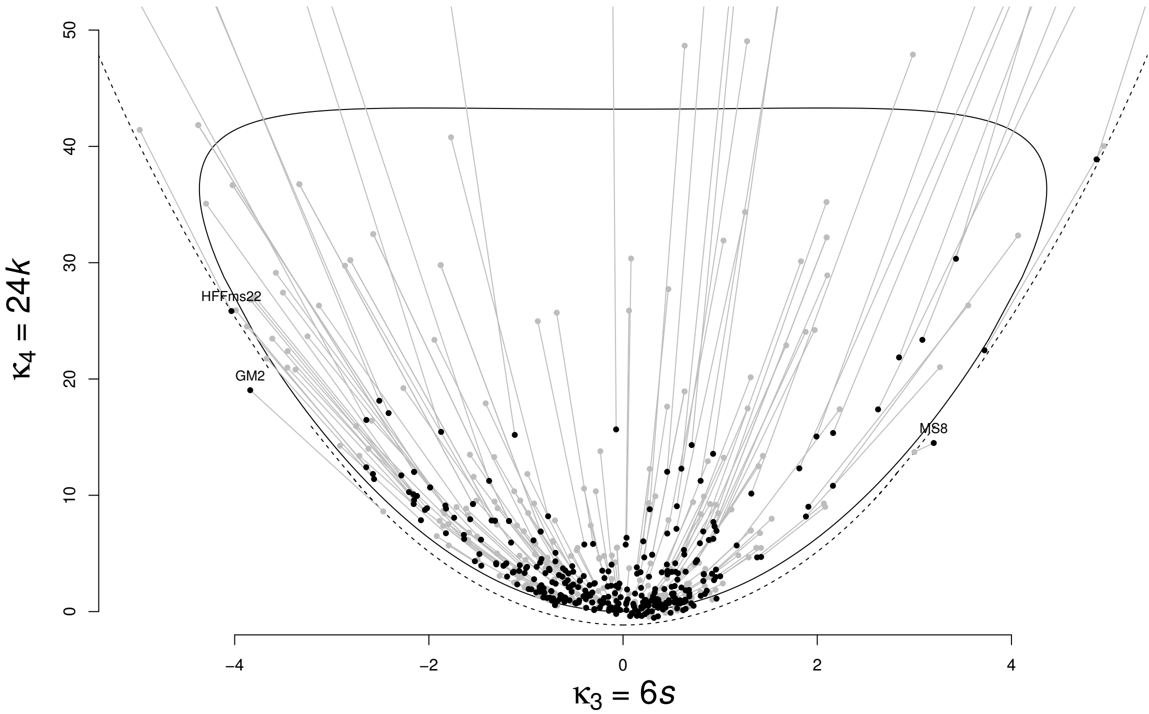

Figure 2: Comparison of Cornish–Fisher (black) and uncorrected (gray) distributions

corresponding points with a gray line to show the effect of not using the correction. The Cornish–Fisher distribution has the same skewness and kurtosis as the data except for GM2, HFFms22, MS8 and the three points we noted above that are out of the range of the chart.

The solid and dashed black lines are described in Section 2.

We omit comparison of mean and variance, because both distributions give the same values. We note that the uncorrected Cornish–Fisher distribution still uses the $\mu_2^{-1/2}$ correction of equation (12) though we do not know of even this correction being used in the published literature.

Although the uncorrected distribution is reasonable for many of the funds, a substantial proportion have greatly exaggerated skewness and kurtosis and sometimes fall outside the region where the distribution is valid.

4.2 Fitting the Cornish–Fisher distribution

We fit Cornish–Fisher distributions for as many of the funds as we can using the R-package in the supplementary material. We do this even if $(\kappa_3, \kappa_4) \notin \hat{R}$ and we compute a bootstrap Anderson–

