VARIANCE REDUCTION TECHNIQUES FOR ESTIMATING VALUE-AT-RISK

RMSC4102 Research Project

Long Hou Tin

ABSTRACT

Estimating value-at-risk of a portfolio is a problem of importance in risk management. This paper dedicate to proposes, implements, and evaluates a method for efficient estimation of portfolio loss probabilities using Monte Carlo simulation with variance reduction techniques. Precise estimation of such loss probabilities is critical to calculating value-at-risk. The method is built on the delta-gamma approximation to the portfolio loss, which is essential to developing effective variance reduction techniques, in particular, control variate, importance sampling and stratified sampling. Numerical results will be displayed to show that an proper combination of importance sampling and stratified sampling could result in the greatest variance reduction.

1. INTRODUCTION

Efficient simulation of value-at-risk (VAR) is of concerned in this paper. VAR, defined as the quantile of the distribution of the loss in portfolio value during a holding period of certain duration, is a critical measure for estimating and managing portfolio risk. Denoted by $V(t)$ the value of the portfolio at time t, and Δt the holding period. Consider the interval $[t, t+\Delta t]$, the portfolio value at time $(t+\Delta t)$ is $V(t+\Delta t)$, and 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 as the $(1-p)th$ quantile of the distribution of the portfolio loss during the holding period, which satisfies the relationship: $\mathrm{P}\{L > x_p\} = p$.

Usually, a holding period of one day or two weeks and $\mathrm{p} \approx 0.01$ is of interest. In practice, Monte Carlo simulation is often used to evaluate p:

$$P\{L > x\} = E\big(I(L > x)\big)$$
$$\hat{P}\{L > x\} = \sum_{i=1}^{n} [I(L_i > x)] \left(\frac{1}{n}\right) \tag{1}$$

Changes in the portfolio's risk factors during the holding period are simulated, the portfolio is re-evaluated at the new values of the risk factors. Such process is repeated for a sufficiently large number of times, and thus the loss distribution may be estimated. Notice that there are two reasons why accurate VAR estimates could be computationally expensive to obtain. First, the portfolio may consist of a large number of instruments, making each evaluation of the portfolio costly. Second, to obtain accurate estimates of the tail probability, a large number of simulation runs (portfolio evaluation) are required. The purpose of this paper is to tackle the second issue by developing variance reduction techniques that offer to reduce the number of simulation runs required to achieve accurate estimates of low probability. These techniques are developed from $[\boldsymbol{Glasserman\ et\ al. 1999a}]$. We will focus on obtaining accurate estimates of $\mathrm{P}\{L > x\} = p$ for x that is close to x_p, which is the key to reducing the variance of an estimator of the VAR x_p.

Our approach is to exploit knowledge of the distribution of the second order Tylor series expansion, or equivalently, the delta-gamma approximation to the portfolio loss to construct more effective and efficient Monte Carlo simulation schemes. Suppose the change in risk factors follow multivariate normal distribution, as is often assumed, then the distribution of the delta-gamma approximation can be computed numerically [**Rouvinez 1997**]. Delta-gamma approximation is not always accurate to provide precise VAR estimate, however, we can make use of the high correlation between such approximation and the actual loss to devise effective variance reduction techniques by using the delta-gamma proximation of the portfolio loss as a control variate, or a basis for importance sampling and stratified sampling.

The rest of the paper is organized as follows. In section 2, we develop the delta-gamma approximation to the portfolio loss, which serve as a guide to the construction of effective variance reduction techniques. Section 3 describes the portfolios that are used to evaluate the effectiveness of such variance reduction techniques. Control variate, importance sampling based on the delta-gamma approximation are introduced in section 4, and 5 respectively. In section 6, stratified sampling and its combination with importance sampling are described. Finally, results are summarized and discussed in section 7.

2. DELTA-GAMMA APPROXIMATION

This section describe the basic model and the quadratic approximation to the portfolio loss. Assume the portfolio value is governed by m risk factors (stock prices, interest rates, currency exchange rates etc.). Denoted by $\mathrm{S}_{(\mathrm{t})} = (S_{1(\mathrm{t})}, \dots, S_{m(\mathrm{t})})$ the value of these risk factors at time t, which is assumed to be known. Define $\Delta \mathrm{S}_{(t)} = \Delta \mathrm{S} = \left[S_{(t+\Delta t)} - S_{(t)}\right]^T$ to be the change in the risk factors during the interval $[t, t+\Delta t]$, which is assume to follow multivariate normal distribution with mean 0 and covariance matrix Σ, e.g., $\Delta S \sim N_m(0\ ,\Sigma)$. Notice that ΔS can be generated by $\Delta S = MZ$, for any square matrix M such that $MM^T = \Sigma$, and Z is a vector of independent standard normal random variables with mean 0 and variance 1, that is, $Z \sim N_m(0, I)$. The distribution of ΔS is analyzed in the **appendix a**.

The portfolio loss during the interval $[t, t+\Delta t]$ is $L_{(\Delta S)} = V_{(S_{(t)})} - V_{(S_{(t)}+\Delta S)}$, And the delta-gamma approximation of L is given by

$$L \approx a_0 + a^T \Delta S + \Delta S^T A \Delta S \equiv a_0 + Q,\ \begin{cases} a_0 = -(\Theta \Delta t),\ \Theta = \dfrac{\partial V}{\partial t} \\ a = -(\delta),\ \delta_i = \dfrac{\partial V}{\partial S_i},\ where\ i = 1, \dots, m \\ A = -\left(\frac{1}{2}\Gamma\right),\ \Gamma_{ij} = \dfrac{\partial^2 V}{\partial S_i \partial S_j},\ where\ i,j = 1,..,m \end{cases}$$

The approximation becomes more accurate as Δt goes to zero. The constant a_0, the vector a, and the symmetric matrix A are known prior to running the simulation, with all partial derivatives evaluated at the initial time t. The payoff function, and the derivation of the Greeks of the options involved in the numerical experiments can be found in **appendix b**. Notice that the methods we described in this paper does not apply to portfolios that are both delta hedged and gamma hedged, that is, $\delta_i = 0\ \forall i\ and\ \Gamma_{ij} = 0\ \forall i,j$.

To facilitate the construction of variance reduction techniques in the latter sections, we express Q as a diagonalized quadratic form by solving the eigen-decomposition problem

$$(\tilde{C}^T A \tilde{C}) = U \Lambda U^T, \begin{cases} \Lambda_{ii} = \lambda_i, \ \ where\ \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \\ U^T U = U U^T = I \end{cases}$$

where C is any square matrix such that $CC^T = \Sigma$, (e.g. such C can be obtained by the Cholesky decomposition of Σ). Set $C = \tilde{C}U$, then we have $CC^T = \tilde{C}UU^T\tilde{C}^T = \Sigma$, it follows that

$$Q = a^T \Delta S + \Delta S^T A \Delta S = (a^T C) Z + Z^T (C^T A C) Z = b^T Z + Z^T \Lambda Z, \begin{cases} \Delta S = CZ, \ \ where\ Z \sim N_m(0, I) \\ b^T = a^T C \end{cases}$$

Hence, $\mathrm{Q} = b^T Z + Z^T \Lambda Z = \sum_{i=1}^{m} (b_i Z_i + \lambda_i Z_i^2)$.

3. PORTFOLIOS OF INTEREST

Section 4 to 6 will be accompanied with numerical results to illustrate the performance of the corresponding estimator in term of variance ratio (e.g. CV refers to the estimated variance of the standard estimator divided by that of the estimator with control variate, IS refers to the estimated variance of the standard estimator divided by that of the estimator with importance sampling, ISSQ refers to the estimated variance of the standard estimator divided by that of the estimator with importance sampling and stratified sampling etc.). The numerical results are drawn from experiments performed on the following 3 sets of portfolios. Unless mentioned otherwise, the portfolios consist of options on 10 underlying assets that are uncorrelated, all assets having an initial value of 100 and an annual volatility of 0.3. For each case, assume 250 trading days per year and a continuously compounded risk free interest rate of 5%, we investigate losses over 10 days (consider $\Delta t = 0.04\ year$). For comparison purpose, in each case we adjust the loss x so that the probability to be estimated is close to 0.01, that is $\mathrm{P}\{L > x\} \approx 0.01$. To standardize results across portfolios, We express x as x_{std} standard deviations above the mean according to the delta-gamma approximation under the original measure $Z \sim N_m(0, I)$:

$$x = \left(\sum_{i=1}^{m} \lambda_i + a_0\right) + x_{std} \sqrt{\sum_{i=1}^{m} b_i^2 + 2 \sum_{i=1}^{m} \lambda_i^2}$$

SET A: The following is the set of portfolios we consider for investigating the effect of negative eigen values and portfolio size, and correlation among assets on the performance of the estimators.

(a.1) 0.5yr ATM, +$\boldsymbol{\lambda}$: Short 10 ATM (at-the-money) calls and 5 ATM puts on each asset, all options having a half-year maturity. All eigenvalues are positive.

