*Mathematical Finance*, Vol. 12, No. 3 (July 2002), 239–269

# PORTFOLIO VALUE-AT-RISK WITH HEAVY-TAILED RISK FACTORS

PAUL GLASSERMAN

*Graduate School of Business, Columbia University*

PHILIP HEIDELBERGER

*IBM Research Division, Yorktown Heights, NY*

PERWEZ SHAHABUDDIN

*IEOR Department, Columbia University*

This paper develops efficient methods for computing portfolio value-at-risk (VAR) when the underlying risk factors have a heavy-tailed distribution. In modeling heavy tails, we focus on multivariate $t$ distributions and some extensions thereof. We develop two methods for VAR calculation that exploit a quadratic approximation to the portfolio loss, such as the delta-gamma approximation. In the first method, we derive the characteristic function of the quadratic approximation and then use numerical transform inversion to approximate the portfolio loss distribution. Because the quadratic approximation may not always yield accurate VAR estimates, we also develop a low variance Monte Carlo method. This method uses the quadratic approximation to guide the selection of an effective importance sampling distribution that samples risk factors so that large losses occur more often. Variance is further reduced by combining the importance sampling with stratified sampling. Numerical results on a variety of test portfolios indicate that large variance reductions are typically obtained. Both methods developed in this paper overcome difficulties associated with VAR calculation with heavy-tailed risk factors. The Monte Carlo method also extends to the problem of estimating the conditional excess, sometimes known as the conditional VAR.

KEY WORDS: value-at-risk, delta-gamma approximation, Monte Carlo, simulation, variance reduction, importance sampling, stratified sampling, conditional excess, conditional value-at-risk

## 1. INTRODUCTION

A central problem in market risk management is estimation of the profit-and-loss distribution of a portfolio over a specified horizon. Given this distribution, the calculation of specific risk measures is relatively straightforward. Value-at-risk (VAR), for example, is a quantile of this distribution. The expected loss and the expected excess loss beyond some threshold are integrals with respect to this distribution. The difficulty in estimating these types of risk measures lies primarily in estimating the profit-and-loss distribution itself, especially the tail of this distribution associated with large losses.

P. Glasserman and P. Shahabuddin were partially supported by NSF NYI Award DMI9457189 and NSF Career Award DMI9625297, respectively.

*Manuscript received August 2000; final revision received January 2001.*

Address correspondence to Paul Glasserman, 403 Uris Hall, Columbia University, New York, NY 10027; pg20@columbia.edu.

All methods for estimating or approximating the distribution of changes in portfolio value rely (at least implicitly) on two types of modeling considerations: assumptions about the changes in the underlying risk factors to which a portfolio is exposed, and a mechanism for translating these changes in risk factors to changes in portfolio value. Examples of relevant risk factors are equity prices, interest rates, exchange rates, and commodity prices. For portfolios consisting of positions in equities, currencies, commodities, or government bonds, mapping changes in the risk factors to changes in portfolio value is straightforward. But for portfolios containing complex derivative securities this mapping relies on a pricing model.

The simplest and perhaps most widely used approach to modeling changes in portfolio value is the variance-covariance method popularized by RiskMetrics (1996). This approach is based on assuming (i) that changes in risk factors are conditionally multivariate normal over a horizon of, say, one day, two weeks, or a month, and (ii) that portfolio value changes linearly with changes in the risk factors. (''Conditionally'' here means conditional on information available at the start of the horizon; the unconditional distribution need not be normal.) Under these assumptions, the portfolio profit-and-loss distribution is conditionally normal; its standard deviation can be calculated from the covariance matrix of the underlying risk factors and the sensitivities of the portfolio instruments to these risk factors. The attraction of this approach lies in its simplicity. But each of the assumptions on which it relies is open to criticism, and research in the area has tried to address the shortcomings of these assumptions.

One line of work has focused on relaxing the assumption that portfolio value changes linearly with changes in risk factors while preserving computational tractability. This includes, in particular, the ''delta-gamma'' methods developed in Britten-Jones and Schaefer (1999), Duffie and Pan (2001), Rouvinez (1997), and Wilson (1999). These methods refine the relation between risk factors and portfolio value to include quadratic as well as linear terms. Methods that combine interpolation approximations to portfolio value with Monte Carlo sampling of risk factors are considered in Jamshidian and Zhu (1997), Picoult (1999), and Shaw (1999). Low variance Monte Carlo methods based on exact calculation of changes in portfolio value are proposed in Cardenas et al. (1999), Glasserman, Heidelberger, and Shahabuddin (2000), and Owen and Tavella (1999).

Another line of work has focused on developing more realistic models of changes in risk factors. It has long been observed that market returns exhibit systematic deviation from normality: across virtually all liquid markets, empirical returns show higher peaks and heavier tails than would be predicted by a normal distribution, especially over short horizons. Early studies along these lines include Mandelbrot (1963), Fama (1965), Praetz (1972), and Blattberg and Gonedes (1974). More recent investigations, some motivated by value-at-risk, include Bouchaud, Sornette, and Potters (1997), Danielsson and de Vries (1997), Eberlein and Keller (1995), Eberlein, Keller, and Prause (1998), Embrechts, McNeil, and Straumann (2001), Hosking, Bonti, and Siegel (2000), Huisman et al. (1998), Koedijk, Huisman, and Pownall (1998), McNeil and Frey (1999), Heyde (1999). Using different approaches to the problem and different sets of data, these studies consistently find high kurtosis and heavy tails. Moreover, most studies find that the tails in financial data are not so heavy as to produce infinite variance (as would be implied by a nonnormal stable distribution), though higher order moments (e.g., fifth and higher) may be infinite.

This paper contributes to both lines of investigation by developing methods for calculating portfolio loss probabilities when the underlying risk factors are heavy

tailed. Most of the literature documenting heavy tails in market data has focused on the univariate case—time series for a single risk factor or in some cases a fixed portfolio. There has been less work on modeling the joint distribution of risk factors with heavy tails (recent work in this direction includes Embrechts et al. 2002 and Hosking et al. 2000). There has been even less work on combining heavy-tailed joint distributions for risk factors with a nonlinear relation between risk factors and portfolio value, which is the focus of this paper.

We model changes in risk factors using a *multivariate t distribution* and some generalizations of this distribution. A univariate $t$ distribution is characterized by a parameter $\nu$, its degrees of freedom. The tails of the $t$ density decay at a polynomial rate of $\nu + 1$, so the parameter $\nu$ determines the heaviness of the tail and the number of finite moments. Empirical support for modeling univariate returns with a $t$ distribution or $t$-like tails can be found in Blattberg and Gonedes (1974), Danielsson and de Vries (1997), Hosking et al. (2000), Huisman et al. (1998), Hurst and Platen (1997), Koedijk et al. (1998), and Praetz (1972). There are many possible multivariate distributions with $t$ marginals. We follow Anderson (1984), Tong (1990), and others in working with a particular class of multivariate distributions having $t$ marginals for which the joint distribution is characterized by a symmetric, positive definite matrix $\Sigma$, along with the degrees of freedom. The matrix $\Sigma$ plays a role similar to that of the covariance matrix for a multivariate normal; this facilitates modeling with the multivariate $t$ and interpretation of the model.

Because it is characterized by the matrix $\Sigma$, the multivariate $t$ shares some attractive properties with the multivariate normal while possessing heavy tails. This is important in combining a realistic model of risk factors with a nonlinear relation between risk factors and portfolio value, which is our goal. We use the structure of the multivaratiate $t$ to develop computationally efficient methods for calculating portfolio loss probabilities capturing heavy tails and without assuming linearity of the portfolio value with respect to changes in risk factors. While it may be possible to find other multivariate distributions that are preferable on purely statistical grounds, the advantage of such a model may be limited if it cannot be integrated with efficient methods for calculating portfolio risk measures. The multivariate $t$ balances tractability with the empirical evidence for heavy tails. Moreover, we will see that some of the methods developed apply to a broader class of distributions of which the multivariate $t$ is a particularly interesting special case.

We develop two methods for estimating portfolio loss probabilities in the presence of heavy-tailed risk factors. The first method uses transform inversion based on a quadratic approximation to portfolio value. It thus extends the delta-gamma approximation developed in the multivariate normal setting. But the analysis here differs from the normal case in several important ways. Because the $t$ distribution has a polynomial tail, it does not have a moment generating function; and whereas uncorrelated multivariate normal random variables are necessarily independent, the same is not true with the multivariate $t$. This means that the characteristic function for a quadratic in $t$'s does not factor into a product of one-dimensional characteristic functions (as it does in the normal case). Indeed, we never explicitly find the characteristic function of a quadratic in $t$'s, which may be intractable. Instead, we use an indirect transform analysis through which we are able to compute the distribution of interest.

This method is fast, but a quadratic approximation to portfolio value is not always sufficiently accurate to produce reliable VAR estimates. We therefore also develop an

efficient Monte Carlo procedure. This method builds on the first; it uses the transform analysis to design highly efficient sampling procedures that are particularly well suited to estimating the tail of the loss distribution. The method combines importance sampling and stratified sampling in the spirit of Glasserman, Heidelberger, and Shahabuddin (1999a, 1999b, 2000). But the methods in these studies assumed a multivariate normal distribution and, as is often the case in importance sampling, they applied an exponential change of measure. An exponential change of measure is inapplicable to a $t$ distribution, again because of the nonexistence of a moment generating function. (Indeed, the successful application of importance sampling in heavy-tailed settings is a notoriously difficult problem; see, e.g., Asmussen and Binswanger 1997 Asmussen, Binswanger, and Højgaard 2002 and Juneja and Shahabuddin 1999.) We circumvent this problem by an indirect approach to importance sampling and stratified sampling. Through careful sampling of market risk factors, we are able to substantially reduce the variance in Monte Carlo estimates of loss probabilities and thus to reduce the number of samples needed to estimate a loss probability to a specified precision. Both a theoretical analysis of the method and numerical examples indicate the potential for enormous gains in computational speed as a result of this approach. This therefore makes it computationally feasible to combine the realism of heavy-tailed distributions and the robustness of Monte Carlo simulation in estimating portfolio loss probabilities.

The rest of this paper is organized as follows. Section 2 describes the multivariate $t$ distribution and an extension of it that allows different marginals to have different degrees of freedom. Section 3 develops the transform analysis for the quadratic (delta-gamma) approximation to portfolio losses. Section 4 builds on the quadratic approximation to develop an importance sampling procedure for efficient Monte Carlo simulation. Section 5 provides an analysis of the importance sampling estimator and discusses adaptations of the importance sampling procedure for estimating the conditional excess. Section 6 extends the Monte Carlo method through stratification of the quadratic approximation. Section 7 presents numerical examples.

## 2. MULTIVARIATE HEAVY TAILS

The univariate $t$ distribution with $\nu$ degrees of freedom (abbreviated $t_\nu$) has density

$$f(x) = \frac{\Gamma\left(\frac{1}{2}(\nu+1)\right)}{\sqrt{\pi}\Gamma\left(\frac{1}{2}\nu\right)}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2}, \qquad -\infty < x < \infty,$$

with $\Gamma(\cdot)$ denoting the gamma function. This distribution is symmetric about 0. It has polynomial tails and the weight of the tails is controlled by the parameter $\nu$: if $X$ has the $t_\nu$ distribution then

$$P(X > x) \sim \text{constant} \times x^{-\nu}, \quad \text{as} \quad x \to \infty.$$

In contrast, if $Z$ has a standard normal distribution then

$$P(Z > x) \sim \text{constant} \times \frac{e^{-x^2/2}}{x}, \quad \text{as} \quad x \to \infty,$$

so the tails are qualitatively different, especially for small values of $\nu$. If $X \sim t_\nu$, then $E[X^r]$ is finite for $0 < r < \nu$ and infinite for $r \geq \nu$. We are mainly interested in values of $\nu$ roughly in the range of 3 to 7 since this seems to be the level of heaviness typical of market data.

As $\nu \to \infty$, the $t_\nu$ distribution converges to the standard normal. Figure 2.1 compares the $t_3$ distribution with a normal distribution scaled to have the same variance. As the figure illustrates, the $t_3$ has a higher peak and its tails decay much more slowly.

For $\nu > 2$, the $t_\nu$ distribution has variance $\nu/(\nu-2)$. One can scale a $t_\nu$ random variable $X$ by a constant to change its variance and translate it by a constant to change its mean. A linear transformation of $t_\nu$ random variable is sometimes said to have a Pearson Type VII distribution (Johnson, Kotz, and Balakrishnan 1994, p. 21).

Following Anderson (1994), Tong (1990), and others, we refer to

$$f_{\nu,\Sigma}(x) = \frac{\Gamma\left(\frac{1}{2}(m+\nu)\right)}{(\nu\pi)^{m/2}\Gamma\left(\frac{1}{2}\nu\right)|\Sigma|^{1/2}}\left(1+\frac{1}{\nu}x'\Sigma^{-1}x\right)^{-\frac{1}{2}(m+\nu)}, \qquad x \in \mathcal{R}^m. \tag{2.1}$$

as a multivariate $t_\nu$ density. Here, $\Sigma$ is a symmetric, positive definite matrix and $|\Sigma|$ is the determinant of $\Sigma$. If $\nu > 2$, then $\nu\Sigma/(\nu-2)$ is the covariance matrix of $f_{\nu,\Sigma}$. Tong's definition requires that the diagonal entries of $\Sigma$ be 1 (making $\Sigma$ the distribution's correlation matrix if $\nu > 2$); in the more general case of (2.1), the marginals are actually scalar multiples of univariate $t_\nu$ random variables. Anderson's (1984) definition allows a general $\Sigma$ and also a nonzero mean vector. This makes each marginal a linear transformation of a $t_\nu$ random variable (and thus a Pearson Type VII random variable). As is customary in estimating risk measures over short horizons, we will assume a mean of zero and thus use (2.1).