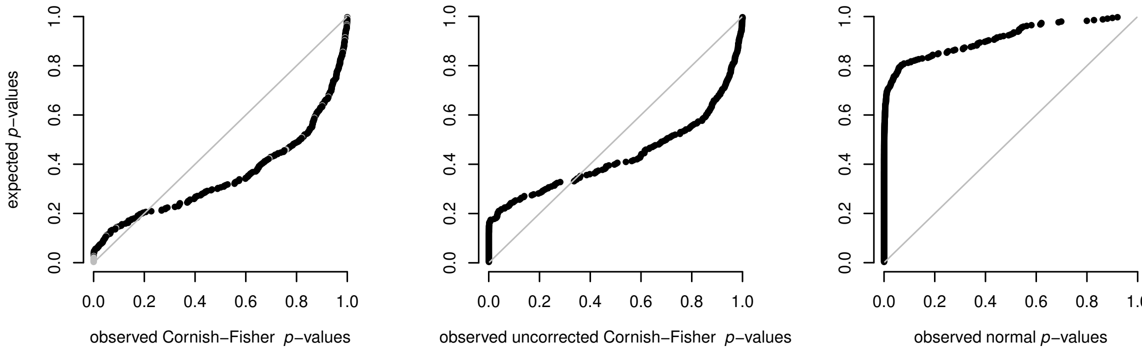

Figure 3: Anderson–Darling p-values for Cornish–Fisher and normal distributions

Darling (Cheng, 2006) test statistic whenever we can. We fail to fit Cornish–Fisher distributions for seven funds: GM2, GM15, HFFms9, HFFms22, LSeq24, MS8 and SF49. All are strongly skewed. GM4, LSeq22 and 13 SF funds also have $(\kappa_3, \kappa_4) \notin \hat{R}$. We find Cornish–Fisher distributions for them when we relax the requirement that the distribution function is defined in its tails. We tried using a Jarque–Bera test in place of Anderson–Darling and found very similar results.

If we assume all fund returns have Cornish–Fisher distributions with known parameters and are independent then the p-values from the Anderson–Darling tests should be uniformly distributed on $[0,1]$. Figure 3 (left) shows the observed and expected p-values under this assumption. Note that Figure 3 does not show p–p plots. We are not comparing observed and expected values of a single distribution. Rather we are comparing the p-values of Anderson–Darling tests of fits of three families of distributions to asset returns under the null hypotheses that each asset fits the family with *some* set of parameter values. The comparison is made under the assumption that all the null hypotheses are true for the given family. Under this assumption we expect to observe p-values uniformly distributed on $[0,1]$. The y coördinates in each plot are uniform on $[0,1]$ and so a family fits well if the points fall roughly on the gray line.

We record a p-value of 0 when we cannot fit a distribution. The funds with $(\kappa_3, \kappa_4) \in \hat{R}$ have black points and the rest gray. For comparison we show the Anderson–Darling p-values if we use uncorrected Cornish–Fisher distributions (center), defined at the end of Section 2, and normal distributions (right). In the center chart we omit the 59 cases where the uncorrected Cornish–Fisher expansion does not give a distribution function.

Both the Cornish–Fisher and uncorrected Cornish–Fisher fits have better than expected p-values for many funds. We think this is largely because the four parameters allow the distribution to fit the data better than its population. We note that the correlation between many pairs of fund returns is too high for them to be plausibly independent.

Figures 4–6 show more detail for some of the fitted Cornish–Fisher distributions. Each chart shows a quantile–quantile plot, the fitted distribution function together with an empirical distribution function, and a histogram together with the fitted density function (solid) and a density function estimated by kernel density estimation. The p shown is the p-value from the bootstrap Anderson–Darling test.

Depending on how accurate a fit is needed, 20–30 funds do not plausibly fit their (corrected) Cornish–Fisher distributions or do not have one. Most of these have one or two extreme values and $(\kappa_3, \kappa_4) \notin \hat{R}$, though ED12 (Figures 1 and 4 left) does not. MS19 (Figure 4 right) and HFFms42

