RC21450(96822)15APR1999 ComputerScience/Mathematics

99A000198

Research Report

Importance Sampling and Stratification for Value-at-Risk

Paul Glasserman
Columbia Business School
Columbia University
New York, NY 10027

Philip Heidelberger
IBM T. J. Watosn Research Center
P. O. Box 218
Yorktown Heights, NY 10598

Perwez Shahabuddin
IEOR Department
Columbia University
New York, NY 10027

Research Division
Almaden - Austin - Beijing - Haifa - T. J. Watson - Tokyo - Zurich

Importance Sampling and Stratification for Value-at-Risk *

Paul Glasserman
Graduate School of Business
Columbia University
New York, NY 10027

Philip Heidelberger
IBM Research Division
T.J. Watson Research Center
Yorktown Heights, NY 10598

Perwez Shahabuddin
IEOR Department
Columbia University
New York, NY 10027

April 1, 1999

Abstract

This paper proposes and evaluates variance reduction techniques for efficient estimation of portfolio loss probabilities using Monte Carlo simulation. Precise estimation of loss probabilities is essential to calculating value-at-risk, which is simply a percentile of the loss distribution. The methods we develop build on delta-gamma approximations to changes in portfolio value. The simplest way to use such approximations for variance reduction employs them as control variates; we show, however, that far greater variance reduction is possible if the approximations are used as a basis for importance sampling, stratified sampling, or combinations of the two. This is especially true in estimating very small loss probabilities.

1 Introduction

Value-at-Risk (VAR) has become an important measure for estimating and managing portfolio risk [11, 13]. VAR is defined as a certain quantile of the change in a portfolio's value during a specified holding period. To be more specific, suppose the current value of the portfolio is $V(t)$, the holding period is Δt, and the value of the portfolio at time $t + \Delta t$ is $V(t + \Delta t)$. The loss in portfolio value during the holding period is $L = -\Delta V$ where $\Delta V = [V(t+\Delta t) - V(t)]$ and the VAR, x_p, associated with a given probability p is defined by the relationship

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

i.e., the VAR x_p is the $(1-p)$'th quantile of the loss distribution. Typically Δt is one day or two weeks, and $p \leq 0.05$; often $p \approx 0.01$ is of interest. To evaluate (1), Monte Carlo simulation is often used; changes in the portfolio's risk factors are simulated, the portfolio is re-evaluated, and the loss distribution is estimated. However, obtaining accurate VAR estimates can be computationally expensive because:

*This paper will appear in *Computational Finance 1999 (Proceedings of the Sixth International Conference on Computational Finance.* Leonard N. Stern School of Business, New York University, January 6-8, 1999). Edited by Yaser S. Abu-Mostafa, Blake LeBaron, Andrew W. Lo, and Andreas S. Weigend. Cambridge, MA: MIT Press, 1999.

1. there may be a large number of instruments in the portfolio thereby making each portfolio evaluation costly, and
2. when p is small, a large number of simulation runs may be required to obtain accurate estimates of the tail probability.

The purpose of this paper is to describe variance reduction techniques that offer the potential to dramatically reduce the number of runs required to achieve a given precision. These techniques build on [5, 6]. The key to reducing the variance of an estimate of the VAR x_p is to obtain accurate estimates of $P\{L > x\}$ for values of x that are close to x_p, and this is the issue that we focus on.

Our approach is to exploit knowledge of the distribution of an approximation to the loss to devise more effective Monte Carlo sampling schemes. The specific loss approximations employed are first and second order Taylor series expansions of L; these are the well known delta and delta-gamma approximations, respectively (see, e.g., [11], [12], and [13]). When the risk factors have a multivariate normal distribution, as is often assumed, the distribution of the delta approximation is known in closed form while the distribution of the delta-gamma approximation can be computed numerically [10] and [12].

These approximations are not always sufficiently accurate to provide precise VAR estimates. Nevertheless, because of correlation between the approximation and the actual loss, knowledge about the approximation can be put to great advantage for the purpose of variance reduction. This correlation is illustrated in Figure 1, which is a scatter plot of the actual loss and the delta-gamma approximation to the loss for one of the portfolios described in Section 4. Clearly, the value of the delta-gamma approximation tells us a great deal about the value of the loss. The most obvious way to try to exploit this correlation for variance reduction is to use the delta-gamma approximation as a control variate. But the approximation can also be used as a basis for importance sampling and stratified sampling. (See, e.g., [8] for background on variance reduction techniques.) In control variates, only the mean of a correlated variable is used to achieve variance reduction whereas importance sampling and stratified sampling can make use of the full distribution.

If one is interested in estimating $P\{L > x\}$, then one can use $I(Q > x)$ as a control variate where Q is the approximation to the loss and $I()$ is the indicator function. Specifically, this means replacing the standard estimator $I(L > x)$ with

$$I(L > x) - \beta[I(Q > x) - P\{Q > x\}]$$

where L and Q are evaluated in the same price scenario and $P\{Q > x\}$ is computed numerically. The coefficient β can be estimated from the simulation to minimize variance, or—as seems preferable in this setting—fixed at 1 to avoid the bias introduced when β is estimated. Independent of our work,

