Saddlepoint approximations for the doubly noncentral t distribution

Broda, Simon ; Paolella, Marc S

DOI: https://doi.org/10.1016/j.csda.2006.11.024

Saddlepoint Approximations for the Doubly Noncentral t Distribution

Simon Broda* Marc S. Paolella†

Swiss Banking Institute, University of Zurich, Switzerland

September 2006

Abstract

Closed-form approximations for the density and cumulative distribution function of the doubly noncentral t distribution are developed based on saddlepoint methods. They exhibit remarkable accuracy throughout the entire support of the distribution and are vastly superior to existing approximations. An application in finance is considered which capitalizes on the enormous increase in computational speed.

Keywords: Financial Econometrics; Analysis of Variance; GARCH; t-statistic

* *E-mail address:* `broda@isb.unizh.ch`

† *E-mail address:* `paolella@isb.unizh.ch`

1 Introduction

This paper derives highly accurate and trivially computed approximations of the probability density function (pdf) and cumulative distribution function (cdf) of a doubly noncentral t random variable. Let $T = X/\sqrt{Y/n}$, where X and Y are independent, X has a normal distribution with mean μ and unit variance, and Y has a noncentral χ^2 distribution with $n \in \mathbb{N}$ degrees of freedom, noncentrality parameter θ, and density

$$f_Y(y) = \frac{1}{2} e^{-(y+\theta)/2} y^{(n-2)/4} \theta^{-(n-2)/4} I_{(n-2)/2}\big(\sqrt{\theta y}\big), \quad y > 0,$$

where

$$I_\nu = \sum_{i=0}^{\infty} \frac{(z/2)^{\nu+2i}}{i!\Gamma(\nu+i+1)}$$

is the modified Bessel function of the first kind of order ν. Random variable T is said to follow a doubly noncentral (Student's) t distribution with n degrees of freedom, numerator noncentrality parameter μ and denominator noncentrality parameter θ. We will write $T \sim t''(n, \mu, \theta)$, with the pdf and cdf of T denoted by $f_{t''}(t; n, \mu, \theta)$ and $F_{t''}(t; n, \mu, \theta)$, respectively. If $\theta = 0$, then T follows a singly noncentral t distribution, denoted $T \sim t'(n, \mu)$. The singly noncentral t is ubiquitous in statistical applications, as it is required for computing the power of a standard t-test. The doubly noncentral t appears quite naturally as the distribution of the t-statistic when the population means are unequal (Robbins, 1948), and in the analysis of variance (Scheffé, 1959, p. 137). Apart from these classical statistical applications, the t-distribution and its noncentral variants have recently been very successfully employed in the modelling of financial returns data; details and references will be given in Section 5 below. In such contexts, the domain of the degrees of freedom parameter is typically extended to contain the entire positive half of the real line.

Letting $\omega_{i,\theta} := \exp(-\theta/2)(\theta/2)^i/i!$ and $s_{i,n} := \sqrt{(n+2i)/n}$, Kocherlakota and Kocherlakota (1991) show that

$$f_{t''}(t; n, \mu, \theta) = \sum_{i=0}^{\infty} \omega_{i,\theta} s_{i,n} f_{t'}(s_{i,n} t; n+2i, \mu) \tag{1}$$

and

$$F_{t''}(t; n, \mu, \theta) = \sum_{i=0}^{\infty} \omega_{i,\theta} F_{t'}(s_{i,n} t; n+2i, \mu), \tag{2}$$

where $f_{t'}$ and $F_{t'}$ refer to the singly noncentral t. The former, for example, can be expressed as

$$\begin{aligned} f_{t'}(t; k, \mu) = e^{-\mu^2/2} & \frac{\Gamma\left(\left(k+1\right)/2\right) k^{k/2}}{\sqrt{\pi}\Gamma\left(k/2\right)} \left(\frac{1}{k+t^2}\right)^{\frac{k+1}{2}} \\ & \times \left(\sum_{i=0}^{\infty} \frac{(t\mu)^i}{i!} \left(\frac{2}{t^2+k}\right)^{i/2} \frac{\Gamma\left(\left(k+i+1\right)/2\right)}{\Gamma\left(\left(k+1\right)/2\right)} \right). \end{aligned} \tag{3}$$

In the doubly noncentral case, computing the doubly infinite sum required for the pdf or cdf of T at a given point t to any specified degree of accuracy is straightforward, but time consuming.

As an example, computation of the 401 pdf values plotted in Figure 1 required 2.2 seconds using (1) on a 2.8 GHz PC, and 0.004 seconds using the approximation developed herein. An example below will demonstrate a situation in which the pdf needs to be evaluated on the order of 10^6 times; use of (1) would be practically impossible.

Approximations to the distribution of T were considered by Krishnan (1967) and by Mudholkar and Chaubey (1976), which are discussed and compared below, while Krishnan (1968) and Nandi and Choudhury (2002) considered series representations of the distribution function. This paper relies on a result of Daniels and Young (1991) to develop highly accurate saddlepoint approximations, hereafter SPA, to both the density and the distribution function of the doubly (and, hence, singly) noncentral t. We demonstrate the outstanding accuracy of the SPA–based method, which holds over the entire support of T, and compare it to existing approximations, which often exhibit errors which are several orders of magnitude larger than obtained by the new approximation. Indeed, in the central Student's t case, the (renormalized) approximation is exact.

In most cases of practical interest, the SPA requires finding the root of an equation or, possibly, solving a nonlinear system of equations for each ordinate t. This differs from use of, say, an Edgeworth expansion, which yields a closed-form expression amenable to fast "vectorized" computation available in modern matrix–based programming languages such as Splus and Matlab. Interestingly, and quite fortunately, the required system of equations in this case can be solved analytically, yielding a completely closed–form expression and obviating multivariate root searching and the potential numerical problems inherently associated with it. Thus, the pdf and cdf approximations are evaluated essentially instantaneously, and never fail.

The remainder of this paper is organized as follows. Section 2 derives the saddlepoint approximation and provides the analytic solution to the requisite system of equations. (That this solution is unique is shown in the appendix.) Section 3 provides an improved approximation to the pdf. Section 4 contains numerical results on the accuracy of the proposed approximations and offers a comparison with alternative approximations. Finally, Section 5 illustrates a useful application of the new method.