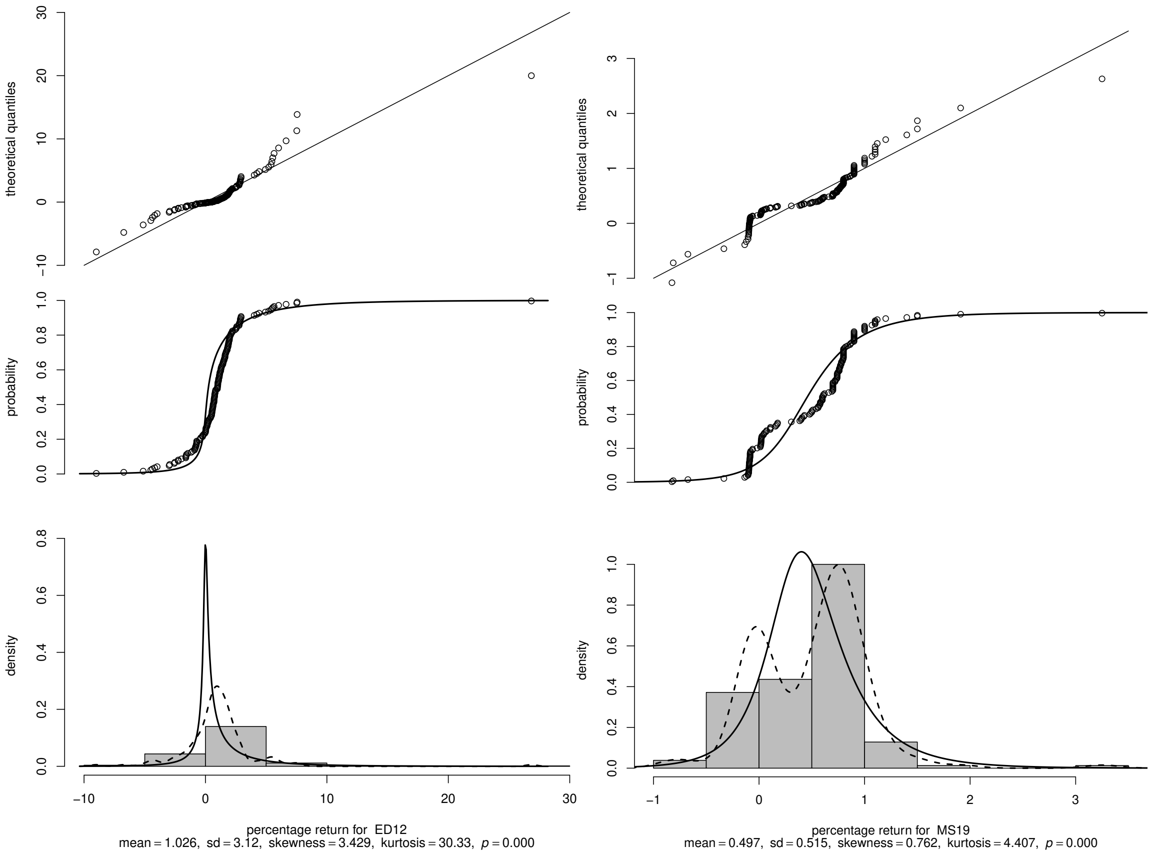

Figure 4: Poor-fit Cornish–Fisher distributions

(not shown) have $(\kappa_3, \kappa_4) \notin \hat{R}$ but density functions that are likely not unimodal.

Figure 5 shows two better-fit distributions. LSeq55 (left) is the 302nd best fit and SF64 (right), the 149th. Figure 6 shows two weaker-fit distribution. GM11 (left) is the 320th best fit. GM4 (right) has the least good fit of the 13 funds with negative kurtosis. We find nine of these funds have Anderson–Darling estimated p-values exceeding 0.9. To see why this might happen, consider the case $s = 0$. The Jacobian has a zero approximately at $k = -0.139$ corresponding to kurtosis about -1.31. And $\xi(u)$ is increasing between its roots at $\pm\sqrt{1-1/(3k)}$. Thus, in the worst case, when $k = -0.139$ the distribution is well defined for approximately 1.84 standard deviations on either side of the mean value. That is, only the extreme tails of the distribution function are undefined. These are not visible on the plot and do not affect the Anderson–Darling test.

The dashed lines on Figure 1 approximate and interpolate skewness and kurtosis corresponding to some of the zeroes of the Jacobian. The lines are not simple. For example, we estimate three closely spaced zeros for skewness about ± 2.7. Nonetheless, we conjecture that the region including R on which G is invertible includes most of the points between $\hat{R}$ and this line. This would account for the very good fit of funds with kurtosis not in $\hat{R}$.

