Efficient Monte Carlo methods for value-at-risk

by Paul Glasserman, Philip Heidelberger and Perwez Shahabuddin

The calculation of value-at-risk (VAR) for large portfolios of complex derivative securities presents a tradeoff between speed and accuracy. The fastest methods rely on simplifying assumptions about changes in underlying risk factors and about how a portfolio's value responds to these changes in the risk factors. Greater realism in measuring changes in portfolio value generally comes at the price of much longer computing times.

The simplest methods – the "variance-covariance" solution popularized by RiskMetrics, and the delta-gamma approximations described by Britten-Jones and Schaefer (1999), Rouvinez (1997) and Wilson 1999) – rely on the assumption that a portfolio's value changes linearly or quadratically with changes in market risk factors. These assumptions limit their accuracy. In contrast, Monte Carlo simulation is applicable with virtually any model of changes in risk factors and any mechanism for determining a portfolio's value in each market scenario. But revaluing a portfolio in each scenario can present a substantial computational burden, and this motivates research into ways of improving the efficiency of Monte Carlo methods for VAR.

Because the computational bottleneck in Monte Carlo estimation of VAR lies in revaluing a portfolio in each market scenario sampled, accelerating Monte Carlo requires either speeding up each revaluation or sampling fewer scenarios. In this article, we discuss methods for reducing the number of revaluations required through strategic sampling of scenarios. In particular, we review methods developed in Glasserman, Heidelberger, and Shahabuddin 2000ab – henceforth referred to as GHS2000a and GHS2000b – that combine *importance sampling* and *stratified sampling* to generate changes in risk factors.

This approach uses the delta-gamma approximation to guide the sampling of market scenarios. Deltas and gammas are routinely calculated for other purposes so we assume their availability, without additional computational overhead, as inputs to the calculation of VAR. We develop sampling methods that are, in a precise sense, close to optimal when the delta-gamma approximation holds exactly. These methods remain attractive so long as the delta-gamma approximation contains useful information about changes in portfolio value, even if the approximation is not accurate enough to replace simulation entirely. Numerical examples indicate that the methods can often reduce by a factor of 20–100 or more the number of scenarios required to achieve a specified precision in estimating a loss probability. Because this means that the number of portfolio revaluations is also reduced by a factor of 20-100 or more, it results in a very large reduction in the computing time required for Monte Carlo estimation of VAR.

The rest of this article is organized as follows. The next section provides some background on Monte Carlo for VAR and on the delta-gamma approximation. After that, we discuss importance sampling and stratified sampling based on the delta-gamma approximation. We then discuss the application of these methods when

Mastering Risk: Application, **3rd ed., ed. Carlo Alexander, vol. 2 (New York: Financial Times/Prentice Hall), 2001.**

volatility is included among the risk factors and portfolio "vegas" are available along with deltas and gammas. Numerical examples are included to illustrate the methods. Throughout this article we assume that changes in risk factors are normally distributed. In Glasserman, Heidelberger, and Shahabuddin 2000c, we develop related methods that apply when changes in risk factors are modeled by heavy-tailed distributions.

Background on Monte Carlo and delta-gamma

Before discussing the new methods developed in GHS2000a and GHS2000b, we briefly review basic Monte Carlo estimation of VAR and the delta-gamma approximation. To give a precise formulation of the problem, we let

S = vector of risk factors
Δt = VAR horizon, (e.g., one day or two weeks)
ΔS = change in risk factors over Δt
L = loss in portfolio value resulting from change ΔS over Δt

The loss L is the difference between the current value of the portfolio and the portfolio value at the end of the VAR horizon Δt if the risk factors move from S to $S + \Delta S$.

There are two closely related problems associated with the tail of the distribution of L. The first is the problem of estimating a loss probability $P(L > x)$ given a loss threshold x. The second is the inverse problem of finding a quantile x_p for which $P(L > x_p) = p$, given a probability p. The estimation of VAR is an instance of the second problem, typically with $p = 1\%$ or 5%. However, calculating loss probabilities is a prerequisite to calculating quantiles so we focus primarily on the first problem. Given values of $P(L > x)$ for several values of x in the vicinity of x_p it is then straightforward to estimate the quantile itself.

Basic Monte Carlo for VAR

The main steps in a basic Monte Carlo approach to estimating loss probabilities are as follows:

1. Generate N scenarios by sampling changes in risk factors $\Delta S^{(1)},\ldots,\Delta S^{(N)}$ over horizon Δt.

2. Revalue portfolio at end of horizon Δt in scenarios $S + \Delta S^{(1)},\ldots, S + \Delta S^{(N)}$; determine losses $L^{(1)},\ldots,L^{(N)}$ by subtracting revaluation in each scenario from current portfolio value.

3. Calculate fraction of scenarios in which losses exceed x: $\mathrm{N}^{-1} \sum_{i=1}^{N} I(L^{(i)} > x)$, where $I(L^{(i)} > x) = 1$ if $L^{(i)} > x$ and 0 otherwise.