2 Main Result

Let $\mathbf{x} = (x_1, x_2)$ be a bivariate random vector having a density and a joint cumulant generating function, and denote the latter as $K(\mathbf{t})$. Denote by $K'(\mathbf{t})$ and $K''(\mathbf{t})$ its gradient and hessian, respectively. Consider a bijection $\mathbf{y} = (y_1, y_2) = g^{-1}(\mathbf{x}) = (g_1^{-1}(\mathbf{x}), g_2^{-1}(\mathbf{x}))'$, so that $\mathbf{x} = g(\mathbf{y}) = (g_1(\mathbf{y}), g_2(\mathbf{y}))'$, and let $\nabla_{y_i} g(\mathbf{y}) = (\partial g_1/\partial y_i, \partial g_2/\partial y_i)', i \in \{1, 2\}$. Daniels and Young (1991) show that saddlepoint approximations to the marginal pdf and cdf of y_1 are given by

$$\hat{f}_{Y_1}(y_1) = \phi(w)/u \tag{4}$$

and

$$\hat{F}_{Y_1}(y_1) = \Phi(w) + \phi(w)\left(\frac{1}{w} - \frac{d}{u}\right), \tag{5}$$

respectively, where ϕ and Φ are the standard normal pdf and cdf, respectively, $w = \sqrt{2\big(\hat{\mathbf{t}}' g(\hat{\mathbf{y}}) - K(\hat{\mathbf{t}})\big)}\,\mathrm{sgn}(y_1 - \alpha)$, $\hat{\mathbf{y}} = (y_1, \hat{y}_2)$, $\alpha = g_1^{-1}(K'(\mathbf{0}))$, $d = \big(\hat{\mathbf{t}}' \nabla_{y_1} g(\hat{\mathbf{y}})\big)^{-1}$,

$$u = \frac{\sqrt{\det\big(K''(\hat{\mathbf{t}})\big)\Big[\nabla_{y_2} g(\hat{\mathbf{y}})' \big(K''(\hat{\mathbf{t}})\big)^{-1} \nabla_{y_2} g(\hat{\mathbf{y}}) + \hat{\mathbf{t}}' \nabla^2_{y_2} g(\hat{\mathbf{y}})\Big]}}{\det\big(\partial g / \partial \mathbf{y}(\hat{\mathbf{y}})\big)},$$

and, for each value of y_1, $\hat{\mathbf{t}}$ and $\hat{y}_2$ solve the system

$$K'(\hat{\mathbf{t}}) = g(\hat{\mathbf{y}})$$
$$\hat{\mathbf{t}}' \nabla_{y_2} g(\hat{\mathbf{y}}) = 0.$$

Paralleling the case of univariate saddlepoint approximations, the pdf approximation can be renormalized by numerically integrating it over its support if additional accuracy is desired.

Now let $x_1 \sim \mathrm{N}(\mu, 1)$, independent of $x_2 \sim \chi^2(n, \theta)$, and let $\mathbf{x} = g(\mathbf{y}) = (y_1 y_2, y_2^2 n)$, so that $(y_1, y_2) = g^{-1}(x_1, x_2) = \big(x_1/\sqrt{x_2/n}, \sqrt{x_2/n}\big)'$ and $y_1 \sim t''(n, \mu, \theta)$. The joint cumulant generating function of (x_1, x_2) is, from independence,

$$K(\mathbf{t}) = K_{x_1}(t_1) + K_{x_2}(t_2) = t_1 \mu + \frac{1}{2} t_1^2 - \frac{n}{2} \log(1 - 2t_2) + \frac{t_2 \theta}{1 - 2t_2},$$

where K_{x_1} and K_{x_2} are the cumulant generating functions of x_1 and x_2, respectively. The saddlepoint $(\hat{\mathbf{t}}, \hat{y}_2) = (\hat{t}_1, \hat{t}_2, \hat{y}_2)$, $\hat{t}_2 < \frac{1}{2}$, $\hat{y}_2 > 0$, solves the system of equations

$$\mu + t_1 = y_1 y_2$$
$$\frac{n}{(1 - 2t_2)} + \frac{\theta}{(1 - 2t_2)^2} = n y_2^2$$
$$t_1 y_1 + 2n t_2 y_2 = 0.$$

Straightforward calculation reveals that

$$\hat{t}_1 = -\mu + y_1 \hat{y}_2, \quad \hat{t}_2 = -\frac{y_1 \hat{t}_1}{2n\hat{y}_2}, \tag{6}$$

and $\hat{y}_2$ solves the cubic $s(y_2) := a_3 y_2^3 + a_2 y_2^2 + a_1 y_2 + a_0 = 0$, where

$$a_3 = y_1^4 + 2n y_1^2 + n^2, \quad a_2 = -2y_1^3 \mu - 2y_1 n \mu, \quad a_1 = y_1^2 \mu^2 - n y_1^2 - n^2 - \theta n,$$

and $a_0 = y_1 n \mu$. Upon defining

$$c_2 = \frac{a_2}{a_3}, \quad c_1 = \frac{a_1}{a_3}, \quad c_0 = \frac{a_0}{a_3}, \quad q = \frac{1}{3} c_1 - \frac{1}{9} c_2^2,$$

$$r = \frac{1}{6}(c_1 c_2 - 3c_0) - \frac{1}{27} c_2^3, \quad m = q^3 + r^2, \quad \text{and} \quad s_{1,2} = (r \pm \sqrt{m})^{1/3},$$

the roots of the cubic are given by

$$z_1 = (s_1 + s_2) - \frac{c_2}{3} \qquad \text{and} \qquad z_{2,3} = -\frac{1}{2}(s_1 + s_2) - \frac{c_2}{3} \pm \frac{i\sqrt{3}}{2}(s_1 - s_2).$$

The saddlepoint solution is always z_1, as proved in the appendix. It can also be expressed as

$$\hat{y}_2 = \sqrt{-4q}\cos\Big(\cos^{-1}\big(r/\sqrt{-q^3}\big)/3\Big) - \frac{c_2}{3}, \tag{7}$$

thus avoiding complex arithmetic.