The density in (2.1) depends on the argument $x$ only through the quadratic form $x'\Sigma^{-1}x$; it is therefore *elliptically contoured*, meaning that the sets on which $f$ is constant are hyperellipsoids. Within the elliptically contoured family this density belongs to the class of scale mixtures of normals and thus has a representation as the distribution of the product of a multivariate normal random vector and a univariate random variable independent of the normal (cf. Fang, Kotz, and Ng 1987). In the case of (2.1), this representation can be expressed as follows: if $(X_1, \ldots, X_m)$ has density $f_{\nu,\Sigma}$, then

$$(X_1, \ldots, X_m) =_{\mathrm{d}} \frac{(\xi_1, \ldots, \xi_m)}{\sqrt{Y/\nu}}, \tag{2.2}$$

where $=_{\mathrm{d}}$ denotes equality in distribution, $\xi = (\xi_1, \ldots, \xi_m)$ has distribution $N(0, \Sigma)$, $Y$ has distribution $\chi^2_\nu$ (chi-square with $\nu$ degrees of freedom), and $\xi$ and $Y$ are

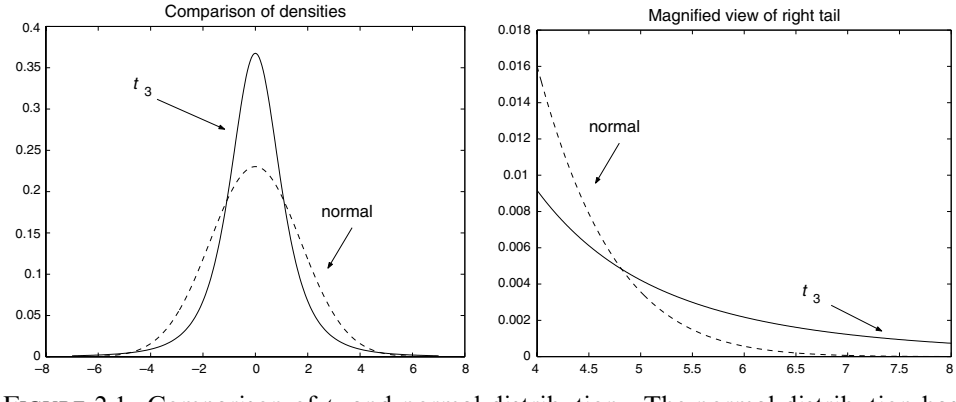


FIGURE 2.1. Comparison of $t_3$ and normal distribution. The normal distribution has been scaled to have variance 3, like the $t_3$ distribution.

independent. This representation is central to our analysis and indeed several of our results hold if $Y$ is replaced with some other positive random variable having an exponential tail. See Chapter 3 of Fang et al. for specific examples of multivariate distributions of the form in (2.2).

From (2.2) we see that modeling changes in risk factors with a multivariate $t$ is similar to assuming a stochastic level of market volatility: given $Y$, the variance of $X_i$ is $\nu\Sigma_{ii}/Y$. It is also clear from (2.2) that $X_i$ and $X_j$ are not independent even if they are uncorrelated—that is, even if $\Sigma_{ij} = 0$. In (2.2), risk factors with little or no correlation may occasionally make large moves together (because of a small outcome of $Y$), a phenomenon sometimes observed in extreme market conditions (see, e.g., Longin and Solnik 1998).

A possible shortcoming of (2.1) and (2.2) is that they require all $X_i$ to share a parameter $\nu$ and thus have equally heavy tails. To extend the model to allow multiple degrees of freedom, we use a *copula*. (For background on copulas see Nelsen 1999; for applications in risk management see Embrechts et al. 2002 and Li 2000.) Let $G_\nu$ denote the cumulative distribution function for the univariate $t_\nu$ density. Let $(X_1, \ldots, X_m)$ have the density in (2.1), assuming for the moment that $\Sigma$ has all diagonal entries equal to 1. Define

$$(\tilde{X}_1, \ldots, \tilde{X}_m) = (G_{\nu_1}^{-1}(G_\nu(X_1)), \ldots, G_{\nu_m}^{-1}(G_\nu(X_m))). \tag{2.3}$$

Each $X_i$ has distribution $t_\nu$ so each $G_\nu(X_i)$ is uniformly distributed on the unit interval and each $G_{\nu_i}^{-1}(G_\nu(X_i)$ has distribution $t_{\nu_i}$; thus, this transformation produces a multivariate distribution whose marginals are $t$ distributions with varying degrees of freedom. This mechanism (as well as the algorithms and proofs in this paper) easily extends to the case where each $X_i$ and $\tilde{X}_i$ is a scalar multiple of a $t$ random variable, but for the sake of simplicity we restrict overselves to the unscaled $t$ whenever we use this copula mechanism.

A limiting special case of this approach takes $\nu = \infty$ in (2.1) and (2.3). This gives $(X_1, \ldots, X_m)$ a normal distribution and thus generates $(\tilde{X}_1, \ldots, \tilde{X}_m)$ through a "normal copula." In practice, we are most interested in values of $\nu_i$ in a relatively narrow range (e.g., 3 to 7); graphical inspection of the joint distributions produced suggests that using (2.3) with $\nu$ close to the values $\nu_i$ of interest is preferable to using a normal copula. For example, Figure 2.2 compares contours of bivariate distributions with $(\nu_1, \nu_2) = (7, 3)$ generated using $\nu = 5$ (left) and $\nu = \infty$ (right). In both cases the correlation in the copula (the correlation between the underlying $X_1$ and $X_2$) is 0.4.

We briefly describe how we envisage the use of (2.3) in modeling market risk factors; see Hosking et al. (2000) for a more thorough discussion and empirical results along these lines. (Hosking et al. use $\nu = \infty$ and call the resulting joint distribution metagaussian; the same approach can be used with a finite $\nu$.) Let $S$ denote an $m$-vector of risk factors (market prices and rates or volatility factors) and let $\Delta S$ denote the change in $S$ over an interval of length $\Delta t$. Think of $\Delta t$ as the horizon over which VAR is to be calculated and thus typically either one day or two weeks. We model each $\Delta S_i$, $i = 1, \ldots, m$, as a scalar multiple of a $t_{\nu_i}$ random variable; assuming $\nu_i > 2$, we can write

$$\Delta S_i = \tilde{\sigma}_i \sqrt{\frac{\nu_i - 2}{\nu_i}} \tilde{X}_i, \qquad \tilde{X}_i \sim t_{\nu_i}. \tag{2.4}$$

This makes $\tilde{\sigma}_i^2$ the variance of $\Delta S_i$. The parameter $\nu_i$ could be estimated using, for example, the methods in Hosking et al. or Johnson, Kotz, and Balakrishnan (1995,

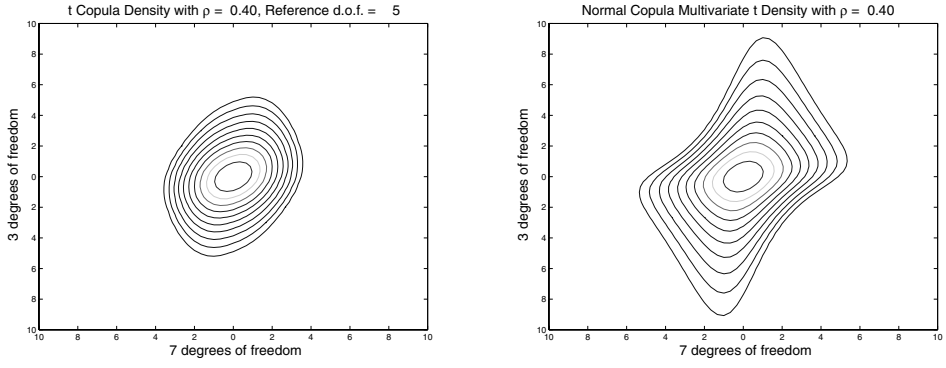


FIGURE 2.2. Comparison of contours of bivariate $t$ distributions with $(\nu_1, \nu_2) = (7, 3)$ generated using a $t$ copula with $\nu = 5$ (left) and a normal copula ($\nu = \infty$, right), both with a correlation of 0.4.

Sec. 28.6). (Alternatively, one might fit a scaled $t$ distribution to the return $\Delta S_i/S_i$. Since we are ultimately interested in the distribution of $\Delta S_i$ conditional on the current $S_i$, the effect of this is to change the definition of $\tilde{\sigma}_i$ in (2.4).)

Once we have chosen marginal distributions as in (2.4), we can define the transformed variables $X_i = G_\nu^{-1}(G_{\nu_i}(\tilde{X}_i))$, $i = 1, \ldots, m$, for some choice of $\nu$. Applying this transformation to historical data, we can then estimate the correlation matrix of $X = (X_1, \ldots, X_m)$. Letting $\Sigma$ denote this correlation matrix and positing that $X$ has the density in (2.1) completes the specification of the model: applying (2.3) to $X$ and then (2.4) to the $\tilde{X}_i$ we recover the $\Delta S_i$. We denote the combined transformation by $\Delta S = K(X)$ with

$$\Delta S_i = K_i(X_i), \qquad K_i(x) = \tilde{\sigma}_i \sqrt{\frac{\nu_i - 2}{\nu_i}} G_{\nu_i}^{-1}(G_\nu(x)). \tag{2.5}$$

This produces a joint distribution for $\Delta S$ that accommodates tails of different heaviness for different marginals and captures some of the dependence among risk factors observed in historical data. Note that $\Sigma$ is not the correlation matrix of $\Delta S$ because (2.3) does not in general preserve correlations. As a monotone transformation, $K(X)$ does however preserve *rank* correlations. For an extensive discussion of dependence properties and the use of copula models in risk management applications, see Embrechts et al. (2002).

## 3. QUADRATIC APPROXIMATION AND TRANSFORM ANALYSIS

This section develops a method for calculating the distribution of the change in portfolio value over a fixed horizon assuming that the changes in underlying risk factors over the horizon are described by a multivariate $t$ distribution and that the change in portfolio value is a quadratic function of the changes in the risk factors. As in the previous section, let $S$ denote an $m$ vector of risk factors to which a portfolio is exposed and let $\Delta S$ denote the change in $S$ from the current time 0 to the end of the horizon $\Delta t$. Fix a portfolio and let $V(t, S)$ denote its value at time $t$ and risk factors $S$. The delta-gamma approximation to the change in portfolio value is

$$V(\Delta t, S+\Delta S) - V(0,\Delta S) \approx \frac{\partial V}{\partial t}\Delta t + \delta'\Delta S + \frac{1}{2}\Delta S'\Gamma\Delta S,$$

with

$$\delta_i = \frac{\partial V}{\partial S_i}, \qquad \Gamma_{ij} = \frac{\partial^2}{\partial S_i \partial S_j}, \qquad i,j = 1,\ldots,m,$$

and all derivatives evaluated at the initial point $(0,S)$.

An important feature of this approximation is that most of the required first- and second-order sensitivites (the deltas, gammas, and time decay) are regularly calculated by financial firms as part of their trading operations. It is therefore reasonable to assume that these sensitivities are available "for free" in calculating VAR and related portfolio risk measures. Although this is an important practical consideration, from a mathematical perspective there is no need to restrict attention to this particular approximation—we will simply assume the availability of some quadratic approximation. Also, we find it convenient to work with the loss $L = V(0,\Delta S) - V(\Delta t, S+\Delta S)$, rather than the increase in value approximated above. Thus, we work with an approximation of the form

$$L \approx a_0 + a'\Delta S + \Delta S' A \Delta S \equiv a_0 + Q, \tag{3.1}$$

with $a_0$ a scalar, $a$ an $m$ vector and $A$ an $m \times m$ symmetric matrix. The delta-gamma approximation has $a = -\delta$ and $A = -\frac{1}{2}\Gamma$. Our interest centers on calculating loss probabilities $P(L > x)$ assuming equality in (3.1), and on the closely related problem of calculating VAR—that is, of finding a quantile $x_p$ for which $P(L > x_p) = p$ with, for example, $p = .01$.