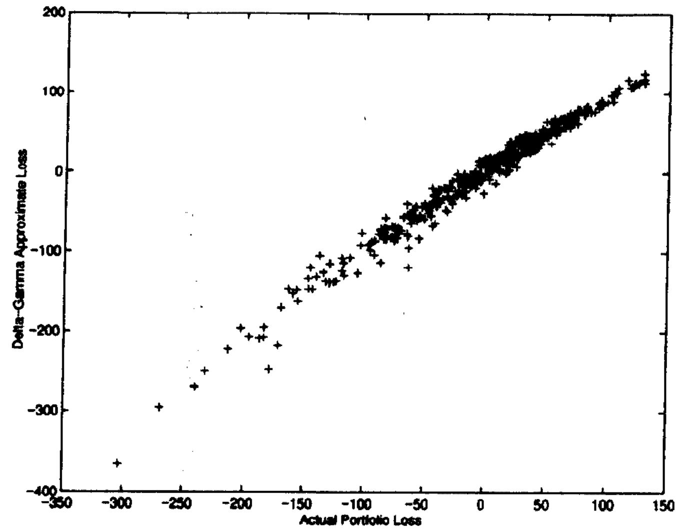

Figure 1: A scatter plot showing delta-gamma approximate losses versus actual portfolio losses. The portfolio is short ten at-the-money calls and puts on each of ten underlying assets. The assets are uncorrelated and each has a volatility of 0.30; the options expire in 0.10 years and the losses are measured over 0.04 years. Negative losses are gains.

[2] have also suggested using the delta-gamma approximation as a control variate. The effectiveness of such a control decreases as we go further out in the tail because the correlation between $I(L > x)$ and $I(Q > x)$ typically decreases as x increases. For example, for the portfolio of Figure 1, the estimated correlation between L and Q is 0.99, but the correlation between $I(L > x)$ and $I(Q > x)$ is 0.85 for $x = 90$ and it drops to 0.50 for $x = 125$. [2] have also suggested using a simple form of stratified sampling in which there are two strata defined by $\{Q \leq x\}$ and $\{Q > x\}$ and in which the fraction of samples drawn from each stratum is proportional to the probability of the stratum. Note that for large x, the fraction of samples drawn from the important region in which $Q \approx x$ remains small.

Examination of Figure 1 reveals that the problem in estimating $P\{L > x\}$ for a large value of x is that very few samples drawn actually have $L > x$. Thus most samples are "wasted" in the sense that $I(L > x) = 0$ with high probability. Importance sampling (IS) [8, 5, 6] is well-suited to such "rare event" simulations. Suppose the joint density of the changes in risk factors is f. Rather than simulating with this density, a different joint density g is used. Write

$$P\{L > x\} = \int I(L > x)f(z)dz = \int I(L > x)\frac{f(z)}{g(z)}g(z)dz = \tilde{E}[I(L > x)\ell(Z)] \qquad (2)$$

where $\tilde{E}$ denotes expectation when sampling is done under density g and $\ell(Z) = f(Z)/g(Z)$ is the likelihood ratio. That is, an unbiased estimate of $P\{L > x\}$ can be obtained by drawing samples with density g and multiplying the output (in this case $I(L > x)$) by the likelihood ratio. The key is picking a good IS density g, and this topic has been the subject of much research in a wide variety of application areas; see, e.g., [5] and [9] and the numerous references therein. Roughly speaking, we want to pick g so as to make the rare event of interest more likely to occur, but extreme care needs to be exercised; see [7] for some cautionary examples. In this paper we describe how the delta and delta-gamma approximations can be used to guide the selection of an effective IS distribution. As described in [5], IS can be effectively combined with stratified sampling to obtain further variance reduction when pricing a class of European-style options. In this paper we consider effective IS-stratification combinations for VAR estimation. In addition, stratified sampling on the approximation, without IS, can also be used to great effect, provided many strata are defined and a greater fraction of samples are drawn from the strata in which the event $\{L > x\}$ is more likely to occur.

The application of these variance reduction techniques does involve some overhead. However, because the per sample cost to evaluate a large portfolio can be enormous, substantial pre-simulation overhead can often be justified so long as it produces even modest variance reduction. Indeed, the overhead of these methods is quite small. Assuming that the delta-gamma approximation is given, the only overheads are:

1. a one time cost to compute an IS change of measure,
2. a one time cost to compute quantities related to the stratified sampling, and
3. a small additional per sample cost to generate risk factors from the required conditional (stratified) distribution and to compute the likelihood ratio; this cost is negligible compared to the cost of evaluating even a modest-sized portfolio.

For the sample portfolios considered in this paper, the variance is typically reduced by more than an order of magnitude, thereby making this overhead well worthwhile.

In this paper, our emphasis is on describing algorithms and providing initial indications of the factors affecting their performance. For these purposes, we use synthetic portfolios of options whose characteristics are easily controlled (e.g., a perfectly delta-hedged portfolio). Theoretical properties of the algorithms and further numerical studies will be described in a subsequent paper.

The rest of the paper is organized as follows. The delta and delta-gamma approximations are reviewed in Section 2. Preliminaries on the basic variance reduction techniques employed in this paper are given in Section 3. Variance reduction techniques based on the delta and delta-gamma approximations are discussed in Sections 4 and 5, respectively. Section 6 considers stratified sampling in further detail (both with and without IS) and in particular focuses on issues related to the effective allocation of samples to strata.

2 The Delta and Delta-Gamma Approximations

In this section, we express the delta and delta-gamma approximations as sums of terms involving independent standard normal random variables (mean zero, variance one). This facilitates the computation of the required quantities for control variates, importance sampling, and stratification. Our development is similar to that of [12], but is included to introduce key notation.

We assume that there are m risk factors, and that $S(t) = (S_1(t), \ldots, S_m(t))$ denotes the value of these factors at time t. Define $\Delta S = [S(t+\Delta t) - S(t)]'$ to be the change in the risk factors during the interval $[t, t+\Delta t]$. The delta-gamma approximation is given by

$$\Delta V = -L \approx \Theta\Delta t + \delta'\Delta S + \tfrac{1}{2}\Delta S'\Gamma\Delta S \tag{3}$$

where $\Theta = \frac{\partial V}{\partial t}$, $\delta_i = \frac{\partial V}{\partial S_i(t)}$, and $\Gamma_{ij} = \frac{\partial^2 V}{\partial S_i(t)\partial S_j(t)}$ (all partial derivatives being evaluated at $S(t)$). The delta approximation is $\Delta V = -L \approx \Theta\Delta t + \delta'\Delta S$.