(a.2) 0.5yr ATM, −$\boldsymbol{\lambda}$: Long 10 ATM calls and 5 ATM puts on each asset, all options having a half-year maturity. All eigenvalues are negative.

(a.3) 0.5yr ATM, ±$\boldsymbol{\lambda}$: Short 10 ATM calls and short 5 ATM puts on the first 5 assets. Long 10 ATM calls and short 5 ATM puts on the next 5 assets. This portfolio has both positive and negative eigenvalues.

(a.4) 0.1yr ATM, +$\boldsymbol{\lambda}$: Same as (a.1) but with a maturity of 0.10 years.

(a.5) 0.1yr ATM, −$\boldsymbol{\lambda}$: Same as (a.2) but with a maturity of 0.10 years.

(a.6) 0.1yr ATM, ±$\boldsymbol{\lambda}$: Same as (a.3) but with maturity of 0.10 years.

(a.7) Delta hedged, +$\boldsymbol{\lambda}$: Same as (a.4) but with number of puts increased so that δ=0.

(a.8) Delta hedged, −$\boldsymbol{\lambda}$: Same as (a.5) but with number of puts increased so that δ=0.

(a.9) Delta hedged, $\pm\boldsymbol{\lambda}$: Short 10 ATM calls on first 5 assets. Long 5 ATM calls on the remaining assets. Long or short puts on each asset so that δ=0.

(a.10) Delta hedged, $\boldsymbol{\lambda_m < -\lambda_1}$: Short 5 ATM calls on first 5 assets. Long 10 ATM calls on next 5 assets. Long or short puts on each asset so that δ=0.

(a.11) index, $+\boldsymbol{\lambda}$: Short 50 ATM and 50 ATM puts on 10 underlying assets, all options expiring in 0.5 years. The initial prices of the assets are taken as (100, 50, 30, 100, 80, 20, 50 ,200, 150, 10). Consider the covariance matrix of 10 international equity indices, which is given in $[Glasserman\ et\ al.\ 1999b]$ and displayed in the following table.

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

(a.12) index, $-\boldsymbol{\lambda}$: Same as (a.11), but now we long 50 ATM calls and 50 ATM puts on the 10 underlying assets.

(a.13) index, $\pm\boldsymbol{\lambda}$: Same as (a.11), but now we short 50 ATM calls and 50 ATM puts on the first 5 assets, and long 50 ATM calls and 50 ATM puts on the next 5 assets.

(a.14) index, $\boldsymbol{\lambda_m < -\lambda_1}$: Same as (a.11), but now we short 50 ATM calls and 50 ATM puts on the first three assets, and long 50 ATM calls and 50 ATM puts on the next seven assets.

(a.15) 100, Block-diagonal: Short 10 ATM calls and 10 ATM 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 if they belong to the same group and is 0 if otherwise. 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.

SET B: The follow is the set of portfolios we consider to investigate the effect of discontinuities in the payoff functions of options on the performance of the estimator. Assume all options being at-the-money and having a maturity of 0.1 year. Notice that to simplify the simulation of knock-out option, we ignore the possibility that the option being knocked out during the interval $[t, t+\Delta t]$, and set the option value to be zero if the simulated asset price at $t+\Delta t$ is blow the barrier.

(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, Delta hedged: Same as (b.3), but the number of puts is adjusted so that δ=0.

(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, Delta hedged: Same as (b.5) but the number of cash-or-nothing puts is adjusted so that δ=0.

(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.

SET C: The following is the set of portfolio we consider to investigate the effect of missing non-diagonal value of the Γ matrix on the performance of the estimator. Since in practice the full matrix Γ might not be available or simply because it is too expensive to compute. Hence, the performance of the estimator applied under two situations are considered. In setting one, the full matrix Γ is known. In setting two, only diagonal entities of Γ is available. For a given portfolio, the same x $(stated\ in\ terms\ of\ x_{srd}\ in\ setting\ 1)$ is used is both setting. For the exchange option, we consider 5 asset pairs where asset i is exchanged for asset $\mathrm{i}+5$, where $\mathrm{i}=1,\ldots,5$.

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

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

(c.3) EO, C & P,+: (c.1) combined with shorting 10 ATM (at-the-money) calls and shorting 5 ATM 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 ATM calls and shorting 5 at-the-money puts on each asset. All off-diagonal elements of Γ are negative.

(c.5) EO, C & P, ±: (c.2) combined with shorting 10 ATM calls and shorting 5 ATM puts on each asset

4. CONTROL VARIATE FOR ESTIMATION OF VAR

Consider $E[H(X)]=E\left[H(X)+c\left(K(X)-\mu_{K(X)}\right)\right]$, the random variable $K(X)$ with known mean $\mu_{K(X)}$, is the control variate. $\left[\overline{H(X)}+c\left(\overline{K(X)}-\mu_{K(X)}\right)\right]$ is an unbiased estimator of $E[H(X)]$. By taking $c^*=\left(-\sigma_{H(X)K(X)}/\sigma^2_{K(X)}\right)$, the covariance of the estimator is minimized,

$$Cov\left[\,\overline{H(X)}+(c^*)\left(\overline{K(X)}-\mu_{K(X)}\right)\right]=\left(\frac{1}{n}\right)\left(\sigma^2_{H(X)}+(c^*)^2\sigma^2_{K(X)}+2(c^*)\sigma_{H(X)K(X)}\right)=\left(\frac{\sigma^2_{H(X)}}{n}\right)(1-\rho^2)$$

where $\rho=cor\left(H(X),K(X)\right)$ is the correlation between $H(X)$ and the control variate $K(X)$. The variance of the estimator with a control variate is never worse than that of the standard estimator, which is $\left(\sigma^2_{H(X)}/n\right)$.

To estimate $\mathrm{P}\{L>x\}=E(I(L>x))$, consider $I(a_0+Q>x)$, or equivalently, $I(Q>x-a_0)$ as a control variate where (a_0+Q) is the quadratic approximation of the portfolio loss L. Specifically, now we consider

$$\overline{I(L>x)}-(c^*)\left[\,\overline{I(Q>x-a_0)}-P\{Q>x-a_0\}\right] \tag{2}$$

instead of the standard estimator. Note that L and Q are evaluated under the same price scenario, and $P\{Q>x-a_0\}$ is computed numerically. Numerical results on three portfolios are displayed in **table 1**. The effectiveness of control variate constructed using the delta-gamma approximation is not satisfactory for two reason. First, there is no significant variance reduction achieved, the variance is reduced by less than one order of magnitude. Second, the correlation between $I(L>x)$ and its control variate $I(a_0+Q>x)$ becomes weaker as p decrease, which in terms cause the effectiveness of the control variate to shrink if the VAR we investigate becomes larger. The effectiveness of such approach also depends on how accurate the delta-gamma approximation is to the portfolio loss, for example, one may expect the variance reduction achieved to be smaller when the portfolios is delta hedged $(\delta=0, thus\ a=0)$, as the delta-gamma approximation to the loss of such portfolios is less accurate. For more numerical results on VAR estimation with control variate, please refers to

table d1 in **appendix d**.

Portfolio	x_{std}	P{L>x}	CV	$\hat{\rho}$
(a.11) index, +λ	2.7	1.9%	4.8	0.89
	3.2	1.1%	4.0	0.86
	3.7	0.6%	3.3	0.84
(b.5) DAO-C & CON-P	2.25	3.1%	1.7	0.65
	2.75	1.1%	1.5	0.58
	3.25	0.3%	1.4	0.52
(c.3) setting 1 EO, C & P,+	2.2	2.6%	6.5	0.92
	2.7	1.1%	5.2	0.9
	3.2	0.4%	4.3	0.87
(c.3) setting 2 EO, C & P,+	2.2	2.6%	2.8	0.8
	2.7	1.1%	2.4	0.77
	3.2	0.4%	2.2	0.74

Table 1: Comparison of the effectiveness of contol variate as p decrease. $\hat{\rho}$ is the sample correlation between $I(L > x)$ and its control variate $I(a_0 + Q > x)$

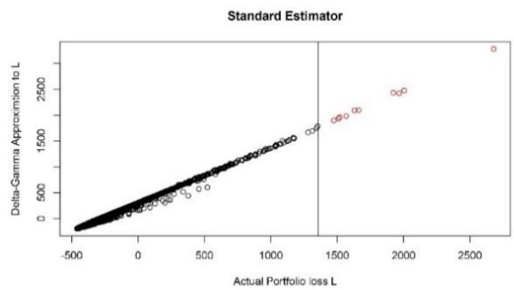

Figure 1

The scatter plot shows actual portfolio loss L versus the delta-gamma approximation to L generated under the original measure. The portfolio considered is (a.11) with x=1357.603. Details of the portfolio can be found in section three. The sample points in black refers to samples that do not satisfy $L > x$ *(990 out of 1000). These samples are consider "wasted" in the sense that* $I(L > x) = 0$.