With

$$\nabla_{y_1} g(\mathbf{y}) = (y_2, 0)', \quad \nabla_{y_2} g(\mathbf{y}) = (y_1, 2ny_2)',$$
$$\nabla^2_{y_2} g(\mathbf{y}) = (0, 2n)', \quad K''(\mathbf{t}) = \operatorname{diag}\big(1, 2n(1-2t_2)^{-2} + 4\theta(1-2t_2)^{-3}\big),$$
$$\det\big[\partial g/\partial \mathbf{y}\big] = \det\big[\nabla_{y_1} g(\mathbf{y}),\, \nabla_{y_2} g(\mathbf{y})\big] = 2ny_2^2$$

and after some simplification, the quantities entering approximations (4) and (5) take the simple form

$$d = (\hat{t}_1\hat{y}_2)^{-1}, \qquad u = \sqrt{(y_1^2 + 2n\hat{t}_2)(2n\nu^2 + 4\theta\nu^3) + 4n^2\hat{y}_2^2}\Big/\Big(2n\hat{y}_2^2\Big),$$
$$w = \sqrt{-\mu\hat{t}_1 - n\log\nu - 2\theta\nu\hat{t}_2}\,\operatorname{sgn}\big(y_1 - \alpha\big), \qquad \alpha = \mu/\sqrt{1+\theta/n},$$

where $\nu = (1-2\hat{t}_2)^{-1}$, and $\hat{\mathbf{t}} = (\hat{t}_1, \hat{t}_2)'$ and $\hat{y}_2$ are given by (6) and (7), respectively. In the singly noncentral case with $\theta = 0$, these reduce to

$$u = \sqrt{(\mu y_1\hat{y}_2 + 2n)/(2n)}/\hat{y}_2, \qquad \text{and} \qquad w = \sqrt{-\mu\hat{t}_1 - 2n\log(\hat{y}_2)}\,\operatorname{sgn}(y_1 - \mu),$$

where d, $\hat{t}_1$ and $\hat{t}_2$ are as before, and

$$\hat{y}_2 = \frac{\mu y_1 + \sqrt{4n(y_1^2 + n) + \mu^2 y_1^2}}{2(y_1^2 + n)}.$$

This is the approximation given in DiCiccio and Martin (1991, p. 897, Eq. (18)). In the central case with $\mu = 0$,

$$\hat{y}_2 = \sqrt{n/(y_1^2 + n)}, \quad d = (y_1\hat{y}_2^2)^{-1}, \quad u = \hat{y}_2^{-1}, \quad w = \sqrt{-2n\log(\hat{y}_2)}\,\operatorname{sgn}(y_1),$$

and the pdf approximation becomes

$$\hat{f}_t(y_1; n) = \frac{1}{\sqrt{2\pi}}\left(\frac{n}{y_1^2 + n}\right)^{\frac{1}{2}(n+1)},$$

which is exact after renormalization.

Note that $\lim_{y_1\to\alpha} w = \lim_{y_1\to\alpha} d^{-1} = 0$, so that at $y_1 = \alpha$, expression (5) is not well defined. This singularity is, however, removable, and a repeated application of l'Hôpital's rule shows that the limiting value is given by

$$\lim_{y_1\to\alpha} \hat{F}_{t''}(y_1, n, \mu, \theta) = \frac{1}{2} - \frac{1}{6\sqrt{\pi}}\frac{\mu\big((n+3\theta)(2\mu^2 + 3n) + 6\theta^2\big)}{\big((n+2\theta)(\mu^2 + 2n) + 2\theta^2\big)^{3/2}}. \tag{8}$$

While knowledge of this limiting value ensures that the saddlepoint cdf is continuous everywhere, numerical inaccuracies may arise in the immediate vicinity of α. In practice, these are most easily circumvented by replacing (5) with

$$\tilde{F}(y_1) = \lim_{y_1 \to \alpha} \hat{F}_{t''}(y_1, n, \mu, \theta) + (y_1 - \alpha) \lim_{y_1 \to \alpha} \hat{F}'_{t''}(y_1, n, \mu, \theta)$$

whenever $|w| < \epsilon$, and where the second limit is given in (10) below. The optimal value of ϵ depends on the arithmetic precision of the machine at hand. The left panel of Figure 2 demonstrates that in doing so, little accuracy is lost, owing to the approximate linearity of the cdf near α.

3 Avoiding Renormalization

As noted before, the saddlepoint approximation to the pdf can be normalized to integrate to unity, resulting in greater accuracy. Also, in the context of maximum likelihood estimation, use of the unnormalized approximation leads to biased estimates. The reason for this is that the constant of integration,

$$k(n, \mu, \theta) = \left[\int_{-\infty}^{\infty} \hat{f}_{t''}(t; n, \mu, \theta) \, \mathrm{d}t \right]^{-1},$$

depends on the parameters of the distribution; e.g., for the doubly noncentral t distribution considered here, our experiments show that the normalizing constant is an increasing function of all the parameters, but is most sensitive to variations in the degrees of freedom n. Thus, if $\hat{n}$ maximizes the normalized likelihood, then evaluating the non-normalized likelihood at a value $n_1 < \hat{n}$ may spuriously result in a higher likelihood value, simply because the densities are not properly normalized. This problem of downward-biased estimates of n is particularly acute when the degrees of freedom parameter is small, which is precisely the case in financial applications, where fitting the singly noncentral t has become commonplace. (Note that the normalizing constant approaches unity as n grows to infinity.)

However, renormalization of the pdf involves numeric integration, and so considerably slows down the otherwise virtually instantaneous computation of the saddlepoint pdf. In an application with a large number of evaluations of the density, such as occurs when conducting maximum likelihood estimation, this factor becomes noticeable. For example, computing the 401 approximate pdf values of Figure 1 takes about 0.056 seconds, or 14 times longer, when using Matlab's built-in `quadl` routine for normalization. (The exact factor by which the computational effort is increased depends on whether the pdf is to be evaluated at many points for the same parameter values, or at a single point for varying values of the parameters. This is due to the fact that the normalizing constant will only have to be evaluated once for each parameter constellation.) This motivates the construction of a pdf approximation which does not require renormalization. We consider two ways.

3.1 Analytic: Differentiating the CDF

Routledge and Tsao (1997) show that differentiating the **?** approximation to the distribution function of the mean of a sample of independent and identically distributed (i.i.d.) random variables gives rise to the same asymptotic expansion as the Daniels (1954) approximation to the corresponding density. Keeping only the terms obtained by differentiating the first-order cdf approximation, a density approximation is obtained that formally has the same error as the first-order pdf approximation, but which integrates to one, thus obviating the need to renormalize.