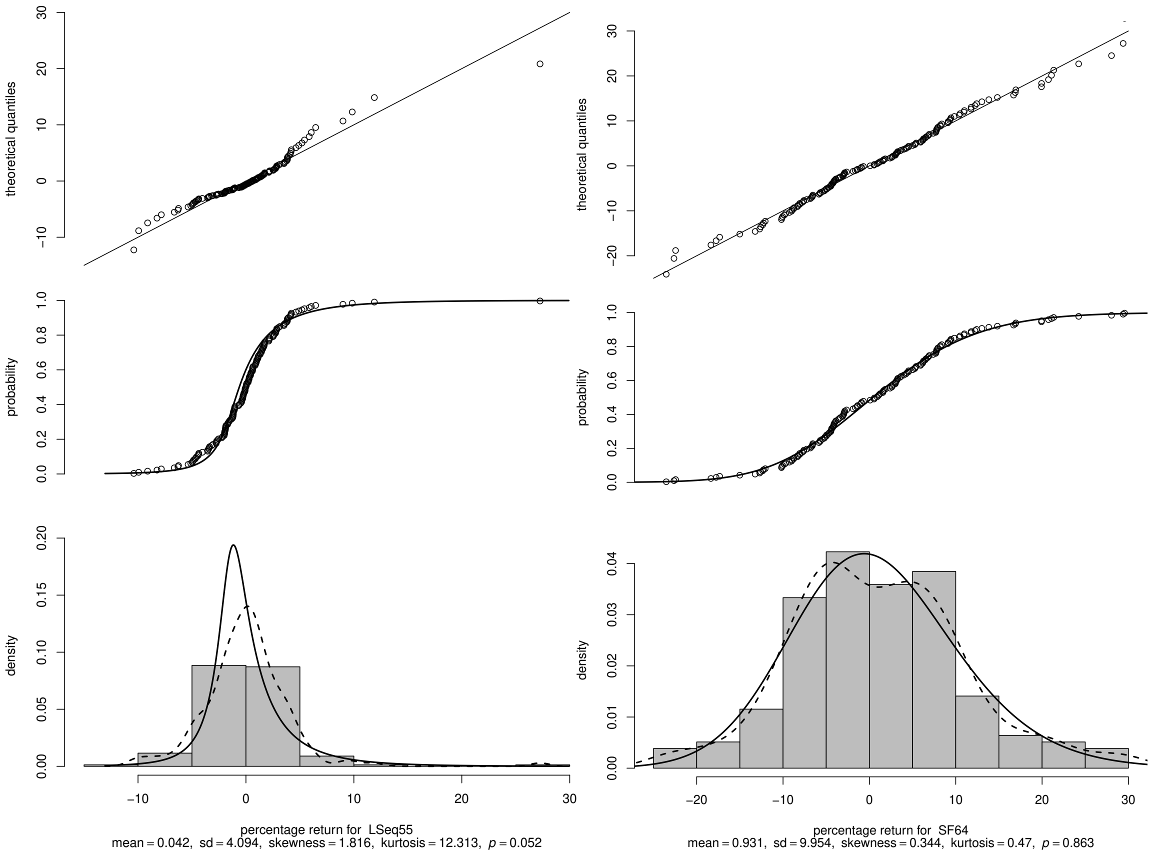

Figure 5: Good-fit Cornish–Fisher distributions

4.3 CVaR estimation

Section 3 showed we can estimate quantiles and tail-means using the Cornish–Fisher distribution. The estimates should be less influenced by outliers than, for example, the method of Acerbi (2007) for estimating CVaR directly from the data.

We estimate CVaR at $\alpha = 0.1$ in four different ways. These are: (i) from the a piecewise-linear empirical distribution estimate (Acerbi, 2007); (ii) using the Cornish–Fisher distribution and equation (14); (iii) using the uncorrected Cornish–Fisher distribution; and (iv) using the Cornish–Fisher distribution together with shrinkage estimators to deal with estimation error.