5. IMPORTANCE SAMPLING FOR ESTIMATION OF VAR

Figure 1 reveal the reason why the standard estimator, equation (1), is inaccurate for estimating $\mathrm{P}\{L > x\}$. Such inaccuracy arise from the fact that very few samples drawn satisfy $L > x$, most samples are considered "wasted" in the sense that $I(L > x)$ =0 with high probability. The estimator with control variate, equation (2), suffer the same problem. To tackle such inefficiency, a natural choice would be importance sampling [$\boldsymbol{Glasserman\ et\ al\ 1997b}$]. Denoted by f the joint density of Z under the original measure, Instead of simulating this density, a different joint density is considered,

$$\mathrm{P}\{L > x\} = \int [I(L > x)]\, f(z)\, dz = \int \left\{[I(L > x)] \frac{f(z)}{g(z)}\right\} g(z)\, dz = \tilde{E}[I(L > x)R(Z)]$$

$\tilde{E}$ is the expectation when sampling of Z is carried out under density g, and $R(Z) = f(z)/g(z)$ is the likelihood ratio. For the estimation of VAR, $\{L > x\}$ is a rare event under the original density $Z \sim N_m(0, I)$, few samples in the region of interest where $L \approx x$ are generated. With effective importance change of measure, a disproportionately large number of samples is expected to be generated in the region of interest.

Recall that the portfolio loss during the interval $[t, t + \Delta t]$ is $L(\Delta S) = V\big(S(t)\big) - V(S(t) + \Delta S)$, where $\Delta S \sim N_m(0\,,\,\Sigma)$. ΔS is generated by $\Delta S = CZ$, where $CC^T = \Sigma$. For standard estimator, we have

$$P\{L > x\} = E[I(L > x)],\ \hat{P}\{L > x\} = \sum_{i=1}^{n} [I(L_i > x)] \left(\frac{1}{n}\right), \text{ where } Z \sim N_m(0, I)$$

For importance sampling, we consider a measure in which the mean of Z is changed from 0 to μ, and the covariance matrix is changed from the identity matrix to B. Instead of the general case, we consider exponential change of measure, which is a distribution shifting technique commonly used in importance sampling. For estimator with importance sampling change of measure, we have

$$P\{L > x\} = E[I(L > x)] = E_\theta[I(L > x)R(Z)]$$

$$\hat{P}\{L > x\} = \sum_{i=1}^{n} [I(L_i > x)R(Z_i)] \left(\frac{1}{n}\right) \qquad (3)$$

where $Z \sim N_m\big(\mu(\theta), B(\theta)\big)$, $B(\theta) = (I - 2\theta\Lambda)^{-1}$, $\mu(\theta) = \theta B(\theta) b$, and θ is the twisting parameter.

With this form of importance sampling, the $Z_i s$ remain independent, while the mean and variance of the individual Z_i are changed to $\sigma_i^2(\theta) = 1/(1 - 2\theta\lambda_i)$, and $\mu_i(\theta) = \theta b_i \sigma_i^2(\theta) = \theta b_i/(1 - 2\theta\lambda_i)$. Notice that in order for the importance sampling change of measure to be valid, θ must satisfy $\sigma_i^2(\theta) > 0\ for\ all\ i$. Hence, we require $(1 - 2\theta\lambda_i) > 0\ for\ all\ i$. Our goal is that, given x, the VAR we want to estimate, choose a θ_x such that under the importance sampling change of measure, $\{L > x\}$ is no longer a rare event and large values of Q are generated with high probability. In fact, latter we will show that under the assumption that the delta-gamma approximation is exact, the mean of $(a_0 + Q)$ equals x. That is, $E_{\theta_x}[a_0 + Q] = E_{\theta_x}[L] = x$, where E_{θ_x} denotes the expectation under importance sampling change of measure with the twisting parameter θ_x.

To motivate the specific choice of the twisting parameter θ, assume the delta-gamma approximation is exact, $L = a_0 + Q$. $P\{L > x\} = E_\theta[I(L > x)R(Z)] = E_\theta[I(Q > x - a_0)exp\{\psi(\theta) - \theta Q\}]$, where $\psi(\theta)$ is the logarithm of the moment generating function of Q [$\boldsymbol{Glasserman\ et\ al\ 1997a}$]. Consider the second moment of a sample taken under importance sampling change of measure,

$$m_2(x, \theta) = E_\theta\{[I(L > x)R(Z)\,]^2\} = E_\theta[I(Q > x - a_0)exp\{2\psi(\theta) - 2\theta Q\}\,]$$

$$\leq exp\{2\psi(\theta) - 2\theta(x - a_0)\}$$

Set $\theta_x = \mathrm{arg}min[exp\{2\psi(\theta) - 2\theta(x - a_0)\}]$, then we have $\psi'(\theta_x) = (x - a_0)$.

Notice that $E_\theta(Q) = E_\theta(\sum_{i=1}^{m}(b_i Z_i + \lambda_i Z_i^2)) = \psi'(\theta)$, it follows that

$$E_{\theta_x}(L) = E_{\theta_x}(a_0 + Q) = a_0 + \psi'(\theta_x) = x \qquad (4)$$

Hence, under this particular importance sampling change of measure, $\{L > x\}$ is no longer a rare event. As suggested by **figure 2** more samples will be generated from the region of interest where $L \approx x$. The derivation of (4) and the exact form of $\psi(\theta)$ and $\psi'(\theta_x)$ are detailed in **appendix c**. Numerical results on 3 portfolios are displayed in table 2. We can see that the variance of all three portfolios are reduced by one order of magnitude, furthermore, the effectiveness of such estimator with importance sampling change of measure increase as the loss probability to be estimated dncrease. For these two reasons, we consider the importance sampling a more effective approach than control variate as a variance reduction technique built on the delta-gamma approximation.

Portfolio	x_{std}	P{L>x}	IS
(a.11) index, +λ	2.7	1.9%	11.3
	3.2	1.1%	18.0
	3.7	0.6%	28.8
(b.5) DAO-C & CON-P	2.25	3.1%	10.2
	2.75	1.1%	21.5
	3.25	0.3%	51.6
(c.3) setting 1 EO, C & P,+	2.2	2.6%	11.9
	2.7	1.1%	23.3
	3.2	0.4%	48.6
(c.3) setting 2 EO, C & P,+	2.2	2.6%	10.3
	2.7	1.1%	18.4
	3.2	0.4%	37.6

Table 2: Comparison of effectiveness of importance sampling as p decrease.

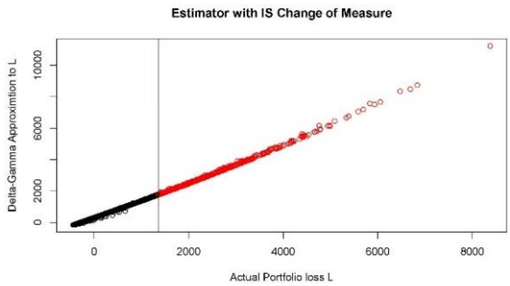

Figure 2

The scatter plot shows actual portfolio loss L versus the delta-gamma approximation to L generated under the importance sampling change of measure. The portfolio considered is (a.11) with x=1357.603. Details of the portfolio can be found in section three. The sample points in black refers to samples that do not satisfy $L > x$ (617 out of 1000), which is considerably smaller than that in figure 1.

An attempt to boost the variance improvement importance sampling achieves by introducing antithetic variate is carried out. Let first recall the general setting of antithetic sampling, consider the random variable $H(X), where\ X \sim p$, then $E[H(X)] = E\big[\big(H(X) + H(X^*)\big)/2\big], where\ X^* \sim p\ as\ well.$ $H(X^*)$ is called the antithetic variable. Consider the following unbiased estimator of $E[H(X)]$ with sample size n, $\overline{H_{anti}(X)} = [1/(n/2)]\left\{\sum_{i=1}^{(n/2)}([H(X_i) + H(X_i^*)]/2)\right\}$. The variance of such estimator is

$$Var\left(\overline{H_{anti}(X)}\right) = \left[\frac{1}{n^2}\right]\left\{\left(\frac{n}{2}\right) Var[H(X_i) + H(X_i^*)]\right\}$$

$$= \left[\frac{1}{2n}\right]\{Var[H(X)] + Var[H(X^*)] + 2Cov[H(X), H(X^*)]\}$$

$$= \left(\frac{\sigma_{H(X)}^2}{n}\right)(1 + \rho)$$

where $\rho = cor[H(X), H(X^*)]$ is the correlation between $H(X)$ and $H(X^*)$. Whether or not the antithetic sampling leads to a variance reduction depend on the sign of ρ. Our goal is to identify X^* such that $\rho = cor\big(H(X), H(X^*)\big) < 0$. For example, if $X \sim U(a, b)$, we may set $X^* = [(b + a) - X]$; if $X \sim N_m(\mu, \Sigma)$, we may set $X^* = [2\mu - X]$, so that the correlation between $X\ and\ X^*$ is negative, e.g., $cor(X, X^*) = (-1)$. However, the correlation between $H(X)$ and $H(X^*)$ is not necessarily negative, the more linear is the function H, the more likely for $cor[H(X), H(X^*)]$ to be negative.

