Variance Reduction Techniques for Estimating Value-at-Risk

Paul Glasserman • Philip Heidelberger • Perwez Shahabuddin
Columbia Business School, Columbia University, New York, New York 10027
IBM Research Division, T.J. Watson Research Center, Yorktown Heights, New York, 10598
IEOR Department, Columbia University, New York, New York 10027

This paper describes, analyzes and evaluates an algorithm for estimating portfolio loss probabilities using Monte Carlo simulation. Obtaining accurate estimates of such loss probabilities is essential to calculating value-at-risk, which is a quantile of the loss distribution. The method employs a quadratic (''delta-gamma'') approximation to the change in portfolio value to guide the selection of effective variance reduction techniques; specifically importance sampling and stratified sampling. If the approximation is exact, then the importance sampling is shown to be asymptotically optimal. Numerical results indicate that an appropriate combination of importance sampling and stratified sampling can result in large variance reductions when estimating the probability of large portfolio losses.

(*Value-At-Risk*; *Monte Carlo*; *Simulation*; *Variance Reduction Technique*; *Importance Sampling*; *Stratified Sampling*; *Rare event*)

1. Introduction

An important concept for quantifying and managing portfolio risk is value-at-risk (VAR) (Jorion 1997, Wilson 1999). VAR is defined as a quantile of the loss in portfolio value during a holding period of specified duration. If the value of the portfolio at time t is $V(t)$, the holding period is Δt, and the value of the portfolio at time $t+\Delta t$ is $V(t+\Delta t)$, then the loss in portfolio value during the holding period is $L = V(t) - V(t+\Delta t)$. For a given probability p, the VAR, x_p, is defined to be the $(1-p)$th quantile of the loss distribution:

$$P\{L > x_p\} = p. \tag{1}$$

Typically, the interval Δt is one day or two weeks and p is close to zero, often $p \approx 0.01$. Monte Carlo simulation is frequently used to estimate the VAR. In such a simulation, changes in the portfolio's ''risk factors'' (e.g., interest rates, currency exchange rates, stock prices, etc.) during the holding period are generated and the portfolio is reevaluated using these new values for the risk factors. This is repeated many times so that the loss distribution may be estimated. We should note that this type of Monte Carlo VAR analysis is often augmented with ''stress'' tests using predetermined stress scenarios. This is done primarily because the underlying distributional assumptions (especially correlations) implicit in the analysis may tend to break down in extreme circumstances.

The computational cost required to obtain accurate Monte Carlo VAR estimates is often enormous. This is because of two factors. First, the portfolio may consist of a very large number of financial instruments. Furthermore, computing the value of an individual instrument may itself require substantial computational effort. Thus each portfolio evaluation may be costly. Second, a large number of runs (portfolio evaluations) are required to obtain accurate estimates of the loss distribution in the region of interest. We focus on this second issue: the development of variance reduction techniques designed to dramatically reduce the number of runs required to achieve accurate estimates of low probabilities. A general discussion

0025-1909/00/4610/1349$05.00
1526-5501 electronic ISSN

on variance reduction techniques may be found in Hammersley and Handscomb (1964). The technique described in this paper builds on the methods of Glasserman et al. (1999a, c), which were developed to reduce the variance when pricing a single instrument. Those methods combine specific implementations of two general purpose variance reduction techniques: importance sampling and stratified sampling. In this paper we also combine these two techniques, but in a way that is tailored to the VAR setting so the method is quite different from that of Glasserman et al. (1999a, c). Preliminary numerical results for the technique described in this paper, and for related techniques, were reported in Glasserman et al. (1999b). We now focus on the most promising approach tried in Glasserman et al. (1999b), provide a rigorous analysis of this approach, and perform more extensive experiments on it.

Our approach is to approximate the portfolio loss by a quadratic function of the underlying risk factors and to use this approximation to design variance reduction techniques. Quadratic approximations are widely used without simulation; indeed the second-order Taylor series approximation is commonly called the ''delta-gamma approximation'' (Britten-Jones and Schaefer 1999, Jorion 1997, Rouvinez 1997, Wilson 1999). While our approach could be combined with other quadratic approximations, many of the first and second derivatives needed for the delta-gamma approximation are routinely computed for other purposes quite apart from the calculation of VAR. One premise of this paper is that these derivatives are thus readily available as inputs to be used in a VAR simulation and do not represent an additional computational burden.

When the change in risk factors has a multivariate normal distribution, as is commonly assumed (and as we will assume), then the distribution of the delta-gamma approximation can be computed numerically (Imhof 1961, Rouvinez 1997). While this approximation is not always accurate enough to provide precise VAR estimates, we describe how it may be used to guide selection of an importance sampling (IS) change of measure for sampling the changes in risk factors. IS is a particularly appropriate technique for ''rare event'' simulations, which corresponds to the VAR problem with a small value of p. See Bucklew (1990), Chen et al. (1993), Glasserman et al. (1999a, c), and Heidelberger (1995) and the references therein for detailed discussions of IS. As the distribution of the quadratic approximation can be computed numerically, it can also be used as either a control variable or for stratified sampling. Numerical results in Glasserman et al. (1999b) showed that while the effectiveness of the control variable decreases as p decreases, the effectiveness of a combination of IS and stratified sampling increases as p decreases. This is the method we focus on in this paper. Independent of our work, Cárdenas et al. (1999) have studied using the delta-gamma approximation as a control variable and in a simple form of stratified sampling (with two strata) without IS.

The rest of the paper is organized as follows. In §2 we develop the quadratic approximation and describe the proposed IS change of measure based on this approximation. When this approximation is exact, we show that the IS is ''asymptotically optimal'' for estimating $P\{L>x\}$ for large x, meaning that the second moment of the estimator decreases at the fastest possible exponential rate as x increases. We also consider asymptotics as the number of risk factors becomes large and establish effectiveness of the method in this limit as well. Stratified sampling and its combination with IS are described in §3. In §4, we show how accurate estimates for $P\{L>x\}$ can translate to accurate estimates of the VAR x_p. First, we show that when the IS is selected so as to optimize estimation of $P\{L>x\}$, it is simultaneously asymptotically optimal for estimating $P\{L>y\}$ for a wide range of ys about x. In addition, we establish a central limit theorem (CLT) for the quantile estimate of the VAR x_p under IS and stratification. The form of the asymptotic variance in this CLT is typical of quantile estimates in other settings and implies that any variance reduction obtained for estimating $P\{L>x\}$ in a neighborhood of x_p carries over to a VAR estimate. The complete algorithm is stated in §5 and numerical results are given for a variety of sample portfolios. In most cases, the variance is reduced by at least one order of magnitude and often more than two orders of magnitude improvement are obtained. For one rather extreme case there is no improvement and further investigation reveals that in this case the delta-gamma approximation is not at all representative of the actual loss in the region of interest. The reader interested only in the practical aspects of the method may skip directly to §5. Results are summarized and some future topics for research are discussed in §6. Proofs are given in the Appendix.

2. Importance Sampling for Value-at-Risk

2.1. Delta-Gamma Approximation and IS

We begin this section with a description of the basic model and the quadratic approximation to the loss. This discussion follows that in Glasserman et al. (1999b). There are m risk factors and $S(t)=(S_1(t), \ldots, S_m(t))$ denotes their values at time t. Let $\Delta S=[S(t+\Delta t)-S(t)]'$ be the change in risk factors during the interval $[t, t+\Delta t]$. The portfolio value at time t is $V(t)\equiv V(S(t), t)$ and the loss over the interval Δt is $L=V(S(t), t)-V(S(t+\Delta t), t+\Delta t)$. We assume that there exists a constant a_0, a vector a, and a symmetric matrix A such that

$$L \approx a_0 + a'\Delta S + \Delta S' A \Delta S \equiv a_0 + Q, \tag{2}$$

and that a_0, a, and A are known prior to running the simulation. As described in Glasserman et al. (1999b), for the delta-gamma approximation $a_0=-\Theta\Delta t$, $a=-\delta$, and $A=-\frac{1}{2}\Gamma$, where $\Theta=\partial V/\partial t$, $\delta_i=\partial V/\partial S_i$, and $\Gamma_{ij}=\partial^2 V/\partial S_i \partial S_j$ (all partial derivatives are evaluated at $(S(t), t)$). Many of the δ_is and Γ_{ij}s (especially the Γ_{ii}s) are routinely computed for other purposes; we thus assume that a quadratic approximation (specifically, the delta-gamma approximation) is available ''for free.'' Some portfolios are both delta and gamma hedged ($\delta=\Gamma=0$); our methods do not apply to such portfolios unless a different quadratic approximation is available.

We assume that ΔS has a multivariate normal distribution with means 0 and covariance matrix Σ. While this might permit negative asset prices, Σ is typically proportional to Δt, and so the chance of this is quite small for the short holding periods of interest. Furthermore, the approach we propose also applies when risk factors follow a lognormal distribution since then $S_i(t+\Delta t)=S_i(t)\exp\{c_i+d_i Y_i\}$ for some constants c_i, d_i, and where the vector Y is multivariate normal. Thus, through the use of the chain rule, (2) can be transformed into a quadratic approximation in Y. Therefore, we may think of ΔS either as the change in risk factors or as the normals that drive the change in risk factors. In addition, while (2) is only an approximation, it typically becomes more accurate as $\Delta t \to 0$.

Samples of ΔS can be generated by setting $\Delta S=\tilde{C}X$ where X is a vector of independent standard normals (mean 0, variance 1) and $\tilde{C}$ is any matrix satisfying $\tilde{C}\tilde{C}'=\Sigma$ (e.g., the one obtained through Cholesky decomposition). For importance sampling, we will find it convenient to express Q as a diagonalized quadratic form. To this end let Λ denote the diagonal matrix of eigenvalues of $\tilde{C}'A\tilde{C}$ and let U be an orthogonal matrix whose columns are the corresponding unit length eigenvectors of $\tilde{C}'A\tilde{C}$, i.e., $\tilde{C}'A\tilde{C}=U\Lambda U'$ and $U'U=I(=UU')$. (Most numerical packages include routines to compute Λ and U.) Let $C=\tilde{C}U$. If Z is a column vector of independent standard normals, then $\Delta S=CZ$ has the $N(0, \Sigma)$ distribution since $CC'=\tilde{C}UU'\tilde{C}'=\tilde{C}\tilde{C}'=\Sigma$. Thus,

$$Q=a'\Delta S+\Delta S'A\Delta S=a'CZ+Z'C'ACZ=a'CZ+Z'\Lambda Z$$
$$=b'Z+Z'\Lambda Z\equiv\sum_{i=1}^{m}(b_i Z_i+\lambda_i Z_i^2) \text{ where } b'=a'C \tag{3}$$

and the Z_is are independent standard normals. In the following discussion, we will assume that the indices are ordered so that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$.