In the present context, differentiating expression (5) for the cdf leads to the adjusted pdf approximation

$$\hat{F}'_{Y_1}(y_1) = \phi(w)\left(\frac{1}{u} - \frac{1}{dw^3} - \frac{d'}{u} + \frac{du'}{u^2}\right), \tag{9}$$

where w, d, and u are as before,

$$u' \equiv \frac{\partial}{\partial y_1}u \qquad \text{and} \qquad d' \equiv \frac{\partial}{\partial y_1}d.$$

The general expressions for u' and d' are lengthy and not particularly insightful. However, in our setting, they simplify markedly, due to the independence of the random variables constituting the doubly noncentral t distribution. Upon defining

$$c_1 \equiv 2n\nu^2 + 4\theta\nu^3, \quad c_2 \equiv 8n\nu^3 + 24\theta\nu^4, \quad c_3 \equiv \hat{t}_1 + y_1\hat{y}_2, \quad c_4 \equiv 2n\hat{t}_2 + y_1^2 + c_1^{-1}(2n\hat{y}_2)^2$$

$$c_5 \equiv c_3/c_4, \quad \text{and} \quad c_6 \equiv 2nc_1^{-1}c_2c_5\hat{y}_2(2n\hat{t}_2 + y_1^2) + 12n^2c_5\hat{y}_2 + 2c_1y_1,$$

we obtain

$$d' = -d^2\Big(\hat{y}_2^2 - c_3c_5\Big) \quad \text{and} \quad u' = -uc_5\Big(c_6(2c_1c_3)^{-1} + 2\hat{y}_2^{-1}\Big).$$

The adjusted pdf approximation, like the saddlepoint cdf, has a removable singularity at the point $\mu/\sqrt{1+\theta/n}$. The limiting value can be found analytically, using a derivation similar to the one which led to (8) for the cdf. Upon defining $\kappa_i = n + i\theta, i \in \{1,2,4\}$, $\omega_1 = 6\kappa_2^2 - \kappa_4$, $\omega_2 = 144\kappa_1^2 - 33\kappa_2$, and $\omega_3 = 4\kappa_1^2 - \kappa_2$, it is given by

$$\lim_{y_1\to\alpha} \hat{F}'_{t''} = \frac{2\mu^6\big[\kappa_2\omega_1 - 10\theta^2\big] + 12\mu^4\big[\omega_1\kappa_1^2 - 7\theta^2\kappa_1 + \theta^3\big] + \mu^2\big[\kappa_1^2\kappa_2\omega_2\big] + 24\big[\kappa_1^4\omega_3\big]}{12\sqrt{n\pi}\kappa_1^{-3/2}(\kappa_2(\mu^2+2n)+2\theta^2)^{7/2}}. \tag{10}$$

3.2 Numeric: Response Surface Fitting

The second way of avoiding numerical integration at run-time is to fit a response surface to pre-computed values of the normalizing constant. In view of the aforementioned ubiquity of the singly noncentral t distribution in financial applications, we restrict our attention to that special case. Our experiments showed that for a given value of μ, $k(n,\mu,0)$ can be parsimoniously and accurately approximated by regressing $k(n,\mu,0)$ on several fractional powers of n and $\log(n)$. This process is repeated for a fine grid of values for μ, and the resulting matrix of regression parameters is stored. The approximate normalizing constant, say $\hat{k}(n,\mu)$, is then determined at run-time by linearly interpolating between the least-squares predictions of $\hat{k}(n, \mu, 0)$ for the nearest values of μ.

This response surface was fitted for the range $0.5 \leq n \leq 20$ and $|\mu| \in \{0, 0.1, 0.2, \ldots, 10\}$, which appears adequate for applications in empirical finance. The resulting approximation is accurate to within 1% of the true values. A program written in Matlab is available from the authors to compute all entertained approximations (renormalized via integration, cdf differentiation, and, for the singly noncentral case, the renormalized pdf using the response surface approximation for the integration constant).

4 Accuracy Assessment

Several approximations for the cdf have been previously proposed. To save space, we only mention the type of approximation; the specific formulae can be found in Broda and Paolella (2006) (and the cited references).

Krishnan (1967) suggests two different approximation to $F_{t''}(t; n, \mu, \theta)$. The first one, to be referred to below as KR1, is based on a scale transformed singly noncentral t distribution with scale parameter and degrees of freedom chosen to equate the first two moments. The second approximation (KR2) is based on the distribution function of the sample correlation coefficient. It involves the numerical evaluation of an integral, where the integrand contains the hypergeometric function. Consequently, the approximation is about as time–consuming to evaluate as the exact t'' cdf, defying the purpose of the approximation. We therefore chose to replace the hypergeometric function with its highly accurate Laplace approximation given in Butler and Wood (2002).

Two closed–form approximations to the t'' cdf are given in Mudholkar and Chaubey (1976). The first one (MC1) is based on Patnaik's (1949) approximation to the noncentral chi square distribution. Their second approximation (MC2) is based on an Edgeworth expansion.

To illustrate the merits of the various approximations, we consider the pdf and cdf of $T \sim t''(n, \mu, \theta)$, for the set of arbitrarily chosen parameters $(n, \mu, \theta) = (5, 2, 5)$, over a range of ± 10 standard deviations around the mean. The exact pdf and cdf values have been computed from (1) and (2), respectively, using Matlab's built-in routine for evaluating the singly noncentral pdf and cdf. Denoting the summands in equations (1) and (2) by S_i^1 and S_i^2, respectively, the infinite sums were truncated at $\bar{i}_j = \inf\{i : i > \theta/2 \;\wedge\; |S_i^j| < 2.3 \times 10^{-16}\}$, as the absolute values of the summands S_i^j assume their maximum at some $i \leq \lceil \theta/2 \rceil$, after which they decrease monotonically.

Figure 1 demonstrates the accuracy of the pdf approximations. Even for the degrees of freedom as low as $n = 5$, the renormalized and adjusted approximations, shown in the left panel, are graphically indistinguishable from the true pdf. The right panel shows that near the mean, the adjusted pdf approximation has relative percentage error, defined as $100 \times (\text{approx} - \text{exact})/\text{exact}$, comparable to the renormalized approximation; this is true in general, and agrees with the findings of Routledge and Tsao (1997) who show that, near the mean, the differentiated Lugannani–Rice formula is a second order approximation. In the limit as $|t| \to \infty$, $\hat{F}'_{t''}(t)/\hat{f}_{t''}(t) = 1$.