Another potential benefit antithetic sampling could bring is that when it is computationally expensive to generate samples of the underlying random variable X, the cost of Monte Carlo simulation with antithetic sampling using n samples would be well smaller than that without antithetic sampling. However, since the underlying random variable of the VAR estimation is multivariate normal which is easy to generate, we do not consider such potential benefit when evaluating the effectiveness of antithetic sampling for estimating VAR.

For VAR estimation, consider $P\{L > x\} = E_{\theta_x}[I(L > x)R(Z)] = E_{\theta_x}[H(Z)], where\ Z \sim N_m(\mu(\theta_x), \mathrm{B}(\theta_x))$ under the importance sampling change of measure. Consider $Z^* = 2\mu(\theta_x) - Z$ so that $cor(Z, Z^*) = (-1)$, take $H(Z^*)$ as the antithetic variable of $H(Z)$. Then we have the following unbiased estimator of $P\{L > x\}$,

$$\hat{P}\{L > x\} = \left[\frac{1}{(n/2)}\right]\left[\sum_{i=1}^{(n/2)} \left(\frac{H(Z_i) + H(Z_i^*)}{2}\right)\right]$$

The numerical results (full table available in **appendix d**) suggest that such approach is nearly rewardless. Further analysis has revealed that, under the importance sampling change of measure, the linearity between

$$H(Z) = I(L > x)R(Z) = I(L > x)exp\{\psi(\theta_x) - \theta_x Q\},$$

$$where\ \mathrm{Q} = b^T Z + Z^T \Lambda Z = \textstyle\sum_{i=1}^{m}(b_i Z_i + \lambda_i Z_i^2)$$

and its antithetic variable $H(Z^*)$ is not strong and not necessarily negative due to the existence of the indicator function and the involvement of quadratic terms Z_i^2 in the exponential function. In particular, for portfolios with continuous payoff that is delta hedged (e.g., $\delta = 0$ and $b_i = 0 \text{ for all i}$), Q becomes the summation of quadratic terms, the correlation $H(Z)$ and $H(Z^*)$ is likely to be positive, and thus resulting in an increment in the variance of the estimator. Positive correlation is also generated for portfolios with discontinuous payoff and for portfolios whose non-diagonal entities of $\Gamma\ \text{matrix}$ is absent in the delta-gamma approximation.

For the other portfolios, the magnitude of negative correlation is small (e.g., $-0.2 < \rho < 0$), and thus the variance reduction achieved is minor. As antithetic variate is ineffective in general for VAR estimation, we would abandon this approach. In next section, we investigate how importance sampling can be combined with stratified sampling to achieve further variance reduction. In particular, the allocation of samples to the strata in stratified sampling and its effect between the importance sampling is of concerned.

6. STRATIFIED SAMPLING FOR VAR AND ITS COMBINATION WITH IMPORTANCE SAMPLING

The essence of stratified sampling is to identify a stratification variable Y whose distribution is known. The stratification variable explains most of the variability of the output of the simulation. If Y is constrained to a small interval, then the output would also be highly constrained, and thereby leading to significant variance reduction. In general, if we want to estimate $E(X)$ with the stratification variable Y, by law of total expectation, we have

$$E(X) = E_Y\left[E_{X|Y}(X|Y \in \mathcal{S})\right] = \textstyle\sum_{j=1}^{k} E(X|Y \in \mathcal{S}_j)p_j, \text{ where } p_j = P(Y \in \mathcal{S}_j).$$

There are k strata $\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_k$, typically intervals. Denoted by $\hat{X}$ the stratified estimator of $E(X)$. Suppose we take n samples in total for the stratified estimate, equivalently, n_j samples from each strata where $n_j = nq_j$, with $\sum_{j=1}^{k} q_j = 1$. Then $\hat{X} = \sum_{j=1}^{k} \overline{X}_j p_j$, where $\overline{X}_j = \sum_{i=1}^{n_j} X_{ij}(1/n_j)$ is the sample average of the n_j samples drawn from strata j. Consider the variance of $\hat{X}$, $Var(\hat{X}) = \sum_{j=1}^{k}[p_j^2(\sigma_j^2/n_j)]$, where $\sigma_j^2 = Var[X|Y \in \mathcal{S}_j]$. It is known that if we set $n_j = np_j$, then a variance reduction can always be achieved compare to the variance of the standard estimator with the same total sample size n, (σ^2/n), where $\sigma^2 = Var[X]$. However, notice that such allocation policy is suboptimal, consider the Lagrange Multiplier $L = [\sum_{j=1}^{k} p_j^2(\sigma_j^2/n_j)] - [\varphi(n - \sum_{j=1}^{k} n_j)]$, and differentiate L with respect to n_j *and* φ, we have

$$n_j = \sqrt{(p_j^2\sigma_j^2/\varphi)},\ \ \sqrt{\varphi} = (1/n)\textstyle\sum_{j=1}^{k}\sqrt{p_j^2\sigma_j^2},$$

it follows that $n_j = (n)[(p_j\sigma_j)/(\sum_{i=1}^{k} p_i\sigma_i)]$ would minimize the variance of $\hat{X}$.

Therefore, given p_j *and* σ_j, setting $n_j = (n)(q_j^*), where\ q_j^* = [(p_j\sigma_j)/(\sum_{i=1}^{k} p_i\sigma_i)]$ is the optimized allocation scheme for stratified sampling in terms of variance reduction. With the optimal sample allocation scheme $n_j = n(q_j^*)$, we have $Var(\hat{X})^* = \left[(\sum_{i=1}^{k} p_i\sigma_i)^2/n\right]$. For any other sample allocation scheme $n_j = n(q_j)$ such that $q_j \neq q_j^*$, we have $Var(\hat{X}) = Var(\hat{X})^*\{\sum_{i=1}^{k}[(q_i^*)^2/q_i]\}$.

For the estimation of VAR, we consider Q as the stratification variable, it follows that

$$\begin{aligned} P\{L > x\} &= E[I(L > x)] \\ &= E_{\theta_x}[I(L > x)R(Z)] \\ &= \textstyle\sum_{j=1}^{k} E_{\theta_x}(I(L > x)R(Z)|Q \in \mathcal{S}_j)\, P(Q \in \mathcal{S}_j) \end{aligned}$$

$$\hat{P}\{L > x\} = \sum_{j=1}^{k}\left[\sum_{i=1}^{n_j} I(L_{ji} > x)R(Z_{ji})\left(\frac{1}{n_j}\right)\right](p_j) \qquad (5)$$

To implement stratified sampling, given p_js, we must first identify the percentiles of Q that define the strata $\mathcal{S}_j s$. Since the transform of Q is known (e.g., $Q = b^T Z + Z^T \Lambda Z = \sum_{i=1}^{m}(b_i Z_i + \lambda_i Z_i^2)$ where $Z \sim N_m(\mu(\theta_x), B(\theta_x))$), we may use numerical transform inversion techniques to calculate the percentiles [***Rouvinez* 1997**]. Next, the generation of n_j samples from strata j is carried out by a “bin tossing” method. We generate a random vector Z under the importance sampling change of measure, then the value Q coreponding to the sample must fall into certain strata. If there are fewer than n_j samples generated for this strata, then this Z is used to evaluate the portfolio, otherwise, discard this sample. Such procedure is repeated unstill the required number of samples n_j is drawn from each strata. The full algorithm for VAR estimation with importance sampling and stratified sampling is detailed in **appendix e**.

Numerical experiments are carried out to examine the additional variance reduction achieved by imposing stratified sampling on the estimator with importance sampling change of measure. The results are displayed in **table 3**. Note that the experiment is conducted under two sample allocation scheme. In the first scheme (ISSQ), we consider equiprobable strata and an equal allocation of samples to strata, that is, $p_j = (1/k)\ and\ n_j = np_j = (n/k)$ for all j. In the second scheme (ISSQO), we consider equiprobable strata and the optimal sample allocation scheme, that is $p_j = (1/k)\ and\ n_j = nq_j$, where $q_j = \{(p_j\sigma_j)/[\sum_{i=1}^{k} p_i\sigma_i]\}$, where σ_is are estimated in advance by generating 10000 samples from each strata.