The reason standard simulation is inaccurate for estimating $P\{L>x\}$ for large x is that few samples are obtained in the important region where $L\approx x$. Effective IS should generate a disproportionately large number of samples in this region. To achieve this goal, we use the approximation $L\approx a_0+Q$ and select an IS technique that generates large values of Q with high probability; in fact under this IS, the mean of a_0+Q equals x. For IS, we consider a change of measure in which the mean of Z is changed from 0 to μ, and the covariance matrix is changed from I (the identity matrix) to B (assumed nonsingular). Then $P\{L>x\}=\tilde{E}[I(L>x)\ell(Z)]$, where $\tilde{E}$ is the expectation under the IS distribution and $\ell(Z)$ is the likelihood ratio (LR), which is given explicitly by

$$\ell(Z)=\frac{\exp(-\frac{1}{2}Z'Z)}{|B|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(Z-\mu)'B^{-1}(Z-\mu)\}}. \tag{4}$$

Here $|B|$ is the determinant of B. With Z sampled from the IS distribution, $I(L>x)\ell(Z)$ is an unbiased estimator of $P\{L>x\}$.

Rather than consider the general case, we restrict the IS so that it corresponds to "exponential twisting" of the quadratic form Q (expressed in the diagonal form of (3)). Exponential twisting frequently arises in the analysis of rare events and corresponding IS procedures (see, e.g., Bucklew 1990 and Heidelberger 1995). Let θ be the twisting parameter (requiring $(1-2\theta\lambda_i)>0$ for all i) and define

$$B(\theta)=(I-2\theta\Lambda)^{-1} \quad \text{and} \quad \mu(\theta)=\theta B(\theta)b. \tag{5}$$

When IS is done by changing the mean vector to $\mu(\theta)$ and the covariance matrix to $B(\theta)$, then the LR of (4) simplifies to

$$\begin{aligned}\ell(Z) &= \exp\{-\theta(b'Z+Z'\Lambda Z')+\psi(\theta)\}\\ &= \exp\{-\theta Q+\psi(\theta)\},\end{aligned} \tag{6}$$

where $\psi(\theta)$ is the logarithm of the moment generating function of Q and is given by

$$\psi(\theta)=\sum_{i=1}^{m}\frac{1}{2}\left(\frac{(\theta b_i)^2}{1-2\theta\lambda_i}-\log(1-2\theta\lambda_i)\right)\equiv\sum_{i=1}^{m}\psi^{(i)}(\theta). \tag{7}$$

With this change of measure, the only random term in the LR is $\exp(-\theta Q)$, which will be used extensively in the analysis of the IS procedure. Note that with this form of IS, the mean and variance of Z_i are changed to $\mu_i(\theta)$ and $\sigma_i^2(\theta)$, where

$$\sigma_i^2(\theta)=1/(1-2\theta\lambda_i), \qquad \mu_i(\theta)=\theta b_i\sigma_i^2(\theta), \tag{8}$$

and the Z_is remain independent. Verification that the LR is given by (6) is a simple exercise in algebra. This is a valid change of measure provided $\sigma_i^2(\theta)>0$ for all i, i.e., we require that $0\le\theta<1/(2\lambda_1)$ if $\lambda_1>0$. (Negative values of θ provide no benefit and need not be considered.)

To motivate the specific θ we choose, suppose the quadratic approximation is exact, i.e., $L=a_0+Q$. Then, because we have $Q>x-a_0$ on $\{L>x\}$, for any $\theta>0$ we obtain the upper bound

$$\begin{aligned}P\{L>x\} &= E_\theta[I(L>x)\ell]=E_\theta[e^{\psi(\theta)-\theta Q}I(a_0+Q>x)]\\ &\le \exp\{\psi(\theta)-\theta(x-a_0)\},\end{aligned} \tag{9}$$

where E_θ denotes expectation under IS with twisting parameter θ. Similarly, the second moment of a sample taken under IS is

$$m_2(x,\theta)=E_\theta[I(L>x)\ell^2]\le\exp\{2\psi(\theta)-2\theta(x-a_0)\}. \tag{10}$$

While finding the value of θ to minimize $m_2(x,\theta)$ is difficult, it is a simple matter to minimize the upper bound in (10). If θ_x is the minimizer, then θ_x satisfies the nonlinear equation

$$\psi'(\theta_x)=x-a_0. \tag{11}$$

The expression for $\psi'(\theta)$ is given in (A5) of the Appendix. As explained in Glasserman et al. (1999b), this equation is simple to solve numerically. Furthermore, since $E_\theta[Q]=\psi'(\theta)$, at θ_x, the mean of Q is $x-a_0$ and the mean of L is x. Thus under IS, $\{L>x\}$ is no longer a rare event. Notice that θ_x also minimizes the upper bound (9) for $P\{L>x\}$.

While the bulk of our analysis concerns the case in which the portfolio is quadratic, consider briefly the situation in which this is not true. Proving efficiency results requires analysis of $E_0[\ell I(L>x)]$, which is difficult if there is no simple relationship between ℓ and L. However, suppose that on $\{L>x\}$, $Q\ge q(x)$ for some deterministic $q(x)$. For example, if $Q\ge L+k$, with k constant, then $q(x)=x+k$. (This is perhaps not such an unreasonable assumption, since European and American put-and-call options are bounded by linear functions of their underlyings; see Hull 1997, pp. 159–165). Then

$$\begin{aligned}m_2(x,\theta) &\le E_0[\ell I(L>x)]\\ &\le \exp\{\psi(\theta)-\theta q(x)\}P\{L>x\}.\end{aligned} \tag{12}$$

The upper bound is minimized by setting θ equal to $\theta_{q(x)}$, i.e., setting the mean of Q equal to $q(x)$. A variance reduction is guaranteed provided $\psi(\theta_{q(x)})-\theta_{q(x)}q(x)<0$ and this variance reduction is exponential if $\psi(\theta_{q(x)})-\theta_{q(x)}q(x)\le -cx+o(x)$ for some constant $c>0$. The bound in (12) indicates that the method will result in good variance reduction provided that Q is large ($q(x)$ is large) whenever $L>x$.

We now return to the assumption that the loss is quadratic and describe two limiting cases in which the IS method is provably effective. The first considers the limit as the loss threshold increases and the second considers the limit as the number of risk factors increases.

2.2. Large Loss Threshold

We assume that the number of risk factors m is fixed. There are two cases to consider: $\lambda_1 > 0$ and $\lambda_1 < 0$. (We don't consider the case $\lambda_1 = 0$, because then the leading term in the quadratic approximation is 0.) Most of our analysis concerns the case when $\lambda_1 > 0$; then Q can grow without bound so in this case we let $x \to \infty$. In practice, of course, one would not literally consider infinitely large x. Examining limits as $x \to \infty$ can nevertheless provide useful insight if x is large but finite. It should also be noted that estimating a quantile x_p typically involves estimating loss probabilities at values of x somewhat larger than x_p, making loss probabilities smaller than 1% relevant to estimation of VAR.

We establish *asymptotic optimality* of our IS estimator, meaning that there is a constant $c > 0$ for which $P\{L > x\} = \exp(-cx + o(x))$ and $m_2(x, \theta_x) = \exp(-2cx + o(x))$. This means that the second moment decreases at twice the exponential rate of the loss probability itself. Nonnegativity of variance implies that this is the fastest possible rate for any unbiased estimator, justifying this notion of asymptotic optimality. Notice also that these rates of decrease imply $m_2(x, \theta_x) \approx P\{L > x\}^2$, whereas the second moment of the standard estimator $I(L > x)$ (without IS) is $P\{L > x\}$.

THEOREM 1. *If $L = a_0 + Q$ and $\lambda_1 > 0$, then*

$$\lim_{x\to\infty} \frac{\log(P\{L > x\})}{x} = -\frac{1}{2\lambda_1}. \tag{13}$$

THEOREM 2. *If $L = a_0 + Q$ and $\lambda_1 > 0$, then importance sampling using exponential twisting with twisting parameter θ_x defined by* (11) *is asymptotically optimal, i.e.,*

$$\lim_{x\to\infty} \frac{\log(m_2(x, \theta_x))}{x} = -\frac{1}{\lambda_1}. \tag{14}$$

Thus $P\{L > x\} = \exp\{-x/(2\lambda_1) + o(x)\}$ and $m_2(x, \theta_x) = \exp\{-x/\lambda_1 + o(x)\}$.

Next, consider the case in which $\lambda_1 < 0$. Now Q is bounded by a constant: $Q \le d \equiv \sum_{i=1}^m -b_i^2/(4\lambda_i)$, which is obtained by completing the square in (3). Letting $x_\varepsilon = a_0 + d - \varepsilon$, we have $P\{L > x_\varepsilon\} \approx P\{Q > d - \varepsilon\} \to 0$ as $\varepsilon \to 0$ and asymptotic results are thus stated as $\varepsilon \to 0$ in which case $x_\varepsilon \to a_0 + d$.

THEOREM 3. *If $L = a_0 + Q$ and $\lambda_1 > 0$, then there exist constants $k_1 > 0$ and $k_2 > 0$ such that*

$$k_1 \le \liminf_{\varepsilon\to 0} \frac{P\{L > x_\varepsilon\}}{\varepsilon^{m/2}} \le \limsup_{\varepsilon\to 0} \frac{P\{L > x_\varepsilon\}}{\varepsilon^{m/2}} \le k_2. \tag{15}$$

Let $\theta_\varepsilon > 0$ solve $\psi'(\theta_\varepsilon) = d - \varepsilon$. Then there exists a positive constant k_3 such that

$$\limsup_{\varepsilon\to 0} \frac{m_2(x_\varepsilon, \theta_\varepsilon)}{\varepsilon^m} \le k_3. \tag{16}$$

Thus the probability being estimated is of order $\varepsilon^{m/2}$ while the second moment of the IS estimate is of order ε^m, the best possible exponent. The bounds in (15) and (16) imply the ''bounded relative error'' property (see Shahabuddin 1994), an even stronger form of asymptotic optimality in which the standard deviation of an estimate divided by its mean remains bounded. Note that as $\varepsilon \to 0$, $\theta_\varepsilon \to \infty$ so that the variance of Z_i under IS approaches 0.

