The Cornish-Fisher-Expansion in the Context of Delta-Gamma-Normal Approximations*

Stefan R. Jaschke[a]

[a]Weierstraß-Institut für Angewandte Analysis und Stochastik, Mohrenstraße 39, 10117 Berlin, Germany, stefan@jaschke-net.de

Version 1.3, July 27, 2001

Abstract. Qualitative and quantitative properties of the Cornish-Fisher-Expansion in the context of Delta-Gamma-Normal approaches to the computation of Value at Risk are presented. Some qualitative deficiencies of the Cornish-Fisher-Expansion – the monotonicity of the distribution function as well as convergence are not guaranteed – make it seem unattractive. In many practical situations, however, its actual accuracy is more than sufficient and the Cornish-Fisher-approximation can be computed faster (and simpler) than other methods like numerical Fourier inversion. This paper tries to provide a balanced view on when and when not to use Cornish-Fisher in this context.

Keywords: Value at Risk, Delta-Gamma-Normal, Cornish-Fisher expansion, Edgeworth series, Gram-Charlier series
JEL Classification: C10

1 Introduction

Financial institutions are facing the important task of estimating and controlling their exposure to market risk, which is caused by changes in prices of equities, commodities, exchange rates and interest rates. A new chapter of risk management was opened when the Basel Committee on Banking Supervision proposed that banks may use internal models for estimating their market risk Basel Committee on Banking Supervision (1995). Its implementation into national laws around 1998 allowed banks to not only compete in the innovation of financial products but also in the innovation of risk management methodology.

Many alternatives exist for the statistical and computational decisions to be made for the computation of Value at Risk (VaR), which is the quantile of a portfolio's loss distribution over a given horizon. One of the more basic model assumptions is that the change in a firm's portfolio value over a specified horizon can be modeled as

$$V = \theta + \Delta^\top X + \frac{1}{2} X^\top \Gamma X, \qquad (1)$$

where X is a multivariate conditionally Gaussian vector and θ, Δ, and Γ are a scalar, a vector, and a matrix of parameters, respectively, derived from the current portfolio positions. This model has been the work-horse for quick, online computations of VaR since its use by RiskMetrics (Longerstaey; 1996), despite doubts about the suitability of the two model assumptions – Gaussian innovations and nearly quadratic price functions – in specific situations.

Several methods have been proposed to compute a quantile of the distribution defined by the model (1), among them Monte-Carlo simulation (Pritsker; 1996), Johnson transformations (Zangari; 1996a; Longerstaey; 1996), Cornish-Fisher expansions (Zangari; 1996b; Fallon; 1996), the Solomon-Stephens approximation (Britton-Jones and Schaefer; 1999), moment-based approximations

*Many thanks to Michael Dziedzina and Peter Nettesheim of Bankgesellschaft Berlin for the real-world sample portfolios. I also benefitted greatly from discussions with Stefan Daske, Matthias Fengler, Yuze Jiang, Peter Mathé, Jörg Polzehl, and Thorsten Sauder.

motivated by the theory of estimating functions (Li; 1999), saddle-point approximations (Rogers and Zane; 1999), and Fourier-inversion (Rouvinez; 1997; Albanese et al.; 2000). What makes the Normal-Delta-Gamma model especially tractable is that the characteristic function of the probability distribution, i.e., the Fourier transform of the probability density, of the quadratic form (1) is known analytically.

Pichler and Selitsch (1999) compare five different VaR-methods: Johnson transformations, Delta-Normal, and Cornish-Fisher-approximations up to the second, fourth and sixth moment. The sixth-order Cornish-Fisher-approximation compares well against the other techniques and is the final recommendation. Mina and Ulmer (1999) also compare Johnson transformations, Fourier inversion, Cornish-Fisher approximations, and Monte Carlo simulation ("Partial Monte Carlo"). Johnson transformations are concluded to be "not a robust choice". Cornish-Fisher is "extremely fast" compared to Partial Monte Carlo and Fourier inversion, but not as robust, as it gives "unacceptable results" in one of the four sample portfolios.

The contribution of this paper is to collect more than anecdotal evidence on the theoretical and empirical properties of the Cornish-Fisher expansion to allow a better decision on when and when not to use Cornish-Fisher in favour of Fourier inversion or Partial Monte Carlo in this context.

Section 2 recalls results on the family of distributions defined by (1). Section 3 recollects the main ideas of the derivation of the Cornish-Fisher expansion. The qualitative properties discussed in section 4 include monotonicity, tail behavior, and convergence. The quantitative results of section 5 include worst-case errors of Cornish-Fisher approximations on a certain subset of the family defined by (1) as well as approximation errors on real-world sample portfolios.

2 Delta-Gamma-Normal Models

Equation (1), $V = \theta + \Delta^\top X + \frac{1}{2} X^\top \Gamma X$, defines the class of Delta-Gamma-Normal models. X is assumed to be (conditionally) Gaussian with mean 0 and covariance matrix Σ. The change in the portfolio value, V, can be expressed as a sum of independent random variables that are quadratic functions of standard normal random variables Y_i by means of the solution of the generalized eigenvalue problem

$$CC^\top = \Sigma,$$
$$C^\top \Gamma C = \Lambda.$$

This implies

$$V = \theta + \sum_{i=1}^{m} (\delta_i Y_i + \frac{1}{2}\lambda_i Y_i^2) \tag{2}$$
$$= \theta + \sum_{i=1}^{m} \left\{ \frac{1}{2}\lambda_i \left(\frac{\delta_i}{\lambda_i} + Y_i \right)^2 - \frac{\delta_i^2}{2\lambda_i} \right\} \tag{3}$$

with $X = CY$, $\delta = C^\top \Delta$ and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_m)$. Packages like LAPACK (Anderson et al.; 1999) contain routines directly for the generalized eigenvalue problem. Otherwise C and Λ can be computed in two steps:

1. Compute some matrix B with $BB^\top = \Sigma$. If Σ is positive definite, the fastest method is Cholesky decomposition. Otherwise an eigenvalue decomposition can be used.

2. Solve the (standard) symmetric eigenvalue problem for the matrix $B^\top \Gamma B$:

   $$Q^\top B^\top \Gamma B Q = \Lambda$$

   with $Q^{-1} = Q^\top$ and set $C := BQ$.