Portfolio	x_{std}	P{L>x}	IS	ISSQ	ISSQO
(a.1) 0.5yr ATM, +λ	2.5	1.0%	30.5	286.4	2875.6
(a.2) 0.5yr ATM, −λ	1.95	1.0%	43.5	253.9	2065.0
(a.3) 0.5yr ATM, ±λ	2.3	1.0%	37.6	349.6	4030.1
(a.4)) 0.1yr ATM, +λ	2.6	1.1%	22.1	68.7	421.5
(a.5) 0.1yr ATM, −λ	1.69	1.0%	42.6	66.9	153.6
(a.6) 0.1yr ATM, ±λ	2.3	0.9%	33.4	135.5	786.9
(a.7) δ hedged, +λ	2.8	1.1%	17.7	30.2	119.2
(a.8) δ hedged, −λ	1.8	1.1%	53.4	126.4	707.8
(a.9) δ hedged, ±λ	2.8	1.1%	15.8	27.8	126.2
(a.10) δ hedged, λm<-λ1	2.0	1.1%	18.1	33.5	143.7
(a.11) index, +λ	3.2	1.1%	18.1	228.2	1411.8
(a.12)) index, −λ	1.02	1.16%	25.8	43.7	101.1
(a.13) index, ±λ	2.5	1.1%	15.3	66.9	358.3
(a.14) index, λm<-λ1	1.65	1.1%	13.9	40.1	213.3
(a.15) 100, Block-diagonal	2.65	1.0%	18.3	28.6	98.1

Table 3A: Comparison of Methods for Portfolios with Continous Payoffs

Portfolio	x_{std}	P{L>x}	IS	ISSQ	ISSQO
(b.1) C	2.55	0.94%	31.6	162.1	1379.8
(b.2) DAO-C	2.45	1.1%	24.8	46.3	133.8
(b.3) DAO-C & P	2.8	1.1%	14.1	16.7	34.8
(b.4) DAO-C & P, δ hedged	4.9	1.0%	7.7	9.1	34.8
(b.5) DAO-C & CON-P	2.75	1.1%	21.6	31.5	94.4
(b.6) DAO-C & CON-P, δ hedged	9	1.1%	0.8	0.8	8.1
(b.7) CON-C and CON-P	2.3	1.0%	23.4	34.8	102.8
(b.8) AON-C and CON-P	2.35	1.0%	22.3	31.7	89.7

Table 3B: Comparison of Methods for Portfolios with Discontinous Payoffs

Portfolio	setting	x_{std}	P{L>x}	IS	ISSQ	ISSQO
(c.1) EO,+	setting 1	2.7	1.1%	26.8	131.9	1030.5
	setting 2	2.7	1.1%	24.7	44.0	144.3
(c.2) EO, ±	setting 1	2.5	1.0%	27.0	134.9	918.0
	setting 2	2.5	1.0%	22.6	33.4	92.9
(c.3) EO, C & P,+	setting 1	2.7	1.1%	23.0	81.3	548.4
	setting 2	2.7	1.1%	19.2	31.8	109.7
(c.4) EO, C & P, −	setting 1	2.45	0.95%	30.1	106.6	576.0
	setting 2	2.45	0.95%	21.6	28.9	70.2
(c.5) EO, C & P, ±	setting 1	2.65	1.1%	22.7	81.3	557.2
	setting 2	2.65	1.1%	19.2	30.3	95.2

Table 3C: Comparison of Methods for Portfolios with Non-diagonal Γ

Large variance reductions are achieved for almost all of the portfolios for at least one order of magnitude with importance sampling alone (IS), and dramatic improvement is brought over IS when stratification on Q is introduced (ISSQ and ISSQO). Since $E_{\theta_x}(L) = x$ is true only with the assumption that $L = a_0 + Q$, more accurate is the delta-gamma approximation to the portfolio loss, the more significant variance reduction would be brought by the importance sampling and stratified sampling.

For table 3A, shorter Δt is compared to the maturity T, the more accurate is the quadratic approximation, hence, the variance reduction achieved by IS for portfolio (a.1), (a.2), and (a.3) are more significant than that of their counterparts with shorter maturity (e.g., 30.5 for portfolio (a.1) versus 22.1 for (a.4)). And such discrepancy amplifies when stratified sampling is introduced. When positive λ (eigen value) is involved, the delta-gamma approximation appears to be less accurate for portfolios that are delta hedged, and the variance reduction generated is thus less significant. On the other hand, if all the eigen value is all negative, the delta-gamma approximation seems to be more accurate for delta hedged portfolios (e.g., 42.5 for (a.5) versus 53.4 for (a.8)). By referring to portfolio (a11) and (a.15), we can see that correlations among assets and portfolios size does not have significant impact on the effectiveness of the estimator in terms of the order of the variance reduction achieved.

It is expected that the delta-gamma approximation to the portfolio loss is less accurate when the payoff of the portfolio is discontinuous as we fit a continuous function $(a_0 + Q)$ to a discontinuous one (L), while the order of magnitude of the variance reduction achieved for IS remain the same with that in table 3A, we can see that the improvement of going from IS to ISSQ and ISSQO is less dramatic by comparing portfolio (b.1) to the rest of the portfolios in table 3B. Diminishing effect arise for portfolio (b.6) for three reasons. First, for $T = 0.01$, $\Delta t = 0.04$ is relatively long. Second, the portfolio purely consist of options with discontinuous payoffs. And third, the portfolio is delta hedged. Since the portfolio is delta hedged (e.g., $\delta = 0$ and $b = aC^T = -(\delta)C^T = 0$), the linear part of the delta-gamma approximation becomes 0 as well, $(a_0 + Q) = a_0 + (b^T Z + Z^T \Lambda Z) = \sum_{i=1}^{m}(b_i Z_i + \lambda_i Z_i^2)$, which in terms require us to fit the pure quadratic part of the delta-gamma approximation to the actual loss L, which is a function with discontinuous first and second derivatives. Therefore, one would expect the delta-gamma approximation to be not at all representative to the portfolio loss at the region of interest $(L \approx x)$ for portfolio (b.6). And introducing importance sampling and stratified sampling based on the delta gamma approximation lead to a slight increase in variance as suggested in the table (0.76 for IS and 0.8 for ISSQ). However, despite all these factor, we managed to brought a variance reduction by employing the optimized samples allocation scheme for stratified sampling (8.15 for ISSQO).

For table 3C, by comparing setting one and setting two of each of the portfolio, we can see that he absence of non-diagonal elements does not have a big impact on the effectiveness of IS. However, the improvement by adding stratified sampling to the estimator with importance sampling change of measure shrink. Consider portfolio (c.1), the variance ratio improve from 26.8 (IS) to 1030.5 (ISSQO) for setting 1, while that of setting 2 improve from 24.7 (IS) to 144.3 (ISSQO). Yet, the effectiveness of the such estimator is still considered satisfactory under both setting.

7. SUMMARY

By using the delta-gamma approximation to the portfolio loss as a basis., this paper has developed variance reduction techniques for the estimation of value-at-risk. The most promising method that results in the greatest variance reduction is constructed by a proper combination of importance sampling and stratified sampling. The first step of the method is to identify the distribution of the risk factors that govern the value of the portfolio, e.g., $\Delta S \sim N_m(0\,,\, \Sigma)$. Then develop the delta -gamma approximation to the portfolio loss, $L \approx a_0 + Q = a_0 + b^T Z + Z^T \Lambda Z$. Next, identify the importance sampling change of measure based on such approximation given x, $Z \sim N_m(\mu(\theta_x), \mathrm{B}(\theta_x))$ where $B(\theta) = (I - 2\theta\Lambda)^{-1}$ *and* $\mu(\theta) = \theta B(\theta) b$. Finally, stratification on Q is combined with the importance sampling to achieve further variance reduction.

For the wild range of test portfolios we considered, the variance is reduced by at least one order of magnitude, and often two order of magnitude reduction is achieved. The sample allocation scheme of stratified sampling has a great impact on the effectiveness of this method. As the numerical results in **table 3** suggest, a tremendous variance reduction is achieved when the stratified sampling is carried out under the optimized sample allocation scheme $n_j = (n)(q_j^*), where\ q_j^* = [(p_j \sigma_j)/(\sum_{i=1}^{k} p_i \sigma_i)]$. However, such approach required prior knowledge on $\sigma_i^2 = Var_{\theta_x}(I(L > x)R(Z)|Q \in \mathcal{S}_i)$, which could be computationally expensive to estimate in practice, e.g., large sample size required for the estimate of σ_i^2 to be accurate, or costly to generate samples of the underlying random variable. To tackle such problem, heuristics approaches that require a much less sample size for the estimation of σ_i^2 is developed $[\boldsymbol{Glasserman\ et\ al.1999c}]$. Based on the observation that q_i^*, as a function of i, appears to have a normal-like shape. We can first calculate the a less accurate estimate of σ_i^2 with a small sample size. Then find a normal curve that best fit q_i, and obtain $\tilde{q}_i$ according to such curve. We can therefore construct near-optimal sample allocation schemes for stratification on Q by setting $n_j = (n)\left(q_j^*\right)$. Such heuristics approaches are proved to be more effective than equal allocation of samples to strata with numerical results in $[\boldsymbol{Glasserman\ et\ al.1999c}]$**.**

APPENDIX

A. The distribution of $\Delta \boldsymbol{S}_{(t)} = \Delta \boldsymbol{S} = \left[\boldsymbol{S}_{(t+\Delta t)} - \boldsymbol{S}_{(t)}\right]^{T}$

Consider $S_{i(\mathrm{t})}$, the price of the underlying asset described by the stochastic differential equation

$$\frac{dS_{i(t)}}{S_{i(t)}} = \mathrm{r}\mathrm{dt} + \sigma_i \mathrm{d}W_{i(t)},\ \ \mathrm{d}ln\left(S_{i(t)}\right) = \left(\mathrm{r} - \frac{1}{2}\sigma_i^2\right)\mathrm{dt} + \sigma_i \mathrm{d}W_{i(t)},$$

$$S_{i(t+\Delta t)} = S_{i(t)}\exp\left\{\left(\mathrm{r} - \frac{1}{2}\sigma_i^2\right)\Delta t + \left(\sigma_i\sqrt{\Delta t}\right)Z_i\right\}, \mathrm{i} = 1,2,\ldots,\mathrm{m}$$

The moment generating function of normal distribution and bivariate normal distribution is essential to determining the value of $\mu\ and\ \Sigma$.