We now return to the situation in which $\lambda_1 > 0$. Our next result shows that asymptotic optimality is preserved if IS is implemented without twisting the Z_is associated with large, negative λ_is. To motivate this case, observe that the contribution to the LR corresponding to Z_i is $\ell_i = \exp\{\psi^{(i)}(\theta) - \theta Q^{(i)}\}$ where $Q^{(i)} \equiv [b_i Z_i + \lambda_i Z_i^2]$. Because $\theta > 0$, this term grows without bound as $|Z_i| \to \infty$ if $\lambda_i < 0$. In fact, $E_\theta[\ell_i^2] = E_0[\ell_i] = k_i E_0[\exp(-\theta Q^{(i)})] = \infty$ with if $\theta|\lambda_i| \ge \frac{1}{2}$, which is possible if $|\lambda_i| > \lambda_1$. In such a situation, twisting Z_i produces an LR with infinite variance. Theorem 2 ensures that the estimator is still asymptotically optimal if the quadratic approximation holds exactly; but it is conceivable that ℓ_i could contribute significantly (if not infinitely) to the variance of the estimator if the quadratic approximation is not exact.

To avoid such situations, we consider an IS estimator in which exponential twisting is applied only to those Z_i with i in a *twisting set* T. The set T is defined by a $\bar{\lambda} \le 0$ such that $\lambda_i \ge \bar{\lambda}$ for all $i \in T$ and $\lambda_i < \bar{\lambda}$ for all $i \notin T$. Using the bound $Q^{(i)} \le -b_i^2/(4\lambda_i)$, for $i \notin T$, and proceeding along lines similar to (10) and (11), the value of θ that minimizes an upper bound on the second moment of the estimator with twisting set T is θ_x^T, the solution to

$$\frac{d}{d\theta}\sum_{i\in T}\psi^{(i)}(\theta_x^T) = x - a_0 + \sum_{i\notin T}\frac{b_i^2}{4\lambda_i}. \tag{17}$$

If $\lambda_i > 0$ for all $i \in T$, then the LR for this twisting procedure is uniformly bounded by a constant; more generally, we have the following result:

COROLLARY 1. *Let $T = \{i\colon\ \lambda_i \geq \bar{\lambda}\}$ be the twisting set. If $\bar{\lambda} \leq 0$, $L = a_0 + Q$, and $\lambda_1 > 0$, then importance sampling using exponential twisting for $i \in T$ with twisting parameter θ_x^T defined by* (17), *is asymptotically optimal, i.e., with $m_2(x, \theta_x^T, T)$ denoting the second moment,*

$$\lim_{x\to\infty}\frac{\log(m_2(x, \theta_x^T, T))}{x} = -\frac{1}{\lambda_1}. \tag{18}$$

The results above on asymptotic optimality assume that the twisting parameter θ satisfies a nonlinear equation, either (11) or (17). Both of these equations involve constants (a_0 or b_i^2/λ_i), however the proof of asymptotic optimality goes through if these constants are replaced by any other constants. Furthermore, Theorem 5 in §4 shows that there is even more latitude in selecting the twisting parameter. While in our experiments we always select the twisting parameter according to (11) or (17), one might obtain better results by solving, for example, $\psi'(\theta) = x - c$ for some constant $c \neq a_0$, i.e., the value of c could be thought of as a tuning parameter of the IS.

2.3. Large Number of Risk Factors

The previous section analyzed the effectiveness of our IS estimator as the loss threshold increases; this section examines its effectiveness as the number of underlying risk factors increases, still under the assumption of a quadratic loss function. Trading portfolios of large financial institutions can easily be exposed to thousands of risk factors, so it is important to ensure that our IS estimator does not deteriorate in this setting. It turns out that under some reasonable assumptions on how the problem changes with size, increasing the number of risk factors can actually improve the effectiveness of the method. Because of the central limit theorem, increasing the number of risk factors tends to lighten the tail of the loss distribution (normalized by the standard deviation), and this can be beneficial for the IS method.

Throughout this section, we use a subscript m to emphasize the number of risk factors. In particular, we have $Q_m = \sum_{i=1}^m (b_i Z_i + \lambda_i Z_i^2)$, with $Z_1, Z_2, \ldots$, independent standard normals, and $\psi_m(\theta) = E[\exp(\theta Q_m)]$. We consider estimation of $P\{Q_m > x_m\}$, allowing the loss threshold x_m to vary with m to keep the problem nondegenerate. Specifically, we fix a $y > 0$ and set

$$x_m = \psi_m'(0) + y\sqrt{\psi_m''(0)} \equiv \sum_{i=1}^m \lambda_i + y\sqrt{\sum_{i=1}^m (b_i^2 + 2\lambda_i^2)},$$

so that for each m the loss threshold is y standard deviations above the expected loss. The condition $\psi'(\theta_m) = x_m$ defines θ_m. How the loss distribution changes with m is determined by the sequence (b_i, λ_i), $i = 1, 2, \ldots$. The interesting case corresponds to bounded (b_i, λ_i) (otherwise, increasingly important risk factors arise as the problem size increases), possibly approaching zero, but not so fast that $\sum_{i=1}^\infty (b_i^2 + \lambda_i^2) < \infty$ (in which case the tail risk factors would be entirely unimportant). These conditions are implied by the hypothesis of the following result.

THEOREM 4. *If*

$$\frac{[\sum_{i=1}^m (|b_i|^3 + |\lambda_i|^3)]^2}{[\sum_{i=1}^m (b_i^2 + \lambda_i^2)]^3} \to 0 \quad as \quad m \to \infty, \tag{19}$$

then $P\{Q_m > x_m\} \to 1 - \Phi(y)$ and the second moment of the IS estimator satisfies $m_2(x_m, \theta_m) \to e^{y^2}(1 - \Phi(2y))$.

The content of this result becomes more transparent if we apply the approximation $1 - \Phi(x) \sim \phi(x)/x$, valid for large x, with ϕ the standard normal density. The comparison between $P\{Q_m > x_m\}$ (the second moment of the standard estimator) and the second moment of the IS estimator then becomes

$$e^{-\frac{1}{2}y^2}/(y\sqrt{2\pi}) \ \ vs.\ \ e^{-y^2}(2y\sqrt{2\pi}).$$

In particular, the exponent on the right is twice as large as the one on the left, which is consistent with asymptotic optimality.

Taylor expansion of $\psi_m'(\theta_m)$ $(= x_m)$ suggests $x_m \approx \psi_m'(0) + \theta_m \psi_m''(0)$, which, through the definition of x_m,

Table 1 Variance Ratios as a Function of Number of Risk Factors m and Loss Threshold y Standard Deviations Above the Mean Loss ($x_m = m + y\sqrt{2m}$ and $\lambda = 1$)

m	$y=1$	$y=1.5$	$y=2$	$y=2.5$	$y=3$
10	2.9	4.7	7.9	14.0	25.9
50	3.2	5.8	11.5	24.0	60.1
5000	3.6	7.3	17.3	50.5	183.1
∞	3.6	7.5	18.3	56.0	217.4

Note. Each entry shows the ratio of the variance of the standard Monte Carlo estimate and the variance of the IS estimator. The case $m = \infty$ is based on the limits in Theorem 4. For example, with $y = 2$, the loss probability p_m decreases from 0.041 to 0.023 as m increases from 10 to ∞.

suggests $\theta_m \approx y/\sqrt{\psi_m''(0)}$. We will show that

$$\theta_m \sqrt{\psi_m''(0)} \to y, \tag{20}$$

and in fact Theorem 4 holds for any sequence θ_m satisfying this property.

As an illustration of this result, consider the case $\lambda_i \equiv \lambda$ and $b_i \equiv 0$ for all $i = 1, 2, \ldots$. This gives Q_m the distribution of λ times a chi-square random variable with m degrees of freedom and thus simplifies calculation of the exact probability $p_m = P\{Q_m > x_m\}$ and the exact second moment $m_2(x_m, \theta_m)$. From these we can calculate the exact variance reduction ratio $(p_m - p_m^2)/(m_2(x_m, \theta_m) - p_m^2)$. This is the ratio of the standard Monte Carlo variance to the IS variance; the larger the ratio the more effective the IS method. Table 1 shows the variance reduction ratio for various values of m and y.

3. Stratified Sampling for Value-at-Risk

In this section we describe how the IS can be combined with stratified sampling to obtain further variance reduction. A similar approach for option pricing was described in Glasserman et al. (1999a, c); in that setting only the means of the underlying normal random variables are changed by the IS and stratification is done on a linear combination of normals. In the current VAR setting, the IS changes both the means and variances of the underlying normals and stratification is done on a quadratic form. The idea in stratified sampling is to identify a stratification variable Y such that the distribution of Y is known and for which Y explains most of the variability of the output. Thus if Y is constrained to a small interval, the simulation's output is also highly constrained thereby resulting in a large variance reduction. In the general setting, suppose we are interested in estimating $E[X]$, then $E[X] = \sum_{j=1}^{k} E[X|Y \in \mathscr{S}_j]p_j$ where there are k strata $\mathscr{S}_1, \ldots, \mathscr{S}_k$ (typically intervals) and $p_j = P\{Y \in \mathscr{S}_j\}$. If one takes n_j samples from stratum j, then the stratified estimate of $E[X]$ is $\hat{X} = \sum_{j=1}^{k} \bar{X}_j p_j$ where $\bar{X}_j$ is the sample average of the Xs from stratum j. The variance of $\hat{X}$ is $\hat{\sigma}^2 = \sum p_j^2 \sigma_j^2/n_j$ where $\sigma_j^2 = \mathrm{Var}[X|Y \in \mathscr{S}_j]$. It is known that if $n_j = p_j n$, then this always produces a variance reduction compared to a standard simulation with the same total sample size n and that the optimal value of n_j is proportional to $p_j\sigma_j$ (Fishman 1996, pp. 297–300).