The characteristic function of a non-central χ_1^2 variate ($(Z+a)^2$, with standard normal Z) is known analytically:

$$\mathrm{E} e^{it(Z+a)^2} = (1-2it)^{-1/2} \exp\left(\frac{a^2 it}{1-2it} \right).$$

This implies the characteristic function for V

$$\mathrm{E} e^{itV} = e^{i\theta t} \prod_j \frac{1}{\sqrt{1 - i\lambda_j t}} \exp\{-\frac{1}{2}\delta_j^2 t^2 / (1 - i\lambda_j t)\}, \tag{4}$$

which can be re-expressed in terms of Γ and B

$$\mathrm{E} e^{itV} = e^{i\theta t} \det(I - itB^\top \Gamma B)^{-1/2}$$
$$\times \exp\{-\frac{1}{2} t^2 \Delta^\top B (I - itB^\top \Gamma B)^{-1} B^\top \Delta\}, \tag{5}$$

or in terms of Γ and Σ

$$\begin{aligned}\mathrm{E}e^{itV} &= e^{i\theta t}\det(I - it\Gamma\Sigma)^{-1/2} \\ &\quad \times \exp\{-\frac{1}{2}t^2\Delta^\top\Sigma(I - it\Gamma\Sigma)^{-1}\Delta\}.\end{aligned} \tag{6}$$

Numerical Fourier-inversion of (4) can be used to compute an approximation to the cumulative distribution function (cdf) F of V. (The α-quantile is computed by root-finding in $F(x) = \alpha$.) The cost of the Fourier-inversion is $\mathcal{O}(N\log N)$, the cost of the function evaluations is $\mathcal{O}(mN)$, and the cost of the eigenvalue decomposition is $\mathcal{O}(m^3)$. The cost of the eigenvalue decomposition dominates the other two terms for accuracies of one or two decimal digits and the usual number of risk factors of more than a hundred. Instead of a full spectral decomposition, one can also just reduce $B^\top\Gamma B$ to tridiagonal form $B^\top\Gamma B = QTQ^\top$. ($T$ is tridiagonal and Q is orthogonal.) Then the evaluation of the characteristic function in (5) involves the solution of a linear system with the matrix $I - itT$, which costs only $\mathcal{O}(m)$ operations. An alternative route is to reduce $\Gamma\Sigma$ to Hessenberg form $\Gamma\Sigma = QHQ^\top$ or to do a Schur decomposition $\Gamma\Sigma = QRQ^\top$. (H is Hessenberg and Q is orthogonal. Since $\Gamma\Sigma$ has the same eigenvalues as $B^\top\Gamma B$ and they are all real, R is actually triangular instead of quasi-triangular in the general case. See (Anderson et al.; 1999).) The evaluation of (6) becomes $\mathcal{O}(m^2)$, since it involves the solution of a linear system with the matrix $I - itH$ or $I - itR$, respectively. Reduction to tridiagonal, Hessenberg, or Schur form is also $\mathcal{O}(m^3)$, so the asymptotics in the number of risk factors m remain the same in all cases. The critical N, above which the complete spectral decomposition + fast evaluation via (4) is faster than the reduction to tridiagonal or Hessenberg form + slower evaluation via (5) or (6) remains to be determined empirically for given m on a specific machine.

The advantage of the Cornish-Fisher approximation is that it is based on the cumulants, which can be computed without any matrix decomposition:

$$\begin{aligned}\kappa_1 &= \theta + \frac{1}{2}\sum_i \lambda_i \\ &= \theta + \frac{1}{2}\operatorname{tr}(\Gamma\Sigma),\end{aligned}$$

and for $r \geq 2$

$$\begin{aligned}\kappa_r &= \frac{1}{2}\sum_i \{(r-1)!\lambda_i^r + r!\delta_i^2\lambda_i^{r-2}\} \\ &= \frac{1}{2}(r-1)!\operatorname{tr}((\Gamma\Sigma)^r) + \frac{1}{2}r!\Delta^\top\Sigma(\Gamma\Sigma)^{r-2}\Delta.\end{aligned}$$

Although the cost of computing the cumulants needed for the Cornish-Fisher approximation is also $\mathcal{O}(m^3)$, this method can be faster then the eigenvalue decomposition for small orders of approximation and relatively small numbers of risk factors[2].

Partial Monte-Carlo (or partial Quasi-Monte-Carlo) costs $O(m^2)$ operations per sample. If Γ is sparse, it may cost even less. The number of samples needed is a function of the desired accuracy. It is clear from the asymptotic costs of the three methods that partial Monte Carlo will be preferable for sufficiently large m.

While Fourier-inversion and partial Monte-Carlo can in principal achieve any desired accuracy, the Cornish-Fisher approximations provide only a limited accuracy as shown in the next sections.

3 Cornish-Fisher-, Gram-Charlier-, and Edgeworth-Expansions

The Cornish-Fisher expansion can be derived in two steps. Let Φ denote some base distribution and ϕ its density function. The generalized Cornish-Fisher expansion (Hill and Davis; 1968) aims to approximate an α-quantile of F in terms of the α-quantile of Φ, i.e., the concatenated function $F^{-1}\circ\Phi$. The key to a series expansion of $F^{-1}\circ\Phi$ in terms of derivatives of F and Φ is Lagrange's inversion theorem. It states that if a function $s \mapsto t$ is implicitly defined by

$$t = c + s \cdot h(t) \tag{7}$$

[2]If a large number r of cumulants is needed, it is better to do a spectral decomposition (of $B^\top\Gamma B$ to diagonal) or a Schur decomposition (of $\Gamma\Sigma$ to triangular) once and then compute the higher cumulants with $\mathcal{O}(m)$ operations each. If r is small but m is large, it is better to do a reduction of $B^\top\Gamma B$ to tridiagonal or a reduction of $\Gamma\Sigma$ to Hessenberg form and again compute the higher cumulants with $\mathcal{O}(m)$ operations each.

and h is analytic in c, then an analytic function $f(t)$ can be developed into a power series in a neighborhood of $s = 0$ ($t = c$):

$$f(t) = f(c) + \sum_{r=1}^{\infty} \frac{s^r}{r!} D^{r-1}[f' \cdot h^r](c), \qquad (8)$$

where D denotes the differentation operator. For a given probability $c := \alpha$, $f := \Phi^{-1}$, and $h := (\Phi - F) \circ \Phi^{-1}$ this yields

$$\Phi^{-1}(t) = \Phi^{-1}(\alpha) + \sum_{r=1}^{\infty} (-1)^r \frac{s^r}{r!} D^{r-1}[((F - \Phi)^r/\phi) \circ \Phi^{-1}](\alpha). \qquad (9)$$