We model the changes in risk factors $\Delta S$ using a multivariate $t$ distribution $f_{\nu,\Sigma}$ as in (2.1) for some symmetric, positive definite matrix $\Sigma$ and a degrees-of-freedom parameter $\nu$. (We consider the more general model in (2.5) at the end of this section.) From the ratio representation (2.2) we know that $\Delta S$ has the distribution of $\xi/\sqrt{Y/\nu}$ with $\xi \sim N(0,\Sigma)$. If $C$ is any matrix for which $CC' = \Sigma$, then $\xi$ has the distribution of $CZ$ with $Z \sim N(0,I)$. Thus, $\Delta S$ has the distribution of $CX$ with $X = Z/\sqrt{Y/\nu}$ (i.e., with $X$ having density $f_{\nu,I}$). It follows that

$$Q =_{\mathrm{d}} (a'C)X + X'(C'AC)X,$$

with $X$ having *uncorrelated* components. We verify in the proof of Theorem 3.1 below that among all $C$ for which $CC' = \Sigma$ it is possible to choose one for which $C'AC$ is diagonal. Letting $\Lambda$ denote this diagonal matrix, $\lambda_1,\ldots,\lambda_m$ its diagonal entries, and $b = a'C$ we conclude that

$$Q =_{\mathrm{d}} b'X + X'\Lambda X = \sum_{j=1}^{m}(b_j X_j + \lambda_j X_j^2). \tag{3.2}$$

At this point we encounter major differences between the normal and $t$ settings. In the normal case ($\nu = \infty$) the $X_j$ are independent because they are uncorrelated. The characteristic function of the sum in (3.2) thus factors as the product of the characteristic functions of the summands. Moreover, each $(b_j X_j + \lambda_j X_j^2)$ is a linear transformation of a noncentral chi-square random variable and thus has a convenient moment generating function and characteristic function (see p. 447 of Johnson et al. 1995). An explicit expression for the characteristic function of $Q$ follows; this can be

inverted numerically to compute probabilities $P(Q > x)$ which can in turn be used to approximate the loss distribution through (3.1). A similar analysis applies if $X$ is a finite mixture of normals.

These simplifying features do not extend to the multivariate $t$. Though uncorrelated, the $X_j$ in (3.2) are no longer independent so the characteristic function does not factor as a product over $j$. Even if it did, the one-dimensional transforms would be difficult to work with: because they are heavy tailed, $X_j$ and $X_j^2$ do not have moment generating functions; each $X_j^2$ has an $F$ distribution, for which the characteristic function is a doubly infinite series (see Johnson et al. 1995, eqn. 27.2). It therefore seems fair to describe the characteristic function of $Q$ in (3.2) as intractable.

Through an indirect approach, we are nevertheless able to calculate the distribution $P(Q \le x)$. Recall the representation $X = Z/\sqrt{Y/\nu}$, define

$$Q_x = (Y/\nu)(Q - x), \tag{3.3}$$

and observe that $P(Q \le x) = P(Q_x \le 0) \equiv F_x(0)$. To compute $P(Q \le x)$ we may therefore find the characteristic function of $Q_x$ and invert it to find $P(Q_x \le 0)$. Note that $Q_x$ is not heavy-tailed and so, unlike $Q$, its moment generating function exists and consequently its characteristic function is more tractable. The necessary steps, starting from (3.1), are provided by the following result. We formulate a more general result by letting $Y$ in the representation $\Delta S = Z/\sqrt{Y/\nu}$ be fairly arbitrary (but positive). Define the moment generating function

$$\phi_Y(\theta) = E\left[e^{\theta Y}\right]$$

and suppose $\phi_Y(\theta) < \infty$ for $0 < \theta < \bar{\theta}_Y$. We specialize to the multivariate $t$ by taking $Y \sim \chi^2_\nu$.

THEOREM 3.1. *Let $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_m$ be the eigenvalues of $\Sigma A$ and let $\Lambda$ be the diagonal matrix with these eigenvalues on the diagonal. There is a matrix $C$ satisfying $CC' = \Sigma$ and $C'AC = \Lambda$. Let $b = a'C$. Then $P(Q \le x) = F_x(0)$, where the distribution $F_x$ has moment generating function (mgf)*

$$\phi_x(\theta) = \phi_Y(\alpha(\theta)) \prod_{j=1}^{m} \frac{1}{\sqrt{1 - 2\theta\lambda_j}} \tag{3.4}$$

*with*

$$\alpha(\theta) = -\frac{\theta x}{\nu} + \frac{1}{2\nu} \sum_{j=1}^{m} \frac{\theta^2 b_j^2}{1 - 2\theta\lambda_j}, \tag{3.5}$$

*provided $\lambda_1\theta < 1/2$ and $\alpha(\theta) < \bar{\theta}_Y$. In the case of the multivariate $t_\nu$,*

$$\phi_x(\theta) = \left(1 + \frac{2\theta x}{\nu} - \sum_{j=1}^{m} \frac{\theta^2 b^2/\nu}{1 - 2\theta\lambda_j}\right)^{-\nu/2} \prod_{j=1}^{m} \frac{1}{\sqrt{1 - 2\theta\lambda_j}}. \tag{3.6}$$

*The characteristic function of $Q_x$ is given by $E[\exp(i\omega Q_x)] = \phi_x(i\omega)$ with $i = \sqrt{-1}$.*

*Proof.* The existence of the required matrix $C$ is the same here as in the normal case (Glasserman et al. 2000); we include the proof because it is constructive and thus useful in implementation. Let $B$ be any matrix for which $BB' = \Sigma$ (e.g., the Cholesky factor

of $\Sigma$). As a symmetric matrix, $B'AB$ has real eigenvalues; these eigenvalues are the same as those of $BB'A = \Sigma A$, namely $\lambda_1, \ldots, \lambda_m$. Moreover, $B'AB = U\Lambda U'$ where $U$ is an orthogonal matrix whose columns are eigenvectors of $B'AB$. It follows that $U'B'ABU = \Lambda$ and $(BU)(BU)' = BB' = \Sigma$, so setting $C = BU$ produces the required matrix.

Given $C$, we can assume $Q$ has the diagonalized form in (3.2) with $X$ having density $f_{v,I}$. To calculate the mgf of $Q_x$, we first condition on $Y$:

$$
\begin{aligned}
(3.7) \qquad E[\exp(\theta Q_x)|Y] &= E[\exp(\theta(Y/v)(Q-x))|Y] \\
&= E\left[\exp\left(\theta\left[\sum_{j=1}^{m}(b_j\sqrt{Y/v}Z_j + Z_j^2 j) - xY/v\right]\right)|Y\right] \\
&= e^{-x\theta Y/v}\prod_{j=1}^{m} E\left[\exp\left(\theta(b_j\sqrt{Y/v}Z_j + \lambda_j Z_j^2)\right)|Y\right],
\end{aligned}
$$

because the uncorrelated standard normal random variables $Z_j$ are in fact independent. As in equation (29.6) of Johnson et al. (1995) we have, for $Z_j \sim N(0,1)$,

$$
E[\exp(u(Z_j + a)^2)] = (1-2u)^{-1/2}\exp\left(\frac{a^2 u}{1-2u}\right), \qquad u < 1/2;
$$

this is the mgf of a noncentral $\chi_1^2$ random variable. Each factor in (3.7) for which $\lambda_j \neq 0$ can be evaluated using this identity by writing

$$
b_j\sqrt{Y/v}Z_j + \lambda_j Z_j^2 = \lambda_j\left(Z_j + \frac{b_j\sqrt{Y/v}}{2\lambda_j}\right)^2 - \frac{b_j^2(Y/v)}{4\lambda_j^2},
$$

If $\lambda_j = 0$, use $E[\exp(uZ_j)] = \exp(u^2/2)$. The product in (3.7) thus evaluates to

$$
e^{-x\theta Y/v}\exp\left(\frac{1}{2}\sum_{j=1}^{m}\frac{\theta^2 b_j^2 Y/v}{1-2\theta\lambda_j}\right)\prod_{j=1}^{m}\frac{1}{\sqrt{1-2\theta\lambda_j}}
$$

which is

$$
(3.8) \qquad e^{\alpha(\theta)Y}\prod_{j=1}^{m}\frac{1}{\sqrt{1-2\theta\lambda_j}}.
$$

Taking the expectation over $Y$ yields (3.4). If $Y \sim \chi_v^2$, then

$$
\phi_Y(\theta) = (1-2\theta)^{-v/2}, \qquad \theta < 1/2,
$$

so the expectation of (3.8) becomes

$$
(1-2\alpha(\theta))^{-v/2}\prod_{j=1}^{m}\frac{1}{\sqrt{1-2\theta\lambda_j}},
$$

which is (3.6).

Finally, from Lukacs (1970, p.11) we may conclude that if the moment generating function is finite in a neighborhood of the origin, then the characteristic function is equal to mgf evaluated at purely imaginary arguments. □

In the specific case of the delta-gamma approximation, the matrix $A$ in (3.1) is $-\frac{1}{2}\Gamma$ and the parameters $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of $-\frac{1}{2}\Sigma\Gamma$. The constant $a_0$ is

$-\Delta t(\partial V/\partial t)$. The delta-gamma approximation to a loss probability is $P(L > x) \approx P(Q > x - a_0)$. We evaluate this approximation using $P(Q > x - a_0) = 1 - P(Q_{x-a_0} \leq 0) = 1 - F_{x-a_0}(0)$. Values of $F_{x-a_0}$ can be found through the inversion integral

$$F_{x-a_0}(t) - F_{x-a_0}(t-y) = \frac{1}{\pi} \operatorname{Re}\left( \int_0^\infty \phi_{x-a_0}(iu) \left[ \frac{e^{iuy} - 1}{iu} \right] e^{-iut} \right) du, \qquad i = \sqrt{-1} \tag{3.9}$$

which is obtained from the standard inversion formula (see, e.g., (3.11) on p. 511 of Feller 1971). This integral can be evaluated numerically; see Abate and Whitt (1992) for a discussion of numerical issues involved in transform inversion. In implementing this method we choose a large value of $y$ for which $F_{x-a_0}(t-y)$ can be assumed to be approximately zero. As the mean and variance of $Q_{x-a_0}$ can be easily computed, Chebychev's Inequality may be used to find a value of $y$ for which $F_{x-a_0}(t-y)$ is appropriately small.

It should be noted that this transform inversion procedure is not quite as computationally efficient as the corresponding approach based on normally distributed risk factors. In the normal case one can evaluate the transform of $Q$ explicitly; a single Fast Fourier Transform inversion then computes $N$ points on the distribution of $Q$ based on an $N$-term approximation to the inversion integral in $O(N \log N)$ time. In our setting, each value $x$ at which the distribution of $Q$ is to be evaluated relies on a separate inversion integral so computing $M$ points of the distribution, each based on an $N$-term approximation to the corresponding integral, requires $O(MN)$ time. Nevertheless, the total computing time of this approach remains modest and the additional effort makes possible the inclusion of realistically heavy tails.

The transform analysis provided by Theorem 3.1 can accommodate different degrees of heaviness in the tails of different risk factors. The result extends to this case through the copula mechanism in (2.3) and the chain rule for differentiation. Suppose we model the changes in risk factors $\Delta S$ as $K(X_1, \ldots, X_m)$ using (3.5) with $(X_1, \ldots, X_m)$ having density $f_{\nu,\Sigma}$ (with the diagonal elements of $\Sigma$ being 1). Then

$$\frac{\partial V}{\partial X_i} = \delta_i \frac{dK_i}{dX_i},$$

$$\frac{\partial^2 V}{\partial X_i \partial X_j} = \Gamma_{ij} \frac{dK_i}{dX_i} \frac{dK_j}{dX_j}, \quad i \neq j, \qquad \frac{\partial^2 V}{\partial X_i^2} = \Gamma_{ii} \left( \frac{dK_i}{dX_i} \right)^2 + \delta_i \frac{d^2 K_i}{dX_i^2},$$

with all derivatives of $K$ evaluated at 0. In this way, the delta-gamma approximation generalizes to a quadratic approximation

$$L \approx a_0 + \tilde{a}' X + X' \tilde{A} X \tag{3.10}$$

in $(X_1, \ldots, X_m)$, the parameters $\tilde{a}$ and $\tilde{A}$ now depending on the derivatives of $K$ as well as the usual portfolio deltas and gammas. The derivatives of $K$ do not depend on the portfolio and are therefore easily computed.

Figure 3.1 shows the accuracy of the delta-gamma approximation for one of the portfolios considered in Section 7, portfolio (a.3) (0.1 yr ATM) consisting of European puts and calls. The exact loss distribution was estimated to a high degree of precision using Monte Carlo simulation. Although the absolute errors (delta-gamma approximation minus exact) are all within 0.01, the relative errors are large (up to 90%) and this translates into a large error in the VAR. This illustrates the need for high-accuracy Monte Carlo techniques.