In this paper, we concentrate on stratifying on Q since experimental results in Glasserman et al. (1999b) indicate that this is typically more effective for VAR estimation than other combinations of IS and stratification, e.g., IS with stratification on Z_1 (the ''ISS-Lin'' method of Glasserman et al. 1999b). Stratification on Q without IS and with just two strata was considered in Cárdenas et al. (1999). Observe that a single observation from the simulation is

$$I(L > x)\exp\{\psi(\theta) - \theta Q\} \approx I(a_0 + Q > x)\exp\{\psi(\theta) - \theta Q\},$$

and thus stratifying on Q removes essentially all of the variability from both the indicator and the LR. In this setting, stratifying on Q is equivalent to stratifying on the LR. In the option pricing setting of Glasserman et al. (1999a, c), the IS changes only the means of the normals to a vector μ, in which case the LR equals a constant times $\exp\{-\mu' Z\}$, and stratification is done on a linear combination $a'Z$. Thus when $a = \mu$, the approach in Glasserman et al. (1999a, c) represents another, often effective, instance of stratifying on the LR.

In this paper, we will use equiprobable strata and an equal allocation of samples to strata, i.e., equal p_js and equal n_js. The potential for boosting the performance of the method by using an optimal, or near optimal allocation is explored in Glasserman et al. (1999d). We also investigate the relationship between IS and adjusting the allocation $\{n_j\}$ in Glasserman et al. (1999d).

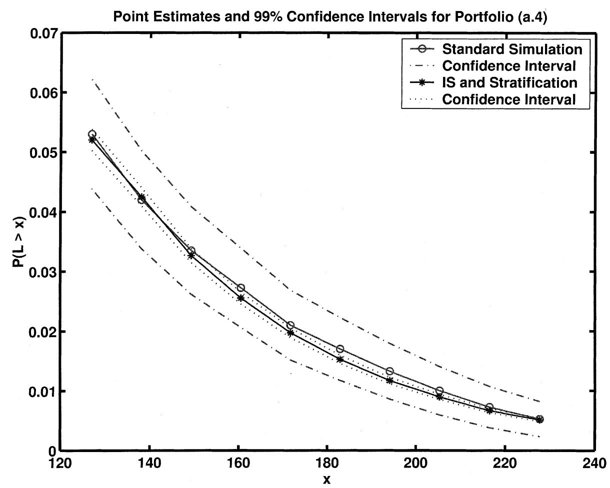
To implement stratified sampling, given strata (intervals) $\mathscr{S}_j$ we must first compute the p_js for the quadratic form Q when (under IS) the mean and variance of Z_i are given in (8). Since the transform of Q is known, numerical transform inversion techniques may be used (Imhof 1961, Rouvinez 1997). Next, given a desired number of samples n_j from stratum j, we need to generate samples $\{Q^{(ij)}, Z^{(ij)}\}$ for $i=1, \ldots, n_j$ such that $Q^{(ij)} \in \mathscr{S}_j$ and such that $Z^{(ij)}$ and $Q^{(ij)}$ have the correct joint distribution, given $Q^{(ij)} \in \mathscr{S}_j$. For this step, we consider a simple "bin tossing" method to perform the stratified sampling.

In the bin tossing method, we simply generate a vector Z of independent normals with the correct mean and variance and compute its corresponding $Q = \sum_i (b_i Z_i + \lambda_i Z_i^2)$. This value of Q must fall in some stratum, say j. Then the pair (Q, Z) has the required joint distribution, given Q is in stratum j. If there are fewer than n_j samples from this stratum, then this Z is used to evaluate the portfolio, otherwise it is discarded. Sampling is continued until the required number of samples is drawn from each stratum. We view the time to generate risk factor changes as being insignificant compared to the time to evaluate the portfolio and our experience has been that this method is reasonably efficient under a broad variety of circumstances. The method's sampling efficiency is analyzed in the Appendix.

4. From Probabilities to Quantiles

In practice, one is typically interested in estimating $P\{L > x\}$ for a range of values of x. For example, one might be interested in the range of xs such that $0.01 \le P\{L > x\} \le 0.05$. In addition, the VAR is defined as a quantile of the loss distribution, so a priori, the appropriate value of x is unknown and must be estimated. However, in IS, a single twisting parameter is selected and setting the twisting parameter to θ_x can be viewed as optimizing the simulation for the estimation of $P\{L > x\}$. Thus it is natural to investigate the robustness of the IS when the twisting parameter θ_x is used to estimate $P\{L > y\}$ for $y \ne x$. The following theorem shows that if the portfolio is quadratic, then the IS using parameter θ_x is also asymptotically optimal for estimating $P\{L > y\}$ for a wide range of ys.

Figure 1 **Point Estimates and 99% Confidence Intervals for the Portfolio (a.4) Using Standard Simulation and Importance Sampling with Stratification**



Theorem 5. *If $L = a_0 + Q$, $\lambda_1 > 0$ and $y_x \to \infty$ in such way that* $\limsup_{x\to\infty} (x/y_x) < \infty$, *then exponential twisting using the twisting parameter θ_x defined by* (11) *is also asymptotically optimal for estimating $P\{L > y_x\}$, i.e.,*

$$\lim_{x\to\infty} \frac{\log(m_2(y_x, \theta_x))}{y_x} = -\frac{1}{\lambda_1}. \tag{21}$$

This robustness is illustrated in Figure 1, which displays the point estimates and 99% confidence intervals for a range of values $P\{L > y\}$ using both standard simulation and IS + stratification for the portfolio (a.4) described in §5. The figure represents the results from a total of 4,000 samples of the loss value. The IS used θ_x for $x = 185$ and the stratified sampling used 40 equiprobable strata with 100 samples/stratum. Some care needs to be taken in interpreting the figure; because the estimates are correlated, it does not represent a true simultaneous confidence region for $P\{L > y\}$ for all values of y in the range. The "bands" for the confidence limits are obtained by connecting the 99% confidence intervals obtained at discrete points y_j. The Bonferroni inequality (Feller 1968, p. 110) may be used to conclude, for example, that with probability at least 0.90, the confidence intervals at y_j for $j=1, \ldots, 10$ are simultaneously true.

As the VAR is a quantile, we now study the properties of a quantile estimate obtained under both IS and stratification. The main result is a central limit theorem for the quantile estimate. The form of the variance in this CLT shows that if the variance of the IS and stratification estimate for $P\{L>x\}$ is reduced for x in a neighborhood of the quantile, then that variance reduction carries over to the quantile estimate. This extends a result in Glynn (1996), in which a CLT for a quantile estimate under IS alone was established.

We assume there are a fixed (finite) number, k, of strata. Let

$$F(x)=P\{L\le x\}=\sum_{j=1}^{k}p_jP\{L\le x|Y\in \mathscr{S}_j\}=\sum_{j=1}^{k}p_j\,F_j(x), \quad (22)$$

where Y is the stratification variable, $\mathscr{S}_j$ is the jth stratum, $p_j=P\{Y\in\mathscr{S}_j\}$, and $F_j(x)=P\{L\le x|Y\in\mathscr{S}_j\}$. Let $n_j=[q_jn]$ be the number of samples in stratum j for fixed constants $q_j>0$ such that $\sum_j q_j=1$. Let L_{ij} be the loss from sample number i in stratum j and let ℓ_{ij} be the corresponding likelihood ratio (we do not rule out $\ell_{ij}=1$, i.e., no importance sampling). For a fixed p, let

$$\alpha=F^{-1}(1-p)=\inf\{x\colon F(x)\ge(1-p)\} \quad (23)$$

be the quantile of interest. Let $f(x)$ denote the derivative of $F(x)$ (when it exists). Define

$$F_{jn}(x)=1-\frac{1}{n_j}\sum_{i=1}^{n_j}\ell_{ij}I(L_{ij}>x), \quad (24)$$

and $F_n(x)=\sum_{j=1}^{k}p_jF_{jn}(x)$. The quantile estimate is

$$\alpha_n=\inf\{x\colon F_n(x)\ge(1-p)\}. \quad (25)$$

Let $\sigma_j^2(x)=E[\ell_{ij}^2I(L_{ij}>x)]-(1-F_j(x))^2$, and let

$$\sigma^2(x)=\sum_{j=1}^{k}p_j^2\sigma_j^2(x)/q_j. \quad (26)$$

THEOREM 6. *If f is positive and continuous in a neighborhood of α, $E[\ell_{ij}^3]<\infty$, and $\sigma(\alpha)>0$, then*

$$\sqrt{n}(\alpha_n-\alpha)\Rightarrow\frac{\sigma(\alpha)}{f(\alpha)}N(0,1). \quad (27)$$

The condition $E[\ell_{ij}^3]<\infty$ is satisfied if we twist only the Z_is associated with nonnegative eigenvalues λ_i as described in §2. The variance in the CLT is inflated by the usual factor $1/f(\alpha)^2$ that appears in other quantile estimation contexts. In particular, for an extreme quantile, $1/f(\alpha)^2$ may be quite large since the density is near zero in the tail of the distribution. However, the impact of this inflation is reduced by making $\sigma^2(\alpha)$ small, which is what the IS and stratification is aimed at doing. Estimation of $f(\alpha)$ is difficult and beyond the scope of this paper. An informal indication of the reduction in variance for the quantile estimate can be obtained from Figure 1. Imagine drawing a horizontal line at, say, $p=0.01$. Consider the points at which this line intersects a method's confidence limits. The width of the interval between the confidence limits gives an indication of the quantile estimate's accuracy. Notice how much narrower this interval is for the IS + stratified estimate than it is for the standard simulation estimate. Strictly speaking, such an interval does not represent a valid confidence interval for the quantile since the bands in Figure 1 are obtained by connecting discrete points and are thus not simultaneously true for all points.

5. Numerical Results

We begin this section by stating the complete algorithm.