Setting $s = 1$ in (7) implies $\Phi^{-1}(t) = F^{-1}(\alpha)$ and with the notations $x := F^{-1}(\alpha)$, $z := \Phi^{-1}(\alpha)$ (9) becomes the formal expansion[3]

$$x = z + \sum_{r=1}^{\infty} (-1)^r \frac{1}{r!} D^{r-1}[((F - \Phi)^r/\phi) \circ \Phi^{-1}](\Phi(z)).$$

With $a := (F - \Phi)/\phi$ this can be written as

$$x = z + \sum_{r=1}^{\infty} (-1)^r \frac{1}{r!} D_{(r-1)}[a^r](z) \qquad (10)$$

with $D_{(r)} = (D + \frac{\phi'}{\phi})(D + 2\frac{\phi'}{\phi}) \ldots (D + r\frac{\phi'}{\phi})$ and $D_{(0)}$ being the identity operator.

(10) is the generalized Cornish-Fisher expansion. The second step is to choose a specific base distribution Φ and a series expansion for a. The classical Cornish-Fisher expansion is recovered if Φ is the standard normal distribution, a is (formally) expanded into the Gram-Charlier series, and the terms are re-ordered as described below.

The idea of the Gram-Charlier series is to develop the ratio of the moment generating function of V ($M(t) = \mathrm{E}e^{tV}$) and the moment generating function of the standard normal distribution ($e^{t^2/2}$) into a power series at 0:

$$M(t)e^{-t^2/2} = \sum_{k=0}^{\infty} c_k t^k. \qquad (11)$$

(c_k are the Gram-Charlier coefficients. They can be derived from the moments by multiplying the power series for the two terms on the left hand side.) Componentwise Fourier inversion yields the corresponding series for the probability density

$$f(x) = \sum_{k=0}^{\infty} c_k (-1)^k \phi^{(k)}(x) \qquad (12)$$

and for the cumulative distribution function (cdf)

$$F(x) = \Phi(x) - \sum_{k=1}^{\infty} c_k (-1)^{k-1} \phi^{(k-1)}(x). \qquad (13)$$

(ϕ und Φ are now the standard normal density and cdf.[4]) Plugging (13) into (10) gives the formal Cornish-Fisher expansion, which is re-grouped as motivated by the central limit theorem.

Assume that V is already normalized ($\kappa_1 = 0$, $\kappa_2 = 1$) and consider the normalized sum of independent random variables V_i with the distribution F, $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i$. The moment generating function of the random variable S_n is

$$M_n(t) = M(t/\sqrt{n})^n = e^{t^2/2} (\sum_{k=0}^{\infty} c_k t^k n^{-k/2})^n.$$

Multiplying out the last term shows that the k-th Gram-Charlier coefficient $c_k(n)$ of S_n is a polynomial expression in $n^{-1/2}$, involving the coefficients c_i up to $i = k$. If the terms in the formal Cornish-Fisher expansion

$$x = z + \sum_{r=1}^{\infty} (-1)^r \frac{1}{r!} D_{(r-1)} \left[\left(- \sum_{k=1}^{\infty} c_k(n) H_{k-1} \right)^r \right] (z) \qquad (14)$$

are sorted and grouped with respect to powers of $n^{-1/2}$, the classical Cornish-Fisher series

$$x = z + \sum_{k=1}^{\infty} n^{-k/2} \xi_k(z) \qquad (15)$$

results. The similarly re-sorted Gram-Charlier series is called Edgeworth series:

$$M_n(t) = e^{t^2/2} \sum_{k=0}^{\infty} n^{-k/2} h_k(t), \qquad (16)$$

[3] Conditions under which $s = 1$ is in the convergence radius of the series (9) will be looked at in section 4.

[4] The derivatives of the standard normal density are $(-1)^k \phi^{(k)}(x) = \phi(x) H_k(x)$, where the Hermite polynomials H_k form an orthogonal basis in the Hilbert space $L^2(\mathbb{R}, \phi)$ of the square integrable functions on $\mathbb{R}$ w.r.t. the weight function ϕ. The Gram-Charlier coefficients can thus be interpreted as the Fourier coefficients of the function $f(x)/\phi(x)$ in the Hilbert space $L^2(\mathbb{R}, \phi)$ with the basis $\{H_k\}$: $f(x)/\phi(x) = \sum_{k=0}^{\infty} c_k H_k(x)$.

where $h_k(t)$ are the Cramér-Edgeworth polynomials in t (of degree $3k$) (compare (Skovgaard; 1999)). Componentwise Fourier inversion yields again the analogous Edgeworth series for the density f_n and the cdf F_n of the sum S_n:

$$f(x) = \sum_{k=0}^{\infty} n^{-k/2} h_k(-\frac{d}{dx})\phi(x).$$

It is a relatively tedious process to express the adjustment terms ξ_k correponding to a certain power $n^{-k/2}$ in the Cornish-Fisher expansion (15) directly in terms of the cumulants κ_r, see (Hill and Davis; 1968). Lee developed a recurrence formula for the k-th adjustment term ξ_k in the Cornish-Fisher expansion, which is implemented in the algorithm AS269 (Lee and Lin; 1992, 1993):[5]

$$\xi_k(H) = a_k H^{*(k+1)} \\ - \sum_{j=1}^{k-1} \frac{j}{k}(\xi_{k-j}(H) - \xi_{k-j}) * (\xi_j - a_j H^{*(j+1)}) * H, \tag{17}$$

with $a_k = \frac{\kappa_{k+2}}{(k+2)!}$. $\xi_k(H)$ is a formal polynomial expression in H with the usual algebraic relations between the summation "+" and the "multiplication" "$*$". Once $\xi_k(H)$ is multiplied out in $*$-powers of H, each H^{*k} is to be interpreted as the Hermite polynomial H_k and then the whole term becomes a polynomial in z with the "normal" multiplication "$\cdot$". ξ_k denotes the scalar that results when the "normal" polynomial $\xi_k(H)$ is evaluated at the fixed quantile z, while $\xi_k(H)$ denotes the expression in the $(+, *)$-algebra.

4 Qualitative Properties of the Cornish-Fisher Expansion

The qualitative properties of the Cornish-Fisher expansion are:

+ If F_m is a sequence of distributions converging to the standard normal distribution Φ, the Edgeworth- and Cornish-Fisher approximations present better approximations (asymptotically for $m \to \infty$) than the normal approximation itself.