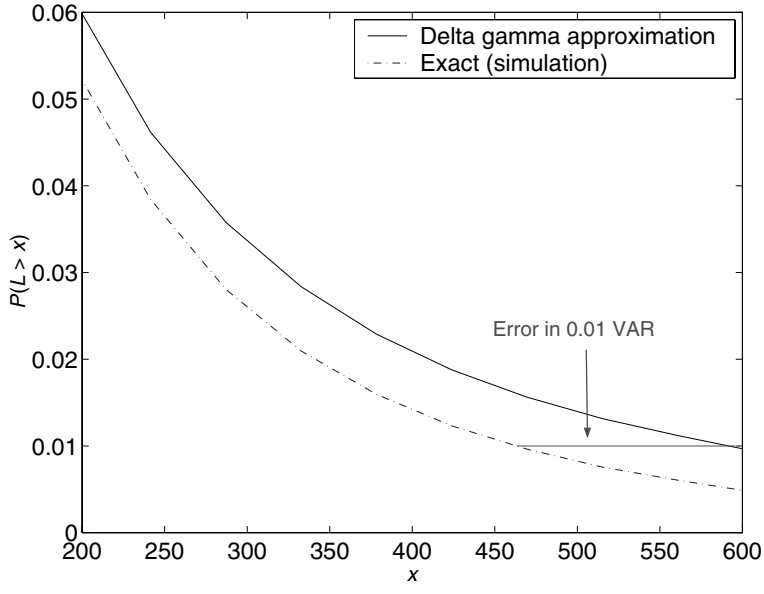


FIGURE 3.1. Comparison of the delta-gamma approximate and exact loss distributions for (a.3), the 0.1 yr ATM portfolio. The exact loss distribution was estimated by Monte Carlo simulation to a high degree of precision.

## 4. IMPORTANCE SAMPLING

The quadratic approximation of the previous section is fast but may not be sufficiently accurate for all applications. In contrast, Monte Carlo simulation does not require any approximation to the relation between changes in risk factors and changes in portfolio value, but it is much more time consuming. The rest of this paper develops methods for accelerating Monte Carlo by exploiting the delta-gamma approximation; these methods thus combine some of the best features of the two approaches.

A generic Monte Carlo simulation to estimate a loss probability $P(L > x)$ consists of the following main steps. Samples of risk-factor changes $\Delta S$ are drawn from a distribution; for each sample, the portfolio is revalued to compute $V(\Delta t, S + \Delta S)$ and the loss $L = V(0, S) - V(\Delta t, S + \Delta S)$; the fraction of samples resulting in a loss greater than $x$ is used to estimate $P(L > x)$. For large portfolios of complex derivative securities, the bottleneck in this procedure is portfolio revaluation. The objective of using a variance reduction technique is to reduce the number of samples (and thus the number of revaluations) needed to achieve a desired precision. We use importance sampling based on the delta-gamma (or other quadratic) approximation to reduce variance, particularly at large loss thresholds $x$.

### 4.1. Exponential Change of Measure

We begin by reviewing an importance sampling method developed in Glasserman et al. (2000) in the case of normally distributed risk factors. Suppose that $Q$ has been diagonalized as in (3.1) but with the $X_j$ replaced (for the moment) with independent standard normals $Z_j$. As in Theorem 3.1, let $\lambda_1$ be the largest of the eigenvalues $\lambda_j$;

suppose now that $\lambda_1 > 0$ (otherwise, $Q$ is bounded above). In Glasserman et al. (2000) we introduced an exponential change of measure by setting, for $0 < \theta < 1/(2\lambda_1)$,

$$dP_\theta = e^{\theta Q - \psi(\theta)} dP, \tag{4.1}$$

with $P$ the original measure under which $Z \sim N(0, I)$ and $\psi(\theta) = \log E[\exp(\theta Q)]$. Moreover, we showed that under $P_\theta$ the components of $Z$ remain independent but with

$$Z_j \sim N\left(\frac{\theta b_i}{1 - 2\theta\lambda_i}, \frac{1}{1 - 2\theta\lambda_i}\right).$$

It is thus a simple matter to sample $Z$ under $P_\theta$ and then to sample $\Delta S$ by setting $\Delta S = CZ$.

By (4.1), the likelihood ratio is given by $e^{-\theta Q + \psi(\theta)}$, so the key identity for importance sampling in the normal setting is

$$P(L > x) = E_\theta\left[e^{-\theta Q + \psi(\theta)} I(L > x)\right],$$

with $I(\cdot)$ denoting the indicator of the event in parentheses. We may thus generate samples under $P_\theta$ and estimate $P(L > x)$ using the expression inside the expectation on the right. This estimator is unbiased and its second moment is

$$E_\theta\left[e^{-2\theta Q + 2\psi(\theta)} I(L > x)\right] = E\left[e^{-\theta Q + \psi(\theta)} I(L > x)\right].$$

If $L \approx a_0 + Q$, then we can expect $e^{-\theta Q}$ to be small when $L > x$, resulting in a reduced second moment, especially for large $x$. In Glasserman et al. (2000) we showed that if $L = a_0 + Q$ holds exactly, then this estimator (with suitable $\theta$) is asymptotically optimal in the sense that its second moment decreases exponentially (as $x \to \infty$) at twice the rate of exponential decrease of $P(L > x)$ itself. This is the best possible rate for the second moment of *any* unbiased estimator. Asymptotic optimality holds with, for example, $\theta = \theta_{x-a_0}$, where $\theta_{x-a_0}$ solves

$$\frac{d}{d\theta}\psi(\theta) = x - a_0.$$

This choice makes $E_\theta[Q] = x - a_0$ and thus makes losses $L$ close to $x$ typical rather than rare. As shown in Glasserman et al. (2000), the method is not very sensitive to the exact choice of $\theta$.

An attempt to apply similar ideas with a multivariate $t$ seems doomed by the failure of (4.1) to generalize to the heavy-tailed setting. In any model in which the $X_i$ (and hence also $Q$) are heavy tailed, one cannot define an exponential change of measure based on $Q$ because $E[\exp(\theta Q)]$ is infinite for any $\theta > 0$. Most successful applications of importance sampling are based on an exponential change of measure; and the extension of light-tailed methods to heavy-tailed models can often give disastrous results, as demonstrated in Asmussen et al. (2000).

As in the transform analysis of Section 3, we circumvent this difficulty by working with $Q_x$ in (3.3) rather than $Q$ itself. We use the notation of Theorem 3.1. Let $\psi_x = \log \phi_x$ and $\psi_Y = \log \phi_Y$. Recall that $\lambda_1 = \max_j \lambda_j$.

THEOREM 4.1. *If $\theta\lambda_1 < 1/2$ and $\alpha(\theta) < \bar{\theta}_Y$, then*

$$dP_\theta = \exp\left(\theta Q_x - \psi_x(\theta)\right) dP$$

*defines a probability measure and*

$$P(L > y) = E_\theta\left[e^{-\theta Q_x + \psi_x(\theta)} I(L > y)\right] = E_\theta\left[e^{-\theta(Y/v)(Q-x)+\psi_x(\theta)} I(L > y)\right]. \tag{4.2}$$

*Under $P_\theta$, $X$ has the distribution of $Z/\sqrt{Y/v}$ where*

$$P_\theta(Y \le u) = E\left[e^{\alpha(\theta)Y - \psi_Y(\alpha(\theta))} I(Y \le u)\right], \tag{4.3}$$

*and conditional on $Y$, the components of $Z$ are independent with*

$$Z_j \sim N(\mu_j(\theta), \sigma_j^2(\theta)), \qquad \mu_j(\theta) = \frac{\theta b_j \sqrt{Y/v}}{1 - 2\theta\lambda_j}, \qquad \sigma_j^2(\theta) = \frac{1}{1 - 2\theta\lambda_j}. \tag{4.4}$$

*In the specific case that the distribution of $X$ under $P$ is multivariate $t_v$ (i.e., the P-distribution of $Y$ is $\chi_v^2$), the distribution of $Y$ under $P_\theta$ is Gamma$(v/2, 2/(1 - 2\alpha(\theta)))$, the gamma distribution with shape parameter $v/2$ and scale parameter $2/(1 - 2\alpha(\theta))$.*

*Proof.* The first assertion follows from the fact that $\psi_x(\theta)$ is finite under the conditions imposed on $\theta$, and (4.2) then follows from the fact that the likelihood ratio $dP/dP_\theta$ is $\exp(-\theta Q_x + \psi_x(\theta))$.

Fix constants $\theta$ and $\alpha$ satisfying $\theta\lambda_1 < 1/2$ and $\alpha < \bar{\theta}_Y$. The probability measure $P_{\alpha,0}$ obtained by *exponentially twisting* $Y$ by $\alpha$ is defined by the likelihood ratio

$$\frac{dP_{\alpha,0}}{dP} = e^{\alpha Y - \psi_Y(\alpha)}.$$

Let $h(z)$ denote the standard normal density $\exp(-z^2/2)/\sqrt{2\pi}$; the density of the $N(\mu, \sigma^2)$ distribution is $h([z - \mu]/\sigma)/\sigma$. The probability measure $P_{\alpha,\theta}$ obtained by exponentially twisting $Y$ by $\alpha$ and, conditional on $Y$, letting the $Z_j$ be independent with the distributions in (4.4) is defined by the likelihood ratio

$$\frac{dP_{\alpha,\theta}}{dP} = e^{\alpha Y - \psi_Y(\alpha)} \times \prod_{j=1}^{m} \frac{h([Z_j - \mu_j(\theta)]/\sigma_j(\theta)])/\sigma_j(\theta)}{h_j(Z_j)}. \tag{4.5}$$

The $j$th factor in the product in (4.5) evaluates to

$$\begin{aligned}\frac{h([Z_j - \mu_j(\theta)]/\sigma_j(\theta)])/\sigma_j(\theta)}{h_j(Z_j)} &= \frac{1}{\sigma_j(\theta)} \exp\left(\frac{1}{2} Z_j^2 \left(1 - \frac{1}{\sigma_j^2(\theta)}\right) + Z_j \frac{\mu_j(\theta)}{\sigma_j^2(\theta)} - \frac{1}{2}\mu_j(\theta)^2\right) \\ &= \sqrt{1 - 2\theta\lambda_j} \exp\left(\theta\lambda_j Z_j^2 + Z_j \theta b_j \sqrt{Y/v} - \frac{\theta_j^2 b_j^2 Y/v}{2(1 - 2\theta\lambda_j)}\right).\end{aligned}$$

The likelihood ratio (4.5) is thus

$$e^{\alpha Y - \psi_Y(\alpha)} \times \prod_{j=1}^{m} \sqrt{1 - 2\theta\lambda_j} \exp\left(\theta \sum_{j=1}^{m} \left(\lambda_j Z_j^2 + Z_j b_j \sqrt{Y/v}\right)\right) \times \exp\left(-\frac{1}{2} \sum_{j=1}^{m} \frac{\theta^2 b_j^2 Y/v}{1 - 2\theta\lambda_j}\right),$$

which we can write as

$$e^{\alpha Y + \theta(Y/v)Q} \left(\prod_{j=1}^{m} \sqrt{1 - 2\theta\lambda_j} e^{-\psi_Y(\alpha)}\right) \exp\left(-Y \frac{1}{2} \sum_{j=1}^{m} \frac{\theta^2 b_j^2/v}{1 - 2\theta\lambda_j}\right).$$

If we choose

$$\alpha = \alpha(\theta) \equiv -\frac{\theta x}{v} + \frac{1}{2}\sum_{j=1}^{m} \frac{\theta^2 b_j^2 / v}{1 - 2\theta\lambda_j},$$

the likelihood ratio simplifies to

$$e^{-\theta x(Y/v)+\theta(Y/v)Q}\left(\prod_{j=1}^{m}\sqrt{1-2\theta\lambda_j}e^{-\psi_Y(\alpha(\theta))}\right) = e^{\theta Q_x - \psi_x(\theta)},$$

in light of the definition of $Q_x$ in (3.3), the definition of $\psi_x$ as $\log \phi_x$, and the expression for $\phi_x$ in (3.4). Since this coincides with the likelihood ratio $dP_\theta/dP$, we conclude that the $P_\theta$-distribution of $(Y, Z_1, \ldots, Z_m)$ is as claimed.

Consider now the multivariate $t_\nu$ case. To find the density of $Y$ under $P_\theta$, we multiply the $\chi^2_\nu$ density by the likelihood ratio to get

$$e^{\alpha y - \psi_Y(\alpha)}\frac{y^{(\nu-2)/2}e^{-y/2}}{2^{\nu/2}\Gamma(\nu/2)} = \left(\frac{2}{1-2\alpha}\right)^{-\nu/2}\frac{y^{(\nu-2)/2}}{\Gamma(\nu/2)}\exp\left(\frac{-y}{2/(1-2\alpha)}\right),$$

using $\exp(-\psi_Y(\alpha)) = (1-2\alpha)^{\nu/2}$. This is the gamma density with shape parameter $\nu/2$ and scale parameter $2/(1-2\alpha)$.

## 4.2. Importance Sampling Algorithm

Embedded in the proof of Theorem 4.1 are the essential steps of a simulation procedure that uses the quadratic approximation to guide the sampling of risk factors. We now make this explicit, detailing the steps involved in estimating a loss probability $P(L > y)$. We assume for now that a value of $\theta > 0$ consistent with the conditions of Theorem 4.1 has been selected. Later we address the question of how $\theta$ should be chosen.

ALGORITHM 4.1. Importance Sampling Estimate of Loss Probability
For each of $n$ independent replications:

1. Generate $Y$ from the distribution in (4.3). (In the multivariate $t_\nu$ model, this means generating $Y$ from the gamma distribution in the theorem.)
2. Given $Y$, generate independent normals $Z_1, \ldots, Z_m$ with parameters as in (4.4).
3. Set $X = Z/\sqrt{Y/\nu}$.
4. Set $\Delta S = CX$ and calculate the resulting portfolio loss $L$ and the quadratic approximation $Q$. Set $Q_x = (Y/\nu)(Q - x)$.
5. Multiply the loss indicator by the likelihood ratio to get

$$e^{-\theta Q_x + \psi_x(\theta)} I(L > y) \tag{4.6}$$

Average (4.6) over the $n$ independent replications.