Assume now that ΔS has a multivariate normal distribution with mean vector zero and covariance matrix Σ. To draw samples of ΔS we can set $\Delta S = CZ$ where Z is a vector of m independent standard normals and C is any matrix for which $CC' = \Sigma$. (Shortly, we make a more specific choice of matrix.) Thus, for the delta approximation,

$$P\{L > x\} \approx P\{b'Z > x + \Theta\Delta t\} \equiv P\{Y > y_x\} \tag{4}$$

where

$$b' = -\delta'C, \qquad Y = b'Z, \qquad \text{and} \qquad y_x = x + \Theta\Delta t. \tag{5}$$

For the delta-gamma approximation, we seek to express $L \approx c + b'Z + Z'\Lambda Z \equiv c + Q$ where the Z's are independent standard normals and Λ is a diagonal matrix. To this end, choose C so that

$$CC' = \Sigma \quad \text{and} \quad \tfrac{1}{2}C'\Gamma C \text{ is diagonal.} \tag{6}$$

To see that such a choice is possible, start with an arbitrary C for which $CC' = \Sigma$ and write

$$\tfrac{1}{2}C'\Gamma C = -U\Lambda U',$$

where U is an orthogonal matrix whose columns are eigenvectors of $\frac{1}{2}C'\Gamma C$ and $-\Lambda$ is a diagonal matrix of eigenvalues of $\frac{1}{2}C'\Gamma C$ (hence also of $\frac{1}{2}\Gamma\Sigma$). Now replace the original C with CU and observe that the new choice satisfies (6). Hence, with (6) in force, we have

$$L \approx -\Theta\Delta t - \delta'CZ + Z'\Lambda Z = -\Theta\Delta t + b'Z + Z'\Lambda Z \tag{7}$$

where $b' = -\delta'C$ and Z is a vector of independent standard normals. Thus

$$P\{L > x\} \approx P\{Q > x + \Theta\Delta t\} = P\{Q > y_x\} \tag{8}$$

where $Q = b'Z + Z'\Lambda Z = \sum(b_i Z_i + \lambda_i Z_i^2)$ is the stochastic part of the quadratic approximation to L. By completing the square, the distribution of Q can be related to that of the sum of noncentral chi-squared random variables. [10] gives expressions that are suitable for numerical integration for computing the distribution function of Q.

3 Variance Reduction Preliminaries

Because the use of control variates is straightforward, we focus on importance sampling, stratified sampling, and combinations of the two. As described in (2) we can write $P\{L > x\} = \tilde{E}[I(L > x)\ell(Z)]$ where $\tilde{E}$ denotes expectation using the IS distribution for Z and $\ell(Z)$ is the likelihood ratio. With Z a vector of standard normals, we consider two types of IS, changing either just the mean of Z or both the mean and covariance matrix. If the mean of Z is changed from 0 to v, then

$$\ell(Z) = \exp\left(\tfrac{1}{2}v'v - v'Z\right), \tag{9}$$

while if the mean is changed from 0 to v and the covariance matrix is changed from I to B, $B > 0$, then

$$\ell(Z) = \frac{\exp(-\frac{1}{2}Z'Z)}{|B|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(Z-v)'B^{-1}(Z-v)\right\}} \tag{10}$$

where $|B|$ is the determinant of B. If n samples $(Z^{(1)}, \ldots, Z^{(n)})$ are drawn and associated losses $(L_1, \ldots, L_n)$ computed, the IS estimate of $P\{L > x\}$ is

$$\hat{P}_x = \frac{1}{n}\sum_{j=1}^{n} I(L_j > x)\ell(Z^{(j)}). \tag{11}$$

It is straightforward to incorporate a control variate with IS. If X is a random variable with known mean, the control variate under IS would be $X\ell(Z)$, which has the same mean.

In stratified sampling, we identify a stratification variable X and partition the range of X into k intervals (strata) $(\mathcal{S}_1, \ldots, \mathcal{S}_k)$. Write

$$P\{L > x\} = \sum_{i=1}^{k} P\{L > x | X \in \mathcal{S}_i\} P\{X \in \mathcal{S}_i\}. \tag{12}$$

We typically have in mind using from $k = 25$ to $k = 100$ equiprobable strata (i.e., $P\{X \in \mathcal{S}_i\} = 1/k$). Draw n_i samples of X from stratum i. Let X_{ij} denote the j'th sample in stratum i, let $Z^{(ij)}$ be a sample of Z having the conditional distribution of Z given $X = X_{ij}$, and let L_{ij} be the portfolio loss corresponding to the sample $Z^{(ij)}$. Then $P\{L > x\}$ is estimated by

$$\hat{P}_x = \sum_{i=1}^{k} P\{X \in \mathcal{S}_i\} \times \frac{1}{n_i}\sum_{j=1}^{n_i} I(L_{ij} > x). \tag{13}$$

We are free to allocate samples to the strata (i.e., choose the n_i's) in an essentially arbitrary manner, and the optimal allocation is well known: for a given total number of samples and equiprobable strata, the optimal n_i is proportional to the standard deviation of $I(L_{ij} > x)$.