– The approximated functions $\tilde{F}$ and $\tilde{F}^{-1} \circ \Phi$ are not necessarily monotone.

– $\tilde{F}$ has the "wrong tail behavior", i.e., the Cornish-Fisher approximation for α-quantiles becomes less and less reliable for $\alpha \to 0$ (or $\alpha \to 1$).

– The Edgeworth- and Cornish-Fisher approximations do not necessarily improve (converge) for a fixed F and increasing order of approximation, k.

Figure 1 shows the true and the approximated quantile functions F^{-1} for the distribution of $-Y^2$, where Y is standard normal. It illustrates the three qualitative deficiencies of the Cornish-Fisher approximation.

Convergence for $F_m \to \Phi$

The most prominent use and motivation of the Edgeworth- and Cornish-Fisher expansions is in the context of the central limit theorem, when F_m is the distribution of the normalized sum of independent random variables. It is clear from (15) and (16) that the Edgeworth and Cornish-Fisher approximations present higher order approximations to F_m than the normal approximation itself. Necessary and sufficient for convergence in the central limit theorem is Lindeberg's condition. I.e., the distribution of V need not converge to normal for increasing number of risk factors if the contribution to the variance of V by a few components $\delta_i Y_i + \frac{1}{2}\gamma_i Y_i^2$ is dominant.

Monotonicity

The Cornish-Fisher expansion approximates the monotone function $F^{-1} \circ \Phi$ by polynomials. It is clear that a necessary condition for monotonicity of F is that the degree of the polynomial is odd, which is the case when the highest order of the Cornish-Fisher expansion, k, is even.

[5] We write the recurrence formula here, because it is incorrect in (Lee and Lin; 1992).

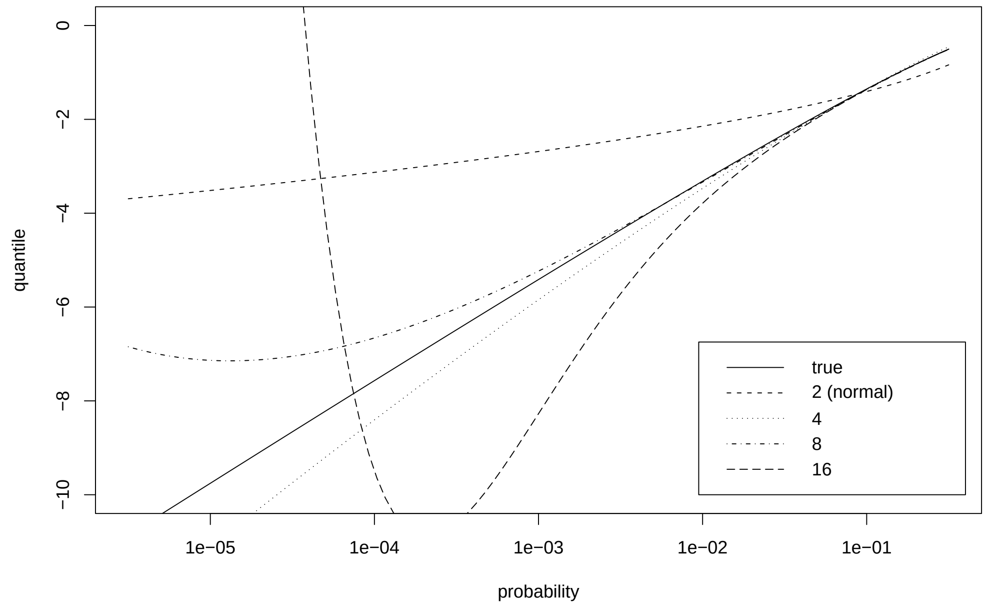

Figure 1: Cornish-Fisher approximations of the quantile function of the negative of a χ_1^2 variate (one risk factor, $\gamma = -1$, $\delta = 0$). The number in the legend is the highest cumulant used.

Tail Behavior

Let p denote the polynomial that approximates $F^{-1} \circ \Phi$, i.e., the random variable at hand is approximated by the random variable $p(Z)$ for a standard normal Z. Assume that p is monotone, so that p^{-1} is well defined. For $z \to \infty$ $p(z)$ behaves like cz^d and $p^{-1}(x)$ like $(x/c)^{1/d}$. Then the probability density of $p(Z)$ is

$$\begin{aligned} \tilde{f}(x) &= \phi(p^{-1}(x))[p^{-1}(x)]' \\ &\sim e^{-\frac{1}{2}(x/c)^{(2/d)}} \frac{1}{cd}(x/c)^{1/d-1} \end{aligned} \tag{18}$$

for $x \to \infty$. For one-dimensional problems with $\gamma_1 > 0$ the density of V has (up to a constant factor) the tail given by (18) with $d = 2$. Clearly, the tail behavior of the approximation deviates more and more from the true tail behavior for $d \to \infty$.

Convergence in the Approximation Order k

The key theorem for the convergence of power series is Cauchy-Hadamard's theorem, which states that a power series $\sum_{k=0}^{\infty} a_k (z - z_0)^k$ converges in the circle around z_0 with the radius

$$r = \frac{1}{\limsup \sqrt[k]{|a_k|}}$$

and diverges outside of that circle. The convergence in the interior of the circle is absolute, that is, it also holds for re-sorted series. If f has a singularity at z_1 and the Taylor series is developed at the point z_0 (i.e., $a_k = \frac{1}{k!} \frac{d^k}{dz^k} f(z_0)$), then the theorem implies $r \le |z_1 - z_0|$, .

Since the moment generating function $M(t)$ of V has poles at $t = 1/\lambda_i$, the convergence radius of the series (11) is at most $1/|\lambda|_{\max}$. Application of the convergence theorem for characteristic functions implies that the Gram-Charlier-series for the cdf (13) cannot converge weakly. (Otherwise (11) should converge uniformly on closed intervals of the imaginary axis.)

The Edgeworth expansion (16) can be interpreted as Tayler series expansion of the function

$$f_t(\tau) = e^{-t^2/2} M(t\tau)^{1/\tau^2}$$

in τ (with $\tau = 1/\sqrt{n}$). Since the moment generating function M has poles at $1/\lambda_i$, the function $\tau \mapsto f_t(\tau)$ has poles at $1/(t\lambda_i)$. The Edgeworth series for $n = 1$ ($\tau = 1$) does not converge if the convergence radius of the Taylor series expansion of $\tau \mapsto f_t(\tau)$ is less than 1, which is the case for $t > 1/|\lambda|_{\max}$. This leads to the following result.