Applying this algorithm with the copula model (2.5) requires changing only the first part of step 4: to sample the change in risk factors, set $\Delta S = K(\tilde{C}Z/\sqrt{Y/\nu})$, where $\tilde{C}$ satisfies $\tilde{C}\tilde{C}' = \Sigma$ and $\tilde{C}'\tilde{A}\tilde{C}$ is diagonal, with $\tilde{A}$ as in (3.10). (Recall that in the setting of (2.5) $\Sigma$ is assumed to have diagonal entries equal to 1 and represents the correlation matrix of the copula variables $(X_1, \ldots, X_m)$ and not of $\Delta S$.) The matrix $\tilde{C}$ can be constructed by following the steps in the proof of Theorem 3.1. Also, $Q = \sum_j b_j X_j + \sum_j \lambda_j^2 X_j^2$ with $(b_1, \ldots, b_m) = \tilde{a}\tilde{C}$ and $\lambda_1, \ldots, \lambda_m$ the diagonal elements of $\tilde{C}'\tilde{A}\tilde{C}$.

Notice that in Algorithm 4.1 we have not applied an exponential change of measure to the heavy-tailed random variables $X_i$. Instead, we have applied an exponential change of measure to $Y$ and then, conditional on $Y$, applied an exponential change of measure to $Z$.

To develop some intuition for this procedure, consider again the diagonalized quadratic approximation in (3.2) and the representation $X = Z/\sqrt{Y/\nu}$ of the transformed (and uncorrelated) risk factors $X$. An objective of any importance sampling procedure for estimating $P(L > y)$ is to make the event $\{L > y\}$ more likely under the importance sampling measure than under the original measure. Achieving this is made difficult by the fact that the actual loss function may be quite complicated and may be known only implicitly through the procedures used to value individual components of a portfolio. As a substitute we use an approximation to $L$ (in particular, the quadratic approximation $Q$) and design the change of measure to make large values of $Q$ more likely.

Consider the change of measure in Theorem 4.1 and Algorithm 4.1 from this perspective. The parameter $\alpha(\theta)$ will typically be negative because $\theta$ is positive and typically small (so $\theta^2 \ll \theta$). In exponentially twisting $Y$ by $\alpha < 0$, we are giving greater probability to smaller values of $Y$ and thus increasing the overall variability of the risk factors, since $Y$ appears in the denominator of $X$. Given $Y$, the mean $\mu_j(\theta)$ in (4.4) is positive if $b_j > 0$ and negative if $b_j < 0$. This has the effect of increasing the probability of positive values of $Z_j$ if $b_j > 0$ and negative values of $Z_j$ if $b_j < 0$; in both cases, the combined effect is to increase the probability of large values of $b_j Z_j$ and thus of $Q$. Similarly, $\sigma_j(\theta) > 1$ if $\lambda_j > 0$ and by increasing the standard deviation of $Z_j$ we make large values of $\lambda_j Z_j^2$ more likely.

This discussion should help motivate the importance sampling approach of Theorem 4.1 and Algorithm 4.1, but it must be stressed that the validity of the algorithm (as provided by (4.2)) is not based on assuming that the quadratic approximation holds exactly. In fact, the procedure above produces unbiased estimates even if the $b_j$ and $\lambda_j$ bear no relation to the true portfolio. Of course, the effectiveness of the procedure in reducing variance will in part be determined by the accuracy of the quadratic approximation.

We close this section by addressing the choice of parameter $\theta$. In fact we also have flexibility in choosing the value $x$ used to define $Q_x$. The choice of $x$ is driven by the approximation $P(L > y) \approx P(Q > x)$; in light of (3.1), it is natural to take $x = y - a_0$. Ideally, we would like to choose $\theta$ to minimize the second moment

$$E_\theta\left[e^{-2\theta Q_x + 2\psi_x(\theta)} I(L > x + a_0)\right] = E\left[e^{-\theta Q_x + \psi_x(\theta)} I(L > x + a_0)\right]. \tag{4.7}$$

Since this is ordinarily intractable, we apply the quadratic approximation and choose a value of $\theta$ that would be attractive with $L$ replaced by $a_0 + Q$. After this substitution, noting that $Q_x > 0$ when $Q > x$, we can bound the second moment using

$$E\left[e^{-\theta Q_x + \psi_x(\theta)} I(Q > x)\right] \le e^{\psi_x(\theta)}.$$

The function $\psi_x$ is convex and $\psi_x(\theta) \to \infty$ as $\theta \uparrow 1/(2\lambda_1)$ (assuming $\lambda_1 > 0$) and as $\theta$ decreases to a point at which $\alpha(\theta) = \bar{\theta}_Y$. Hence, this upper bound is minimized by the point $\theta_x$ solving

$$\frac{d}{d\theta}\psi_x(\theta) = 0. \tag{4.8}$$

The root of this equation is easily found numerically.

The value $\theta_x$ determined by (4.8) has a second interpretation that makes it appealing. The function $\psi_x$ is the cumulant generating function (the logarithm of the moment generating function) of the random variable $Q_x$. A standard property of exponential families implies that at any $\theta$ for which $\psi_x(\theta) < \infty$, we have $E_\theta[Q_x] = d\psi_x(\theta)/d\theta$. By choosing $\theta_x$ as the root of (4.8) we are choosing $E_{\theta_x}[Q_x] = 0$. This may be viewed as centering the distribution of $Q_x$ near 0, which is equivalent to centering $Q$ near $x$. Thus, by sampling under $P_{\theta_x}$ we are making values of $Q$ near $x$ typical, whereas they may have been rare under the original probability measure.

Equation (4.8) provides a convenient and effective means of choosing $\theta$. In our numerical experiments we find that the performance of the importance sampling method is not very sensitive to the exact choice of $\theta$ and a single parameter can safely be used for estimating $P(L > y)$ at multiple values of $y$. These observations are consistent with the theoretical properties established in the next section.

## 5. ANALYSIS OF THE ESTIMATOR

In this section we provide theoretical support for the effectiveness of the importance sampling method of the previous section. We do this by analyzing the second moment of the estimator at large loss thresholds (and thus small loss probabilities). We carry out this analysis under the hypothesis that the quadratic approximation (3.1) holds exactly and interpret the results as ensuring the effectiveness of the method whenever the quadratic approximation is sufficiently informative. The results of this section are specific to the multivariate $t_\nu$ distribution though similar results should hold under appropriate conditions on the distribution of $Y$.

### 5.1. Bounded Relative Error and Asymptotic Optimality

Consider any unbiased estimator $\hat{p}$ of the probability $P(Q > x)$ and let $m_2(\hat{p})$ denote its second moment. The variance of the estimator is $m_2(\hat{p}) - (P(Q > x))^2$. Since the variance must be nonnegative, the second moment can never be smaller than $(P(Q > x))^2$. As in Shahabuddin (1994), we say that an estimator has *bounded relative error* if

$$\limsup_{x\to\infty} \frac{m_2(\hat{p})}{(P(Q > x))^2} < \infty. \tag{5.1}$$

In Lemma 5.1 we provide conditions under which $P(Q > x)$ is of the order of $x^{-\nu/2}$. When this holds, we must also have

$$m_2(\hat{p}) \geq \text{constant} \times x^{-\nu}.$$

In this case, bounded relative error becomes equivalent to the requirement that there be a constant $c$ such that

$$m_2(\hat{p}) \leq cx^{-\nu}, \quad \text{for all sufficiently large } x. \tag{5.2}$$

The bounded relative error property ensures that the number of replications required to achieve a fixed relative accuracy (confidence interval halfwidth of the estimator divided by the quantity that is being estimated) remains bounded as $x \to \infty$, unlike standard simulation where this can be shown to tend to infinity. This property is stronger than the standard notion of asymptotic optimality used in much of the rare

event simulation literature (see, e.g., the discussion in Glasserman et al. 1999a) where the number of replications may also tend to infinity but at a much slower rate than in standard simulation. It is also worth noting that (5.1) and (5.2) apply to the degenerate (and best possible) estimator $\hat{p} \equiv P(Q > x)$, which corresponds to knowing the quantity being estimated. The second moment of this estimator is simply $(P(Q > x))^2$, and from Lemma 5.1, below, we know that this decays at rate $x^{-\nu}$. Conditions (5.1) and (5.2) may thus be interpreted as stating that an estimator with bounded relative error is, up to a constant factor, as good as knowing the answer, at large values of $x$.

As indicated by this discussion, the first step in analyzing the relative error of our estimator is analyzing the tail of $Q$ in (3.1). As explained in the discussion leading to (3.2) we may assume that the $X_i$ are uncorrelated, with density $f_{\nu,I}$. We begin by noting that the quadratic form $Q$ is bounded above by a constant if $\lambda_i < 0$ for all $i$; that is, $P(Q > x) = 0$ for large enough $x$ in this case. To avoid such trivial cases, we assume $\lambda_1 > 0$ (recall that $\lambda_1$ is the largest of the $\lambda_i$'s).

LEMMA 5.1. *If $\lambda_1 > 0$, there are positive constants $c_1, c_2$ such that for all sufficiently large $x$*

$$P(Q > x) \le c_1 x^{-\nu/2} \tag{5.3}$$

*and if $\lambda_j > 0$, $j = 1, \ldots, m$,*

$$P(Q > x) \ge c_2 x^{-\nu/2}. \tag{5.4}$$

*Proof.* Recall from the definition of $Q_x$ in (3.3) that $P(Q > x) = P(Q_x > 0)$. For any $\theta > 0$ at which $\psi_x(\theta) < \infty$ we have

$$P(Q_x > 0) \le E\left[e^{\theta Q_x} I(Q_x > 0)\right] \le E\left[e^{\theta Q_x}\right] = e^{\psi_x(\theta)} = \phi_x(\theta).$$

From (3.6) we see that

$$\phi_x(\theta) = \frac{a_1}{(a_2 + a_3 x)^{\nu/2}} \le c_1 x^{-\nu/2}, \tag{5.5}$$

for some $a_1$, $a_2$, $a_3$, and $c_1 > 0$ (the $a_i$ depending on $\theta$). This proves (5.3).

For the second claim, let $d_j = b_j/(2\lambda_j)$, $j = 1, \ldots, m$, so

$$\begin{aligned} P(Q > x) &= P\left(\sum_{j=1}^{m} (\lambda_j X_j^2 + b_j X_j) > x\right) \\ &= P\left(\sum_{j=1}^{m} \lambda_j (X_j + d_j)^2 - d_j^2/\lambda_j > x\right) \\ &\ge P\left(\lambda_1 (X_1 + d_1)^2 > x + \sum_{j=1}^{m} (d_j^2/\lambda_j)\right) \\ &\ge P\left(X_1 > -d_1 + \sqrt{\left[x + \sum_{j=1}^{m} (d_j^2/\lambda_j)\right]/\lambda_1}\right) \\ &\ge P(X_1 > c_3 \sqrt{x}) \end{aligned} \tag{5.6}$$

for some constant $c_3$ and all sufficiently large $x$. But because $X_1 \sim t_\nu$, we have $P(X_1 > u) \geq c_4 u^{-\nu}$ for some $c_4$ and all sufficiently large $u$. Applying this to (5.6) proves (5.4). □

We now use this result and the ideas surrounding (5.1) and (5.2) to examine our importance sampling estimator applied to $P(Q > x)$, namely

$$e^{-\theta Q_x + \psi_x(\theta)} I(Q > x) \tag{5.7}$$

sampled under $P_\theta$. This coincides with our estimate of $P(L > x + a_0)$ under the hypothesis that the quadratic approximation is exact. Let

$$m_2(\theta, x) = E_\theta\left[e^{-2\theta Q_x + 2\psi_x(\theta)} I(Q > x)\right] = E\left[e^{-\theta Q_x + \psi_x(\theta)} I(Q > x)\right] \tag{5.8}$$

denote the second moment at parameter $\theta$.

THEOREM 5.1. *If $\lambda_1 > 0$, for any fixed $\theta > 0$ at which $\psi_x(\theta) < 0$ there is a constant $c(\theta)$ for which*

$$m_2(\theta, x) \leq c(\theta) P(Q > x) x^{-\nu/2} \tag{5.9}$$

for all sufficiently large $x$; if $\lambda_j > 0$, $j = 1, \ldots, m$, the estimator (5.7) of $P(Q > x)$ has bounded relative error. If $\theta_x$ denotes the solution to (4.8) and $\lambda_1 > 0$, then there is a constant $d$ for which

$$m_2(\theta_x, x) \leq P(Q > x) x^{-\nu/2} d \tag{5.10}$$

for all sufficiently large $x$; if $\lambda_j > 0$, $j = 1, \ldots, m$, the estimator based on $\theta_x$ also has bounded relative error.