When combining IS and stratified sampling, we can think of applying either stratified sampling or importance sampling first. Applying stratified sampling first, write

$$\begin{aligned} P\{L > x\} &= \sum_{i=1}^{k} P\{L > x | X \in \mathcal{S}_i\} P\{X \in \mathcal{S}_i\} \\ &= \sum_{i=1}^{k} \tilde{E}[I(L > x)\ell(Z) | X \in \mathcal{S}_i\} P\{X \in \mathcal{S}_i\}, \end{aligned} \tag{14}$$

i.e., X is drawn from its original distribution and then Z is drawn from the IS distribution, given X. The estimator associated with (14) is

$$\hat{P}_x = \sum_{i=1}^{k} P\{X \in \mathcal{S}_i\} \times \frac{1}{n_i} \sum_{j=1}^{n_i} I(L_{ij} > x)\ell(Z^{(ij)}). \tag{15}$$

Applying importance sampling first, write

$$P\{L > x\} = \tilde{E}[I(L > x)\ell(Z)] = \sum_{i=1}^{k} \tilde{E}[I(L > x)\ell(Z) | X \in \mathcal{S}_i\} \tilde{P}\{X \in \mathcal{S}_i\}, \tag{16}$$

where $\tilde{P}$ denotes the probability under IS. In this method, both X and Z (given X) are drawn from the IS distribution. The estimator associated with (16) is

$$\hat{P}_x = \sum_{i=1}^{k} \tilde{P}\{X \in \mathcal{S}_i\} \times \frac{1}{n_i} \sum_{j=1}^{n_i} I(L_{ij} > x)\ell(Z^{(ij)}). \tag{17}$$

To see the distinction between (14) and (16), suppose each method employs k equiprobable strata (under P and $\tilde{P}$, respectively). If $X = b'Z$ and $L \approx X$, then using (14), the mean of X is 0 and only a few strata will have positive indicators, $I(L > x)$, for a large x. If, using (16), the mean of Z is shifted from 0 to v where $b'v = x$, then the mean of X is x and approximately half the strata will have positive indicators. Hence, when we combine IS and stratification, we apply IS first, i.e., we use (16) and (17).

An efficient sampling scheme for a linear stratification variable $X = b'Z$ is described in [5]. Briefly, let $v = b/\sqrt{(b'b)}$, let Φ^{-1} denote the inverse of the standard normal distribution; then if U is uniformly distributed over the unit interval and ξ is an m-vector of independent standard normals,

$$Z = \Phi^{-1}(U)v + (I - vv')\xi$$

is also a vector of independent normals. Replacing independently sampled U's with a stratified sample (using subintervals of the unit intervals as strata) has the effect of stratifying $v'Z$ and hence also $b'Z$. See [5] for details.

4 Variance Reduction Based on the Delta Approximation

The most obvious variance reduction technique based on the delta approximation is to use the tail probability of the delta approximation itself as a control variate. Specifically, (4) suggests using $I(Y > y_x)$ as a control variate where $Y = b'Z$ and $y_x = x + \Theta\Delta t$. The mean of this control variate is easily calculated. However, as discussed earlier the effectiveness of this approach diminishes as x increases. Note also that there may be some other $y' \neq y_x$ for which the control variate $I(Y > y')$ yields more variance reduction.

Stratified sampling on $Y = b'Z$ uses more information about the delta approximation and avoids some of the difficulties involved in applying control variates (e.g., selecting the best y' and estimating the optimal control variate multiplier, which can introduce bias and even lead to negative estimates of the probability). However, unless the strata allocation $\{n_i\}$ is designed so as to assign most samples to "promising" strata, most samples will result in $L < x$.

In estimating a rare event probability such as $P\{L > x\}$, one heuristic for choosing an IS distribution is based on a large deviations analysis that states that (under appropriate technical conditions) the probability of a rare event is approximately equal to the probability of "the most likely path" to the rare event (see, e.g., [1]). With this heuristic, the IS distribution is selected so as to make this most likely path to the rare event the most likely path selected under IS. For multivariate Gaussian distributions this approach has been studied in [3] and developed for option pricing in [5, 6]. To apply this heuristic we change the means of the Gaussian random variables from 0 to $\mu = (\mu_1, \ldots, \mu_m)$ where μ is the point that maximizes the probability of the rare event. Using the delta approximation in (4), we find μ by solving the optimization problem:

$$\max \; -\tfrac{1}{2}\sum_{i=1}^{m} z_i^2 \qquad \text{such that} \qquad \sum_{i=1}^{m} b_i z_i \geq y_x \tag{18}$$

The solution to this optimization problem is $\mu = by_x/(b'b)$.

The appropriate likelihood ratio is given by (9) with $v = \mu$. The per sample second moment of this estimator is $\tilde{E}[I(L > x)\ell(Z)^2] = E[I(L > x)\ell(Z)]$, so a sufficient condition for variance reduction is $\ell(Z) \leq 1$ for all points Z such that $L > x$. The condition $\ell(Z) \leq 1$ is equivalent to $\frac{1}{2}\mu'\mu - \mu'Z \leq 0$, which, by completing the square, is equivalent to the condition that every point $Z \in \{L > x\}$ is closer to μ than it is to 0. In fact, we obtain a variance reduction by shifting the mean to any vector v (not just μ) provided each point in $\{L > x\}$ is closer to v than it is to 0. Similarly, if $\ell(Z) \leq f$ for all Z in $\{L > x\}$, then the second moment of the IS estimate is reduced by at least a factor of f.

The variance of the IS estimator could further be reduced by using a control variate, most obviously $I(Y > y_x)\ell(Z)$.