Proposition 1 *The Edgeworth series for the moment generating function (16) (with $n = 1$) converges pointwise on the imaginary axis and the corresponding Edgeworth series for the distribution function converges weakly for a distribution F from the family defined by (1) if and only if F is a normal distribution ($\Gamma = 0$). The same holds for the Gram-Charlier series (11) and (13).*

The Cornish-Fisher expansion for a given normal quantile z and for a distribution F depends on the value and all derivatives of the Edgeworth approximation for F at the point z. Since the Edgeworth expansion does not converge for all non-normal F from the delta-gamma-normal family, it is plausible that the Cornish-Fisher expansion also fails to converge for a large subclass of the family. (A precise characterization of the set of convergence seems difficult because of the two-step derivation of the Cornish-Fisher expansion.)

Even a converging series $\tilde{a}_k \to a = (F - \Phi)/\phi$ instead of the Edgeworth expansion may not lead to a converging generalized Cornish-Fisher expansion (10), however. (10) does not converge if the convergence radius of the power series (9) is less than one. For a fixed probability level α, (9) is a Taylor series expansion of the inverse of the function $s(x) = \frac{\Phi(x)-\alpha}{\Phi(x)-F(x)}$ at $s = 0$ ($x = \Phi^{-1}(\alpha)$). Since $\Phi(x) - F(x)$ usually changes sign one or more times, s is not globally invertable. Let s^{-1} denote the inverse of s in that neighborhood of $x = \Phi^{-1}(\alpha)$ where s is monotone. Among the reasons, why the generalized Cornish-Fisher expansion may not converge, are:

1. The neighborhood where s^{-1} is defined may not contain the interval $(-1, 1]$. This is the case for $\alpha = 0.25$ and the normalized χ_1^2 distribution ($\gamma_1 = 1, \delta_1^2 = \frac{1}{2}, \theta_1 = -\frac{1}{2}$), for example.

2. If all eigenvalues λ_i are non-zero, the cdf F of V has a singularity at

$$x_0 := \sum_{i=1}^{m} (\theta_i - \frac{1}{2}\delta_i^2/\lambda_i).$$

(F is C^∞ except in x_0, where the highest continuous derivative has order $[(m-1)/2]$. Consequently, the generalized Cornish-Fisher expansion cannot converge if $s^{-1}(-1, 1) \ni x_0$. This is the case for $\alpha < \Phi(-0.75)$ and the normalized χ_1^2 distribution, for example.

5 Quantitative Properties of the Cornish-Fisher Expansion

5.1 Worst-Case Errors

Consider an approximation method $(\alpha, \theta, \Delta, \Gamma) \mapsto Q(\alpha, \theta, \Delta, \Gamma)$, where α is the probability level and $Q(\alpha, \ldots)$ the corresponding approximated quantile. The criterion considered here is the "worst-case error relative to the standard deviation of the portfolio":

$$\begin{aligned} e(\alpha) := \sup \{ & |Q(\alpha, \theta, \Delta, \Gamma) - q_\alpha(\theta, \Delta, \Gamma) \\ & \text{s.t.} \quad \mu(\theta, \Delta, \Gamma) = 0, \sigma(\theta, \Delta, \Gamma) = 1 \}, \end{aligned} \tag{19}$$

where $q_\alpha(\ldots)$ is the true quantile, $\mu(\ldots)$ the expectation, and $\sigma(\ldots)$ the standard deviation of the distribution of V with parameters (θ, Δ, Γ). The alternative criterion "relative error" has the problem that the true quantile may be close to zero (due to a well-hedged portfolio, for instance).

The worst-case view may seem exaggerated and far from practice, as (1) the Cornish-Fisher approximations in fact achieve much higher accuracy near the normal distribution than at the worst case and (2) well-diversified firm-wide portfolios typically are relatively close to the normal distribution. Risk-management systems, however, are usually applied at all levels of aggregation, i.e., also at the trading desk level. At these levels, a few risk factors may dominate the picture. We argue, moreover, that the essence of risk measurement is to provide *estimates of bounds on what can go wrong*. In this sense, approximation methods for risk measurement should be judged based on what accuracy they can *very likely guarantee*.

The distribution of V is close to normal if in the decomposition (2)

1. δ_i is large compared to λ_i, for all i, since then each individual of the independent random variables is close to normal, or

2. there is a large number of i where λ_i is large compared to δ_i, but all those λ_i are approximately of the same size, because then the central limit theorem applies.

Since the Cornish-Fisher approximations are especially good near the normal distribution, the worst-case will likely occur when the exposure to one or two risk factors dominates all others. Thus the worst-case errors on the classes of one- and two-dimensional problems provide interesting lower bounds on the worst-case error on the whole family of distributions.[6]

The analysis of the one-dimensional sub-family is simplified by the fact that the density and cdf of a non-central χ_1^2 are known analytically:

$$P\{(Z+a)^2 \leq x\} = \Phi(\sqrt{x}-a) - \Phi(-\sqrt{x}-a) \quad (20)$$

for $x \geq 0$, and 0 otherwise. (Z is standard-normal.) The density of the non-central χ_1^2 with non-centrality parameter a is consequently

$$\begin{aligned} f(x;a) &= \frac{1}{2\sqrt{x}}(\phi(\sqrt{x}-a) + \phi(-\sqrt{x}-a)) \\ &= \frac{1}{\sqrt{2\pi x}} e^{-(x+a^2)/2} \cosh(a\sqrt{x}) \end{aligned} \quad (21)$$

for $x \geq 0$ and 0 otherwise.