To estimate VAR, the last step can be repeated for multiple values of x; the required quantiles can then be estimated by, for example, interpolating between the estimated loss probabilities.

The first step requires some assumptions about market data. In *historical simulation*, the $\Delta S^{(i)}$ are the changes observed (or are obtained from the *percentage* changes observed) in market data over N past periods of length Δt. This implicitly assumes that future changes in risk factors will look like samples from past changes. Alternatively, a statistical model uses historical data to select a distribution with estimated parameters to describe future changes. A simple and widely used assumption is that, conditional on past data, the change ΔS over a short horizon Δt is described by a multivariate normal distribution $N(0,\Sigma_S)$. The conditional covariance matrix Σ_S is commonly estimated from past changes (or returns) using a sample covariance matrix, using an exponentially weighted moving average, or using a GARCH forecast – see Alexander (1998) or Jorion (1997) for a discussion of this issue. We will focus primarily on the case of normally distributed changes in risk factors, but touch on alternative models in our concluding remarks.

Given a covariance matrix Σ_S and the assumption of normally distributed changes in risk factors, it is a simple matter to generate the samples of Δ_S required in the simulation above. We factor the covariance matrix to find a matrix C for which $CC' = \Sigma_S$ (the prime denoting transpose) and then set

$$\Delta S = CZ, \tag{1}$$

where Z is a vector of independent, standard (i.e., mean 0, variance 1) normal random variables. For example, assuming Σ_S is positive definite, Cholesky factorization produces the unique lower triangular matrix C for which $CC' = \Sigma_S$.

The only difficult step in the Monte Carlo algorithm above is the second one – revaluing the portfolio in each scenario. For a large portfolio of complex derivative securities, each revaluation may be very time-consuming, with individual instruments requiring execution of numerical pricing routines or even separate Monte Carlo pricing estimates. The time required to revalue a portfolio is the limiting factor in determining the number of scenarios that can be generated.

The delta-gamma approximation

An alternative to full portfolio revaluation is to use an approximation to how changes in risk factors determine changes in portfolio value. Assuming a linear relation between risk factors and portfolio value leads to the "variance-covariance" method associated with RiskMetrics; assuming a quadratic relation leads to the delta-gamma approximation. In both cases, the approximation makes it possible to find the loss distribution numerically, without Monte Carlo simulation.

The delta-gamma approximation assumes the availability of (i) the vector δ of first partial derivatives of portfolio value with respect to the components of the vector S of risk factors, (ii) the matrix γ of the corresponding second partial derivatives, and (iii) a scalar θ giving the partial derivative of portfolio value with respect to time. From these we obtain the Taylor approximation

$$L \approx a_0 - \delta'\Delta S - \tfrac{1}{2}\,\Delta S'\Gamma\Delta S,$$

where $a_0 = -\Theta\Delta t$. The derivatives appear with minus signs in this approximation because the loss L is the negative of the increase in portfolio value.

Through a change of variables and some matrix algebra, we can rewrite this approximation in the form

$$\begin{aligned} L &\approx a_0 + b'Z + Z'\Lambda Z \\ &\equiv a_0 + Q, \end{aligned} \tag{2}$$

where Z is a vector of independent standard normal random variables and Λ is a diagonal matrix,

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{pmatrix},$$

with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ the eigenvalues of $\frac{1}{2}\quad\Gamma\Sigma_S$. This is accomplished by choosing C in (1) to satisfy

$$CC' = \Sigma_S \text{ and } -\tfrac{1}{2}C'\Gamma C = \Lambda. \tag{3}$$

(Calculation of C will be discussed later.) The vector b in the linear term of (2) is then given by $b' = -\delta'C$.

This transformation accomplishes two important simplifications: it replaces the correlated changes in risk factors ΔS with the uncorrelated elements of Z, and it diagonalizes the quadratic term in the approximation. The vector ΔS is recovered from Z through (1), so one may think of the elements of Z as (hypothetical) primitive underlying risk factors driving the market changes ΔS. Notice that the diagonal matrix Λ captures information about both the portfolio (through Γ) and the distribution of risk factors (through Σ_S).

With these simplifications it becomes relatively straightforward to find the characteristic function (Fourier transform) of the delta-gamma approximation – more precisely, of the quadratic Q in (2). Define

$$\psi(\theta) = \sum_{i=1}^{m} \tfrac{1}{2}\left(\frac{(\theta b_i)^2}{1 - 2\theta\lambda_i} - \log(1 - 2\theta\lambda_i) \right); \tag{4}$$

then $E[\exp(\sqrt{-1}\omega Q)] = \exp(\psi(\sqrt{-1}\omega))$. Transform inversion can now be used to calculate values of the distribution $P(Q < x)$. In light of (2), the loss distribution can be approximated using $P(L < x) \approx P(Q < x - a_0)$.