The right-hand panel of Figure 2 shows the relative percentage error, now defined as $100 \times (\text{approx} - \text{exact})/\min(\text{exact}, 1 - \text{exact})$, of the various cdf approximations. Although the graph refers only to a particular parameter set, it is to a large extent representative of the general picture: Approximations MC1 and MC2 tend to perform acceptably near the mean, but break down for larger deviations. At least for the latter approximation, this behavior is expected, as it is derived as an Edgeworth expansion; these are well-known to deteriorate in the tails of the distribution, see, e.g., Daniels (1954, p. 631).

Approximations KR1 and KR2, on the other hand, are useful over the entire real line; however, their relative error is generally several orders of magnitude higher than that of the SPA. It should also be noted that for approximations KR1 and KR2 to exist, it is required that $n > 2$ and $n > 4$, respectively, which may not be fulfilled in practice, especially in the context of empirical finance, which we consider below. Also, approximation KR1 is not defined if $\mu = 0$. Finally, for some combinations of n, μ and θ, the KR2 approximation is invalid (in Krishnans (1967) notation, his parameter ρ will exceed one). The SPA, on the other hand, is well-defined over the entire support of the distribution.

To get an idea how the error from the SPA behaves as a function of the parameters n, μ and θ, Figure 3 plots, for several values of θ and as a function of n, the relative percentage error of the saddlepoint approximation of the cdf evaluated at the ordinate t such that $F_{t''}(t; n, \mu, \theta) = p$, where $p = 0.95$ was used. The quantiles for each constellation of points shown were computed by root search using the SPA to the cdf. From the left panel, for which $\mu = 0$, we see that, beyond four degrees of freedom, the approximation has less than one percent error. The right panel is the same, but uses $\mu = 10$, which performs very similarly except for the—not overly practical—case of less than one degree of freedom. A variety of similar runs using different p and μ confirm that Figure 3 is quite typical: for larger than four degrees of freedom, the relative error is under one percent, while as n approaches zero, the error increases without bound. Not apparent from the graphs is that, as θ increases past six, the accuracy increases (for example, with $p = 0.95$, $\mu = 0$ and $n = 1$ degree of freedom, the error is already under one percent for $\theta = 12$, and continues to decrease as θ increases).

5 Empirical Application

We illustrate the use of the new computational method in the context of modelling the daily returns on the financial index NASDAQ which, like essentially all returns on financial assets observed at weekly or higher frequencies, are characterized by high excess kurtosis relative to the normal distribution ("fat tails"), mild skewness, and strong volatility clustering. The interest in modelling the returns on financial assets has grown enormously over the past decade. From the viewpoint of risk managers in financial institutions, short-term out-of-sample quantile prediction

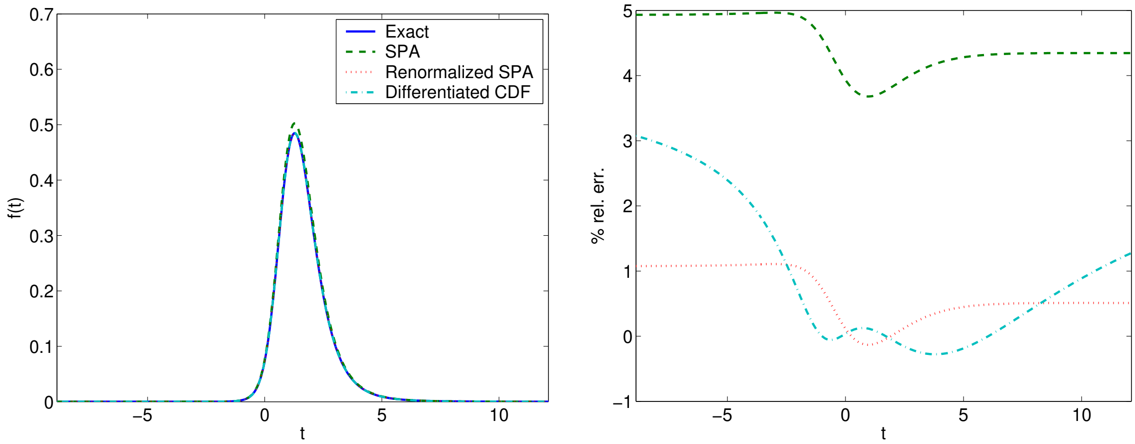

Figure 1: Left panel: Density of $T \sim t''(5, 2, 5)$. **Right panel**: Relative percentage errors.

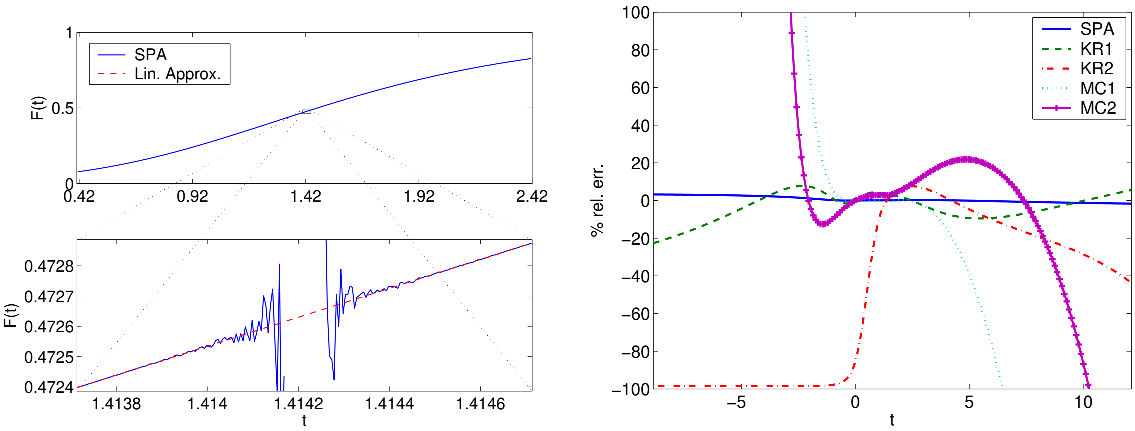

Figure 2: Left Panel: Illustration of numerical inaccuracies near α, $T \sim t''(5, 2, 5)$. **Right Panel**: Percentage relative errors of approximations to the cdf of $T \sim t''(5, 2, 5)$.