This implies the cdf for a one-dimensional delta-gamma-normal-variate (excluding the trivial case where both δ and λ are zero):

$$P\{\theta + \delta Z + \frac{1}{2}\lambda Z^2 \leq x\} =$$

$$\begin{cases} \Phi(y-a) - \Phi(-y-a) & \lambda > 0, x \geq \theta - \frac{\delta^2}{2\lambda} \\ 1 - \Phi(y-a) + \Phi(-y-a) & \lambda < 0, x \geq \theta - \frac{\delta^2}{2\lambda} \\ \Phi(\frac{x-\theta}{\delta}) & \lambda = 0 \\ 0 & \text{otherwise} \end{cases} \quad (22)$$

with $y := \sqrt{\frac{2}{\lambda}(x - \theta + \frac{\delta^2}{2\lambda})}$ and $a := \delta/\lambda$. The density for the one-dimensional delta-gamma-normal-variate is

$$f(x;\theta,\delta,\lambda) = \begin{cases} \frac{2}{|\lambda|} f(y^2;a) & \lambda \neq 0, x \geq \theta - \frac{\delta^2}{2\lambda} \\ \phi(\frac{x-\theta}{\delta})/\delta & \lambda = 0 \\ 0 & \text{otherwise.} \end{cases} \quad (23)$$

We parameterize the family of probability distributions by $\lambda \in [-\sqrt{2}, \sqrt{2}]$. $\theta := -\lambda/2$ and $\delta := \sqrt{1 - \lambda^2/2}$ ensure mean 0 and standard deviation 1. The true 0.01-quantile increases monotonically from about -3.984σ to -0.707σ. This means that the worst-case error (on the one-dimensional sub-family) of the normal-quantile approximation – taking the normal-quantile but computing the variance from δ and λ - is about 1.658σ, realized for a short-gamma position.

Figure 2 shows the approximation error of the Cornish-Fisher approximations using up to the second, fourth, eighth, and sixteenth cumulant, respectively. It shows that the higher order approximations have increasing accuracy near the normal distribution, but become less reliable far from the normal distribution.

Figure 3 shows the worst-case error on the one- and two-dimensional sub-families for increasing order of approximation. The one-dimensional sub-family obviously is not rich enough to expose the weaknesses of the Cornish-Fisher approximation.

5.2 A Real-World Example

The data provided by the Bankgesellschaft Berlin contain

[6] The numerical computation of the worst case error $e(\alpha)$ for higher numbers of risk factors $m >> 1$, a given probability level α, and Cornish-Fisher approximation Q appears to be intrinsically difficult, as the function $Q(\alpha,\theta,\delta,\Lambda) - q_\alpha(\theta,\delta,\Lambda)$ has many local optima.

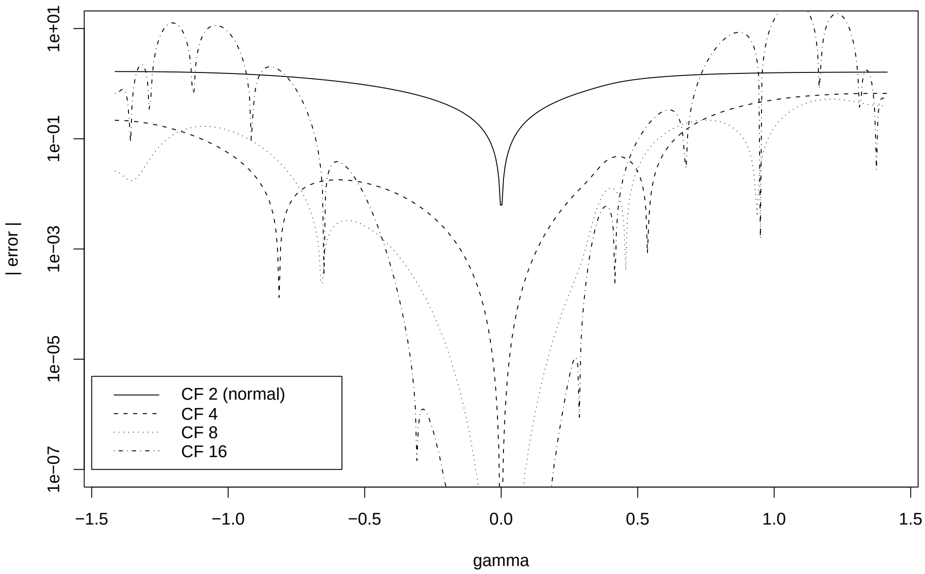

Figure 2: The approximation error for the 1%-quantile on the one-dimensional sub-family of distributions. The number in the legend is the highest cumulant used. $\gamma = 0$ is the normal distribution. “CF 2” is the normal approximation.

- volatilities (standard deviations) and correlations of daily risk factor changes and
- aggregated sensitivities (first and second derivatives of the portfolio value function w.r.t. the risk factors) for two portfolios

on two dates. 928 risk factors are in use. Before doing the eigenvalue decomposition, empty (zero) rows and columns in Γ are eliminated in order to reduce the dimension. The last two columns of table 1 contain the setup costs for the Fourier inversion and Cornish-Fisher approximation, respectively, using standard methods of the statistical software package **R** (Ihaka and Gentleman; 1996, development version May 2001) on an Athlon with 750MHz. Both computations are suboptimal, so the times are to be taken as an upper bound on what can be achieved.

The $\mathcal{O}(m^3)$-contributions to the cost of the Fourier inversion are two matrix multiplications ($B^\top \Gamma B$, BLAS routine DGEMM[7]) and a reduction to a tridiagonal matrix ($B^\top \Gamma B = QTQ^\top$, LAPACK routine DSYTRD). The $\mathcal{O}(m^3)$-operations needed for the Cornish-Fisher approximation up to the $k-$th cumulant are either k matrix multiplications or one matrix multiplication and one reduction to Hessenberg form (LAPACK routine DGEHRD). Table 2 shows that the computation of the first four cumulants is *not* significantly faster than the initial decomposition needed for the Fourier inversion.[8]

The final table 3 shows the 99%-VaR (after normalization to $\sigma = 1$) for the four cases, computed with the Cornish-Fisher approximation using up to the fourth cumulant as well as a Fourier inversion. The numbers for skewness and curtosis suggest that

[7]DSYMM is not significantly faster than DGEMM.

[8]Most vector-vector (BLAS1) and matrix-vector (BLAS2) routines are memory-bound instead of CPU-bound on current machines. Some (blocked versions of) algorithms can benefit better from cache hierarchies than others, which explains why the algorithm with the highest operations count actually is the fastest on this machine.