Importance sampling based on the delta-gamma approximation

The main virtue of the delta-gamma approximation is that it can be computed quickly. However, the accuracy of the approximation may not always be satisfactory. Monte Carlo simulation is more accurate but much more time-consuming. Our objective is to use the information contained in the delta-gamma approximation to accelerate Monte Carlo simulation and thus exploit the best features of two methods.

The simplest way to use the delta-gamma approximation in a simulation is to implement it as a *control variate* In estimating a loss probability $P(L > x)$, this produces an estimator of the form

$$\tfrac{1}{N} \sum_{i=1}^{N} I(L^{(i)} > x) - \beta\left[\tfrac{1}{N} \sum_{i=1}^{N} I(Q^{(i)} > x - a_0) - P(Q > x - a_0) \right].$$

Here, the $L^{(i)}$ are actual losses calculated in the N simulated scenarios and the $Q^{(i)}$ are the quadratic approximations (see (2)) computed *in the same scenarios*. The true probability $P(Q > x - a_0)$ is computed through transform inversion. The term in square brackets is thus the observed simulation error in the delta-gamma approximation; this observed error is used to adjust the simulation estimate of the true portfolio loss. The coefficient β can be chosen to try to minimize the variance of the combined estimator. Fixing β at 1 should yield most of the benefit of the control variate and avoids issues that arise in estimating an optimal β.

This method was proposed independently in Cardenas et al. (1999) and GHS2000a. It can provide reasonable variance reduction in some examples; but as observed in GHS2000a, its effectiveness diminishes at larger loss thresholds x. Notice that the control variate method uses the delta-gamma approximation to adjust the standard estimator "after the fact" – in particular, the scenarios used are generated in the usual way (i.e., as in our discussion above of basic Monte Carlo). In contrast, the method we describe next uses the delta-gamma approximation *before* any scenarios are generated; it uses the approximation to guide the sampling of scenarios.

Importance sampling: preliminaries

Through (1), the problem of sampling changes ΔS in market risk factors is transformed into a problem of sampling the vector Z of underlying normal random variables. In importance sampling (IS), we change the distribution from which underlying variables are generated in order to generate more samples from "important" regions. We will focus on IS methods that change the distribution of Z from $N(0,I)$ (the standard multivariate normal) to $N(\mu,\Sigma)$ (the multivariate normal with mean vector μ and covariance matrix Σ).

The key identity we need for importance sampling is

$$P(L > x) = \mathrm{E}_{\mu,\Sigma}[l(Z)I(L > x)]. \tag{5}$$

In subscripting the expression on the right by μ and Σ, we are indicating that the expectation is taken with Z sampled from $N(\mu,\Sigma)$ rather than its original distribution $N(0,I)$. To correct for this change of distribution, we must weight the loss indicator $I(L > x)$ by the *likelihood ratio*

$$l(Z) = |\Sigma|^{1/2} e^{-\frac{1}{2}\mu'\Sigma^{-1}\mu_e - \frac{1}{2}[Z'(I-\Sigma^{-1})Z - 2\mu'\Sigma^{-1}Z]} \tag{6}$$

which is simply the ratio of the $N(0,I)$ and $N(\mu,\Sigma)$ densities evaluated at Z. On both sides of (5), the loss L is computed from the market changes ΔS which are in turn calculated from Z through (1). Through (5) we are free to sample Z from any $N(\mu,\Sigma)$ and still obtain an unbiased estimate

$$l(Z)I(L > x) \tag{7}$$

of the loss probability.

How should μ and Σ be chosen to produce an estimator with lower variance (and thus greater precision)? Since changing μ and Σ does not change the resulting expectation, comparing variances is equivalent to comparing second moments. The second moment of (7) is

$$E_{\mu,\Sigma}\left[(l(Z)I(L > x))^2\right] = E\left[l(Z)I(L > x)\right], \tag{8}$$

the expectation on the right taken with respect to the original $N(0,I)$ distribution. protect From this we see that the key to reducing variance is *making the likelihood ratio small when* $L > x$. Equivalently, we would like to choose μ and Σ to make scenarios with $L > x$ more likely under $N(\mu,\Sigma)$ than under $N(0,I)$.

Using the approximation

Unfortunately, the expression in (6) provides little insight into what choice of μ and Σ might accomplish this objective. However, we can use the delta-gamma approximation to get a sense for which scenarios tend to produce large losses and use this information in the selection of μ and σ.

We can write (2) more explicitly as

$$L \approx a_0 + \sum_i b_i Z_i + \sum_i \lambda_i Z_i^2$$

and now ask, what values of Z will tend to make the (approximate) loss expression large? Inspection of this formula suggests that large losses result from

- large positive values of Z_i for those i with $b_i > 0$;
- large negative values of Z_i for those i with $b_i < 0$;
- large values of Z_i^2 for those i with $\lambda_i > 0$.

This describes the regions that should be given greater probability under the IS distribution than under the original distribution. It suggests that we should

- increase the mean of Z_i for those i with $b_i > 0$;
- decrease the mean of Z_i for those i with $b_i < 0$;
- increase the variance of Z_i for those i with $\lambda_i > 0$;

and perhaps

- decrease the variance of Z_i for those i with $\lambda_i < 0$.