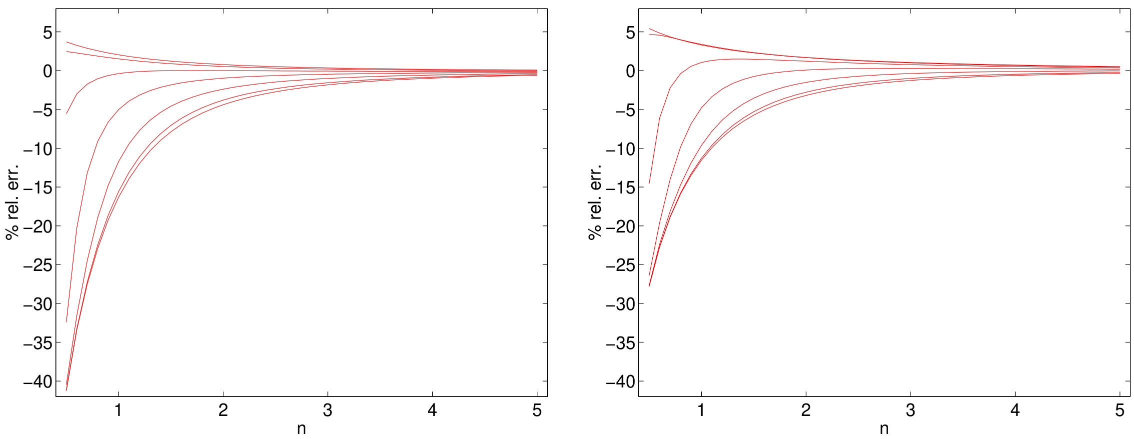

Figure 3: Left panel: The relative percentage error of the SPA to $F_{t''}(t; n, 0, \theta)$ as a function of the degrees of freedom parameter n, where t is chosen such that $F_{t''}(t; n, \mu, \theta) = 0.95$, $\mu = 0$, and $\theta = 0, 1, \ldots, 6$ (from bottom to top). **Right panel**: Same, but using $\mu = 10$.

is of extreme importance (most notably the accurate estimation of the Value-at-Risk, or VaR, measure, which is currently the most popular and important measure of risk used in risk management; see Dowd (2005) and Kuester et al. (2006) for detailed overviews of the rampantly growing literature in this field). Failure to account for the non-normality of the data invariably results in biased VaR forecasts; therefore, unless this bias is corrected a posteriori, as, e.g., in Hartz et al. (2006), a different, non-Gaussian innovations distribution has to be used.

Because the (central) Student's t distribution can assume fat–tailed behavior at least as extreme as the Cauchy and also nests (in a limiting sense) the normal, it has become the de facto standard choice of innovations distribution for both unconditional (in which the returns are treated as being i.i.d.) and conditional (in which a stochastic process for the time-varying location and, more importantly, scale term, is used) modelling of financial returns data. Two recent examples are Ané (2006), who considers model (11) below, and Chen et al. (2006); see also Palm (1997) for a detailed overview. Numerous empirical papers have demonstrated that the Student's t can accommodate the fat–tailed nature of the data very well, but, obviously, not the skewness. To account for this, several asymmetric extensions of the Student's t distribution have been proposed and demonstrated to be highly effective in a risk forecasting context. These include the singly noncentral t, as first advocated in this context by Harvey and Siddique (1999).

In the conditional modelling case, by far the most popular method for capturing the strong stochastic volatility component inherent in financial returns is use of a GARCH–type model, which we briefly discuss. Denote the return at time t by R_t, $t = 1, \ldots, T$, which we model as $R_t = a_0 + Z_t\sigma_t$, where the Z_t are assumed to be i.i.d.. To capture the evolution of the scale parameter σ_t, we use the popular Asymmetric Power ARCH, or A-PARCH, model proposed by Ding, Granger and Engle (1993), given by

$$\sigma_t^\delta = c_0 + \sum_{i=1}^{r} c_i \left(|\epsilon_{t-i}| - \gamma_i \epsilon_{t-i}\right)^\delta + \sum_{i=1}^{s} d_i \sigma_{t-i}^\delta, \quad \epsilon_t = Z_t\sigma_t, \tag{11}$$

with $c_i > 0$, $d_i \geq 0$, $\delta > 0$, and $|\gamma_i| < 1$. As detailed in Mittnik, Paolella and Rachev (1998), correct estimation of (11) and handling of the initial $\max(r, s)$ values of σ_t and ϵ_t necessitate knowing $\mathrm{E}\big[\,(|Z| - \gamma Z)^\delta\,\big]$. This can be effectively and nearly instantly approximated by simple numeric integration when using the SPA for the singly or doubly noncentral t.

In our (and most GARCH) applications, choosing $r = s = 1$ and setting $\delta = 1$ suffices. The model (11), when coupled with the singly noncentral t distribution for the Z_t, i.e., $Z_t \overset{\text{iid}}{\sim} t'(n, \mu)$, has been shown to be very successful in a risk prediction context. We thus require

$$\kappa := \mathbb{E}\big[\,(|Z| - \gamma Z)\,\big] = \mathbb{E}\big[\,|Z|\,\big] - \gamma\mathbb{E}\big[Z\big], \quad Z \sim t'(n, \mu),$$

which can be analytically determined: an expression for the first moment of $T \sim t''(n, \mu, \theta)$ is given in Krishnan (1967) as

$$\mathbb{E}[T] = \mu \left(\frac{n}{2}\right)^{1/2} \frac{\Gamma\big((n-1)/2\big)}{\Gamma\big(n/2\big)} \, {}_1F_1\left(\frac{1}{2}, \frac{n}{2}, -\frac{\theta}{2}\right), \tag{12}$$

while one for $\mathbb{E}[|T|]$ does not appear to have been considered in the literature. Straightforward calculation shows that

$$\mathbb{E}(|T|^r) = n^{r/2} \frac{\Gamma\left((n-r)/2\right)\Gamma\left((r+1)/2\right)}{\Gamma\left(n/2\right)\Gamma\left(1/2\right)} {}_1F_1\left(\frac{r}{2}, \frac{n}{2}, -\frac{\theta}{2}\right) {}_1F_1\left(-\frac{r}{2}, \frac{1}{2}, -\frac{\mu^2}{2}\right), \quad r < n. \tag{13}$$

The confluent hypergeometric functions ${}_1F_1$ appearing in (12) and (13) could be quickly approximated using the approximation given in Butler and Wood (2002). Based on (12) and (13), κ can be computed.