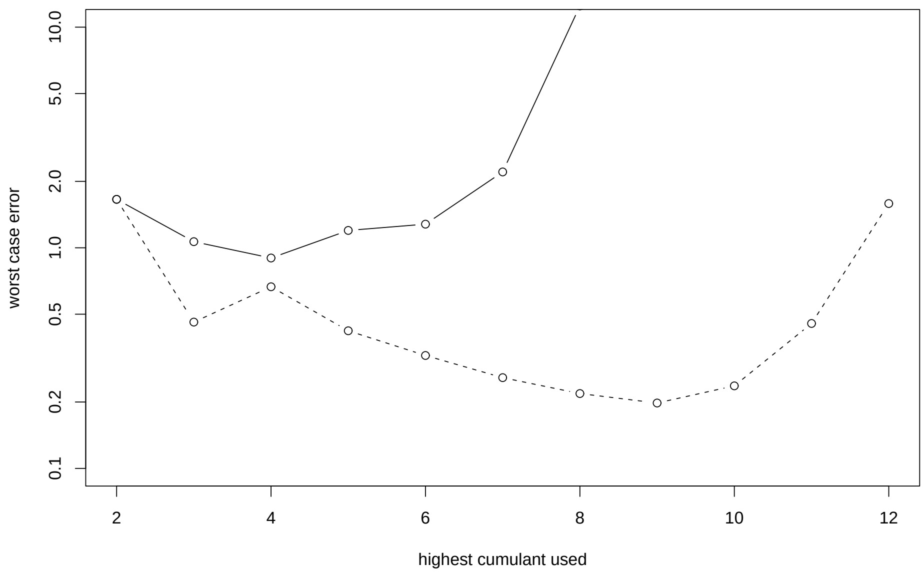

Figure 3: The worst-case error for the 1%-quantile on the one- (dotted) and two-dimensional (solid) sub-families of distributions for increasing order of approximation.

case	relevant risk factors	nonzero gammas	computing time in seconds spectral decomp.	4 cumulants
1	113	731	0.05	0.03
2	111	697	0.05	0.03
3	218	650	0.30	0.14
4	209	607	0.25	0.13

Table 1: Dimensions and actual computing times of the four real-world sample portfolios. The third column contains computing times for the matrix multiplication $B^\top \Gamma B$ and the eigenvalue decomposition of the matrix $B^\top \Gamma B$. Since the estimate of Σ usually only changes once per day, the decomposition $BB^\top = \Sigma$ can be done offline and is not counted towards the initial costs of the Fourier inversion. The computation of the four cumulants (fourth column) uses four matrix multiplications (instead of a reduction to Hessenberg form).

problem	floating point operations	MFLOPS	time in nanoseconds
DGEMM	$2\,m^3$	800	$2.5m^3$
DSYTRD	$4/3\,m^3$	200	$6.7m^3$
DGEHRD	$10/3\,m^3$	250	$13.3m^3$
FI (2 DGEMM + 1 DSYTRD)	$5.3m^3$	457	$11.7m^3$
CF4 (1 DGEMM + 1 DGEHRD)	$5.3m^3$	336	$15.8m^3$
CF4 (4 DGEMM)	$8m^3$	800	$10.0m^3$
CF4 (4 SGEMM), 3DNow!	$8m^3$	1850	$4.3m^3$

Table 2: Estimated computing times using ATLAS (Whaley et al.; 2000) version 3.2.1 and LAPACK (Anderson et al.; 1999) version 3.0 on an Athlon with 750MHz. The first three lines contain the operation counts and timing for the building block routines. "FI" denotes the initial cost for the Fourier inversion and "CF4" the initial cost for computing the first four cumulants. The last line is not really comparable, as "3DNow!" yields only single precision and does not fully support IEEE arithmetic.

case	skewness	curtosis	VaR_{FI}	$\mathrm{VaR}_{CF4} - \mathrm{VaR}_{FI}$	$\mathrm{VaR}_{CF4} - \mathrm{VaR}_{CF4'}$
1	0.093	0.012	2.238191	1.663802e-06	4.174439e-14
2	0.092	0.012	2.238394	1.841259e-06	4.396483e-14
3	0.017	0.001	2.309548	-1.618836e-06	-3.552714e-15
4	0.019	0.001	2.306958	-2.322726e-06	-4.440892e-16

Table 3: Skewness, Curtosis, 99%-VaR, and Differences. Column 3 contains the 99%-VaR, normalized to $\sigma = 1$. The difference between the Fourier inversion and the Cornish-Fisher approximation is in column 5. The last column contains the difference between the Cornish-Fisher approximations when the cumulants are computed from (Δ, Γ, Σ) and (δ, λ), respectively. It indicates the size of the error introduced by the eigenvalue decomposition.

the distributions are very close to normal. A QQ-Plot against normal confirms this. The actual accuracy of about $2 \cdot 10^{-6}$ is obviously more than sufficient.

6 Conclusion and Open Questions

In order to put the errors of the Cornish-Fisher approximations into perspective, look at the different error sources in the context of Delta-Gamma-Normal approaches to the computation of VaR:

1. random fluctuations that influence the estimate of the covariance matrix Σ,
2. deviations from model assumptions of (conditional) Gaussian risk factor changes,
3. differences between the real price function and its quadratic approximation, and
4. approximation errors of the Delta-Gamma-Normal method (Cornish-Fisher, Fourier inversion, partial Monte Carlo, ...).

Simple Monte-Carlo simulation shows that the error in the 99%-VaR to expect from fluctuations in the estimate of Σ is about 0.1σ for the "perfect case": $V = \theta + \Delta^\top X$ and X is normally distributed. It is about 0.3σ for the specific "Delta-Gamma-Normal case" $V = 0.5 + \frac{1}{2}\sqrt{2}X - \frac{1}{2}X^2$ where X is normal. (This is for the equally weighted covariance estimator with 250 trading days horizon and the square root of the expected squared error.) It makes no sense to strive for an accuracy in the fourth step that is much higher than 2 decimal digits, if the expected error in the first step is already 0.1σ, even in the best case. According to figure 2, the Cornish-Fisher approximation (up to the fourth cumulant) achieves an accuracy of 0.1σ on a relatively large neighborhood of the normal distribution.

Deviations from normality can in principal lead to large errors in the second step, but do not in practice, since risk factors are chosen by the modeler (risk manager). If a certain derivative security is by market convention expressed as a function of some underlying with very non-normal increments, some other, "more normal" risk factor is usually chosen by the risk manager and the nonlinearity is put into the price function (the "mapping"). In the case that risk factor innovations are in fact t-distributed with three degrees of freedom, but assumed normal, the error in the 99%-VaR is about 0.3σ (for a linear price function).