We accomplish this in two steps. We first reduce the choice of μ and Σ to the choice of a scalar parameter θ, and then specify the value of this parameter. For any $\theta > 0$ (and $\theta < 1/(2\lambda_1)$ if $\lambda_1 > 0$)

$$\Sigma(\theta) = (I - 2\theta\lambda)^{-1}, \qquad \mu(\theta) = \theta\Sigma(\mu\theta)\, b. \tag{9}$$

With these parameters, Z_i becomes normal with mean and variance

$$\mu_i(\theta) = \frac{\theta\, b_i}{1 - 2\theta\lambda_i}, \quad \sigma_i^2(\theta) = \frac{1}{1 - 2\theta\lambda_i}, \tag{10}$$

and the Z_i remain independent of each other. Note that with this type of IS, the sampling distribution of Z_i is as suggested; for example, if $\lambda_i > 0$, then the variance

of Z_i is increased, resulting in more samples with large values of Z_i^2. The key observation is that with this change of distribution the likelihood ratio (6) collapses to

$$l(Z) = e^{-\theta Q + \psi(\theta)}. \tag{11}$$

Here, ψ is precisely the function introduced in (4) and may be interpreted as a normalization constant. The remarkable feature of this expression is that the likelihood ratio – which in general could depend on the entire vector Z, as in (6) – now has the scalar Q as its only stochastic element. The estimator associated with this IS distribution is

$$e^{-\theta Q + \psi(\theta)} I(L > x),$$

where the Z used to compute L and Q is now generated using (10). It must be stressed that this estimator is unbiased (in light of (5)) for the exact loss probability $P(L > x)$, even though it involves the delta-gamma approximation.

Recall from the discussion surrounding (8) that an effective importance sampling distribution makes the likelihood ratio small in those scenarios for which $L > x$. Based on (2), we can expect that when $L > x$ we will often have $Q > x - a_0$; in particular, Q will typically be large when L is and in this case the likelihood ratio (11) will indeed tend to be small when $L > x$.

It remains to specify the parameter θ. A consequence of the specification in (9) is that

$$\frac{d}{d\theta} \psi(\theta) = E_{\mu(\theta),\Sigma(\theta)}[Q]. \tag{12}$$

(In statistical terminology, (9) defines an *exponential family* of distributions with *cumulant generating function* ψ; (12) is a special case of a standard property of exponential families.) We may paraphrase (12) as stating that the derivative of ψ at θ gives the expected delta-gamma approximate loss when Z is drawn from $N(\mu(\theta), \Sigma(\theta))$. Since our objective is to estimate $P(L > x) \approx P(Q > x - a_0)$, we choose θ to be θ_x, the solution to

$$\frac{d}{d\theta} \psi(\theta_x) = E_{\mu(\theta x),\Sigma(\theta x)}[Q] = x - a_0.$$

If we sample Z from $N(\mu(\theta_x), \Sigma(\theta_x))$, scenarios in which $L > x$, which were previously rare, should now be "typical," since the expected value of the approximate loss $a_0 + Q$ is now x.

This choice of parameter θ is shown in GHS2000b to minimize an upper bound on the second moment of the estimator, providing further support for the approach. In addition, both experimental and theoretical results in GHS2000b indicate that the effectiveness of the IS procedure is not very sensitive to the choice of θ. Consequently, we may use a single IS distribution $N(\mu(\theta), \Sigma(\theta))$ to estimate the loss probability $P(L > x)$ for multiple levels of x.

The procedure

We now summarize the importance sampling procedure. We assume the availability

of the portfolio delta vector (δ) and gamma matrix (Γ), which would also be required for the delta-gamma approximation.

1. Compute C satisfying (3):
   (a) Find any matrix A satisfying $AA' = \sum_S$ (e.g., the Cholesky factor).
   (b) Find V, an orthogonal matrix ($VV' = I$) whose columns are eigenvectors of $-\frac{1}{2}A'\Gamma A$ and Λ, a diagonal matrix of associated eigenvalues (so $-\frac{1}{2}A'\Gamma A = V\Lambda V'$).
   (c) Set $C = AV$ and $b = -\delta'C$.
2. Set $\theta = \theta_x$, the solution to (13).
3. Set $\sum(\theta) = (I - 2\theta\lambda)^{-1}$ and $\mu(\theta) = \theta\sum(\theta)b$.
4. Simulate:
   (a) Generate $Z^{(1)},\ldots, Z^{(N)}$ independently from $N(\mu(\theta), \sum(\theta))$.
   (b) Set $\Delta S^{(i)} = CZ^{(i)}$, $i = 1,\ldots,N$.
   (c) Calculate portfolio losses $L(i)$ resulting from scenarios $\Delta S^{(i)}$, $i = 1,\ldots,N$.
   (d) Calculate $Q^{(i)}$ for each $Z^{(i)}$, $i = 1,\ldots, N$, as in (2).
   (e) Return estimate

$$\frac{1}{N}\sum_{i=1}^{N} e^{-\theta Q^{(i)} + \psi(\theta)} I(L^{(i)} > x). \tag{14}$$