*Proof.* From (5.8) and the fact that $\theta$ is positive we get

$$m_2(\theta, x) \leq E\left[e^{\psi_x(\theta)} I(Q > x)\right] = \phi_x(\theta) P(Q > x).$$

From (5.5) we get (5.9). If all $\lambda_j$ are positive then (5.4) holds and

$$\frac{m_2(\theta, x)}{(P(Q > x))^2} \leq \frac{\phi_x(\theta)}{P(Q > x)} \leq \frac{c_1 x^{-\nu/2}}{c_2 x^{-\nu/2}} = \frac{c_1}{c_2},$$

for some positive constants $c_1$, $c_2$ and all sufficiently large $x$. This establishes the bounded relative error property.

For (5.10), we claim that

$$0 < \lim_{x \to \infty} \theta_x < 1/(2\lambda_1) \tag{5.11}$$

with $\lambda_1$ the largest of the $\lambda_j$. Once we establish that (5.11) holds, it follows from (3.6) that $\phi_x(\theta_x) x^{\nu/2}$ is bounded by a constant $d$ as $x \to \infty$. This implies (5.10) by the same argument used for (5.9). Similarly, bounded relative error using $\theta_x$ again follows once (5.4) holds.

It remains to verify (5.11). We can write the derivative of $\psi_x$ as

$$\psi_x'(\theta) \equiv \frac{d}{d\theta} \psi_x(\theta) = \sum_{j=1}^{m} \frac{\lambda_j}{1 - 2\theta\lambda_j} - \frac{\nu}{2} \frac{d\alpha(\theta)/d\theta}{1 + \alpha(\theta)}$$

with

$$\frac{d}{d\theta}\alpha(\theta) = \frac{2x}{\nu} - \frac{2}{\nu}\sum_{j=1}^{m}\frac{\theta b_j^2(1+\lambda_j\theta)}{1-2\theta\lambda_j}.$$

From this we see that the limit $g(\theta) = \lim_{x\to\infty}\psi_x'(\theta)$ exists for all $0 < \theta < 1/(2\lambda_1)$ and is given by

$$g(\theta) = -\frac{\nu}{2\theta} + \sum_{j=1}^{m}\frac{\lambda_j}{1-2\theta\lambda_j}.$$

The function $g$ is increasing in $\theta$ with $g(\theta) \to -\infty$ as $\theta \downarrow 0$ and $g(\theta) \to \infty$ as $\theta \uparrow 1/(2\lambda_1)$. It follows that there is a unique point $\beta$, in $(0, 1/(2\lambda_1))$ at which $g(\beta) = 0$. The claim (5.11) holds if we can show that $\theta_x \to \beta$. For this, choose $\epsilon > 0$ sufficiently small that $\beta - \epsilon > 0$ and $\beta + \epsilon < 1/(2\lambda_1)$. Then $g(\beta-\epsilon) < 0$ and $g(\beta+\epsilon) > 0$. For all sufficiently large $x$ we therefore also have $\psi_x'(\beta-\epsilon) < 0$ and $\psi_x'(\beta+\epsilon) > 0$, which implies $\beta - \epsilon < \theta_x < \beta + \epsilon$ for all sufficiently large $x$. Since $\epsilon$ is arbitrary, we conclude that $\theta_x \to \beta$. □

This result indicates that we can expect the importance sampling procedure to be effective at large loss thresholds $x$ if $Q$ provides a good approximation to $L$ (more precisely, to $L - a_0$). It also indicates that we should have quite a bit of freedom in choosing the parameter $\theta$. In our numerical experiments, we choose $\theta = \theta_x$. In fact, the constant $d$ in the upper bound (5.10) on the second moment when using $\theta_x$ can be made as small as the best constant $c(\theta)$ in the upper bound (5.9) when using a fixed value of $\theta$. This follows since, in the notation of (5.5), $x^{\nu/2}\phi_x(\theta) \to a_1(\theta)/a_3(\theta)^{\nu/2}$ and $\theta_x \to \beta$; simple algebra shows that $\beta$ also minimizes the function $a_1(\theta)/a_3(\theta)^{\nu/2}$.

Since we may be interested in estimating multiple points on the loss distribution from a single set of simulations, it is worth considering whether importance sampling using $\theta_x$ remains effective in estimating $P(Q > y)$ with $y \neq x$. Let $m_2(\theta, x, y)$ be the second moment in (5.8) but with the indicator $I(Q > x)$ replaced by $I(Q > y)$. Arguing as in the proof of Theorem 5.1, we find that if $y > x$ and $y/x \to \gamma$, then

$$\limsup_{x\to\infty}\frac{y^{\nu/2}m_2(\theta_x, x, y)}{P(Q > y)} \leq \gamma^{\nu/2}d$$

with $d$ the same constant as in (5.10). In particular, if $\gamma = 1$ (i.e., $y = x + o(x)$) then we get the same upper bound using $\theta_x$ as we would using the "optimal" value $\theta_y$. This suggests that we can optimize the parameter for some loss level $x$ and still obtain good estimates for a moderate range of values $y$, $y > x$.

## 5.2. Estimating the Conditional Excess

A common criticism of VAR as a measure of risk is that it is insensitive to the magnitude of losses beyond a certain percentile. An alternative type of measure sometimes proposed (Artzner et al. 1999; Bassi, Embrechts, and Kafetzaki 1998; Uryasev and Rockafellar 2000) is the conditional excess

$$\eta = \eta(y) = E[L|L > y]. \tag{5.12}$$

Unlike VAR, the conditional excess weights large losses by their magnitudes. The threshold $y$ in the definition (5.12) may be a fixed loss level or else the VAR itself.

We examine the effectiveness of our importance sampling procedure in estimating $\eta$. Using ordinary Monte Carlo, one generates independent replications $L_1, \ldots, L_n$, all having the distribution of $L$, and estimates $\eta(y)$ using

$$\hat{\eta}_n = \frac{\sum_{k=1}^n L_k I(L_k > y)}{\sum_{k=1}^n I(L_k > y)}. \tag{5.13}$$

Applying the law of large numbers to both numerator and denominator shows that this estimator is consistent—though, being a ratio estimator, it is biased for finite $n$. Under importance sampling, the estimator is

$$\hat{\eta}_{\theta,n} = \frac{\sum_{k=1}^n \ell_k L_k I(L_k > y)}{\sum_{k=1}^n \ell_k (L_k > y)}, \tag{5.14}$$

where $\ell_k$ denotes the likelihood ratio on the $k$th replication.

The following result compares these estimators based on their asymptotic (as $n \to \infty$) variances. Let $p = P(L > y)$ and $R_k = L_k I(L_k > y) - \eta I(L_k > y)$. Define $\sigma^2 = E[R_k^2]$ and $\sigma_\theta^2 = E_\theta[(\ell_k R_k)^2)]$. In the following, $\Rightarrow$ denotes convergence in distribution.

PROPOSITION 5.1. *If $0 < \sigma^2 < \infty$, then as $n \to \infty$,*

$$\frac{\sqrt{n}(\hat{\eta}_n - \eta)}{\sigma/p} \Rightarrow N(0,1),$$

*and if $0 < \sigma^2(\theta) < \infty$*

$$\frac{\sqrt{n}(\hat{\eta}_{\theta,n} - \eta)}{\sigma_\theta/p} \Rightarrow N(0,1).$$

*If there is a constant $\epsilon$ such that $\ell_k \le \epsilon$ whenever $L_k > y$, then $\sigma_\theta^2 \le \epsilon\sigma^2$.*

The first two statements in this proposition are instances of the usual central limit theorem for ratio estimators (see, e.g., Serfling 1980) and the last statement follows from the definition of $\sigma_\theta^2$. An upper bound on the likelihood ratio, as required for this result, holds if $Q > x$ whenever $L > y$ (e.g., if the quadratic approximation (3.1) is exact and $x = y - a_0$). In this case, as in Lemma 5.1, the likelihood ratio is bounded by $\phi_x(\theta)$ and, also as in Lemma 5.1, this bound becomes smaller than a constant times $x^{-\nu/2}$ for sufficiently large $x$. Thus, in this case, the $\epsilon$ in Proposition 5.1 can be made quite small if $x$ is large. This suggests that our importance sampling method should be similarly effective for estimating the expected excess as for estimating a loss probability.

## 6. STRATIFYING THE LIKELIHOOD RATIO

In this section we further exploit the delta-gamma (or other quadratic) approximation and the structure of the multivariate $t$ distribution to further reduce variance in Monte Carlo estimates of portfolio loss probabilities. Inspection of the importance sampling estimator (4.6) suggests that to achieve greater precision we should reduce the sampling variability associated with the likelihood ratio $\exp(-\theta Q_x + \psi_x(\theta))$. This general approach to improving importance sampling estimators proved effective in the multivariate normal settings treated in Glasserman et al. (1999a, 1999b, 2000).

For the estimator in (4.6), reducing sampling variability in the likelihood ratio is equivalent to reducing it in $Q_x$ as defined in (3.3). We accomplish this through stratified

sampling of $Q_x$: we partition the real line into intervals (these are the strata) and generate samples of $Q_x$ so that the desired number of samples falls in each stratum. Two issues need to be addressed in developing this method. First, we need to have a way of defining strata with known probabilities, and this requires being able to compute the distribution of $Q_x$ under $P_\theta$. Second, we need a way of generating samples within strata which ensures that the $(Y, Z_1, \ldots, Z_m)$ generated have the correct conditional distribution given the stratum in which $Q_x$ falls.

To find the distribution of $Q_x$ under $P_\theta$ we extend the transform analysis of Section 3. In particular, we find the characteristic function of $Q_x$ under $P_\theta$ through the following simple observation.

LEMMA 6.1. *The characteristic function of $Q_x$ under $P_\theta$ is given by $\phi_{\theta,x}(\sqrt{-1}\omega)$, where $\phi_{\theta,x}(s) = \phi_x(\theta + s)/\phi_x(\theta)$ and $\phi_x$ is as in (3.4).*

*Proof.* The moment generating function of $Q_x$ under $P_\theta$ is

$$\begin{aligned}\phi_{x,\theta}(s) &= E_\theta\left[e^{sQ_x}\right] \\ &= E\left[e^{\theta Q_x - \psi_x(\theta)} e^{sQ_x}\right] \\ &= E\left[e^{(\theta+s)Q_x}\right]/e^{\psi_x(\theta)} = \phi_x(\theta + s)/\phi_x(\theta).\end{aligned}$$

As in the proof of Theorem 3.1, the characteristic function is the moment generating function evaluated at a purely imaginary argument. □

Using this result and the inversion integral (3.9) applied to $\phi_{\theta,x}$, we can compute $P_\theta(Q_x \le a)$ for arbitrary $a$. Given a set of probabilities $p_1, \ldots, p_N$ summing to 1, we can use the transform inversion iteratively to find points $-\infty = a_0 < a_1 < \cdots < a_N < a_{N+1} = \infty$ such that $P_\theta(Q_x \in (a_{i-1}, a_i)) = p_i$, $i = 1, \ldots, N$. The intervals $(a_{i-1}, a_i)$ form the strata for stratified sampling. We often use equiprobable strata ($p_i \equiv 1/N$) but this is by no means necessary. Alternatively, if the $a_i$'s are given, then the $p_i$'s can be found via transform inversion.

Given $N$ strata and a budget of $n$ samples, we allocate $n_i$ samples to stratum $i$, with $n_1 + \cdots + n_N = n$. For example, we may choose a proportional allocation with $n_i \approx np_i$; this choice guarantees a reduction in variance. Let $Q_x^{(ij)}$ denote the $j$th sample from stratum $i$, $j = 1, \ldots, n_i$ and let $L^{(ij)}$ denote the corresponding portfolio loss for that scenario. The combined importance sampling and stratified sampling estimator of the loss probability $P(L > y)$ is

$$\sum_{i=1}^{N} \frac{p_i}{n_i} \sum_{j=1}^{n_i} e^{-\theta Q_x^{(ij)} + \psi_x(\theta)} I(L^{(ij)} > y). \tag{6.1}$$

This estimator is unbiased for any set of positive stratum probabilities and positive allocations. This is true for any $\theta$ at which $\psi_x(\theta) < \infty$ (e.g., $\theta = \theta_x$). With the loss threshold $y$ specified we would typically use $x = y - a_0$ as suggested by (3.1).

It remains to specify the mechanism for sampling the $Q_x^{(ij)}$ so that $n_i$ samples fall in stratum $i$. Recall from Algorithm 4.1 that we do not sample $Q_x$ directly. Rather, we generate $Y$ from its exponentially twisted distribution and then generate $(Z_1, \ldots, Z_m)$ according to (4.4). Given $(Y, Z) \equiv (Y, Z_1, \ldots, Z_m)$, we can then calculate $X = Z/\sqrt{Y/\nu}$, $\Delta S$, $L$, and $Q_x$.

To implement stratified sampling, we apply a "bin-tossing" method developed in Glasserman et al. (2000). Keeping count of how many samples have produced values of $Q_x$ in each stratum, we repeatedly generate $(Y, Z)$ as in Algorithm 4.1. For each $(Y, Z)$ we compute $Q_x$ and check which stratum it falls in. If $Q_x$ falls in stratum $i$ and we have previously generated $j < n_i$ samples with $Q_x$ in stratum $i$, then the newly generated $(Y, Z)$ becomes the $(j+1)$th sample for the stratum. If we already have $n_i$ samples for stratum $i$, we discard $(Y, Z)$ and generate a new sample. We repeat this procedure until the number of samples for each stratum reaches the allocation for the stratum. This method is somewhat crude, but it is fast and easy to implement; see Glasserman et al. (2000) for an analysis of its computational requirements.