Importance Sampling and Stratification Algorithm for VAR

Goal: Estimate $P\{L>x\}$ assuming $\Delta S\stackrel{\mathrm{d}}{=}N(0,\Sigma)$
Given: Σ, a_0, a, and A (Note: $L\approx a_0+Q=a_0+a'\Delta S+\Delta S'A\Delta S$)

1. Express $Q=b'Z+Z'\Lambda Z$ where $Z\stackrel{\mathrm{d}}{=}N(0,I)$, Λ diagonal:
   - Find $\tilde{C}\tilde{C}'=\Sigma$
   - Solve the eigenvalue problem: $\tilde{C}'A\tilde{C}=U\Lambda U'$, U orthonormal
   - Set $C=\tilde{C}U$, $b'=a'C$
2. Identify the IS distribution:
   - Set $\theta=\theta_x$ where $\psi'(\theta_x)=(x-a_0)$ (see (A5))
   - For each $i=1,\ldots,m$, define $\sigma_i^2(\theta)=1/(1-2\theta\lambda_i)$, $\mu_i(\theta)=\theta b_i\sigma_i^2(\theta)$

3. Define k strata: given probabilities p_j with $\sum_{j=1}^{k} p_j = 1$, find s_j such that $P_\theta\{Q \le s_j\} = \sum_{i=1}^{j} p_i$, $j = 1, \ldots, k-1$ (for equiprobable bins, $p_j \equiv 1/k$)
4. Perform the simulation: For $j = 1, \ldots, k$, $i = 1, \ldots, n_j$
   - With parameter θ, generate $Z^{(ij)}$, $Q^{(ij)}$ in stratum j and set $\Delta S^{(ij)} = CZ^{(ij)}$
   - Using $\Delta S^{(ij)}$, evaluate the loss L_{ij} and LR $\ell_{ij} = \exp\{\psi(\theta) - \theta Q^{(ij)}\}$
5. Estimate $P\{L > x\}$ by $\hat{P} = \sum_{j=1}^{k} p_j (1/n_j) \sum_{i=1}^{n_j} I(L_{ij} > x)\ell_{ij}$

We perform experiments with IS alone and IS with stratification (ISS-Q) on a set of portfolios. A preliminary set of experiments was done in Glasserman et al. (1999b) where portfolios consisting of standard European calls and puts were investigated. Only portfolios for which the λ_is (elements of Λ) were positive were considered. In this paper we expand on the study in Glasserman et al. (1996b) by also considering portfolios that have negative λ_is, portfolios that have discontinuous payoffs, and portfolios consisting of instruments whose payoffs are functions of more than one risk factor. Quadratic approximations are based on the delta-gamma approximation.

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). In order to limit the number of cases considered, most experiments use portfolios with instruments based on ten uncorrelated assets with all assets having an initial value of 100 and an annual volatility of 0.3; unless we mention otherwise we will assume that this is the case. In two cases we also consider correlated assets and in one of these the portfolio involves 100 assets with different volatilities.

For comparison purposes, in each case we adjust the loss x so that the probability to be estimated is close to 0.01. To standardize results across portfolios, as in Glasserman et al. (1999b), we specify x as x_{std} standard deviations above the mean according to the delta-gamma approximation:

$$x = \left(\sum_i \lambda_i + a_0\right) + x_{\text{std}} \sqrt{\sum_i b_i^2 + 2\sum_i \lambda_i^2}.$$

We begin by considering portfolios for which some or all of the λ_is are negative; in particular portfolios in which at least the smallest eigenvalue $\lambda_m < 0$. For comparison, we also consider some cases in which all the λ_is are positive. The following is the set of portfolios we consider for investigating the effect of negative λ_is and portfolio size.

(a.1) *0.5yr 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) *0.5yr ATM*, $-\lambda$: Long 10 at-the-money calls and 5 at-the-money puts on each asset, all options having a half-year maturity. All eigenvalues are negative.

(a.3) *0.5yr ATM*, $\pm\lambda$: Short 10 at-the-money calls and short 5 at-the-money puts on the first 5 assets. Long 10 at-the-money calls and short 5 at-the-money puts on the next 5 assets. This portfolio has both positive and negative eigenvalues.

(a.4) *0.1yr ATM*: Same as (a.1) but with a maturity of 0.10 years.

(a.5) *0.1yr ATM*, $-\lambda$: Same as (a.2) but with a maturity of 0.10 years.

(a.6) *0.1yr ATM*, $\pm\lambda$: Same as (a.3) but with maturity of 0.10 years.

(a.7) *Delta hedged*: Same as (a.4) but with number of puts increased so that $\delta = 0$.

(a.8) *Delta hedged*, $-\lambda$: Same as (a.5) but with number of puts increased so that $\delta = 0$.

(a.9) *Delta hedged*, $\pm\lambda$: Short 10 at-the-money calls on first 5 assets. Long 5 at-the-money calls on the remaining assets. Long or short puts on each asset so that $\delta = 0$.

(a.10) *Delta hedged*, $\lambda_m < -\lambda_1$: Short 5 at-the-money calls on first 5 assets. Long 10 at-the-money calls on next 5 assets. Long or short puts on each asset so that $\delta = 0$.

(a.11) *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 for the asset prices was downloaded from the RiskMetrics$^{\text{TM}}$ website and is given in Glasserman et al. (1999b). The initial asset prices are taken as (100, 50, 30, 100, 80, 20, 50, 200, 150, 10).

(a.12) *Index*, $-\lambda$: Same as (a.11) but now we long 50 at-the-money calls and 50 at-the-money puts on the 10 underlying assets.

(a.13) *Index*, $\pm\lambda$: Same as (a.11), but now we long 50 at-the-money calls and 50 at-the-money puts on the

Table 2 Comparison of Methods for Portfolios with Continuous Payoffs

	Portfolio	x_{std}	$P\{L > x\}$	Variance Ratios IS	ISS-Q
(a.1)	0.5yr ATM	2.5	1.0%	30	270
(a.2)	0.5yr ATM, $-\lambda$	1.95	1.0%	43	260
(a.3)	0.5yr ATM, $\pm\lambda$	2.3	1.0%	37	327
(a.4)	0.1yr ATM	2.6	1.1%	22	70
(a.5)	0.1yr ATM, $-\lambda$	1.69	1.0%	43	65
(a.6)	0.1yr ATM, $\pm\lambda$	2.3	0.9%	34	132
(a.7)	Delta hedged	2.8	1.1%	17	31
(a.8)	Delta hedged, $-\lambda$	1.8	1.1%	52	124
(a.9)	Delta hedged, $\pm\lambda$	2.8	1.1%	16	28
(a.10)	Delta hedged, $\lambda_m < -\lambda_1$	2.0	1.1%	19	34
(a.11)	Index	3.2	1.1%	18	124
(a.12)	Index, $-\lambda$	1.02	1.0%	28	48
(a.13)	Index, $\pm\lambda$	2.5	1.1%	15	65
(a.14)	Index, $\lambda_m < -\lambda_1$	1.65	1.1%	14	45
(a.15)	100, Block-diagonal	2.65	1.0%	18	28

first 5 assets, and long 50 at-the-money calls and 50 at-the-money puts on the next 5 assets.

(a.14) *Index*, $\lambda_m < -\lambda_1$: Same as (a.12), but now we short 50 at-the-money calls and 50 at-the-money puts on the first three assets, and long 50 at-the-money calls and 50 at-the-money puts on the next seven assets.

(a.15) 100, *Block-diagonal*: Short 10 at-the-money calls and 10 at-the-money puts on 100 underlying assets, all options expiring in 0.10 years. The assets are divided into 10 groups of 10 assets each. The correlation is 0.2 between assets in the same group and is 0 between assets if they belong to different groups. The assets in the first three groups have a volatility of 0.5, those in the next four groups have a volatility of 0.3, and those in the last three groups have a volatility of 0.1.

Results are given in Table 2. The table lists the estimated variance ratios for both IS and ISS-Q, i.e., the (estimated) variance using standard Monte Carlo divided by the (estimated) variance using IS (or ISS-Q); this is an estimate of the computational speed-up that can be obtained using a method. In all experiments, unless otherwise mentioned, the variance ratios are estimated from a total of 80,000 replications; the stratified estimator uses 40 strata with 2,000 samples per stratum. These sample sizes are much larger than what would likely be used in practice, but they are needed to obtain precise estimates of variance ratios and thus a reliable comparison of methods. These variance ratios provide accurate estimates of the variance reduction that would be obtained for any total sample size that was divided equally among the 40 strata.

Large variance reductions are obtained for all the portfolios. For the case where not all λ_is are negative, the method is somewhat less effective for the delta-hedged portfolios (e.g., compare (a.7) to (a.4)); for this case the quadratic approximation appears to be more accurate for portfolios which are not delta-hedged. Correlations between assets and portfolio size have little impact on the order of the variance improvement. As explained before in Corollary 1, if some of the λ_is are negative one has to be careful about large (even infinite) variance in twisting all the Z_is, since asymptotic optimality is shown only for the quadratic approximation and not for the actual loss function. However, this problem did not surface in our experiments. Furthermore, for these portfolios, we found that twisting only Z_is associated with nonnegative eigenvalues resulted in diminished performance; e.g., for portfolio (a.14) the variance ratio for ISS-Q decreases from 45 to 12.

Next we investigate the effect of discontinuities in the payoff functions of options. We consider the following set of portfolios, with all options being at-the-money and having a maturity of 0.1 years. We refer the reader to Hull (1997) for a description of these options and the respective option pricing formulas.

(b.1) *C*: Short 10 (standard European) calls on each asset (done for comparison purposes).

(b.2) *DAO-C*: Short 10 down-and-out calls on each asset. The barrier in the down-and-out calls was set at 95.

(b.3) *DAO-C* & *P*: short 10 down-and-out calls and short 5 (standard European) puts on each asset.

(b.4) *DAO-C & P*, *Hedged*: Same as above, but the number of puts adjusted to delta hedge.

(b.5) *DAO-C & CON-P*: Same as (b.3) except that now we replace (standard European) puts by cash-or-nothing puts. The cash value is set to be equal to the strike price.

(b.6) *DAO-C* & *CON-P*, *Hedged*: Same as above but the number of cash-or-nothing puts adjusted to delta hedge.