For the daily NASDAQ returns from June 1993 to June 2001 (about 2000 observations), the maximum likelihood estimates using the exact singly noncentral t density are

$$\hat{a}_0 = 0.185\,(0.046), \quad \hat{c}_0 = 0.0187\,(0.0057), \quad \hat{c}_1 = 0.112\,(0.016), \quad \hat{d}_1 = 0.884\,(0.017),$$
$$\hat{\gamma}_1 = 0.283\,(0.078), \quad \hat{n} = 9.15\,(1.7), \quad \hat{\mu} = -0.0870\,(0.049).$$

As is common in this and related statistical applications (see, e.g., the detailed discussion in Morgan, 2000, Chapters 3 and 4), the approximate standard errors, given in parentheses, have been obtained from the numerical Hessian, evaluated at the estimates. The estimation required about 60 iterations using a quasi-Newton method for multivariate optimization; this corresponds to about 550 evaluations of the likelihood, each of which required computing the pdf for each of the 2000 data points, i.e., over one million evaluations of the singly noncentral t density are required. Use of the SPA with renormalization via numeric integration yields

$$\hat{a}_0 = 0.188\,(0.044), \quad \hat{c}_0 = 0.0191\,(0.0055), \quad \hat{c}_1 = 0.112\,(0.016), \quad \hat{d}_1 = 0.883\,(0.017),$$
$$\hat{\gamma}_1 = 0.281\,(0.078), \quad \hat{n} = 8.97\,(1.7), \quad \hat{\mu} = -0.0919\,(0.047).$$

The point estimates based on the SPA barely differ from those using the exact pdf calculation, and their differences are negligible with respect to the reported standard errors. Using the SPA with the response surface approximation to the constant of integration yields

$$\hat{a}_0 = 0.191\,(0.034), \quad \hat{c}_0 = 0.0143\,(0.0054), \quad \hat{c}_1 = 0.104\,(0.015), \quad \hat{d}_1 = 0.895\,(0.016),$$
$$\hat{\gamma}_1 = 0.279\,(0.076), \quad \hat{n} = 8.73\,(1.6), \quad \hat{\mu} = -0.0947\,(0.039),$$

which are also not significantly different from either the values based on the exact pdf or the renormalized-via-integration SPA.

With regard to the (statistically insignificant and practically negligible) parameter difference based on the renormalized SPA and true density, it is understood in such empirical finance applications that the model innovations are certainly not really distributed as a noncentral t; it is just an excellent parametric approximation because of its flexible skewness and kurtosis (but does offer the plausible interpretation as a continuous mixture of normals; see Praetz, 1972). As such, in this context, the fact that the renormalized SPA is not exact is fully irrelevant; being a proper density in its own right, its use is justified just as well as that of the true density.

To illustrate the importance of renormalization, use of the SPA without renormalization gave highly different parameters; for example, $\hat{n} = 1.18$, which would misleadingly indicate that the GARCH innovations could follow a Cauchy distribution and thus highly overestimate the risk of a long or short position in the asset. Use of the adjusted pdf (9) performed between the two extremes: the estimated degrees of freedom in this case is $\hat{n} = 3.90$, which is too low. The explanation of this disappointing performance is that the amount of skewness in this data set (and most financial return series) is not great, so that the distribution is very close to being the regular Student's t, for which the renormalized SPA is *exact*.

It is well-known in the empirical finance literature that the distributional assumption on the conditional innovation sequence of a GARCH model is far more decisive for the quality of in-sample fit and out-of-sample forecasting ability than the form of the GARCH recursion for the conditional volatility (see, e.g., the extensive results in Mittnik and Paolella, 2000, and Bao et. al, 2003). As such, it might be expected that use of the doubly noncentral t distribution (with the additional parameter θ viewed as another shape parameter) in model (11) would provide a better description of the data generating process than use of the singly noncentral t. To the best of the authors' knowledge, this has never been attempted, owing, in all likelihood, to the computational complexity of the t'' pdf. If, however, the exact pdf is replaced by its SPA, the estimation of such a model becomes feasible, requiring in fact no more time than with the singly noncentral t distribution.

For the NASDAQ data set, the estimate of $\hat{\theta}$ was insignificantly different from zero, based on the asymptotic normality the MLE (see Straumann, 2005, Chapter 6.3). Nevertheless, given the plethora of different kinds of financial data sets and their well-reported distributional asymmetries (see, e.g., Cappuccio et. al, 2004, Premaratne and Bera, 2005, Lisi, 2005, and the references therein), it seems likely that the additional flexibility of the doubly noncentral t could be of potentially great value in this modelling context.

A Appendix

We first prove that the discriminant $m < 0$, i.e., all three roots are real. The discriminant can be written as

$$\begin{aligned} m = &- \frac{n}{108\left(y_1^2+n\right)^6}\left(-8y_1^2\mu^2\theta^2 n + 4\theta^3 n^2 + 4y_1^4\mu^4\theta\right) - \\ &- \frac{n}{108\left(y_1^2+n\right)^6}\left(ny_1^6\mu^2 + 12n^2y_1^4\theta + 2n^2y_1^4\mu^2 + 12n^2y_1^2\theta^2 + 24n^3y_1^2\theta + y_1^2\mu^2n^3 + \right. \\ &\left. + 4n^2y_1^6 + 12n^3y_1^4 + 12n^4y_1^2 + 12\theta^2n^3 + 12\theta n^4 + 4n^5 + 20ny_1^4\mu^2\theta + 20y_1^2\mu^2\theta n^2\right). \end{aligned}$$