The combined simulation algorithm using both importance sampling and stratified sampling follows. We formulate the algorithm to estimate a specific loss probability $P(L > y)$, though multiple $y$'s could be considered simultaneously.

ALGORITHM 6.1. Importance Sampling and Stratified Sampling Estimate of Loss Probability

1. Set $x = y - a_0$ and find $\theta_x$ solving $\psi'_x(\theta_x) = 0$ as in (4.8). Set $\theta = \theta_x$.
2. Numerically invert the characteristic function of $Q_x$ under $P_\theta$ to find stratum boundaries $a_1, \ldots, a_N$ for which $P_\theta(a_{i-1} < Q_x < a_i) = p_i$, for given $p_1, \ldots, p_N$.
3. Use the bin-tossing method to generate $(Y^{(ij)}, Z^{(ij)})$, $j = 1, \ldots, n_i$, $i = 1, \ldots, N$, so that each $Q_x^{(ij)}$ calculated from $(Y^{(ij)}, Z^{(ij)})$ falls in stratum $i$.
4. Set $X^{(ij)} = Z^{(ij)}/\sqrt{Y^{(ij)}/\nu}$ and $\Delta S^{(ij)} = CX^{(ij)}$ with $C$ as in Algorithm 4.1. Compute the portfolio loss $L^{(ij)} = V(0, S) - V(\Delta t, S + \Delta S^{(ij)})$.
5. Return the estimator in (6.1).

This is also applicable with the copula specification in (2.5). As in Algorithm 4.1, only the sampling of the values of $\Delta S$ changes. The required modification of step 4 of Algorithm 6.1 is exactly as described immediately following Algorithm 4.1.

## 7. NUMERICAL EXAMPLES

We perform experiments with the transform inversion routine of Section 3, the importance sampling procedure of Section 4, and the combination with stratified sampling in Section 6. We use a subset of the portfolios that were considered in Glasserman et al. (2000), but with the light-tailed Gaussian assumptions of that paper replaced by the heavy-tailed assumptions of this paper. The portfolios in Glasserman et al. (2000) were chosen so as to incorporate a wide variety of characteristics, such as portfolios that have all eigenvalues $\lambda_i$ positive, portfolios that have some negative $\lambda_i$'s, portfolios that have all $\lambda_i$'s negative, portfolios with discontinuous payoffs (e.g., cash-or-nothing puts and barrier options), and portfolios with block diagonal correlation matrices. In the subset of those portfolios that we consider in this paper, we have tried to give sufficient representation to most of these characteristics. We have, in particular, included both the best and worst performing cases of Glasserman et al. (2000), where we experimented with diagonal and nondiagonal correlation matrices and found that this had little effect on performance. To limit the number of cases, here we mainly consider uncorrelated risk factors. Also, we focus on estimating loss probabilities and the conditional excess; issues specific to estimating a quantile rather than a loss probability were addressed in Glasserman et al. (2000).

In our numerical experiments we value the options in a portfolio using the Black–Scholes formula and its extensions. For the implied volatility of $S_i$ we use $\tilde{\sigma}_i/S_i\sqrt{\Delta t}$ with $\tilde{\sigma}_i$ as in (2.5); in other words, we make the implied volatility consistent with the standard deviation of $\Delta S_i$ over the VAR horizon $\Delta t$. There is still an evident inconsistency in applying Black–Scholes formulas when price changes follow a $t$ distribution, but option pricing formulas are commonly used this way in practice. Moreover, it seems reasonable to expect that this simple approach to portfolio revaluation gives a good indication of the variance reduction that would be obtained from our Monte Carlo method even if more complex revaluation procedures were used. The greatest computational gains from reducing the number of Monte Carlo samples required would in fact be obtained in cases where revaluation is most time consuming, such as when revaulation requires finite-difference methods, lattices and trees, and possibly even simulation.

As in Glasserman et al. (2000), we assume 250 trading days in a year and a continuously compounded risk-free rate of interest of 5%. For each case we investigate losses over 10 days ($\Delta t = 0.04$ years). Most of the test porfolios we consider are based on 10 uncorrelated underlying assets having an initial value of 100 and an annual volatility of 0.3 (i.e., $\tilde{\sigma}_i = 0.3S_i\sqrt{\Delta t}$). In three cases we also consider correlated assets and in one of these the portfolio involves 100 assets with different volatilities. Detailed descriptions are given in Table 7.1. For comparison purposes, in each case we adjust the loss threshold $x$ so that the probability to be estimated is close to 0.01.

In the first set of experiments, we assume all the marginals to be $t$ distributions with degree of freedom (d.o.f) 5. Results are given in Table 7.2, which lists the quadratic approximation and the estimated variance ratios using importance sampling (IS) and IS combined with stratified sampling (ISS-Q)—that is, the estimated variance using standard Monte Carlo divided by the estimated variance using IS (or ISS-Q). This variance ratio indicates how many times as many samples would be required using ordinary Monte Carlo to achieve the same precision achieved with the corresponding variance reduction technique; it is thus an estimate of the computational speedup that can be obtained using a method. In all experiments, unless otherwise mentioned, the variance ratios are estimated from a total of 40,000 replications; the stratified estimator uses 40 (approximately) equiprobable strata with 1,000 samples per stratum. In practice, fewer replications are usually used; the high number we use is to get accurate estimates of the variances and thus the computational speedups.

We achieve at least double-digit variance reduction in all cases. It is also encouraging that the variance ratios obtained for the 100 asset example (a.9) are comparable to the best variance ratios obtained for the other much smaller 10 asset examples. The effectiveness of the method is further illustrated in Figure 7.1, which compares standard simulation to importance sampling with stratification for the 0.1 yr ATM portfolio. The figure plots point estimates and 99% confidence for $P(L > x)$ over a range of $x$ values; a total of 4,000 replications were used for each method to simultaneously estimate $P(L > x)$ for the set of $x$ values indicated in the figure. The importance sampling uses a single value of the parameter $\theta$, chosen to be $\theta_x$ for an $x$ in the middle of the range. Notice how much narrower the confidence intervals are for the ISS-Q method over the entire range of $x$'s.

In the next set of experiments, for (a.1) to (a.11), we assume that the first five marginals have d.o.f 3 and the next five have d.o.f. 7. The "reference d.o.f." was taken to be 5 (i.e., we use the copula method described earlier with $\nu = 5$ in (2.5)).

TABLE 7.1
Test Portfolios for Numerical Results

| Portfolio | Description |
|---|---|
| (a.1) *0.5 yr ATM* | Short 10 at-the-money calls and 5 at-the-money puts on each asset, all options having a half-year maturity. All eigenvalues are positive. |
| (a.2) *0.5 yr ATM, $-\lambda$* | Long 10 at-the-money calls and 5 at-the-money puts on each asset, all options having half a year maturity. All eigenvalues are negative. |
| (a.3) *0.1 yr ATM* | Same as (a.1) but with a maturity of 0.10 years. |
| (a.4) *0.1 yr ATM, $-\lambda$* | Same as (a.2) but with a maturity of 0.10 years. |
| (a.5) *Delta hedged* | Same as (a.3) but with number of puts increased so that $\delta = 0$. |
| (a.6) *Delta hedged, $\pm\lambda$* | Short 10 at-the-money calls on first five assets. Long 5 at-the-money calls on the remaining assets. Long or short puts on each asset so that $\delta = 0$. This has both negative and positive eigenvalues. |
| (a.7) *DAO-C* | Short 10 down-and-out calls on each asset with barrier at 95. |
| (a.8) *DAO-C & CON-P* | Short 10 down-and-out calls with barrier at 95, and short 5 cash-or-nothing puts on each asset. The cash payoff is equal to the strike price. |
| (a.9) *DAO-C & CON-P, Hedged* | Same as (a.8) but the number of puts is adjusted so that $\delta = 0$. |
| (a.10) *Index* | Short 50 at-the-money calls and 50 at-the-money puts on 10 underlying assets, all options expiring in 0.5 years. The covariance matrix was downloaded from the RiskMetrics™ web site for international equity indices. The initial asset prices are (100, 50, 30, 100, 80, 20, 50, 200, 150, 10). |
| (a.11) *Index, $\lambda_m < -\lambda_1$* | Same as (a.10) but short 50 at-the-money calls and 50 at-the-money puts on the first three assets, long 50 at-the-money calls and 50 at-the-money puts on the next seven assets. This has both negative and positive eigenvalues with the absolute value of the minimum eigenvalue greater than that of the maximum. |
| (a.12) *100, Block-diagonal* | Short 10 at-the-money calls and 10 at-the-money puts on 100 underlying assets, all expiring in 0.10 years. Assets are divided into 10 groups of 10. The correlation is 0.2 between assets in the same group and 0 across groups. Assets in the first three groups have volatility 0.5, those in the next four have volatility 0.3, and those in the last three groups have volatility 0.1. |

TABLE 7.2
Comparison of Methods for Estimating $P(L > x)$ for Test Portfolios. All the Marginals are $t$ with 5 Degrees of Freedom

| | | | | Variance ratios | |
|---|---|---|---|---|---|
| Portfolio | $x$ | $P\{L > x\}$ | $P\{Q + c > x\}$ | IS | ISS-Q |
| (a.1) 0.5 yr ATM | 311 | 1.02% | 1.17% | 53 | 333 |
| (a.2) 0.5 yr ATM, $-\lambda$ | 145 | 1.02% | 1.33% | 35 | 209 |
| (a.3) 0.1 yr ATM | 469 | 0.97% | 1.56% | 46 | 134 |
| (a.4) 0.1 yr ATM, $-\lambda$ | 149 | 0.97% | 0.86% | 21 | 28 |
| (a.5) Delta hedged | 617 | 1.07% | 1.69% | 42 | 112 |
| (a.6) Delta hedged, $\pm\lambda$ | 262 | 1.02% | 1.70% | 27 | 60 |
| (a.7) DAO-C | 482 | 0.91% | 0.52% | 58 | 105 |
| (a.8) DAO-C & CON-P | 835 | 0.97% | 1.19% | 18 | 20 |
| (a.9) DAO-C & CON-P, Hedged | 345 | 1.09% | 0.36% | 17 | 25 |
| (a.10) Index | 2019 | 1.04% | 1.22% | 26 | 93 |
| (a.11) Index, $\lambda_m < -\lambda_1$ | 426 | 1.02% | 1.16% | 18 | 48 |
| (a.12) 100, Block-diagonal | 5287 | 0.95% | 1.58% | 61 | 287 |

For (a.12), we assume that the marginals in the first three groups have d.o.f. 3, the marginals in the second four groups have d.o.f. 5, and the marginals in the last three groups have d.o.f. 7; the reference d.o.f. was again taken to be 5. Results for all these cases are given in Table 7.3. Note that the performance of IS remains roughly the same (except for (a.9)), but the performance of ISS-Q decreases substantially. This is to be expected as the transformation from the $t$ distribution with the reference d.o.f. to the $t$ distribution of the marginal introduces further nonlinearity in the relation between the underlying variables and the portfolio value. Case (a.9) was also the worst performing case in Glasserman et al. (2000; case (b.6) in that paper); in this particular case, the delta-gamma approximation gives a poor approximation to the true loss.

Finally, we estimate the conditional excess for all the portfolios described above. Table 7.4 gives results using IS and ISS-Q. We again compare each case with standard simulation, where by standard simulation we mean the estimator given by (5.13). In particular, for IS we estimate the ratio $\sigma^2/\sigma_\theta^2$ where $\sigma^2$ and $\sigma_\theta^2$ have been defined in Proposition 5.1; expressions that may be used to estimate these quantities are given in Serfling (1980). One can similarly estimate the variance ratios for the ISS-Q.

## 8. CONCLUDING REMARKS

This paper develops efficient computational procedures for approximating or estimating portfolio loss probabilities in a model that captures heavy tails in the joint distribution of market risk factors. The first method is based on transform inversion of a quadratic approximation to portfolio value. The second method uses the first to develop Monte Carlo sampling procedures that can greatly reduce variance compared with ordinary Monte Carlo.

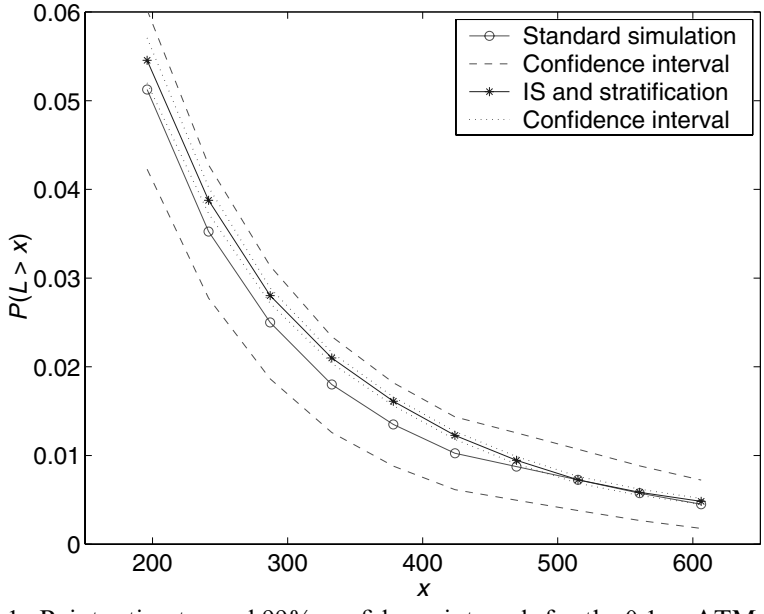


Figure 7.1. Point estimates and 99% confidence intervals for the 0.1 yr ATM portfolio using standard simulation and importance sampling with stratification. The estimates are from a total of 4,000 replications and 40 strata.