Table 3 Comparison of Methods for Portfolios with Discontinuous Payoffs

	Portfolio	x_{std}	$P\{L>x\}$	Variance Ratios IS	ISS-Q
(b.1)	C	2.55	1.0%	31	166
(b.2)	DAO-C	2.45	1.1%	25	46
(b.3)	DAO-C & P	2.8	1.1%	14	16
(b.4)	DAO-C & P, Hedged	4.9	1.0%	7.7	9.1
(b.5)	DAO-C & CON-P	2.75	1.1%	21	31
(b.6)	DAO-C & CON-P, Hedged	9	1.0%	0.7	0.8
(b.7)	CON-C & CON-P	2.3	1.0%	24	36
(b.8)	AON-C & CON-P	2.35	1.0%	22	32

(b.7) *CON-C and CON-P*: Short 5 cash-or-nothing calls and short 10 cash-or-nothing puts on each asset.

(b.8) *AON-C and CON-P*: Short 5 asset-or-nothing calls and short 10 cash-or-nothing puts on each asset.

To simplify the simulation, in pricing barrier options we ignore the possibility of being knocked-out in the interval $(t, t+\Delta t)$; we use the underlying asset value at time $t+\Delta t$ to revalue the option at time $t+\Delta t$, assuming it was not previously knocked-out (and give the option a value of 0 if the underlying's value at time $t+\Delta t$ is below the barrier). Results are given in Table 3.

One would expect the delta-gamma approximation to be less accurate in this setting, leading to IS and ISS-Q performing worse than in the case of continuous payoff options considered before. Leaving out (b.6) for the moment, the variance improvement with IS seems to be of the same order as before. However as can be seen by comparing cases (b.1) and (b.2), the boost in performance in going from IS to ISS-Q is less than before.

Portfolio (b.6) is a difficult and extreme cases: All options have discontinuous payoffs, the portfolio is delta hedged (requiring the pure quadratic part to accurately fit a function with discontinous first and second derivatives), and $\Delta t = 0.04$ years is relatively large compared to the expiration data of 0.10 years for all of the options. For this portfolio, the method actually results in a slight increase in variance. (A total of 800,000 replications with 40 strata and 20,000 samples per stratum were required to get accurate variance estimates.) The delta-gamma quadratic gives a poor approximation to the true loss in this example and, in particular, underestimates the loss in the region of interest. Because the approximation underestimates the true loss, it overestimates the amount of twisting needed to sample the region $\{L>x\}$ effectively; in this example, we obtained better results using a smaller value of θ. We have also found that an alternative quadratic approximation based on a least-squares fit improves the variance ratio for ISS-Q to 9.5. However, identifying such a fit would in general require substantial additional computational effort.

Finally we consider portfolios in which there are nonzero off-diagonal elements in the matrix Γ of second derivatives. Since in practice the full matrix Γ may not be available and may be too expensive to compute, we consider the performance when applying the method in two different situations: one in which the full matrix Γ is available (Setting 1) and the other in which only diagonal entries of Γ are known (Setting 2). Such portfolios may be constructed using, e.g., *exchange options* where the option is to exchange one asset for another. If S_1 and S_2 are values of the asset at the time of maturity of the option, then the payoff from an exchange option, where Asset 1 may be exchanged for Asset 2, is $(S_2 - S_1)^+$. We consider 5 asset pairs where asset number i, $1 \leq i \leq 5$, is paired with asset number $i+5$. We consider the following portfolios, in which all options have a maturity of 0.1 year:

(c.1) *EO*, +: Short 10 exchange options on each of these asset pairs. All off-diagonal elements of Γ are positive.

(c.2) *EO*, $\pm$: Short 10 exchange options on three of the pairs and long 10 exchange options on two of the pairs. Γ has both positive and negative off-diagonal elements.

(c.3) *EO*, *C* & *P*, +: (c.1) combined with shorting 10 at-the-money calls and shorting 5 at-the-money puts on each asset. This lessens the importance of the off-diagonal entries.

(c.4) *EO*, *C* & *P*, $-$: Long 10 exchange options on each of these asset pairs. Combine this with shorting 10 at-the-money calls and shorting 5 at-the-money puts on each asset. All off-diagonal elements of Γ are negative.

(c.5) *EO*, *C* & *P*, $\pm$: (c.2) combined with shorting 10 at-the-money calls and shorting 5 at-the-money puts on each asset.

Table 4 Comparison of Methods for Portfolios with Exchange Options

Portfolio	Setting	x_{std}	$P\{L>x\}$	Variance Ratios IS	ISS-Q
(c.1)	Setting 1	2.7	1.1%	26	128
	Setting 2		1.1%	24	45
(c.2)	Setting 1	2.5	1.1%	28	132
	Setting 2		1.1%	23	34
(c.3)	Setting 1	2.7	1.1%	23	79
	Setting 2		1.1%	19	31
(c.4)	Setting 1	2.45	1.1%	28	105
	Setting 2		1.1%	20	27
(c.5)	Setting 1	2.65	1.1%	23	80
	Setting 2		1.1%	19	33

For a given portfolio, we use the same x in both settings (which is stated in terms of the x_{std} for Setting 1).

The results are presented in Table 4. Note that the performance of IS is about the same in the two settings. The effectiveness of ISS-Q is reduced in Setting 2 compared to Setting 1, but still gives significant improvement over IS.

Results similar to those in above tables are obtained when the risk factors are assumed to have a lognormal distribution. For example, the ISS-Q variance ratios for portfolios (a.4) and (b.5) are 91 and 29, respectively.

6. Summary

This paper has analyzed and experimented with a variance reduction technique for estimating value-at-risk. The first step is to approximate the portfolio loss by a quadratic function, specifically the delta-gamma approximation that is commonly available in practice. Then, based on the approximation, an importance sampling change of measure is determined. If the approximation is exact, then the method was shown to be asymptotically optimal. The importance sampling can then be combined with stratified sampling (on the quadratic) to obtain further variance reduction. The method was applied to a variety of portfolios with different characteristics that can affect its efficiency. For a wide range of portfolios, the method produces large variance reductions; more than one order of magnitude and often more than two orders of magnitude improvements are obtained. As described in Glasserman et al. (1999d), further improvements are possible when samples are better allocated to strata. The method produced no improvement only in one rather extreme situation: when the method's underlying assumption that the loss is well approximated by the delta-gamma quadratic failed.

This paper considered the situation in which the change in risk factors ΔS has a multivariate normal, or lognormal, distribution, with a fixed covariance matrix Σ. However, Σ might itself represent the current value in a random time series model such as GARCH (although we assume Σ remains fixed during the holding period). In addition, while we typically think of S as representing market prices and rates, it could also include volatility factors (e.g., implied volatilities). Our analysis applies to this case provided the necessary sensitivities are available.

There is currently much interest in considering other situations in which the distribution of ΔS has heavier tails. One way to model heavier tails is using a mixture of normals in which, e.g., $\Delta S \stackrel{d}{=} N(\mu_1, \Sigma_1)$ with probability q and $\Delta S \stackrel{d}{=} N(\mu_2, \Sigma_2)$ with probability $(1-q)$. Such situations can be handled directly by our methodology by applying it separately to each case and appropriately combining the cases. Our current research focuses on extending these techniques to other classes of heavy-tailed distributions, such as the multivariate t distribution.[1]

[1] The work of the first author was partially supported by NSF Young Investigator Award DMI 9457189, and that of the third author by NSF Career Award DMI 9625297. The authors would like to thank three anonymous referees and the area editor for valuable comments that significantly improved this paper.

Appendix A Proofs

Proof of Theorem 1. Because $L = Q + a_0$, it suffices to show that $\log(P\{Q>x\})/x \to -1/(2\lambda_1)$. We first show that $\liminf_{x\to\infty} \log(P\{Q>x\})/x \geq -1/(2\lambda_1)$. Using the diagonalized form of Q as in (3), and setting $\varepsilon = 1/x$, we have

$$P\{Q>x\} = k \times I(\varepsilon)$$

$$\equiv k \times \int \exp(-\tfrac{1}{2}\sum z_i^2) I\left(\sum \lambda_i z_i^2 + b_i z_i \geq \frac{1}{\varepsilon}\right) dz. \quad \text{(A1)}$$

Changing variables, $y_i = \sqrt{\varepsilon} z_i$, we have

$$I(\varepsilon) = (1/\sqrt{\varepsilon})^m \int \exp(-\tfrac{1}{2}\sum y_i^2/\varepsilon) I(\sum \lambda_i y_i^2 + \sqrt{\varepsilon} b_i y_i \geq 1) dy \quad \text{(A2)}$$

$$\geq (1/\sqrt{\varepsilon})^m \int_{\{|y_i| \leq B\}} \exp(-\tfrac{1}{2}\sum y_i^2/\varepsilon) I(\sum \lambda_i y_i^2 + \sqrt{\varepsilon} b_i y_i \geq 1) dy \quad \text{(A3)}$$

for any fixed constant B. Write $\sum \lambda_i y_i^2 + \sqrt{\varepsilon} b_i y_i = \sum \lambda_i y_i^2 + R(y, \varepsilon) \equiv Y + R(y, \varepsilon)$. On the set $\{|y_i| \leq B\}$, given any $\delta > 0$ we have $-\delta \leq R(y, \varepsilon) \leq \delta$ for all y and ε sufficiently small. So, if $Y - \delta \geq 1$, then $Y + R(y, \varepsilon) \geq 1$, i.e., $I(Y + R(y, \varepsilon) \geq 1) \geq I(Y - \delta \geq 1)$. Thus

$$I(\varepsilon) \geq (1/\sqrt{\varepsilon})^m \int_{\{|y_i| \leq B\}} \exp(-\tfrac{1}{2}\sum y_i^2/\varepsilon) I(Y - \delta \geq 1) dy. \quad \text{(A4)}$$

This integral is now in a suitable form for a version of Laplace's method (in particular, Theorem 4.3.1 of Dembo and Zeitouni (1998), as applied in Glasserman et al. (1999a)) and $\liminf \varepsilon \log(I(\varepsilon)) \geq -\frac{1}{2}\sum \alpha_i^2$ where $(\alpha_1, \ldots, \alpha_m)$ is any point in the domain of integration in (A4) for which the integrand is positive. In particular, we can take $\alpha_1^2 = (1+\delta)/\lambda_1$ and $\alpha_i = 0$ for $i \geq 2$. Thus $\liminf \varepsilon \log(I(\varepsilon)) \geq -(1+\delta)/(2\lambda_1)$. Since δ is arbitrary, $\liminf \varepsilon \log(I(\varepsilon)) \geq -1/(2\lambda_1)$.