- For $X \sim N(\mu_X, \sigma_X^2)$, $M_X(t) = E(e^{tX}) = exp\left\{t\mu_X + \frac{1}{2}t^2\sigma_X^2\right\}$
- For $\binom{X}{Y} \sim N_2\left(\binom{\mu_X}{\mu_Y}, \begin{pmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2\end{pmatrix}\right)$,

  $M_{X,Y}(t_1,t_2) = E(e^{t_1X+t_2Y}) = exp\left\{t_1X + t_2Y + \frac{1}{2}(t_1^2\sigma_X^2 + t_2^2\sigma_Y^2 + 2\rho t_1t_2\sigma_X\sigma_Y)\right\}$

Consider the distribution of $\Delta S = \left[S_{(\Delta t)} - S_{(0)}\right]^T$, consider $X_i = ln\left(\frac{S_{i(\Delta t)}}{S_{i(0)}}\right) \sim N\left(\left(\mathrm{r} - \frac{1}{2}\sigma_i^2\right)(\Delta t), \left(\sigma_i\sqrt{\Delta t}\right)^2\right)$,

- $\mathrm{E}(\Delta S_i) = \mathrm{E}\left(S_{i(\Delta t)} - S_{i(0)}\right) = S_{i(0)}[\mathrm{E}(exp\{X_i\}) - 1]$

  $= S_{i(0)}\left[exp\left\{\left(\mathrm{r} - \frac{1}{2}\sigma_i^2\right)(\Delta t) + \frac{1}{2}\left(\sigma_i\sqrt{\Delta t}\right)^2\right\} - 1\right] = S_{i(0)}[exp\{\mathrm{r}\Delta t\} - 1] \approx 0$ for small Δt

- $\mathrm{Cov}\left(\Delta S_i, \Delta S_j\right) = \mathrm{Cov}\left(S_{i(\Delta t)}, S_{j(\Delta t)}\right) = S_iS_j\mathrm{Cov}\left(exp\{X_i\}, exp\{X_j\}\right)$

  $= S_iS_j\left\{E\left(e^{X_i+X_j}\right) - \mathrm{E}(e^{X_i})\mathrm{E}(e^{X_j})\right\} = S_iS_j\exp\{2r\Delta t\}\left(exp\{\rho_{ij}\sigma_i\sigma_j\Delta t\} - 1\right)$,

  where $\rho_{ij} = cor\left(X_i, X_j\right)$ is the correlation between the annual log return of the assets.

Hence, for $\Delta S \sim N_m(\mu, \Sigma)$ with $\mathrm{r} = 0.03, \Delta\mathrm{t} = 0.04$,

we have $\begin{cases} \mu_i \approx 0,\ i = 1,2,..,m \\ \Sigma_{ij} = S_iS_j\exp\{2r\Delta t\}\left(exp\{\rho_{ij}\sigma_i\sigma_j\Delta t\} - 1\right),\ i,j = 1,2,\ldots,m \end{cases}$

B. The payoff functions of the options and their respective Greeks

Theta: derivative of an option with respect to time, $\Theta = (\partial C/\partial t)$

Delta: derivative of an option with respect to price, $\delta = (\partial C/\partial S)$

Gamma: second derivative of an option with respect to price, $\Gamma = (\partial^2 C/\partial S^2)$

Relationship of Θ, δ, Γ of a portfolio with value Π consisting of derivatives dependent on non-dividend paying stock:

$(\Theta) + \mathrm{rS}(\delta) + (1/2)(\sigma^2 S^2)(\Gamma) = \mathrm{r}\Pi$ $[\boldsymbol{Hull}\ \mathbf{2012}]$

b1. Standard European Call Option and Standard European Put Option with $\boldsymbol{q = 0}$

$C_{call} = Se^{-q(T-t)}\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2) = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2)$

$C_{put} = -Se^{-q(T-t)}\Phi(-d_1) + Ke^{-r(T-t)}\Phi(-d_2) = -S\Phi(-d_1) + Ke^{-r(T-t)}\Phi(-d_2)$

where $d_1 = [ln(S/K) + (r + \sigma^2/2)(T-t)]/(\sigma\sqrt{T-t})$, $d_2 = d_1 - (\sigma\sqrt{T-t})$

consider $\phi(d_2) = \phi(d_1 - \sigma\sqrt{T-t})$, we have $\mathrm{S}\phi(d_1) = Ke^{-r(T-t)}\phi(d_2)$

$\Theta_{call} = (\partial C_{call}/\partial t) = \mathrm{S}(-\sigma/2\sqrt{T-t})\ \phi(d_1) + \mathrm{K}(-re^{-r(T-t)})\ \Phi(d_2)$

$\Theta_{put} = (\partial C_{put}/\partial t) = \mathrm{S}(-\sigma/2\sqrt{T-t})\ \phi(d_1) + \mathrm{K}(re^{-r(T-t)})\ \Phi(-d_2)$

$\delta_{call} = (\partial C_{call}/\partial S) = \Phi(d_1)$

$\delta_{put} = (\partial C_{put}/\partial S) = \Phi(d_1) - 1$

$\Gamma_{call} = (\partial^2 C_{call}/\partial S^2) = \phi(d_1)/S\sigma\sqrt{T-t}$

$\Gamma_{put} = (\partial^2 C_{put}/\partial S^2) = \phi(d_1)/S\sigma\sqrt{T-t}$

b2. Barrier Option with $\boldsymbol{q = 0}$

A full description of the pricing formulas for barrier options is available in $[Zhang\ 1998]$. Here, we only consider the case of Down-and-Out Call Option and Down-and-In Call Option. A down-and-out call is one type of knock-out option. It is a regular option that ceases to exist if the asset price reaches a certain barrier level H, where $\mathrm{H} < S_0$. Its corresponding knock-in option is down-and-in call is option, which is a regular option that comes into exist if the asset price reaches the barrier level H, where $\mathrm{H} < S_0$. Notice that the value of a regular call option equals the value of a down-and-in call option plus the value of a down-and-in call option.

For $\mathrm{H} < \mathrm{K}$,

$C_{daic} = Se^{-q(T-t)}(H/S)^{2\lambda}\Phi(y) - Ke^{-r(T-t)}(H/S)^{(2\lambda-2)}\Phi(y - \sigma\sqrt{T-t})$

$= S(H/S)^{2\lambda}\Phi(y) - Ke^{-r(T-t)}(H/S)^{(2\lambda-2)}\Phi(y - \sigma\sqrt{T-t})$

$C_{daoc} = C_{call} - C_{daic}$

where $\lambda = [r - q + (\sigma^2/2)]/\sigma^2 = [r + (\sigma^2/2)]/\sigma^2$,

$\mathrm{y} = ln(H^2/SK)/\sigma\sqrt{T-t} + \lambda\sigma\sqrt{T-t}$

Consider approximating the Greeks of the down-and-out call option with **finite difference method (FDM)**:

$\delta_{daoc} = (\partial C_{daoc}/\partial S) \approx [C_{daoc}(S + \Delta S) - C_{daoc}(S)]/\Delta S$

$\Gamma_{daoc} = (\partial^2 C_{daoc}/\partial S^2) \approx [C_{daoc}(S + \Delta S) - 2C_{daoc}(S) + C_{daoc}(S - \Delta S)]/(\Delta S)^2$

$\Theta_{daoc} = \mathrm{r}C_{daoc} - \mathrm{rS}(\delta_{daoc}) - (1/2)(\sigma^2 S^2)(\Gamma_{daoc})$

b3. Binary Option with $q = 0$

b3.1 (refers to b.1 for the exact form of d_1 and d_2)

A cash-or-nothing call option pays off nothing if the price of the underlying asset S is below the strike price K at maturity and pays a fixed amount Q if it is above the strike price. A cash-or-nothing put option pays a fixed amount Q if S is below K at maturity and pays off nothing if S is above K.