Shrinkage estimators correct estimation error, which underestimates the smallest and overestimate the largest statistic in multivariate samples (Herold and Maurer, 2006). We use the method of Jorion (1986) for the mean value and that of Ledoit and Wolf (2004) for the covariance matrix. The means are estimated as

$$\bar{\mathbf{r}}^* \approx 0.234\,\bar{r}\,\mathbb{1}_n + 0.766\,\bar{\mathbf{r}},$$

where the usual estimates of the n mean returns is $\bar{\mathbf{r}}$, $\mathbb{1}_n$ is a vector of n 1s and $\bar{r}$ is the mean of $\bar{\mathbf{r}}$. The covariance matrix estimate is

$$S^* \approx 0.073\,m I_n + 0.927\,S_n, \tag{16}$$

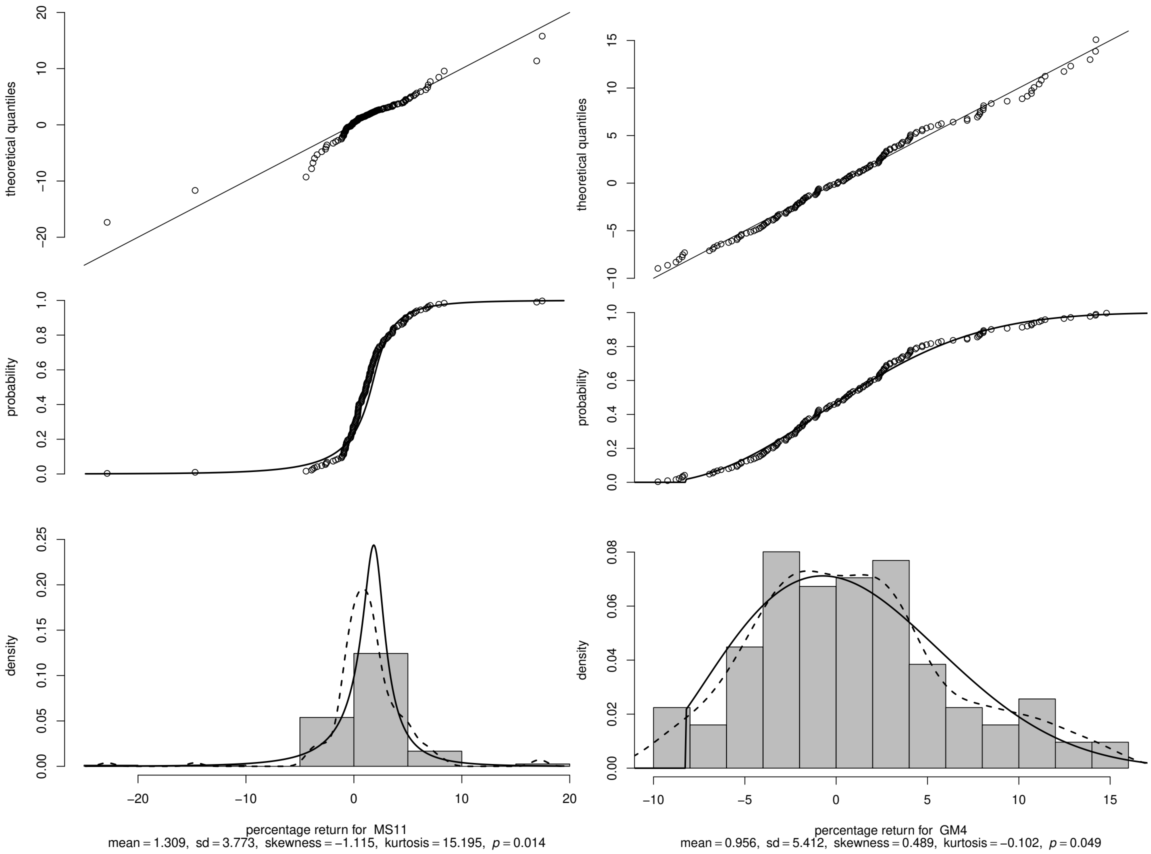

Figure 6: Weak-fit Cornish–Fisher distributions

where S_n is the unbiased estimator of the covariance matrix, $m \approx 22.81$ is the average variance of the $n = 339$ funds and I_n is the $n \times n$ identity matrix. Supplement D shows the details of the calculation. We do not discuss specific choices of shrinkage estimators but note they shrink the mean and variance towards the overall mean, and the correlations towards zero. We expect them to increase small and reduce large values of CVaR.