To combine IS with stratification, one could stratify upon virtually any random variable. For normal distributions, it is particularly convenient to stratify upon a linear combination of the Z's. Note that when the mean of the IS distribution is μ, then $\ell(Z) = c_1 \exp(-\mu' Z) = c_1 \exp(-c_2 b' Z)$ for some constants c_1 and c_2. This strongly suggest stratifying upon $\mu' Z$ (equivalently $b' Z$), since this simultaneously removes essentially all the variability in the likelihood ratio and much of the variability in the indicator $I(L > x)$, provided the delta approximation is close to the loss L.

Numerical Examples

We now illustrate the performance of the estimators described above. Our examples consist of portfolios of options on ten underlying assets. Even in this restricted setting there are far too many possible variations to attempt an exhaustive comparison here. Instead, we choose simple examples illustrating general principles. We keep the properties of the underlying assets particularly simple to make the effect of the portfolio structure more transparent. Thus, all ten assets have an initial value of 100 and an annual volatility of 0.30, and all pairs of distinct assets have a common correlation of ρ, with either $\rho = 0$ or $\rho = 0.2$. We consider three portfolios: (1) a portfolio short ten at-the-money calls on each underlying asset, each call having an expiry of 0.5 years; (2) a portfolio short ten calls struck at 100 and long ten calls struck at 105 (a "bear spread") on each underlying asset, each call having an expiry of 0.1 years; (3) a portfolio short ten at-the-money calls and five at-the-money puts on each underlying asset, each option having an expiry of 0.1 years. We assume 250 trading days in a year and use a continuously compounded risk-free interest rate of 5%.

Table 1 compares three methods on these portfolios. For each case we estimate losses over ten days (.04 years). Since the relevant magnitude of losses varies widely across models, we specify the loss threshold x as x_{std} standard deviations above the mean loss according to the delta-gamma approximation and vary x_{std}:

$$x = (\sum_i \lambda_i - \Theta \Delta t) + x_{std} \sqrt{\sum_i b_i^2 + 2 \sum_i \lambda_i^2}.$$

This makes it easier to compare loss thresholds across models. In the examples of Table 1 we chose values of x_{std} that would result in loss probabilities close to 1% or 5% with $\rho = 0$ and then used the same values of x_{std} with $\rho = 0.2$.

We evaluate the performance of the methods by estimating the ratio of the variance using standard Monte Carlo (no variance reduction) to the variance obtained with each method based on an equal number of samples. Thus, the larger the ratio the greater the variance reduction. Assuming roughly equal computing time per sample with and without variance reduction (which is the case in these examples), the variance ratio is a measure of computational speed-up: a method

Table 1: Comparison of variance reduction methods. Variance ratios are estimated from 120,000 replications; the stratified estimator uses 40 strata and 3000 samples per strata. Variance ratios are estimates of the computational speed-up relative to standard Monte Carlo.

				Variance Ratios		
	Portfolio	x_{std}	$P\{L > x\}$	δ-Γ CV	δ-IS	δ-IS-Strat.
$\rho = 0$	ATM Calls	1.7	4.7%	16.1	7.9	23.4
		2.5	.8%	10.2	40.0	96.5
	Spreads	1.7	4.8%	3.1	8.6	19.8
		2.4	1.1%	2.1	30.8	58.9
	Calls & Puts	1.7	5.3%	5.1	0.6	0.5
		2.5	1.3%	3.0	0.4	0.6
$\rho = 0.2$	ATM Calls	1.7	5.0%	22.7	10.0	62.3
		2.5	1.0%	10.6	38.0	220.0
	Spreads	1.7	5.3%	5.1	8.8	35.3
		2.4	1.0%	3.0	33.4	86.2
	Calls & Puts	1.7	5.9%	7.5	5.9	10.2
		2.5	2.1%	4.5	10.4	18.6

with a variance ratio of 10, say, produces as precise an estimate as standard Monte Carlo in 1/10 as much computing time.

The last three columns of the table report variance ratios using a delta-gamma control variate, using IS based on the delta approximation and combining IS with stratification in the direction determined by the delta approximation. (We could have used a delta control variate but we wanted to compare the IS methods with the best available control variate.) The first portfolio, consisting solely of calls far from expiration, has a strong linear component and, not surprisingly, all three methods result in notable variance reduction; but the combined IS-stratification method substantially outperforms the control variate, particularly at small loss probabilities. The second portfolio is far less linear because of the form of the payoff on each asset (it is neither convex nor concave) and because the options are much closer to expiration. The control variate is much less effective in this case, but the other two methods remain very effective. Losses in this portfolio occur only in one direction for each underlying asset, and the IS scheme takes advantage of this property even in the absence of linearity. This property fails to hold in the third portfolio—losses can now occur from large movements in the underlying assets in either direction. The IS scheme makes losses from the calls more likely (since they have greater weight) but in so doing makes losses from the puts rarer and thus fails to reduce variance effectively. Indeed, the rarity of losses from the puts results in a highly skewed distribution for the estimator which can therefore appear to underestimate the loss probability in small samples. These effects are less pronounced at $\rho = 0.2$, where the common correlation across assets turns out to magnify the linear term in the portfolio value.

The results in Table 1 suggest that the potential gains from the combination of IS and stratification are very substantial, but that the delta approximation alone does not provide a consistently reliable basis for implementation—particularly when movements of an underlying risk factor in more than one direction can result in portfolio losses. The delta-gamma based methods in the next section address this shortcoming.