An important feature of this method is that it can be wrapped around an existing implementation of Monte Carlo. The core of the algorithm – the calculation of portfolio losses in each scenario – is exactly the same here as in the basic Monte Carlo method presented earlier in this article. After some preprocessing steps 1–3, the importance sampling algorithm differs only in how it generates scenarios and in how it weights scenarios in (14). As with the basic Monte Carlo method, (14) could easily be calculated for multiple values of the loss threshold x, all based on a single value of θ. If we plan to estimate loss probabilities at large thresholds $x_1 < x_2 < \ldots < x_k$, we would probably fix θ at θ_{x1}.

A theoretical analysis of this IS method is reported in GHS2000b. We show there that the method is provably effective, in the sense of substantially reducing variance, as either the loss threshold or the number of risk factors increase. These results are established under the hypothesis that the relation $L = a_0 + Q$ holds exactly rather than merely as an approximation. We interpret these results as evidence that the method should remain effective whenever $a_0 + Q$ provides a reasonable approximation to L, even if it is not sufficiently accurate to replace simulation altogether. The importance of reducing variance in the simulation estimate is that it reduces the number of scenarios required to achieve a desired precision. This can result in substantial reductions in computing times, because revaluing a portfolio in each scenario is the most time-consuming step in estimating loss probabilities through Monte Carlo.

Stratified sampling

Inspection of (14) suggests that to further reduce variance we should reduce variability in the sampling of the quadratic approximation Q. Indeed, if we had $L = a_0 + Q$, then eliminating the variance due to Q would eliminate all the variance in (14). If $a_0 + Q$ only approximates L, reducing the variability from Q should nevertheless result in further overall variance reduction.

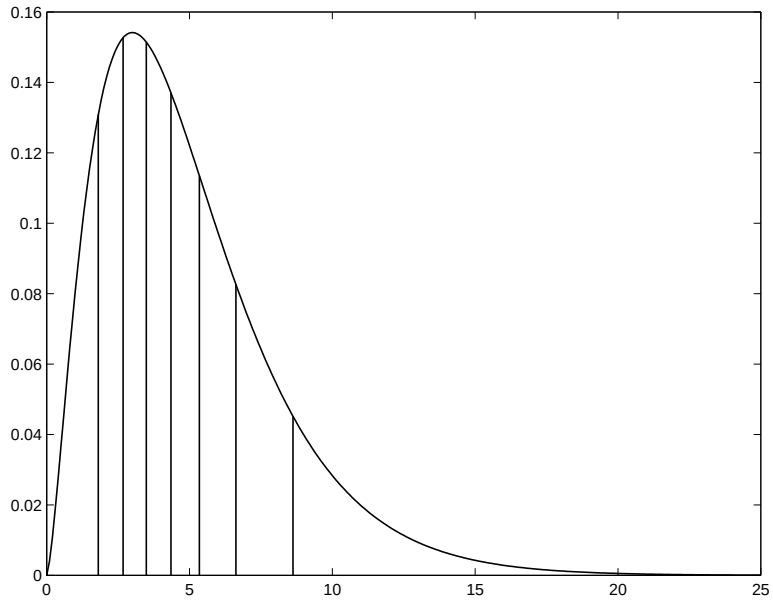

Figure 1: Illustration of equiprobable strata

We implement this idea through *stratified* sampling of Q. This mechanism is best explained through reference to Figure 1. The figure shows a hypothetical density for Q. (It is in fact the chi-square density with five degrees of freedom and thus a special case of the density of Q in (2).) More precisely, this should be interpreted as the density of Q under the importance sampling distribution, which is to say with Z drawn from $N(\mu(\theta), \Sigma(\theta))$. The $Q^{(i)}$ used in the algorithm above are independent samples from the density of Q under the IS distribution.

In stratified sampling, rather than drawing the $Q^{(i)}$ randomly and independently we ensure that fixed fractions of the samples fall within specified ranges. For example, the vertical lines in Figure 1 define eight equiprobable bins or strata: the area under the curve between each consecutive pair of lines is 1/8. If we generate samples $Q^{(i)}$ independently, we cannot expect that exactly 1/8th of the samples will fall in each of the strata; because the sampling mechanism is random, some strata will end up with too many samples, some with too few. In contrast, using stratified sampling we ensure that exactly 1/8th of the generated samples do indeed fall in each of the strata. In practice, we typically use 40 equiprobable strata and ensure that 1/40th of the samples fall within each stratum. With 40 strata, much of the variance in the estimated loss probability due to sampling variability in Q is eliminated.