Figure 7 compares the various CVaR estimates for the 339 funds. Section 3 notes that we can sometimes estimate CVaR when $(\kappa_3, \kappa_4) \notin \hat{R}$. For example, for GM4 (see Figure 6 right) ξ is increasing between about -0.917 and 1.428 so that the Cornish–Fisher expansion defines a quantile function except in the extreme (< 0.005) tails. So we compute CVaR in all 339 cases including the seven of Section 4 where the fit fails. Circles show funds with skewness and kurtosis in R (see Figure 1). Squares show funds with skewness and kurtosis in $\hat{R}$. And triangles show the worst-behaved funds, including the seven where the fit fails. As expected, Cornish–Fisher CVaR fails in most of seven the worst cases. The left graph shows good fit between CVaR fitted with empirical and Cornish–Fisher distributions. Except in extreme cases we can attribute variations to the smoothing effect of the Cornish–Fisher distribution.

The center graph shows the problem of using the uncorrected Cornish–Fisher expansion. Often CVaR is underestimated. This usually happens when $(\kappa_3, \kappa_4) \notin R$.

The right graph illustrates why we should consider estimation error. Otherwise we tend to

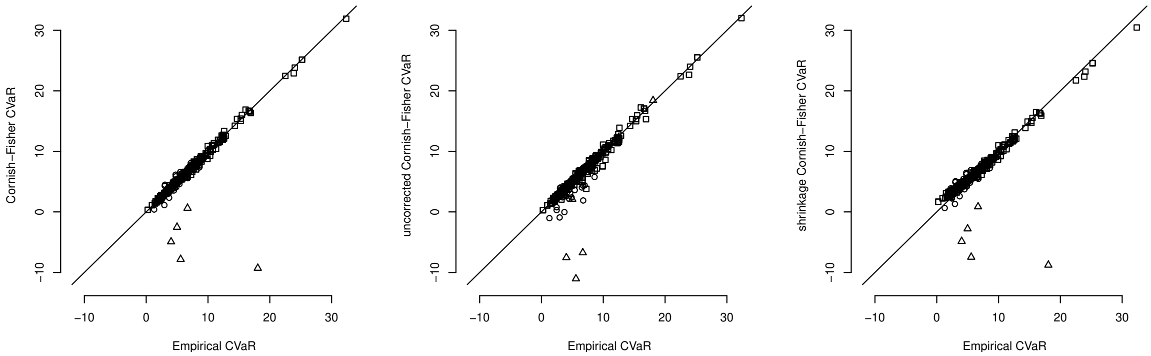

Figure 7: Comparison of CVaR estimates

underestimate CVaR when it is small and overestimate it when it is large. Note especially that the empirical CVaR estimate can be less than 50% of shrinkage CVaR or even negative when CVaR is small.

Note that the shrinkage CVaR estimates here use shrinkage only for mean and variance. We expect (see Figure 2) better estimators of skewness and kurtosis to shrink the more extreme estimates of these values. Section 4.2 notes that the worst-fit funds tend to have extreme skewness and kurtosis and so are likely to fit better if we can find such estimators. In addition, we expect negative kurtosis to be reduced, leading to risk estimators that are valid even in the more extreme tails of the Cornish–Fisher distribution. Martellini and Ziemann (2010) discuss possible shrinkage estimators for the co-cumulants. These can be used in the univariate case and we discuss them further in the next section.

4.4 Multivariate fitting

Section 3.2 describes how we can fit multivariate Cornish–Fisher distribution using a mean vector, covariance matrix and skewness and kurtosis tensors. We cannot illustrate every example of multivariate portfolio fit and, in general, we expect randomly chosen portfolios to fit as well as the funds above. So, we illustrate with an example of a problematic portfolio and show how Cornish–Fisher can help with risk estimation.

The portfolio contains 11 of the 339 funds with weights between 0 and 0.313. The charts of Figure 8 show the empirical distribution function as circles. We choose this portfolio because it has multiple returns slightly greater than -1 and giving empirical estimate of VaR at 10% of 1. It is evident that this likely underestimates the risk of future portfolio returns.

Figure 8 (left) shows the (corrected) Cornish–Fisher distribution fitted to the empirical data (that is using data estimates for means, covariances, co-skewnesses and co-kurtoses). The bootstrap Anderson–Darling p-value confirms an evident poor fit. In this case the Cornish–Fisher fit gives us a worse estimate of VaR but tells us the empirical estimate is also likely unreliable.