5 Variance Reduction Based on the Delta-Gamma Approximation

We now show how the full delta-gamma approximation can be used to derive variance reduction techniques. First, (8) suggests using $I(Q > y_x)$ as a control variate, or using Q as a stratification variable. To stratify on Q, we must be able to sample Q and also sample Z given Q. Since $P\{Q \in \mathcal{S}_i\}$ can be computed numerically, one can in principle sample Q using the inverse transform method. However, we do not know of a direct method to sample Z given the quadratic form Q. (A method based on acceptance-rejection will be described elsewhere.) However, a simple method for generating Z's from the correct conditional distribution is as follows. First, generate a vector Z of independent standard normals and compute Q. If $Q \in \mathcal{S}_i$, then this Z has the distribution of Z given $Q \in \mathcal{S}_i$. If there are fewer than n_i samples from stratum i, then use this Z to evaluate the portfolio, otherwise discard it. Continue sampling until there are the required number of samples from each stratum. Like a rejection method, this wastes some samples. However, our experience has been that except for the most skewed $\{n_i\}$, this sampling overhead is modest (especially compared to the cost of evaluating the portfolio). Analysis of this overhead will be described elsewhere.

Stratification upon other variables is also possible. First, suppose $\lambda_1 \gg \lambda_i (i > 1)$ and $\lambda_1 > 0$. Then much of the variability in the positive part of Q is explained by Z_1^2. This suggests stratifying on Z_1^2, which has a chi-square distribution (or Z_1, which has a normal distribution). This distribution can be easily sampled using the inverse transform. If the λ_i's are all approximately equal, then much of the variability in Q is explained by $R^2 = Z_1^2 + \cdots + Z_m^2$ which suggests stratifying upon R^2 ("radial stratification"). Note that R^2 has a chi-square distribution with m degrees of freedom. To accomplish the stratified sampling, first draw R^2 using the inverse transform. Now draw independent standard normals $X_1, \ldots, X_m$ and set $Z_i = X_i\sqrt{R}/\sqrt{X_1^2 + \ldots + X_m^2}$ (see page 234 of [4]). The primary advantage of stratifying on R^2 rather than Q is this direct method of sampling R^2 and then Z given R^2. However, stratifying on Q is typically more effective than stratifying on R^2.

We can use the delta-gamma approximation in a variety of ways to select an IS distribution. First, as in Section 4, we could apply the "most likely path" approach in which only the mean vector is changed. In this approach, we would solve the same optimization problem as in (18), except that the constraint is $\sum(b_i z_i + \lambda_i z_i^2) \geq y_x$ which is derived from the quadratic approximation rather than the linear (delta) approximation. The optimal mean vector now has the form $\mu_i = \beta b_i/(1-2\beta\lambda_i)$ for some normalization constant β. However, this approach will suffer from the same problem as that described in Section 4: poor results will be obtained unless the event $\{L > x\}$ is dominated by the point μ.

We now examine IS changes of measure in which both the mean and the covariance matrix of Z are changed. We restrict ourselves to a particular form of the mean and covariance matrix, which are arrived at by considering "exponential twisting" of the quadratic form Q. Such exponential twisting arises frequently in the study of rare events and associated IS procedures (see, e.g., [1]). Let θ be a twisting parameter and define

$$B_\theta = (I - 2\theta\Lambda)^{-1} \qquad \text{and} \qquad \mu_\theta = \theta B_\theta b. \tag{19}$$

When IS is done by setting the mean vector to μ_θ and the covariance matrix to B_θ, then the likelihood ratio of (9) simplifies to

$$\ell(Z) = \exp\{-\theta(b'Z + Z'\Lambda Z') + \psi(\theta)\} = \exp\{-\theta Q + \psi(\theta)\} \tag{20}$$

where

$$\psi(\theta) = \frac{1}{2}\sum_{i=1}^{m}\left(\frac{(\theta b_i)^2}{1-2\theta\lambda_i} - \log(1-2\theta\lambda_i)\right) \tag{21}$$

is the logarithm of the moment generating function of the random variable Q. Notice that when importance sampling is done this way, the only random term in the likelihood ratio is Q. How should θ be chosen? Suppose that the delta-gamma approximation is exact. Then the per sample second moment of the estimator is

$$\begin{aligned} E_\theta[\ell(Z)^2 I(Q > y_x)] &= E_\theta[\exp(-2\theta Q + 2\psi(\theta)) I(Q > y_x)] \\ &\leq \exp(-2\theta y_x + 2\psi(\theta)) \end{aligned} \tag{22}$$

where E_θ denotes expectation under IS when the twisting parameter is θ. Picking $\theta = \theta_x$ to minimize the right side of (22) yields the best possible upper bound for the second moment (although not necessarily the smallest second moment). While there is no closed form expression for θ_x, it may be easily found by solving the nonlinear equation $\psi'(\theta_x) = y_x$. (The function ψ is strictly convex with $|\psi'(\theta)| \to \infty$ as $|\theta|$ increases, ensuring that a unique solution exists and is easily found numerically.) With this value of θ_x, the mean of Q under IS is equal to y_x. Furthermore, as will be shown in a subsequent paper, selecting the twisting parameter in this fashion yields an "asymptotically optimal" IS technique as $x \to \infty$, provided the delta-gamma approximation is exact. Roughly speaking, "asymptotically optimal" means that the second moment goes to zero at twice the rate that the first moment goes to zero, which is the best possible rate since the variance is non-negative.

Although simulation would not even be required if the delta-gamma approximation were exact, this analysis indicates that the IS procedure should be very effective in practice.

Let the λ_i's be ordered so that $\lambda_1 \geq \cdots \geq \lambda_m$. To avoid unininteresting cases, we assume $\lambda_1 > 0$. Then it may be shown that $0 < \theta_x < \lambda_1/2$. Under IS, the variance of Z_i is changed from 1 to $1/(1-2\theta_x\lambda_i)$. Therefore, if $\lambda_i > 0$, the variance of Z_i is increased, but if $\lambda_i < 0$, the variance of Z_i is decreased. Thus, under IS, stochastic variables that increase the quadratic part of delta-gamma are made more variable while those that decrease the quadratic part are made less variable.