Using the Markov inequality, it can be shown that the ratio between the 1%-quantile and the standard deviation of a distribution with mean 0 can maximally be 10. A 99%-VaR of about 9.95σ actually appears in the following sample portfolio. A portfolio with normally distributed fluctuations is held. Additionally, digital put options with expiry date at the VaR horizon on an independent, normally distributed risk factor are sold:

$$V = w_1(p - \mathbf{1}_{\{X_1 \le K\}}) + w_2 X_2.$$

p is the premium for the digital put option. K is the strike of the digital put and assumed to be the standard normal p-quantile. X_1 and X_2 are independent standard normal. $w_1^2 p(1-p) + w_2^2 = 1$ ensures that the standard deviation of V is 1. Letting tend p to 0.01 from above and $w_2 \to 0$, the $99\% - \mathrm{VaR}$ approaches $\sqrt{99}\sigma$. This shows that – in the worst-case view – this error source is the most critical and is about one magnitude higher than the worst-case error of the Cornish-Fisher approximation.

The conclusion is that despite its qualitative shortcomings the Cornish-Fisher approximation is a competitive, and probably underrated, technique, which achieves a sufficient accuracy potentially faster than the other numerical techniques (mainly Fourier inversion and Partial Monte-Carlo) over a certain range of practical cases.

If one takes the worst-case view and cares about the corner cases – as we believe one should in the field of risk management – the potential errors from the quadratic approximation are much larger than the errors from the Cornish-Fisher expansion. Hence a full-valuation Monte-Carlo technique should be used anyway to frequently check the suitability of the quadratic approximation. This will also take care of the "bad" cases for the Cornish-Fisher approximation.

From a more theoretical point of view, there are several open questions. (1) Although we collected evidence that the two steps leading to the Cornish-Fisher expansion do not converge in many cases, the exact characterization of the set of parameters (θ, Δ, Γ) for which the Cornish-Fisher approximation converges, is open. (2) The worst-case errors on the one- and two-dimensional sub-families provide only lower bounds for the worst-case error on the whole family. It would be nice to have an upper bound. (3) The reason for the non-convergence of the Edgeworth expansion is that the tails of the considered probability densities are much "fatter" than the tail of the normal distribution. A generalized Cornish-Fisher expansion with a base distribution that has comparable tail behavior ("semi-heavy tails") could potentially lead to a converging expansion. (4) There are many alternative techniques described in the probability and statistics literature, like Ruben's series expansion in terms of Gamma distributions (Mathai and Provost; 1992), alternative series representations (Abate and Whitt; 1999b), saddlepoint approximations (Rogers and Zane; 1999; Daniels; 1987), and continued fractions (Abate and Whitt; 1999a).

References

Abate, J. and Whitt, W. (1999a). Computing Laplace transforms for numerical inversion via continued fractions, *INFORMS J. Comput.* **11**(4): 394–405.

Abate, J. and Whitt, W. (1999b). Infinite-series representations of Laplace transforms of probability density functions for numerical inversion, *J. Oper. Res. Soc. Japan* **42**(3): 268–285.

Albanese, C., Jackson, K. and Wiberg, P. (2000). Fast convolution method for VaR and VaR gradients, http://www.math-point.com/fconv.ps.

Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Croz, J. D., Greenbaum, A., Hammarling, S., McKenney, A. and Sorensen, D. (1999). *LAPACK Users' Guide*, third edn, SIAM. http://www.netlib.org/lapack/lug/.

Basel Committee on Banking Supervision (1995). An internal model-based approach to market risk capital requirements, http://www.bis.org/publ/bcbsc224.pdf.

Britton-Jones, M. and Schaefer, S. (1999). Non-linear Value-at-Risk, *European Finance Review* **2**: 161–187.

Daniels, H. E. (1987). Tail probability approximations, *International Statistical Review* **55**(1): 37–48.

Fallon, W. (1996). Calculating Value at Risk, http://wrdsenet.wharton.upenn.edu/fic/wfic/papers/96/9649.pdf. Wharton Financial Institutions Center Working Paper 96-49.

Hill, G. W. and Davis, A. W. (1968). Generalized asymptotic expansions of Cornish-Fisher type, *Ann. Math. Statist.* **39**: 1264–1273.

Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics, *Journal of Computational and Graphical Statistics* **5**(3): 299–314. http://www.r-project.org.

Lee, Y. S. and Lin, T. K. (1992). Higher-order Cornish Fisher expansion, *Applied Statistics* **41**: 233–240.

Lee, Y. S. and Lin, T. K. (1993). Correction to algorithm AS269 : Higher-order Cornish Fisher expansion, *Applied Statistics* **42**: 268–269.

Li, D. (1999). Value at Risk based on the volatility, skewness and kurtosis, http://www.riskmetrics.com/research/working/var4mm.pdf. RiskMetrics Group.

Longerstaey, J. (1996). RiskMetrics technical document, *Technical Report fourth edition*, J.P.Morgan. originally from http://www.jpmorgan.com/RiskManagement/RiskMetrics/, now http://www.riskmetrics.com.

Mathai, A. M. and Provost, S. B. (1992). *Quadratic forms in random variables*, Marcel Dekker Inc., New York.

Mina, J. and Ulmer, A. (1999). Delta-gamma four ways, http://www.riskmetrics.com.

Pichler, S. and Selitsch, K. (1999). A comparison of analytical VaR methodologies for portfolios that include options, http://www.tuwien.ac.at/E330/Research/paper_var.pdf. Working Paper TU Wien.

Pritsker, M. (1996). Evaluating Value at Risk methodologies: Accuracy versus computational time, http://wrdsenet.wharton.upenn.edu/fic/wfic/papers/96/9648.pdf. Wharton Financial Institutions Center Working Paper 96-48.

Rogers, L. and Zane, O. (1999). Saddle-point approximations to option prices, *Annals of Applied Probability* **9**(2): 493–503. http://www.bath.ac.uk/~maslcgr/papers/.

Rouvinez, C. (1997). Going greek with VaR, *Risk* **10**(2): 57–65.

Skovgaard, I. (1999). Asymptotic methods in statistical inference, http://www.dina.kvl.dk/~ib/asymp/plan.html. lecture notes.

Whaley, R. C., Petitet, A. and Dongarra, J. J. (2000). Automated empirical optimizations of software and the ATLAS project, http://www.netlib.org/atlas/.

Zangari, P. (1996a). How accurate is the delta-gamma methodology?, *RiskMetrics Monitor* **1996**(third quarter): 12–29.

Zangari, P. (1996b). A VaR methodology for portfolios that include options, *RiskMetrics Monitor* **1996**(first quarter): 4–12.