$C_{conc} = Qe^{-r(T-t)}\Phi(d_2),\ C_{conp} = Qe^{-r(T-t)}\Phi(-d_2)$

$\Theta_{conc} = Qre^{-r(T-t)}\Phi(d_2) + Qe^{-r(T-t)}\phi(d_2)\left\{\left(\frac{ln\left(\frac{S}{K}\right)}{2\sigma}\right)(T-t)^{(-3/2)} - \left(\frac{r-\frac{\sigma^2}{2}}{2\sigma}\right)(T-t)^{(-1/2)}\right\}$

$\Theta_{conp} = Qre^{-r(T-t)}\Phi(-d_2) - Qe^{-r(T-t)}\phi(-d_2)\left\{\left(\frac{ln\left(\frac{S}{K}\right)}{2\sigma}\right)(T-t)^{(-3/2)} - \left(\frac{r-\frac{\sigma^2}{2}}{2\sigma}\right)(T-t)^{(-1/2)}\right\}$

$\delta_{conc} = \left[\frac{(Qe^{-r(T-t)})\phi(d_2)}{(S\sigma\sqrt{T-t})}\right],\ \delta_{conp} = \left[\frac{(-Qe^{-r(T-t)})\phi(-d_2)}{(S\sigma\sqrt{T-t})}\right]$

$\Gamma_{conc} = [\mathrm{r}C_{conc} - \Theta_{conc} - \mathrm{rS}(\delta_{conc})]\left(\frac{2}{\sigma^2 S^2}\right),\ \Gamma_{conp} = [\mathrm{r}C_{conp} - \Theta_{conp} - \mathrm{rS}(\delta_{conp})]\left(\frac{2}{\sigma^2 S^2}\right)$

b3.2 (refers to b.1 for the exact form of d_1 and d_2)

A asset-or-nothing call option pays off nothing if the price of the underlying asset S is below the strike price K at maturity and pays the asset price if it is above the strike price. A asset-or-nothing put option pays the asset price if S is below K at maturity and pays off nothing if S is above K.

$C_{aonc} = Se^{-q(T-t)}\Phi(d_1) = S\Phi(d_1),\ C_{aonp} = Se^{-q(T-t)}\Phi(-d_1) = S\Phi(-d_1)$

$\Theta_{aonc} = S\phi(d_1)\left\{\left(\frac{ln\left(\frac{S}{K}\right)}{2\sigma}\right)(T-t)^{(-3/2)} - \left(\frac{r+\frac{\sigma^2}{2}}{2\sigma}\right)(T-t)^{(-1/2)}\right\}$

$\Theta_{aonp} = (-S)\phi(-d_1)\left\{\left(\frac{ln\left(\frac{S}{K}\right)}{2\sigma}\right)(T-t)^{(-3/2)} - \left(\frac{r+\frac{\sigma^2}{2}}{2\sigma}\right)(T-t)^{(-1/2)}\right\}$

$\delta_{aonc} = \Phi(d_1) + \left[\frac{\phi(d_1)}{(\sigma\sqrt{T-t})}\right],\ \delta_{aonp} = \Phi(-d_1) - \left[\frac{\phi(-d_1)}{(\sigma\sqrt{T-t})}\right]$

$\Gamma_{aonc} = [\mathrm{r}C_{aonc} - \Theta_{aonc} - \mathrm{rS}(\delta_{aonc})]\left(\frac{2}{\sigma^2 S^2}\right),\ \Gamma_{aonp} = [\mathrm{r}C_{aonp} - \Theta_{aonp} - \mathrm{rS}(\delta_{aonp})]\left(\frac{2}{\sigma^2 S^2}\right)$

b3.3

Consider the European exchange option that give up an asset U and receive in return asset V at maturity. The payoff of this exchange option is $max(V_T - U_T, 0)$.

$V_{exvu} = Ve^{-q_V(T-t)}\Phi(d_1) - Ue^{-q_U(T-t)}\Phi(d_2)$

where $d_1 = [ln(V/U) + (q_U - q_V + \hat{\sigma}^2/2)(T-t)]/(\hat{\sigma}\sqrt{T-t})$, $d_2 = d_1 - (\hat{\sigma}\sqrt{T-t})$,

$\hat{\sigma} = \sqrt{\sigma_U^2 + \sigma_V^2 - 2\rho\sigma_U\sigma_V}$, ρ is the instantaneous correlation between U and V.

As the portfolios we concern in portfolio set C consist of uncorrelated assets with same volatility and zero dividend yield. The payoff of the exchange option can be simplified to $V_{exvu} = V\Phi(d_1) - U\Phi(d_2)$, with $\hat{\sigma} = \sqrt{2}\sigma$, $d_1 = [ln(V/U) + (\sigma)(T-t)]/(\sqrt{2}\sigma\sqrt{T-t})$, $d_2 = d_1 - (\sqrt{2}\sigma\sqrt{T-t})$.

Consider $\phi(d_2) = \phi(d_1 - \sqrt{2}\sigma\sqrt{T-t})$, we have $V\phi(d_1) = U\phi(d_2)$

$$\Theta_{exvu} = \left(\frac{\partial C_{exvu}}{\partial t}\right) = \left(\frac{-\sigma}{\sqrt{2}}\right)(T-t)^{(-1/2)}[V\phi(d_1)]$$

$$\delta_V = \left(\frac{\partial C_{exvu}}{\partial V}\right) = \Phi(d_1),\ \delta_U = -\Phi(d_2)$$

$$\Gamma_V = \left(\frac{\partial^2 C_{exvu}}{\partial V^2}\right) = \left[\frac{\phi(d_1)}{V\sqrt{2}\sigma\sqrt{T-t}}\right],\ \Gamma_U = \left[\frac{\phi(d_2)}{U\sqrt{2}\sigma\sqrt{T-t}}\right],\ \Gamma_{VU} = \left(\frac{\partial^2 C_{exvu}}{\partial V \partial U}\right) = \left[\frac{-\phi(d_2)}{V\sqrt{2}\sigma\sqrt{T-t}}\right],\quad \Gamma_{UV} = \left[\frac{-\phi(d_1)}{U\sqrt{2}\sigma\sqrt{T-t}}\right]$$

C. Derivation of $E_{\theta_x}(L) = x$ under the assumption that the delta-gamma approximation is exact $L = a_0 + Q$

$Delta - Gamma\ approxiamtion: L \approx a_0 + a^T \Delta S + \Delta S^T A \Delta S \equiv a_0 + Q,$

$$where\ \mathrm{Q} = b^T Z + Z^T \Lambda Z = \sum_{i=1}^{m}(b_i Z_i + \lambda_i Z_i^2)$$

$$P\{L > x\} = E[I(L > x)], Z \sim N_m(0, I)$$

$$= E_\theta[I(L > x)R(Z)],\ Z \sim N_m(\mu(\theta), B(\theta))\ with\ B(\theta) = (I - 2\theta\Lambda)^{-1},\ \mu(\theta) = \theta B(\theta) b$$

$$R(Z) = \frac{f(Z)}{g(Z)} = \frac{(2\pi)^{-\left(\frac{m}{2}\right)}|I|^{-\left(\frac{1}{2}\right)} exp\left\{\left(\frac{-1}{2}\right) Z^T I^{-1} Z\right\}}{(2\pi)^{-\left(\frac{m}{2}\right)}|B|^{-\left(\frac{1}{2}\right)} exp\left\{\left(\frac{-1}{2}\right)(Z-\mu)^T B^{-1}(Z-\mu)\right\}}$$

$$= |B|^{-\left(\frac{1}{2}\right)} exp\left\{\left(\frac{-1}{2}\right)[Z^T Z - (Z-\mu)^T B^{-1}(Z-\mu)]\right\}$$

$$= \left[\prod_{i=1}^{m}(1 - 2\theta\lambda_i)^{-1}\right]^{\left(\frac{1}{2}\right)} exp\left\{-\theta Q + \left[\sum_{i=1}^{m}\left(\frac{1}{2}\right)\left(\frac{(\theta b_i)^2}{1-2\theta\lambda_i}\right)\right]\right\} \quad \text{by (c1) and (c2)}$$

$$= exp\left\{-\theta Q + \sum_{i=1}^{m}\left(\frac{1}{2}\right)\left[\left(\frac{(\theta b_i)^2}{1-2\theta\lambda_i}\right) - ln(1 - 2\theta\lambda_i)\right]\right\} = exp\{\psi(\theta) - \theta Q\} \quad \text{by (c3)}$$