Because the likelihood ratio takes the form $\ell(Z) = \exp(-\theta_x Q + \psi(\theta_x))$, stratification on the quadratic form Q is particularly attractive. It removes essentially all of the variability due to the likelihood ratio, and much of the variability in $I(L > x)$ provided the delta-gamma approximation (essentially Q) is close to the loss L.

Combining IS with stratification on Z_1 (or Z_1^2) or with radial stratification (on R^2) is also possible, although these methods are usually not as effective as stratification on Q. IS can also be combined with stratification on a linear combination of the Z's, most obviously upon $b'Z$.

Numerical Examples

We test the performance of some of the methods described above on a variety of portfolios. As in the previous examples, we mainly consider portfolios of options on ten underlying assets; the assets all have annual volatilities of 0.30, and in all but two cases we take them to be uncorrelated. Our experience (as well as some of the experiments in this paper) indicate that varying the covariance structure does not have a marked effect on the relative performance of the methods beyond the differences that can be observed by varying the structure of the portfolio; varying the portfolio makes the qualitative features more transparent. One of the correlated cases uses a covariance matrix of 10 international equity indices downloaded from the RiskMetrics™ web site, but the others are purely hypothetical.

Table 3 shows results for the following portfolios:

1. *0.5yr ATM*: short ten at-the-money calls and five at-the-money puts on each asset, all options having a half-year maturity;

2. *0.1yr ATM*: same as previous but with maturity of 0.10 years;

3. *Delta hedged:* same as previous but with number of puts on each asset increased to result in a delta of zero;

4. *0.25yr OTM*: short ten calls struck at 110 and ten puts struck at 90, all expiring in 0.25 years;

5. *0.25yr ITM*: same as previous but with calls struck at 90, puts at 110;

Table 2: The covariance matrix used for Portfolio 10 (rounded to three decimal places).

0.289	0.069	0.008	0.069	0.084	0.085	0.081	0.052	0.075	0.114
0.069	0.116	0.020	0.061	0.036	0.088	0.102	0.070	0.005	0.102
0.008	0.020	0.022	0.013	0.009	0.016	0.019	0.016	0.010	0.017
0.069	0.061	0.013	0.079	0.035	0.090	0.090	0.051	0.031	0.075
0.084	0.036	0.009	0.035	0.067	0.055	0.049	0.029	0.022	0.062
0.085	0.088	0.016	0.090	0.055	0.147	0.125	0.073	0.016	0.112
0.081	0.102	0.019	0.090	0.049	0.125	0.158	0.087	0.016	0.127
0.052	0.070	0.016	0.051	0.029	0.073	0.087	0.077	0.014	0.084
0.075	0.005	0.010	0.031	0.022	0.016	0.016	0.014	0.143	0.033
0.114	0.102	0.017	0.075	0.062	0.112	0.127	0.084	0.033	0.176

6. *Large* λ_1: same as "Delta hedged" but with number of calls and puts on first asset increased by a factor of 10;

7. *Linear* λ: same as "Delta hedged" but with number of calls and puts on ith asset increased by a factor of i, $i = 1, \ldots, 10$;

8. *100,* $\rho = 0.0$: short ten at-the-money calls and ten at-the-money puts on 100 underlying assets, all options expiring in 0.10 years;

9. *100,* $\rho = 0.2$: same as previous but with correlations of distinct assets set to 0.20.

10. *Index*: short fifty at-the-money calls and fifty at-the-money puts on 10 underlying assets, all options expiring in 0.5 years. We use the covariance matrix in Table 2. The initial asset prices are taken as (100, 50, 30, 100, 80, 20, 50, 200, 150, 10).

The first two portfolios are similar to one considered in Section 4. Shortening the time to expiration as in the second portfolio increases the quadratic terms relative to the linear terms. In the third portfolio, the linear terms are eliminated completely. The next two are similar to the second portfolio but with all options out-of and in-the-money, respectively. The sixth portfolio has one dominant eigenvalue and the seventh portfolio has linearly increasing eigenvalues. The eighth and ninth portfolios are designed to test the effect of increasing the number of underlying assets. As mentioned before, the last portfolio is designed to test the effect of using a real covariance matrix.

Table 3 compares five methods in estimating loss probabilities over a 10-day horizon. Their performance is indicated by the variance ratios in the last five columns of the table: "δ-Γ CV" is the delta-gamma control variate, "IS" is importance sampling as described in (19)–(22), "ISS-Lin" combines IS with stratification along the eigenvector of the quadratic approximation associated with the largest eigenvalue, "ISS-Rad" combines IS with stratification of the radius R, "ISS-Q" combines IS with stratification of Q. (All the stratification methods use equi-probable strata and an

Table 3: Comparison of variance reduction methods based on delta-gamma approximations. Variance ratios are estimated from 120,000 replications; the stratified estimator uses 40 strata and 3000 samples per strata. Variance ratios are estimates of the computational speed-up relative to standard Monte Carlo.