As $n > 0$ and $\theta \geq 0$, all terms in the second parentheses are positive. The first term in parentheses can be factored as $-8y_1^2\mu^2\theta^2 n + 4\theta^3 n^2 + 4y_1^4\mu^4\theta = 4\theta(\theta n - y_1^2\mu^2)^2$, so that m cannot be positive, whence all three roots are real and, as shown in Butler and Paolella (2002), ordered as $z_2 < z_3 <$

z_1. Furthermore, as $a_3 > 0$ and $\operatorname{sgn}(a_2) = -\operatorname{sgn}(y_1\mu) = -\operatorname{sgn}(a_0)$, the coefficients of $s(y_2)$ can have at most two variations in sign, and thus, from Descartes' rule of signs, $s(y_2)$ can have at most two positive roots. Similarly, the coefficients of $s(-y_2)$ can have at most two variations in sign, and therefore $s(y_2)$ has at least one positive root. Consequently,

$$z_2 < 0 \tag{14}$$

and $z_1 > 0$. Also, we require that $\hat{t}_2 < 1/2$, or, plugging in $\hat{t}_1$ and $\hat{t}_2$ from (6),

$$\hat{y}_2 > \frac{y_1\mu}{y_1^2 + n} =: y_2^*. \tag{15}$$

Next observe that, as $a_3 > 0$ and $m < 0$,

$$s(y_2) \begin{cases} < 0, & y_2 < z_2, \\ > 0, & z_2 < y_2 < z_3, \\ < 0, & z_3 < y_2 < z_1, \\ > 0, & y_2 > z_1, \end{cases} \tag{16}$$

as well as that $s(0) = y_1\mu n$ and $s(y_2^*) = -y_1\mu n\theta / \left(y_1^2 + n\right)$. If $y_2^* > 0$, then $s(y_2^*) < 0$, and it follows from (14) and (16) that $z_3 < y_2^* < z_1$, so that the saddlepoint solution must be z_1 from (15). If $y_2^* \leq 0$, then $s(0) \leq 0$ and, from (14) and (16), $z_3 \leq 0 < z_1$, implying that the saddlepoint solution must be $z_1 > 0$.

Acknowledgements

Part of the research of M. Paolella has been carried out within the National Centre of Competence in Research "Financial Valuation and Risk Management" (NCCR FINRISK), which is a research program supported by the *Swiss National Science Foundation*.

References

Ané, T., 2006. An analysis of the flexibility of asymmetric power garch models. Computational Statistics & Data Analysis 51, 1293–1311.

Bao, Y., Lee, T.-H., Saltoglu, B., 2003. A test for density forecast comparison with applications to risk management, working paper, University of California, Riverside.

Broda, S., Paolella, M. S., 2006. Saddlepoint approximations for the doubly noncentral t distribution, working paper, University of Zurich.

Butler, R. W., Paolella, M. S., 2002. Calculating the density and distribution function for the singly and doubly noncentral F. Statistics and Computing 12, 9–16.

Butler, R. W., Wood, A. T. A., 2002. Laplace approximation for hypergeometric functions of matrix argument. Annals of Statistics 30, 1155–1177.

Cappuccio, N., Lubian, D., Raggi, D., 2004. Asymmetry and heavy tails in financial returns: Evidence from international stock markets, working paper, Department of Economics, University of Padova.

Chen, C. W. S., Gerlach, R., So, M., 2006. Comparison of non-nested asymmetric heteroskedastic models. Computational Statistics & Data Analysis 51, 2164–2178.

Daniels, H. E., 1954. Saddlepoint approximation in statistics. Ann. Math. Statist. 25, 631–650.

Daniels, H. E., Young, G. A., 1991. Saddlepoint approximation for the studentized mean, with an application to the bootstrap. Biometrika 78, 169–179.

DiCiccio, T. J., Martin, M. A., 1991. Approximations of marginal tail probabilities for a class of smooth functions with applications to bayesian and conditional inference. Biometrika 78, 891–902.

Ding, Z., Granger, C. W., Engle, R. F., 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83–106.

Dowd, K., 2005. Measuring Market Risk, 2nd Edition. Wiley, New York.

Hartz, C., Mittnik, S., Paolella, M. S., 2006. Accurate value–at–risk forecasting based on the normal–garch model. Computational Statistics & Data Analysis 51, 2295–2312.

Harvey, C. R., Siddique, A., 1999. Autoregressive conditional skewness. Journal of Financial and Quantitative Analysis 34 (4), 465–487.

Kocherlakota, K., Kocherlakota, S., 1991. On the doubly noncentral t distribution. Communications in Statistics: Simulation and Computation 20, 23–31.

Krishnan, M., 1967. The moments of a doubly noncentral t-distribution. Journal of the American Statistical Association 62, 278–287.

Krishnan, M., 1968. Series representations of the doubly noncentral t-distribution. Journal of the American Statistical Association 63, 1004–1012.

Kuester, K., Mittnik, S., Paolella, M. S., 2006. Value–at–risk prediction: A comparison of alternative strategies. Journal of Financial Econometrics 4, 53–89.

Lisi, F., 2005. Testing asymmetry in financial time series, working paper, Department of Statistical Sciences, University of Padua.

Mittnik, S., Paolella, M. S., 2000. Conditional density and value-at-risk prediction of asian currency exchange rates. Journal of Forecasting 19, 313–333.

Mittnik, S., Paolella, M. S., Rachev, S. T., 1998. Unconditional and conditional distributional models for the nikkei index. Asia-Pacific Financial Markets 5 (2), 99–128.

Morgan, B. J. T., 2000. Applied Stochastic Modelling. Arnold, London.

Mudholkar, G. S., Chaubey, Y. P., 1976. A simple approximation for the doubly noncentral-t distribution. Communications in Statistics: Simulation and Computation 5, 85–92.

Nandi, S. B., Choudhury, S., 2002. Series representation of doubly noncentral t distribution by use of the mellin integral transform. Communications in Statistics: Theory and Methods 31, 2139–2151.

Patnaik, P. B., 1949. The noncentral χ^2 and F-distributions and their applications. Biometrika 36, 202–232.

Praetz, P. D., 1972. The distribution of share price changes. The Journal of Business 45 (1), 49–55.

Premaratne, G., Bera, A., 2005. A test for symmetry with leptokurtic financial data. Journal of Financial Econometrics 3, 169–187.

Robbins, M., 1948. The distribution of student's t when the population means are unequal. Annals of Mathematical Statistics 19, 406–410.

Routledge, R., Tsao, M., 1997. On the relationship between two asymptotic expansions for the distribution of sample mean and its applications. Annals of Statistics 25 (5), 2200–2209.

Scheffé, H., 1959. The Analysis of Variance. John Wiley & Sons, Inc., New York.

Straumann, D., 2005. Estimation in Conditionally Heteroscedastik Time Series Models. Springer, Berlin, Heidelberg.