TABLE 7.3
Comparison of Methods for Estimating $P(L > x)$ for Different Portfolios Where All the Marginals Are $t$'s with Different Degrees of Freedom

| | | | | Variance ratios | |
|---|---|---|---|---|---|
| Portfolio | $x$ | $P\{L > x\}$ | $P\{Q + c > x\}$ | IS | ISS-Q |
| (a.1) 0.5 yr ATM | 322 | 1.05% | 0.82% | 37 | 48 |
| (a.2) 0.5 yr ATM, $-\lambda$ | 143 | 0.99% | 1.25% | 35 | 136 |
| (a.3) 0.1 yr ATM | 475 | 1.01% | 1.16% | 38 | 55 |
| (a.4) 0.1 yr ATM, $-\lambda$ | 149 | 1.08% | 0.91% | 17 | 20 |
| (a.5) Delta hedged | 671 | 0.98% | 1.11% | 39 | 57 |
| (a.6) Delta hedged, $\pm\lambda$ | 346 | 0.95% | 0.18% | 27 | 34 |
| (a.7) DAO-C | 447 | 1.16% | 0.50% | 28 | 32 |
| (a.8) DAO-C & CON-P | 777 | 1.23% | 1.15% | 16 | 18 |
| (a.9) DAO-C & CON-P, Hedged | 333 | 1.26% | 0.29% | **1.1 | **3.1 |
| (a.10) Index | 1979 | 1.12% | 1.02% | 23 | 39 |
| (a.11) Index, $\lambda_m < -\lambda_1$ | 442 | 0.99% | 0.27% | **3.7 | **4.0 |
| (a.12) 100, Block-diagonal | 5690 | 0.96% | 0.86% | 79 | 199 |

** indicates that we used 400,000 replications (instead of 40,000) for these cases but the variance estimates and thus the variance ratio estimates still did not stabilize.

TABLE 7.4
Comparison of Methods for Estimating $E[L|L > x]$ for Different Portfolios Where All the Marginals Are $t$'s with Varying Degrees of Freedom

| | | | Variance ratios | |
|---|---|---|---|---|
| Portfolio | $x$ | $E(L\|L > x)$ | IS | ISS-Q |
| (a.1) 0.5yr ATM | 322 | 564 | 33 | 35 |
| (a.2) 0.5yr ATM, $-\lambda$ | 143 | 160 | 39 | 59 |
| (a.3) 0.1yr ATM | 475 | 785 | 50 | 53 |
| (a.4) 0.1yr ATM, $-\lambda$ | 149 | 156 | 20 | 21 |
| (a.5) Delta hedged | 671 | 638 | 40 | 63 |
| (a.6) Delta hedged, $\pm\lambda$ | 346 | 1106 | 44 | 55 |
| (a.7) DAO-C | 447 | 680 | 46 | 50 |
| (a.8) DAO-C & CON-P | 777 | 969 | 17 | 20 |
| (a.9) DAO-C & CON-P, Hedged | 333 | 548 | **8.4 | **10 |
| (a.10) Index | 1979 | 3583 | 23 | 24 |
| (a.11) Index, $\lambda_m < -\lambda_1$ | 442 | 841 | **2.2 | **4.6 |
| (a.12) 100, Block-diagonal | 5690 | 9300 | 93 | 236 |

** indicates that we used 400,000 replications (instead of 40,000) for these cases but the variance estimates and thus the variance ratio estimates still did not stabilize.

Our results are based on modeling the joint distribution of risk factors using a multivariate $t$ and some extensions of it. This may be viewed as a reduced-form approach to modeling changes in risk factors, in the sense that we have not specified a continuous-time process for the evolution of the risk factors. Though it is possible to construct a process with a multivariate $t$ distribution at a fixed time $\Delta t$, we know of no process having this distribution at all times. So the model used here requires fixing a time interval $\Delta t$. (The same is true of most time-series models, including GARCH, for example.) This is in contrast to Lévy process models considered in Barndorff-Nielsen (1998), Eberlein et al. (1998), and Madan and Seneta (1990); but the distributions in those models have exponential tails and are thus not as heavy as the distributions considered here. Some of the distributions that arise in these models admit representations through the normal distribution similar to (2.2) so the methods developed here may be applicable to them as well.

## REFERENCES

ABATE, J., and W. WHITT (1992): The Fourier-series Method for Inverting Transforms of Probability Distributions, *Queueing Sys.* 10, 5–88.

ANDERSON, T. W. (1984): *An Introduction to Multivariate Statistical Analysis*, 2nd ed. New York: Wiley.

ARTZNER, P., F. DELBAEN, J.-M. EBER, and D. HEATH (1999): Coherent Measures of Risk, *Math. Finance* 9, 203–228.

ASMUSSEN, S., and K. BINSWANGER (1997): Simulation of Ruin Probabilities for Subexponential Claims, *ASTIN Bull.* 27, 297–318.

ASMUSSEN, S., K. BINSWANGER, and B. HØJGAARD (2000): Rare Events Simulation for Heavy Tailed Distributions, *Bernovlli* 6, 303–322.

BARNDORFF-NIELSEN, O. E. (1998): Processes of Normal Inverse Gaussian Type, *Finance Stoch.* 2, 41–68.

BASSI, F., P. EMBRECHTS, and M. KAFETZAKI (1998): Risk Management and Quantile Estimation; in *A Practical Guide to Heavy Tails*. eds. R. J. Adler, R. Feldman and M. Taqqu. Boston: Birkhäuser, 111–130.

BLATTBERG, R. C., and N. J. GONEDES (1974): A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices, *J. Business* 47, 244–280.

BOUCHAUD, J. P., D. SORNETTE, and M. POTTERS (1997): Option Pricing in the Presence of Extreme Fluctuations, in *Mathematics of Derivative Securities*, eds. M. A. H. DEMPSTER and S. PLISKA. New York: Cambridge University Press.

BRITTEN-JONES, M., and S. M. SCHAEFER (1999): Non-linear Value-at-Risk, *European Finance Rev.* 2, 161–187.

CARDENAS, J., E. FRUCHARD, J.-F., PICRON, C. REYES, K. WALTERS, and W. YANG (1999): Monte Carlo Within a Day, *Risk* 12(2), 55–59.

DANIELSSON, J., and C. G. de VRIES (1997): Tail Index and Quantile Estimation with Very High Frequency Data, Working paper, Tinbergen Insitute, Rotterdam, The Netherlands.

DUFFIE, D., and J. PAN (2001): Analytical Value-at-Risk with Jumps and Credit Risk, *Finance and Stochastics* 5(2), 155–180.

EBERLEIN, E., and U. KELLER (1995): Hyperbolic Distributions in Finance, *Bernoulli* 1, 281–299.

EBERLEIN, E., U. KELLER, and K. PRAUSE (1998): New Insights Into Smile, Mispricing, and Value-at-Risk: The Hyperbolic Model, *J. Business* 71, 371–406.

EMBRECHTS, P., A. MCNEIL, and D. STRAUMANN (2002): Correlation and Dependence Properties in Risk Management: Properties and Pitfalls, *Extremes and Integrated Risk Management* London: Risk Book.

FAMA, E. F. (1965): The Behavior of Stock Market Prices, *J. Business* 38, 34–105.

FANG, K.-T., S. KOTZ, and K. W. NG (1987): *Symmetric Multivariate and Related Distributions*. London: Chapman & Hall.

FELLER, W. (1971): *An Introduction to Probability Theory and its Applications*, Vol. II, 2nd ed. New York: Wiley.

GLASSERMAN, P., P. HEIDELBERGER, and P. SHAHABUDDIN (1999a): Asymptotically Optimal Importance Sampling and Stratification for Pricing Path-Dependent Options, *Math. Finance* 9, 117–152.

GLASSERMAN, P., P. HEIDELBERGER, and P. SHAHABUDDIN (1999b): Importance Sampling in the Heath-Jarrow-Morton Framework, *J. Derivatives* 7(1), 32–50.

GLASSERMAN, P., P. HEIDELBERGER, and P. SHAHABUDDIN (2000): Variance Reduction Techniques for Estimating Value-at-Risk, *Mgmt. Sci.* 46, 1349–1364.

HEYDE, C. C. (1999): A Risky Asset Model with Strong Dependence Through Fractal Activity Time, *J. Appl. Probab.* 36, 1234–1239.

HOSKING, J. R. M., G. BONTI, and D. SIEGEL (2000): Beyond the Lognormal, *Risk* 13(5), 59–62.

HUISMAN, R., K. KOEDIJK, C. KOOL, and F. PALM (1998): The Fat-Tailedness of FX Returns. Working paper, Limburg Institute of Financial Economics, Maastricht University, The Netherlands. Available at www.ssrn.com.

HURST, S. R., and E. PLATEN (1997): The Marginal Distributions of Returns and Volatility, in *$L_1$-Statistical Procedures and Related Topics*, IMS Lecture Notes, Monograph Series Vol. 31. Hayward, CA: IMS, 301–314.

JAMSHIDIAN, F., and Y. ZHU (1997): Scenario Simulation: Theory and Methodology, *Finance Stoch.* 1, 43–67.

JOHNSON, N., S. KOTZ, and N. BALAKRISHNAN (1994): *Continuous Univariate Distributions*, Vol. 1. New York: Wiley.

JOHNSON, N., S. KOTZ, and N. BALAKRISHNAN (1995): *Continuous Univariate Distributions*, Vol. 2. New York: Wiley.

JUNEJA, S., and P. SHAHABUDDIN (1999): Simulating Heavy Tailed Processes Using Delayed Hazard Rate Twisting. To appear in *ACM TOMACS*.

KOEDIJK, K., R. HUISMAN, and R. POWNALL (1998): VaR-x: Fat Tails in Financial Risk Management, *J. Risk* 1, 47–62.

LI, D. (2000): On Default Correlation: A Copula Function Approach, *J. Fixed Income* 9, 43–54.

LONGIN, F., and B. SOLNIK (2001): Correlation Structure of International Equity Markets During Extremely Volatile Periods, *J. Finance* 56, 649–676.

LUKACS, E. (1970): *Characteristic Functions*, 2nd ed. New York: Hafner Publishing.

MADAN, D., and E. SENETA (1990): The Variance Gamma Model for Share Market Returns, *J. Business* 63, 511–524.

MANDELBROT, B. (1963): The Variation of Certain Speculative Prices, *J. Business* 36, 394–419.

MATACZ, A. (2000): Financial Modeling and Option Theory with the Truncated Levy Processes, *Inter. J. Theo. Appl. Fin.* 3, 143–160.

MCNEIL, A. J., and R. FREY (1999): Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach. Working paper, RiskLab, ETH Zurich.

NELSEN, R. B. (1999): *An Introduction to Copulas*. New York: Springer.

OWEN, A., and D. TAVELLA (1999): Scrambled Nets for Value-at-Risk Calculations; in *Monte Carlo: Methodologies and Applications for Pricing and Risk Management*, ed. B. Dupire. London: Risk Publications, 255–263.

PICOULT, E. (1999): Calculating Value-at-Risk with Monte Carlo Simulation; in *Monte Carlo: Methodologies and Applications for Pricing and Risk Management*, ed., B. Dupire. London: Risk Publications, 209–229.

PRAETZ, P. D. (1972): The Distribution of Share Price Changes, *J. Business* 45, 49–55.

RiskMetrics Group (1996): *RiskMetrics Technical Document*, www.riskmetrics.com/research/techdocs.

ROUVINEZ, C. (1997): Going Greek with VAR, *Risk* 10(2), 57–65.

RYDBERG, T. H. (1999): Generalized Hyperbolic Diffusion Processes with Applications in Finance, *Math. Finance* 9, 183–201.

SERFLING, R. J. (1980): *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

SHAHABUDDIN, P. (1994): Importance Sampling for the Simulation of Highly Reliable Markovian Systems, *Mgmt. Sci.* 40, 333–352.

SHAW, J. (1999): Beyond VAR and Stress Testing; in *Monte Carlo: Methodologies and Applications for Pricing and Risk Management*, ed., B. Dupire. London: Risk Publications, 231–244.

TONG, Y. L. (1990): *The Multivariate Normal Distribution*. New York: Springer.

URYASE, S., and R. T. ROCKAFELLAR (2000): Optimization of Conditional Value-At-Risk, *J. Risk* 2, 21–41.

WILSON, T. (1999): Value at Risk; in *Risk Management and Analysis*, Vol 1, ed., C. Alexander. Chichester, England: Wiley, 61–124.