Figure 8 (right) shows the (corrected) Cornish–Fisher distribution fitted to the data with the most extreme left and right tail values removed. Notice the substantial changes in skewness and kurtosis and the much more plausible fit. This time the Cornish–Fisher distribution gives us a more conservative and plausible estimate of VaR than the empirical estimate.

If one's interest is to estimate risk in individual cases, then censoring extreme data points may

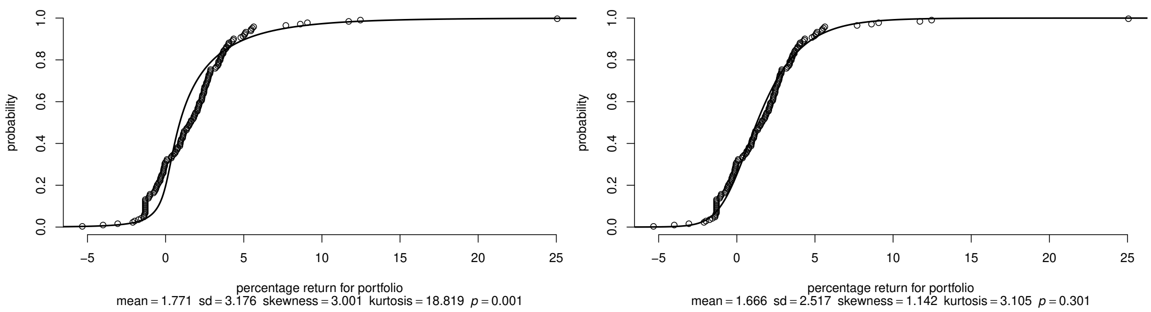

Figure 8: Fitting a portfolio from the multivariate distribution

be sufficient and will, for example, improve the fits in Figures 4 and 6. More generally, it is likely that co-cumulant shrinkage estimators such as those of Martellini and Ziemann (2010) can be combined with the multivariate Cornish–Fisher distribution to give good risk estimators for most portfolios.

5 CONCLUSION

Extending and correcting an idea of Maillard (2012), we show how to fit a Cornish–Fisher distribution with specified mean, variance, skewness and kurtosis. We show the distribution has desirable properties of unimodality and normal-like tails. We also demonstrate we can fit the distribution for a range of skewness and kurtosis big enough so that we can fit a multivariate Cornish–Fisher distribution by excluding only the rarest of asset returns.

Risk measures like CVaR and VaR are based on the tail of a distribution. This makes estimates from data very sensitive to extremes. Estimating from our Cornish–Fisher distribution has advantages. First the distribution shape – flat-tailed, smooth and unimodal – is likely for asset returns. Second, we can use it together with methods to reduce estimation risk. This makes it more promising as we develop improved shrinkage estimators. In the univariate case we suggest using the Cornish–Fisher distribution to help identify where estimating CVaR or VaR may be problematic. In the multivariate case we recommend using shrinkage estimators to reduce estimation error in estimating risk. The Cornish–Fisher distribution allows us to do this for CVaR and VaR and is particularly important in helping us avoid underestimating risk in lower-risk assets.

Proposition 3.2 tells us that risk measures like CVaR will be twice continuously differentiable as a function of the coefficients of the linear combination. This makes minimizing risk over a set of assets easier, especially when we use risk measures like CVaR that are convex functions of the coefficients of a linear combination of assets.

Scope and space have limited our exploration of two issues. First, we have ignored time series effects to simplify Section 4. We can include these effects in principle (Gabrielsen et al., 2015). It would be useful to see what happens in practice.

Second, we have limited our discussion of estimation error and shrinkage estimators to some well-known methods for mean, variance and covariance (Herold and Maurer, 2006; Ledoit and Wolf, 2004). This could be further explored. One benefit of the Cornish–Fisher distribution is that it allows the possibility of developing shrinkage estimators for skewness and kurtosis. Just as Ledoit and Wolf (2004) shrink not just the variances but the whole covariance matrix, such methods could shrink the co-skewess and co-kurtosis tensors. Then the Cornish–Fisher distribution is likely to be useful: it is more likely to fit. Moreover, while we can change the mean and variance of data without affecting any other cumulants, changing the covariance is problematic and we know of no way to change its skewness or kurtosis without affecting other cumulants. Shrinking skewness and kurtosis would prevent us estimating risk measures such as VaR and CVaR empirically. But the Cornish–Fisher distribution estimates would remain.