				Variance Ratios			
Portfolio	x_{std}	$P\{L > x\}$	δ-Γ CV	IS	ISS-Lin	ISS-Rad	ISS-Q
0.5yr ATM	1.65	5.3%	10.3	7.8	15.3	8.0	86.0
	2.5	1.0%	4.7	29.5	52.1	29.9	271.0
	2.8	0.5%	4.1	54.1	94.0	56.6	454.0
0.1yr ATM	1.75	5.0%	5.3	7.3	8.8	7.6	30.0
	2.6	1.1%	2.5	21.9	25.7	22.1	69.9
	3.3	0.3%	1.5	27.1	29.2	29.4	173.0
Delta hedged	1.9	4.7%	3.0	6.0	6.8	13.3	13.8
	2.8	1.1%	1.9	17.6	17.8	30.6	30.3
	3.2	0.5%	1.9	28.5	29.3	45.4	48.1
0.25yr OTM	2.7	1.1%	2.5	23.0	23.0	24.2	60.2
0.25yr ITM	2.7	1.1%	2.5	23.0	23.0	24.2	60.3
Large λ_1	3.5	1.2%	2.5	9.6	28.1	9.8	22.8
Linear λ	3.0	1.0%	2.1	17.3	18.3	17.9	29.2
100, $\rho = 0.0$	2.5	1.0%	2.6	26.9	27.7	43.5	45.4
100, $\rho = 0.2$	2.5	1.0%	2.0	10.3	20.5	10.3	23.4
Index	3.2	1.1%	3.1	18.3	54.0	18.3	119.0

equal number of samples per stratum.) Stratifying R is equivalent to stratifying Q when $b = 0$ and the λ_i are all equal (e.g., Portfolio 3); otherwise, it is a simple but potentially crude approximation to stratifying Q.

The results in Table 3 suggest some consistent patterns. At a loss probability near 5% the first four methods give broadly similar improvements, but as the loss probability decreases the effectiveness of the importance sampling and stratified sampling methods can increase dramatically whereas the control variate becomes less effective. In the presence of a strong linear component (especially the first portfolio) or a dominant eigenvalue (the sixth portfolio), linear stratification can produce substantial improvement beyond IS. When the quadratic terms dominate and are symmetric (as in the third and eighth portfolios), radial stratification provides substantial benefit. In the absence of symmetry in the quadratic terms (due to variations in the number of options in portfolios six and seven and due to correlation in the ninth portfolio), radial stratification is much less effective. Stratifying the approximation Q gives consistently impressive results and appears to be the best method overall, achieving the best variance reduction in all but one case. Finally, the eighth and ninth portfolios suggest that increasing the number of underlying assets or risk factors does not in itself entail a loss of effectiveness of the methods.

6 Effective Allocation of Samples to Strata

Suppose there are k strata and a limit n on the total number of samples that can be drawn ($n = n_1 + \cdots + n_k$). As is well known, there is an easily derived optimal allocation of samples to strata: for equiprobable strata, the optimal n_i is proportional to σ_i, the standard deviation of a stratum i sample. Whereas our earlier experiments used a uniform allocation (i.e., equal n_i's), in this section we briefly explore what further gains can be obtained by using an optimal, or near optimal allocation $\{n_i\}$. We limit this study to two methods: stratification on Q with and without the asymptotically optimal IS (i.e., the mean and covariance are given by (19) with $\theta = \theta_x$). To estimate the appropriate σ_i's, pilot studies are performed (with and without IS, respectively). Then the portfolio is simulated using the estimated optimal allocations.

We experimented with these methods on Portfolio 10 of the last section. The pilot runs used 100 samples for each of the 40 strata. From these we estimated optimal allocations and then used these allocations for a full run of 120,000 samples for each case, as in the previous numerical results. Using the estimated optimal allocations without IS reduced variance by a factor of 70, compared with a factor of 119 reported for ISS-Q in Table 3. But using the estimated optimal allocations with IS reduced variance by a factor of 1397, about 12 times greater that that achieved with ISS-Q.

7 Summary

We have proposed a variety of ways of using delta-gamma approximations as a basis for variance reduction in Monte Carlo estimation of portfolio loss probabilities. The simplest of these methods uses the delta-gamma approximation as a control variate. Numerical results suggest, however, that far greater variance reduction can be achieved by using the delta-gamma approximation with importance sampling and stratified sampling. An exponential change of measure based on the delta-gamma approximation together with stratified sampling of the approximation appears to be especially effective.

References

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

[2] Juan Cárdenas, Emmanuel Fruchard, Jean-François Picron, Cecilia Reyes, Kristen Walters, and Weiming Yang. Monte Carlo within a day. *Risk*, 12 (Feb):55 – 59, February 1999.

[3] J.-C. Chen, D. Lu, J.S. Sadowsky, and K.Yao. On importance sampling in digital communications - part i: Fundamentals. *IEEE J. Selected Areas in Communications*, 11:289 – 299, 1993.

[4] George S. Fishman. *Monte Carlo Concepts, Algorithms, and Applications.* Springer, New York, 1996.

[5] Paul Glasserman, Philip Heidelberger, and Perwez Shahabuddin. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Mathematical Finance*, 9:117–152, 1999.

[6] Paul Glasserman, Philip Heidelberger, and Perwez Shahabuddin. Importance sampling in the Heath-Jarrow-Morton framework. Technical report, IBM Research Report RC 21367, Yorktown Heights, NY, 1999.

[7] Paul Glasserman and Y. Wang. Counterexamples in importance sampling for rare event probabilities. *The Annals of Applied Probability*, 7:731–746, 1997.

[8] J.M. Hammersley and D.C. Handscomb. *Monte Carlo Methods.* Methuen & Co., Ltd., London, 1964.

[9] Philip Heidelberger. Fast simulation of rare events in queueing and reliability models. *ACM Transacations on Modeling and Computer Simulation*, 5:43 –85, 1995.

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

[11] Philippe Jorion. *Value at Risk: The New Benchmark for Controlling Derivatives Risk.* McGraw-Hill, New York, 1997.

[12] Christophe Rouvinez. Going Greek with VAR. *Risk*, 10 (Feb):57 – 65, February 1997.

[13] Thomas Wilson. Value at risk. In C.O. Alexander, editor, *Risk Management and Analysis, Vol 1*, pages 61 – 124. Wiley, Chichester, England, 1999.