The first step in implementing this method is to define the strata. In order to define k equiprobable strata, we need to find points $y_1,..., y_{k-1}$ such that

$$P_\theta(\leq y_i) = i/k, \quad i = 1,..., k-1.$$

We have subscripted the probability by θ to emphasize that this should hold under the IS distribution. Under the IS distribution, Q remains a quadratic function in

normal random variables, so the transform analysis outlined in our discussion of the delta-gamma approximation (after (4)) is still applicable. Using this method we can solve for the required y_i. The intervals $(y_i, y_i + 1)$, $i = 0,..., k - 1$, $(y_0 \equiv -\infty, y_k \equiv \infty)$ then form k equiprobable bins. As discussed in GHS2000b, one could just as easily define strata with any other fixed set of probabilities, but here we focus on the case of equal probabilities for simplicity.

Having defined the strata, it remains to define a sampling mechanism under which an equal fraction of the $Q^{(i)}$ generated fall in each stratum. For this we use a simple if somewhat crude approach. Suppose we want to generate n samples from each stratum for a total sample size of nk. We generate a large number of independent samples Z from $N(\mu(\theta), \Sigma(\theta))$; for each Z generated we evaluate Q and check which stratum it falls in; if we have not already generated n samples for that stratum, we keep the Z generated, otherwise we discard it. We repeat this procedure until we have the required number of samples for each stratum.

Let $Q^{(ij)}$ denote the jth sample from stratum i and let $Z^{(ij)}$ denote the draw from $N(\mu(\theta), \Sigma(\theta))$ that produced this sample. From $Z^{(ij)}$ we get $\Delta S^{(ij)} = CZ^{(ij)}$ as before and compute the corresponding portfolio loss $L^{(ij)}$. The resulting estimator is

$$\sum_{i=1}^{k} \frac{1}{nk} \sum_{j=1}^{n} e^{-\theta Q^{(ij)} + \psi(\theta)} I(L^{(ij)} > x).$$

A bit more generally, if we define strata with probabilities $p_1,..., p_k$ and allocate n_i samples to stratum i, $i = 1,..., k$, the estimator is

$$\sum_{i=1}^{k} \frac{p_i}{n_i} \sum_{j=1}^{n_i} e^{-\theta Q^{(ij)} + \psi(\theta)} I(L^{(ij)} > x).$$

This does not require that the allocations ni be proportional to the stratum probabilities p_i. Various strategies for choosing the allocations $\{n_i\}$ are investigated in Glasserman et al. (1999). A very simple form of stratified sampling based on the delta-gamma approximation – using just two strata, proportional allocation, and no importance sampling – was proposed independently in Cardenas et al. (1999).

Numerical illustration

Extensive numerical experiments using control variates and a variety of importance sampling and stratified sampling methods have been reported in Glasserman et al. (1999, 2000ab).

Here we reproduce one table of results from GHS2000b for illustration.

The results in Table 1 apply to test portfolios defined in GHS2000b, which should be consulted for detailed descriptions. Briefly, each of portfolios (a.1)–(a.14) consists of 150–1,000 standard calls and puts distributed over 10 underlying assets; (a.15) has 20 options on each of 100 underlying assets. The options in (a.1)–(a.3) have expirations of 0.5 years; those in (a.4)–(a.6) have expirations of 0.1 years and thus comparatively larger gammas. Portfolios (a.7)–(a.10) are delta hedged. The underlying assets in (a.11)–(a.15) are correlated whereas those in (a.1)–(a.10) are not. All results are based on a VAR horizon Δt of 10 days.

The second column of Table 1 specifies the loss threshold x as x_{std} standard deviations of Q above the mean of $a_0 + Q$. The associated portfolio loss probabilities (all close to 1 percent) are indicated in the third column.

The last two columns of the table are estimates of the ratio of variances in the estimated loss probabilities using standard Monte Carlo and using importance sampling (IS) or importance sampling with stratification (ISS-Q). (The ISS-Q results use 40 equiprobable strata.) These variance ratios indicate how many times more scenarios would have to be generated using standard Monte Carlo to achieve the same precision obtained with the indicated variance reduction technique. Since the bulk of the computational effort in using Monte Carlo with complex portfolios lies in revaluing the portfolio in each scenario, these variance ratios are estimates of the computational speed-up obtained through variance reduction. The results clearly indicate the potential for enormous speed-ups using the methods reviewed here.

Further results and details of the experiments can be found in GHS2000b. The only test portfolios for which we have found results substantially inferior to those in Table 1 are portfolios of digital and barrier options combined to achieve a net delta of 0. Given the nature of these portfolios, it is perhaps unsurprising that a Taylor approximation turns out to be not very informative.