SUPPLEMENTARY MATERIAL

Making Cornish–Fisher distributions fit: detailed derivations: the file *CornishFisherSupplement.pdf* provides detailed derivations of some of the results used in this article (portable document file format).

R-package for CornishFisher functions: the file *CornishFisher 1.0.tar.gz* contains code to compute the CornishFisher expansions, distribution functions, quantile functions and bootstrap AndersonDarling test described in the article. (gnu zipped tar file).

REFERENCES

Acerbi, C., 2002. Spectral measures of risk: A coherent representation of subjective risk aversion. Journal of Banking and Finance 26 (7), 1505–1518.

Acerbi, C., 2007. Coherent measures of risk in everyday market practice. Quantitative Finance 7 (4), 359–364.

Bali, T. G., Gokcan, S., Liang, B., 2007. Value at risk and the cross-section of hedge fund returns. Journal of Banking & Finance 31 (4), 1135–1166.

Cheng, R. C. H., 2006. Validating and comparing simulation models using resampling. Journal of Simulation 1 (1), 53–63.

Cornish, E. A., Fisher, R. A., 1938. Moments and cumulants in the specification of distributions. Revue de l'Institut International de Statistique 5, 307–322.

Gabrielsen, A., Kirchner, A., Liu, Z., Zagalia, P., 2015. Forecasting value-at-risk with time-varying variance, skewness and kurtosis in an exponential weighted moving average framework. Annals of Financial Economics 10, 29 pages.

Herold, U., Maurer, R., 2006. Portfolio Choice and Estimation Risk: A Comparison Of Bayesian To Heuristic Approaches. ASTIN Bulletin 36 (1), 135–160.

Hill, G. W., Davis, A. W., 1968. Generalized asymptotic expansions of Cornish–Fisher type. The Annals of Mathematical StatisticsStatistics 39 (4), 1264–1273.

Hitaj, A., Martellini, L., Zambruno, G., 2012. Optimal hedge fund allocation with improved estimates for coskewness and cokurtosis parameters. Journal of Alternative Investments 14 (3), 6–16.

Jondeau, E., Jurczenko, E., Rockinger, M., 2017. Moment Component Analysis: An Illustration With International Stock Markets. Journal of Business and Economic Statistics, 1–23, DOI: 10.1080/07350015.2016.1216851.

Jorion, P., 1986. Bayes-Stein Estimation for Portfolio Analysis. The Journal of Financial and Quantitative Analysis 21 (3), 279–292.

Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis 88 (2), 365–411.

Liang, B., Park, H., 2007. Risk Measures for Hedge Funds—a Cross-sectional Approach. European Financial Management 13 (2), 333–370.

Maillard, D., 2012. A User's Guide to the Cornish Fisher Expansion. SSRN Electronic Journal, 1–19.
URL http://ssrn.com/paper=1997178

Maillard, D., 2013. More on Cornish Fisher: Distribution Density and Boundary Conditions. SSRN Electronic Journal, 1–11.
URL http://ssrn.com/paper=2236338

Martellini, L., Ziemann, V., 2010. Improved Estimates of Higher-Order Comoments and Implications for Portfolio Selection. The Review of Financial Studies 23 (4), 1467–1502.

Moreno, D., Rodríguez, R., 2009. The value of coskewness in mutual fund performance evaluation. Journal of Banking and Finance 33 (9), 1664–1676.

Morningstar, 2014. Morningstar – Independent Investment Research.
URL http://www.morningstar.com/

Rigby, R. A., Stasinopoulos, D. M., 2004. Smooth centile curves for skew and kurtotic data modelled using the Box–Cox power exponential distribution. Statistics in Medicine 23, 3053–3076.

Tee, K.-H., 2009. The effect of downside risk reduction on UK equity portfolios included with Managed Futures Funds. International Review of Financial Analysis 18 (5), 303–310.