For the upper bound, let α_x be the twisting value so that the mean of Q under twisting, which is $\psi'(\alpha_x)$, equals x. Since as in (9) $P\{Q > x\} \leq \exp(\psi(\alpha_x) - \alpha_x x)$, it suffices to show that $\alpha_x \to 1/(2\lambda_1)$ and that $\psi(\alpha_x)/x \to 0$. Differentiating (7) yields

$$\psi'(\theta) = \sum_{i=1}^m \frac{\theta b_i^2(1-\theta\lambda_i)}{(1-2\theta\lambda_i)^2} + \frac{\lambda_i}{1-2\theta\lambda_i}. \quad \text{(A5)}$$

At $\theta = \alpha_x$, the LHS of (A5) equals x, so as $x \to \infty$, the RHS of (A5) must also $\to \infty$, which implies that $\alpha_x \to 1/(2\lambda_1)$. Evaluating (A5) at α_x and dividing by x shows that if $b_i \neq 0$ for any i for which $\lambda_i = \lambda_1$, then $\lim_{x\to\infty} x(1-2\alpha_x\lambda_1)^2$ exists and is finite. Since the dominant term in $\psi(\alpha_x)$ from (7) involves $1/(1-2\alpha_x\lambda_1)$, we have $\psi(\alpha_x)/x \to 0$. If $b_i = 0$ for all i for which $\lambda_i = \lambda_1$, a similar argument shows that $\lim_{x\to\infty} x(1-2\alpha_x\lambda_1)$ exists. Now the dominant term in $\psi(\alpha_x)$ involves $\log(1-2\alpha_x\lambda_1)$ in which case we again have $\psi(\alpha_x)/x \to 0$. □

Proof of Theorem 2. Since $m_2(x, \theta_x) \geq P\{L > x\}^2$, it suffices to show that $\limsup \log(m_2(x, \theta_x))/x \leq -1/\lambda_1$. From (10) the result will follow provided $\theta_x \to 1/(2\lambda_1)$ and $\psi(\theta_x)/x \to 0$. The proofs of these results are virtually identical to the arguments given for α_x in the proof of Theorem 1. □

Proof of Theorem 3. Write $Q = \sum Q_i + d$ where $Q_i = \lambda_i(Z_i - c_i)^2 \leq 0$ and $c_i = -b_i/(2\lambda_i)$.

Then

$$P\{Q > d - \varepsilon\} \geq P\{Q_i > -\varepsilon/m \ \forall i\}$$

$$= \prod_{i=1}^m P\{|Z_i - c_i| < \sqrt{-\varepsilon/(m\lambda_i)}\},$$

which, by the mean value theorem, is sufficient to show the lower bound in (15). For the upper bound in (15), write $\psi(\theta) = \psi_1(\theta) + \psi_2(\theta)$ where $\psi_2(\theta) = -\frac{1}{2}\sum \log(1-2\theta\lambda_i)$. As in (9), for $\theta \geq 0$,

$$P\{Q > d - \varepsilon\} \leq \exp\{\psi(\theta) - \theta(d-\varepsilon)\}$$

$$= \exp\{\psi_1(\theta) - \theta d + \theta\varepsilon + \psi_2(\theta)\}. \quad \text{(A6)}$$

Simple algebra shows that $(\psi_1(\theta) - \theta d)$ converges to a constant as $\theta \to \infty$. Now let α_ε minimize $\psi_2(\theta) + \theta\varepsilon$; α_ε satisfies

$$\varepsilon = -\sum_{i=1}^m \lambda_i/(1-2\alpha_\varepsilon\lambda_i). \quad \text{(A7)}$$

We consider the RHS of (A6) evaluated at $\theta = \alpha_\varepsilon$. As $\varepsilon \to 0$, the LHS of (A7) approaches zero implying that $\alpha_\varepsilon \to \infty$. Furthermore, by multiplying both sides of (A7) by α_ε, we have $\alpha_\varepsilon\varepsilon \to m/2$. For the last term on the RHS of (A6),

$$\frac{\exp(\psi_2(\alpha_\varepsilon))}{\varepsilon^{m/2}} = \prod_{i=1}^m \frac{1}{\sqrt{\varepsilon}\sqrt{1-2\alpha_\varepsilon\lambda_i}}, \quad \text{(A8)}$$

which remains bounded since, again, $\alpha_\varepsilon\varepsilon \to m/2$. This establishes the upper bound in (15). A similar proof shows (16) since the upper bound on the second moment as in (10) is simply $\exp\{2\psi(\theta) - 2\theta(d-\varepsilon)\}$. The value of this bound using the optimal value θ_ε is less than that when using the suboptimal α_ε. □

Proof of Theorem 4. The convergence $P\{Q_m > x_m\} \to 1 - \Phi(y)$ follows from the Liapounov central limit theorem, as in Theorem 7.1.1 of Chung (1974). The second claim follows from the following key steps: we establish the representation

$$m_2(x_m, \theta_m) = e^{\psi_m(\theta_m) + \psi_m(-\theta_m)} P_{-\theta_m}\{Q_m > x_m\}, \quad \text{(A9)}$$

for all sufficiently large m, we show that

$$\psi_m(\theta_m) + \psi_m(-\theta_m) \to y^2, \quad \text{(A10)}$$

and

$$P_{-\theta_m}\{Q_m > x_m\} \to 1 - \Phi(2y). \quad \text{(A11)}$$

In proving these steps, we make use of the following observations. Condition (19) implies that the b_i, λ_i are bounded, and therefore that $\delta^* \equiv 1/(2\sup_i |\lambda_i|)$ is positive. For any $0 < \delta < \delta^*$, there are constants c_1, c_2 such that

$$c_1 \cdot \psi_m''(0) \leq \inf_{0 \leq \theta \leq \delta} |\psi_m''(\theta)| \leq \sup_{0 \leq \theta \leq \delta} |\psi_m''(\theta)| \leq c_2 \cdot \psi_m''(0), \quad \text{(A12)}$$

and

$$\sup_{0 \leq \theta \leq \delta} |\psi_m'''(\theta)| \leq c_2 \cdot \psi_m''(0). \quad \text{(A13)}$$

Both (A12) and (A13) follow from the boundedness of the b_i, λ_i and inspection of the derivatives of ψ_m.

As a first consequence of these observations we argue that $\theta_m \to 0$. Fix any $0 < \delta < \delta^*$; if $\delta < \theta_m < \delta^*$, then from the definition of y, the convexity of ψ_m, the mean value theorem, and (A12), we get

$$y = \frac{\psi'_m(\theta_m) - \psi'_m(0)}{\sqrt{\psi''_m(0)}} \geq \frac{\psi'_m(\delta) - \psi'_m(0)}{\sqrt{\psi''_m(0)}}$$
$$\geq \delta \frac{\inf_{0\leq\theta\leq\delta} \psi''_m(\theta)}{\sqrt{\psi''_m(0)}} \geq \delta c_1 \sqrt{\psi''_m(0)}.$$

But (19) implies that $\psi''_m(0)$ increases without bound, so we must have $\theta_m \leq \delta$ for all sufficiently large m.

Next we establish (20). From the definition of θ_m and the mean value theorem we have $\sqrt{\psi''_m(0)}\theta_m = y\psi''_m(0)/\psi''_m(a_m)$ for some $0 \leq a_m \leq \theta_m$. We need to show that $\psi''_m(a_m)/\psi''_m(0) \to 1$. As $\theta_m \to 0$, we have $a_m < \delta$ for all sufficiently large m and any $\delta > 0$. A further application of the mean value theorem and (A12) thus yields

$$\left|\frac{\psi''_m(a_m)}{\psi''_m(0)} - 1\right| \leq \frac{a_m \sup_{0\leq\theta\leq a_m} |\psi'''_m(\theta)|}{\psi''_m(0)} \leq a_m \cdot c_2$$

for all sufficiently large m, and this upper bound approaches 0 as m increases.

To show (A9), note that $\psi_m(-\theta_m) < \infty$ if $0 < \theta_m < \delta^*$; because $\theta_m \to 0$, we thus have $\psi_m(-\theta_m) < \infty$ for all sufficiently large m. For all such m, two changes of measure now yield

$$\begin{aligned} m_2(x_m, \theta_m) &= E_{\theta_m}[I(Q_m > x_m)e^{-2\theta_m Q_m + 2\psi_m(\theta_m)}] \\ &= E_0[I(Q_m > x)e^{-\theta_m Q_m + \psi_m(\theta_m)}] \\ &= E_{-\theta_m}[I(Q_m > x)e^{-\theta_m Q_m + \psi_m(\theta_m)}e^{\theta_m Q_m + \psi_m(-\theta_m)}] \\ &= e^{\psi_m(\theta_m) + \psi_m(-\theta_m)} P_{-\theta_m}\{Q_m > x_m\}. \end{aligned}$$

Next we turn to (A10). Using R_m to denote the remainder in a second-order Taylor approximation, we have

$$\psi_m(\theta_m) = \psi_m(0) + \theta_m \psi'_m(0) + \tfrac{1}{2}\theta_m^2 \psi''_m(0) + R_m(\theta_m)$$
$$\psi_m(-\theta_m) = \psi_m(0) - \theta_m \psi'_m(0) + \tfrac{1}{2}\theta_m^2 \psi''_m(0) + R_m(-\theta_m).$$

Adding these equations and invoking (20) establishes (A10) provided the remainder terms vanish as m increases. But for any $0 < \delta < \delta^*$ and all sufficiently large m,

$$\begin{aligned} |R_m(\theta_m)| &\leq |\theta_m|^3 \sup_{0\leq\theta\leq\theta_m} |\psi'''_m(\theta)| \\ &\leq |\theta_m|^3 \sup_{0\leq\theta\leq\delta} |\psi'''_m(\theta)| \leq |\theta_m|^3 \cdot c_2 \cdot \psi''_m(0), \end{aligned}$$

by (A13). From (20) we conclude that $|R_m(\theta_m)| \to 0$. The same argument applies to $|R_m(-\theta_m)|$.