Table 1: Variance reduction estimates for test portfolios

			Variance ratios	
Portfolio	$x$std	P(L > x)	IS	ISS-Q
(a.1)	2.5	1.0%	30	270
(a.2)	1.95	1.0%	43	260
(a.3)	2.3	1.0%	37	327
(a.4)	2.6	1.1%	22	70
(a.5)	1.69	1.0%	43	65
(a.6)	2.3	0.9%	34	132
(a.7)	2.8	1.1%	17	31
(a.8)	1.8	1.1%	52	124
(a.9)	2.8	1.1%	16	28
(a.10)	2.0	1.1%	19	34
(a.11)	3.2	1.1%	18	124
(a.12)	1.02	1.0%	28	48
(a.13)	2.5	1.1%	15	65
(a.14)	1.65	1.1%	14	45
(a.15)	2.65	1.0%	18	28

Including volatility as a risk factor

Thus far, we have interpreted S as a vector of market prices and rates. However, S could also include risk factors associated with levels of volatility rather than prices or rates. The methodology above continues to apply.

We develop this idea through a simple formulation of the problem. Our intent is to illustrate how volatility can be incorporated rather than to propose a specific model. We interpret some of the components of S as asset prices and some as implied volatilities for those assets. For simplicity, we do not incorporate a volatility skew or smile: we assume all options in a portfolio on the same underlying asset have the same implied volatility. In contrast to the previous setting, we now allow the level of implied volatility to change over the VAR horizon. We assume that correlations among prices, among implied volatilities, and between prices and

implied volatilities are unchanged over the VAR horizon. We impose this assumption solely for notational simplicity.

Partition the vector S as $(\widetilde{S}, \widetilde{\sigma})$ with $\widetilde{\sigma}_i$ the implied volatility of $\widetilde{S}_i$. We assume that the changes $(\Delta\widetilde{S}, \Delta\widetilde{\sigma})$ over the VAR horizon are conditionally normally distributed, given the current history of prices and implied volatilities, with a conditional mean of 0 and a known conditional covariance matrix. We assume the availability of a vector of "vegas" υ, with υ_i the partial derivative of a portfolio's value with respect to $\widetilde{\sigma}_i$. We continue to assume the availability of the usual Δ and Γ with respect to the prices $\widetilde{S}$. It seems less likely that second derivatives involving $\widetilde{\sigma}_i$ would be available as these are not routinely computed for other purposes. We therefore assume these are unavailable and arbitrarily set their values at 0. The quadratic approximation thus takes the form

$$L \approx a_0 - (\delta' \;\; \upsilon') \begin{pmatrix} \Delta\widetilde{S} \\ \Delta\widetilde{\sigma} \end{pmatrix} - \tfrac{1}{2} \;\; (\Delta\widetilde{S}' \;\; \Delta\widetilde{\sigma}') \begin{pmatrix} \Gamma & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta\widetilde{S} \\ \Delta\widetilde{\sigma} \end{pmatrix}.$$

From here on, the analysis proceeds exactly as before.

We tested this method on portfolio (a.1) (0.5 year at-the-money options), (a.4) (0.1 year at-the-money options), and (a.7) (a delta-hedged version of (a.4)). All underlying assets have an initial volatility of 0.30. We also tested the method on a new portfolio that is both delta and gamma hedge. On each of 10 underlying assets with a spot price of 100, the portfolio is short 4.5 calls struck at 110, long 4 calls struck at 105, long 2 puts struck at 95, and short 2.1 puts struck at 90, with all options expiring in 0.5 years. This combination results in deltas and gammas very close to 0.

With each portfolio we consider two levels of the volatility of volatility: 20 percent ("High") and 10 percent ("Low"). We also consider two possible cases for the correlation structure: uncorrelated, and correlated with

$$\mathrm{Corr}[\Delta\widetilde{S}_i, \Delta\widetilde{S}_j] = 0.20,\ \mathrm{Corr}[\Delta\widetilde{S}_i, \Delta\widetilde{\sigma}_i] = -0.25,\ \mathrm{Corr}[\Delta\widetilde{S}_i, \Delta\widetilde{\sigma}_j] = 0,\ \mathrm{Corr}[\Delta\widetilde{\sigma}_i, \Delta\widetilde{\sigma}_j] = 0.6.$$

The interpretation of this case is as follows. All assets are affected by a common "market level" factor inducing a positive correlation in price changes; each asset has a negative correlation with its own implied volatility (so volatility goes up when prices drop); and all implied volatilities are affected by a common "volatility level" factor inducing a positive correlation in volatility changes.

The results are summarized in Table 2. The variance ratios for portfolios (a.1), (a.4), and (a.7) are similar to what we found in the case of constant volatility. We see less variance reduction for the portfolio that is both delta and gamma hedged. For this portfolio the IS and ISS-Q methods rely entirely on vega information as we do not assume the availability of second derivatives involving volatilities. The delta-gamma-vega approximation is therefore less informative in this case than in the others.

Table 2: Variance reduction estimates with volatility as a risk factor

	Portfolio	$x$std	P(L > x)	Variance ratios IS	ISS-Q
(a.1)	Uncorrelated, High	2.5	1.0%	30	244
	Uncorrelated, Low		1.0%	30	291
	Correlated, High	2.6	1.1%	28	281
	Correlated, Low		1.2%	26	285
(a.4)	Uncorrelated, High	2.6	1.1%	23	65
	Uncorrelated, Low		1.1%	23	74
	Correlated, High	3.0	1.1%	19	84
	Correlated, Low		1.1%	19	96
(a.7)	Uncorrelated, High	2.8	1.1%	17	28
	Uncorrelated, Low		1.1%	17	29
	Correlated, High	3.2	1.0%	12	20
	Correlated, Low		1.1%	11	18
δ–Γ hedged	Uncorrelated, High	3.3	1.0%	9	8
	Uncorrelated, Low		0.9%	10	13
	Correlated, High	2.7	1.1%	14	21
	Correlated, Low		1.2%	11	19

Conclusion

The methods reviewed in this article attempt to combine the best features of two approaches to calculating VAR: the speed of the delta-gamma approximation and the accuracy of Monte Carlo simulation. We use the delta-gamma approximation not as a substitute for simulation but rather as an aid. By using the delta-gamma approximation to guide the sampling of scenarios – through a combination of importance sampling and stratified sampling – we can greatly reduce the number of scenarios needed in a simulation to achieve a specified precision.

For simplicity, in this article we have restricted attention to methods based on modeling changes in market risk factors over the VAR horizon using a normal distribution. But empirical studies consistently find that market returns exhibit greater kurtosis and heavier tails than can be captured with a normal distribution. In Glasserman, Heidelberger, and Shahabuddin 2000c, we extend the methods discussed here to certain heavy-tailed distributions, including multivariate t distributions. This setting poses interesting new theoretical questions as well as having practical relevance. Numerical results indicate that our methods are generally at least as effective in the heavy-tailed setting as in the normal case.

Summary

The calculation of value-at-risk for large portfolios presents a tradeoff between speed and accuracy, with the fastest methods relying on rough approximations and the most realistic approach – Monte Carlo simulation – often too slow to be practical. This article describes methods that use the best features of both approaches. The methods build on the delta-gamma approximation, but they use the approximation not as a substitute for simulation but rather as an aid to it. **Paul Glasserman**, **Philip Heidelberger** and **Perwez Shahabuddin** use the delta-gamma approximation to guide the sampling of market scenarios through a combination of importance sampling and stratified sampling. This can greatly reduce the number of scenarios required in a simulation to achieve a desired precision. The authors also describe an extension of the method in which "vega" terms are included in the approximation to capture changes in the level of volatility.

Suggested further reading

Alexander, C. (1998) "Volatility and correlation: methods, models, and applications" in *Risk Management and Analysis*, Vol. 1 pp. 125–172, C. Alexander, ed., Wiley, Chichester, England.

Britten-Jones, M. and Schaefer, S.M. (1999) "Non-linear value-at-risk," *European Finance Review*, 2, pp. 161–187.

Cardenas, J, Fruchard, E., Picron, J.-F., Reyes, C., Walters, K. and Yang, W. (1999) "Monte Carlo within a day," *Risk*, 12:2, pp. 55–59.

Glasserman, P., Heidelberger, P. and Shahabuddin, P. (1999) "Stratification issues in estimating value-at-risk" in *Proceedings of the 1999 Winter Simulation Conference*, pp. 351–358, IEEE Computer Society Press, Piscataway, New Jersey.

Glasserman, P., Heidelberger, P. and Shahabuddin, P. (2000a) "Importance sampling and stratification for value-at-risk" in *Computational Finance 1999 (Proceedings of the Sixth International Conference on Computational Finance)*, Y.S. Abu-Mostafa, B. LeBaron, A.W. Lo and A.S. Weigend, eds, pp. 7–24, MIT Press, Cambridge, Mass.

Glasserman, P., Heidelberger, P. and Shahabuddin, P. (2000b) "Variance reduction techniques for estimating value-at-risk," *Management Science*.

Glasserman, P., Heidelberger, P. and Shahabuddin, P. (2000c) "Portfolio value-at-risk with heavy-tailed risk factors," IBM Research Report RC 21817, Yorktown Heights, New York. Available at www.research.ibm.com/people/b/berger/papers.html.

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

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

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

Orthogonal GARCH

by Professor Carol Alexander

The univariate generalized autoregressive conditional heteroscedasticity (GARCH) models that were introduced by Engle (1982) and Bollerslev (1986) have been very successful for short term volatility forecasting in financial markets. The mathematical foundation of GARCH models compares favourably with some of the alternatives used by financial practitioners, and this mathematical coherency makes GARCH models easy to adapt to new financial applications.

There is also evidence that GARCH models generate more realistic long-term forecasts than exponentially weighted moving averages. This is because the GARCH volatility and correlation term structure forecasts will converge to the long-term average level, which may be imposed on the model, whereas the exponentially weighted moving average model forecasts average volatility to be the same for all risk horizons (see Alexander, 1998). As for short-term volatility forecasts, statistical results are mixed – for example, see Andersen and Bollerslev (1998) Alexander and Leigh (1997), Brailsford and Faff (1996), Cumby, Figlewski and Hasbrouck (1993), Dimson and Marsh (1990), Figlewski (1997), Frennberg and Hansson (1996), and