We now turn to (A11). Under $P_{-\theta}$, $Z_1, Z_2, \ldots$ remain independent and Z_i has distribution $N(\mu_i(-\theta), \sigma_i^2(-\theta))$, with μ_i, σ_i as in (8). Consequently, the distribution of Q_m under $P_{-\theta}$ is the same as that of

$$\begin{aligned} Q_m(-\theta) &= \sum_{i=1}^m \{\lambda_i(\mu_i(-\theta) + \sigma_i(-\theta)Z_i)^2 + b_i(\mu_i(-\theta) + \sigma_i(-\theta)Z_i)\} \\ &\equiv \sum_{i=1}^m \{\lambda_i(-\theta)Z_i^2 + b_i(-\theta)Z_i + c_i(-\theta)\} \end{aligned}$$

under P_0. Straightforward algebra shows that for any $0 < \delta < \delta^*$ and all $|\theta| < \delta$,

$$|\lambda_i(\theta) - \lambda_i| \leq k|\theta\lambda_i|, \quad |b_i(\theta) - b_i| \leq k|\theta b_i|, \quad |c_i(\theta)| \leq k|\theta|$$

for some constant k. Consequently, for m large enough that $\theta_m < \delta$,

$$\left|\frac{Q_m(-\theta_m) - \psi'_m(-\theta_m)}{Q_m(0) - \psi'(0)} - 1\right| \leq k'|\theta_m|$$

for some constant k'. It follows from the central limit theorem for $Q_m(0)$ and the Corollary on p. 93 of Chung (1974) that the distribution (under P_0) of $(Q_m(-\theta_m) - \psi'_m(-\theta_m))/\sqrt{\psi''(0)}$ converges to the standard normal. Arguments similar to those used previously show that

$$\lim_{m\to\infty} \frac{\psi'_m(-\theta_m) - \psi'_m(0)}{\sqrt{\psi''_m(0)}} = \lim_{m\to\infty} \frac{-\theta_m \psi''_m(0)}{\sqrt{\psi''_m(0)}} = -y.$$

We therefore have

$$\begin{aligned} P_{-\theta_m}\{Q_m > x_m\} &= P_0\{Q_m(-\theta_m) > \psi'_m(0) + y\sqrt{\psi''_m(0)}\} \\ &= P_0\left\{\frac{Q_m(-\theta_m) - \psi'_m(-\theta_m)}{\sqrt{\psi''(0)}} + \frac{\psi'_m(-\theta_m) - \psi'_m(0)}{\sqrt{\psi''_m(0)}} > y\right\} \\ &\to 1 - \Phi(2y). \quad \square \end{aligned}$$

Proof of Theorem 5. Using (10)

$$\begin{aligned} \frac{\log(m_2(y_x, \theta_x))}{y_x} &\leq \frac{2\psi(\theta_x) - 2\theta_x(y_x - a_0)}{y_x} \\ &= \frac{2\psi(\theta_x)}{x} \times \frac{x}{y_x} - 2\theta_x \frac{(y_x - a_0)}{y_x}. \end{aligned} \quad \text{(A14)}$$

The result then follows since $\psi(\theta_x)/x \to 0$, $\limsup(x/y_x) < \infty$, and $\theta_x \to 1/(2\lambda_1)$. $\square$

Proof of Theorem 6. With $q = 1 - p$, then as in Equation (4) of Glynn (1996),

$$\begin{aligned} P\{\sqrt{n}(\alpha_n - \alpha) \leq x\} &= P\{q \leq F_n(\alpha + x/\sqrt{n})\} \\ &= P\Big\{q \leq \sum_j p_j F_{jn}(\alpha + x/\sqrt{n})\Big\} \\ &= P\Big\{\sqrt{n}(q - F(\alpha + x/\sqrt{n})) \\ &\leq \sum_j p_j \sqrt{n}(F_{jn}(\alpha + x/\sqrt{n}) - F_j(\alpha + x/\sqrt{n}))\Big\} \\ &\equiv P\Big\{c_n(x) \leq \sum_j p_j Y_{jn}(x)\Big\} \\ &= P\Big\{0 \leq -c_n(x) + \sum_j p_j Y_{jn}(x)\Big\}. \end{aligned} \quad \text{(A15)}$$

Under the stated assumptions on f, $c_n(x) \to -xf(\alpha)$. Furthermore, by the argument used in Equations (5) and (6) of Glynn (1996), $Y_{jn}(x) \Rightarrow (\sigma_j(\alpha)/\sqrt{q_j})Y_j$ where the Y_js are independent standard normal random variables. Thus

$$-c_n(x) + \sum_j p_j Y_{jn}(x) \Rightarrow xf(\alpha) + \sum_j \frac{p_j \sigma_j(\alpha)}{\sqrt{q_j}} Y_j \stackrel{d}{=} xf(\alpha) + N(0, \sigma^2(\alpha)). \tag{A16}$$

Combining (A15) and (A16) yields $P\{\sqrt{n}(\alpha_n - \alpha) \le x\} \to P\{0 \le xf(\alpha) + N(0, \sigma^2(\alpha))\}$ which is equivalent to (27). □

Analysis of Bin Tossing Method. To analyze the method's overhead, model the sampling process as follows. Let $\{N_j(t)\}$ be a Poisson process with rate p_j and let $\{N(t)\}$ be the superposition of these processes: $\{N(t)\}$ is a Poisson process with Rate 1. Furthermore, the joint distribution of the $\{N_j(t)\}$s at the jump times of $\{N(t)\}$ has the correct multinomial distribution. If $\sum_j n_j = n$, then $E[N(n)] = n$ represents the expected time to draw n samples (without discarding any). Consider a time $t_n > n$. Then (in the model) the method is done sampling by time t_n provided $N_j(t_n) \ge n_j$ for all j. Thus if T_n denotes the random time at which the method is done sampling, then $P\{T_n \le t_n\} = \prod_{j=1}^k P\{N_j(t_n) \ge n_j\}$, which is easily computed. For example, with 100 equiprobable strata and 20 samples/stratum, then $P\{T_n \le 1.9 \times n\} = 0.95$, i.e., the probability of requiring more than $1.9 \times n$ samples to generate the required n stratified samples in the correct strata is about 0.05. Fox (1999) states that the expected time to obtain (n/k) samples from each of k equiprobable bins is approximately $n \log(k)$. This Poisson model is quite accurate and, furthermore, the bin tossing method is reasonably efficient even for moderately skewed allocations which attempt to further reduce variance by allocating more samples to those strata with the highest variances.

References

Britten-Jones M., S. M. Schaefer. 1999. Non-linear value-at-risk. *Euro. Finance Rev.* **2** 161–187.

Bucklew, J. A. 1990. *Large Deviation Techniques in Decision, Simulation and Estimation*. John Wiley & Sons, Inc., New York.

Cárdenas, J., E. Fruchard, J.-F. Picron, C. Reyes, K. Walters, W. Yang. 1999. Monte Carlo within a day. *Risk* **12** (February) 55–59.

Chen, J.-C., D. Lu, J. Sadowsky, K. Yao. 1993. On importance sampling in digital communications—Part I: Fundamentals. *IEEE J. Selected Areas Comm.* **11** 289–299.

Chung, K. 1974. *A Course in Probability Theory*, 2nd ed. Academic Press, New York.

Dembo, A., O. Zeitouni. 1998. *Large Deviations Techniques and Applications*, 2nd ed. Springer-Verlag, New York.

Feller, W. 1968. *An Introduction to Probability Theory and its Applications*, Volume I, 3rd ed. John Wiley & Sons, Inc., New York.

Fishman, G. S. 1996. *Monte Carlo Concepts, Algorithms, and Applications*. Springer, New York.

Fox, B. 1999. Generating random normal vectors for Quasi-Monte-Carlo. Available from author at⟨bfox@carbon.cudenver.edu⟩

Glasserman, P., P. Heidelberger, P. Shahabuddin. 1999a. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Math. Finance* **9** 117–152.

——, ——, ——. 1999b. Importance sampling and stratification for Value-at-Risk. Y. S. Abu-Mostafa, B. LeBaron, A. W. Lo, A. S. Weigend, eds. *Computational Finance 1999 (Proc. of the Sixth Internat. Conf. Comput. Finance)* MIT Press, Cambridge, MA, 7–24.

——, ——, ——. 1999c. Importance sampling in the Heath-Jarrow-Morton framework. *J. Derivatives* **7**(1) 32–50.

——, ——, ——. 1999d. Stratification issues in estimating value-at-risk. *Proc. 1999 Winter Simulation Conf.*, IEEE Computer Society Press, Phoenix, AZ, 351–358.

Glynn, P. W. 1996. Importance sampling for Monte Carlo estimation of quantiles. *Mathematical Methods in Stochastic Simulation and Experimental Design: Proc. 2nd St. Petersburg Workshop on Simulation*. Publishing House of Saint Petersburg University, St Petersburg, Russia, 180–185.

Hammersley, J., D. Handscomb. 1964. *Monte Carlo Methods*. Methuen & Co., Ltd., London, U.K.

Heidelberger, P. 1995. Fast simulation of rare events in queueing and reliability models. *ACM Trans. Modeling and Comput. Simulation* **5** 43–85.

Hull, J. C. 1997. *Options, Futures and Other Derivatives*, 2nd ed. Prentice Hall, Upper Saddle River, NJ.

Imhof, J. 1961. Computing the distribution of quadratic forms in normal variables. *Biometrika* **48** 419–426.

Jorion, P. 1997. *Value at Risk*. McGraw-Hill, New York.

Rouvinez, C. 1997. Going Greek with VAR. *Risk* **10** (February) 57–65.

Shahabuddin, P. 1994. Importance sampling for the simulation of highly reliable Markovian systems. *Management Sci.* **40** 333–352.

Wilson, T. 1999. Value at risk. C. Alexander, ed., *Risk Management and Analysis, Vol 1.* Wiley, Chichester, England, 61–124.

Accepted by Phelim P. Bogle; received October 1999. This paper was with the authors 2 months for 1 revision.