- (c1): $|B|^{\left(\frac{1}{2}\right)} = \left(\prod_{i=1}^{m}(1 - 2\theta\lambda_i)^{-1}\right)^{\left(\frac{1}{2}\right)}$
- (c2): $Z^T Z - (Z-\mu)^T B^{-1}(Z-\mu)$

$$= Z^T Z - [(Z^T B^{-1} Z) - (Z^T B^{-1}\mu) - (\mu^T B^{-1} Z) + (\mu^T B^{-1}\mu)]$$

$$= Z^T Z - [(Z^T Z) - 2\theta(Z^T \Lambda Z)] + [\theta(Z^T b)] + [\theta(b^T Z)] - [(\theta^2)(b^T(I - 2\theta\Lambda)b)]$$

$$= (2\theta)(b^T Z + Z^T \Lambda Z) - (\theta^2)(b^T(I - 2\theta\Lambda)b)$$

$$= (2\theta)Q - \left[\sum_{i=1}^{m}\left(\frac{(\theta b_i)^2}{1-2\theta\lambda_i}\right)\right], \text{wher } \mathrm{Q} = b^T Z + Z^T \Lambda Z = \sum_{i=1}^{m}(b_i Z_i + \lambda_i Z_i^2)$$

- (c3): $ln\left(M_Q(\theta)\right) = \sum_{i=1}^{m}\left(\frac{1}{2}\right)\left[\left(\frac{(\theta b_i)^2}{1-2\theta\lambda_i}\right) - ln(1 - 2\theta\lambda_i)\right] = \psi(\theta) \quad [\boldsymbol{Glasserman\ et\ al\ 1997a}]$

Consider the second moment of a sample taken under importance sampling change of measure,

$$m_2(x, \theta) = E_\theta\{[I(L > x)R(Z)]^2\} = E_\theta[I(L > x) exp\{2\psi(\theta) - 2\theta Q\}]$$

$$\leq exp\{2\psi(\theta) - 2\theta(x - a_0)\} \text{ under the assumption that } L = a_0 + Q$$

Set $\theta_x = \mathrm{argmin}[exp\{2\psi(\theta) - 2\theta(x - a_0)\}]$ to minimize the upper bound of the second moment of a sample taken under importance sampling change of measure, then $\psi'(\theta_x) = (x - a_0)$, with $\psi'(\theta_x) = \sum_{i=1}^{m}\left[\frac{\theta_x b_i^2(1-\theta_x\lambda_i)}{(1-2\theta_x\lambda_i)^2} + \frac{\lambda_i}{1-2\theta_x\lambda_i}\right]$.

For a general choice of θ, we have $E_\theta(Q) = E_\theta(\sum_{i=1}^{m}(b_i Z_i + \lambda_i Z_i^2)) = \sum_{i=1}^{m}[b_i E_\theta(Z_i) + \lambda_i E_\theta(Z_i^2)]$

$$= \sum_{i=1}^{m}\left[b_i\left(\frac{\theta b_i}{1-2\theta\lambda_i}\right) + \lambda_i\left[\left(\frac{\theta b_i}{1-2\theta\lambda_i}\right)^2 + \left(\frac{1}{1-2\theta\lambda_i}\right)\right]\right] = \psi'(\theta)$$

By choosing $\theta = \theta_x$, we have $E_{\theta_x}(Q) = \psi'(\theta_x) = (x - a_0)$

it follows that $E_{\theta_x}(L) = E_{\theta_x}(a_0 + Q) = a_0 + (x - a_0) = x$ under that assumption that $L = a_0 + Q$

D. Additional Numerical Results

d1. numerical results on estimator with control variate alone

Portfolio		x_{std}	P{L>x}	CV	$\hat{\rho}$
(a.1) 0.5yr ATM, +λ		2.5	0.01023	5.8163	0.9104
(a.2) 0.5yr ATM, −λ		1.95	0.01024	3.901	0.8637
(a.3) 0.5yr ATM, ±λ		2.3	0.00962	11.377	0.9563
(a.4)) 0.1yr ATM, +λ		2.6	0.01108	3.4665	0.8438
(a.5) 0.1yr ATM, −λ		1.69	0.009932	2.4133	0.7686
(a.6) 0.1yr ATM, ±λ		2.3	0.00862	5.1544	0.897
(a.7) δ hedged, +λ		2.8	0.01103	2.5553	0.7761
(a.8) δ hedged, −λ		1.8	0.01056	1.4507	0.5533
(a.9) δ hedged, ±λ		2.8	0.01094	2.5752	0.7782
(a.10) δ hedged, λm<-λ1		2.0	0.01119	2.4581	0.7703
(a.11) index, +λ		3.2	0.01061	3.9533	0.8637
(a.12) index, −λ		1.02	0.01155	1.4445	0.05414
(a.13) index, ±λ		2.5	0.01119	4.0307	0.8646
(a.14) index, λm<-λ1		1.65	0.01083	3.6622	0.8497
(a.15) 100, Block-diagonal		2.65	0.009612	2.2884	0.7437
(b.1) C		2.55	0.00952	5.6144	0.9081
(b.2) DAO-C		2.45	0.01129	3.6971	0.8489
(b.3) DAO-C & P		2.8	0.01085	2.1247	0.7263
(b.4) DAO-C & P, δ hedged		4.9	0.01037	1.03909	0.1845
(b.5) DAO-C & CON-P		2.75	0.01077	1.529	0.5837
(b.6) DAO-C & CON-P, δ hedged		9	0.01054	0.9999	0.004351
(b.7) CON-C and CON-P		2.3	0.01018	2.2667	0.7458
(b.8) AON-C and CON-P		2.35	0.009972	2.0469	0.7165
(c.1) EO,+	setting 1	2.7	0.01074	8.6247	0.9402
	setting 2	2.7	0.01071	2.03663	0.711
(c.2) EO, ±	setting 1	2.5	0.009875	8.4649	0.9402
	setting 2	2.5	0.009941	2.0148	0.713
(c.3) EO, C & P,+	setting 1	2.7	0.01106	5.2264	0.9001
	setting 2	2.7	0.01107	2.4056	0.7692
(c.4) EO, C & P, −	setting 1	2.45	0.009551	5.4449	0.907
	setting 2	2.45	0.009496	2.05115	0.725
(c.5) EO, C & P, ±	setting 1	2.65	0.01072	5.701	0.9041
	setting 2	2.65	0.0177	2.363	0.7596

Table d1: numerical results of the estimator with control variate (CV), $\hat{\rho}$ is the sample correlation between $I(L > x)$ and the control variate $I(Q > x - a_0)$.

d2. numerical results on estimator with importance change of measure and antithetic variate

Portfolio		x_{std}	P{L>x}	IS	ISAN	$\hat{\rho}$
(a.1) 0.5yr ATM, +λ		2.5	0.01015	30.03491	34.2134	-0.1173
(a.2) 0.5yr ATM, –λ		1.95	0.01027	43.8885	39.8989	0.08194
(a.3) 0.5yr ATM, ±λ		2.3	0.009635	38.0794	42.8622	-0.1413
(a.4)) 0.1yr ATM, +λ		2.6	0.01106	22.2902	24.443	-0.05737
(a.5) 0.1yr ATM, –λ		1.69	0.009895	42.9545	35.8517	0.2169
(a.6) 0.1yr ATM, ±λ		2.3	0.008666	34.2164	38.8884	-0.1122
(a.7) δ hedged, +λ		2.8	0.01103	17.7567	12.5839	0.4256
(a.8) δ hedged, –λ		1.8	0.01057	53.0152	36.8206	0.411
(a.9) δ hedged, ±λ		2.8	0.01091	16.2326	11.5905	0.4249
(a.10) δ hedged, λm<-λ1		2.0	0.01116	19.2491	12.8909	0.4798
(a.11) index, +λ		3.2	0.01063	17.8414	21.8688	-0.1859
(a.12) index, –λ		1.02	0.01161	25.6953	16.9483	0.5357
(a.13) index, ±λ		2.5	0.01125	14.7373	17.8643	-0.1664
(a.14) index, λm<-λ1		1.65	0.01082	14.541	16.4326	-0.1562
(a.15) 100, Block-diagonal		2.65	0.009632	17.7193	17.2132	0.02953
(b.1) C		2.55	0.009446	30.9754	35.4785	-0.1254
(b.2) DAO-C		2.45	0.0113	26.1268	22.5214	0.1523
(b.3) DAO-C & P		2.8	0.01086	13.8324	13.9668	0.002743
(b.4) DAO-C & P, δ hedged		4.9	0.01038	7.5696	7.7469	-0.01977
(b.5) DAO-C & CON-P		2.75	0.01073	21.5229	15.3944	0.3778
(b.6) DAO-C & CON-P, δhedged		9	0.01055	0.7447	0.7413	-0.004332
(b.7) CON-C and CON-P		2.3	0.01022	24.8074	21.2752	0.1463
(b.8) AON-C and CON-P		2.35	0.010027	23.01758	21.6283	0.0466
(c.1) EO,+	setting 1	2.7	0.01075	26.7651	30.6242	-0.1356
	setting 2	2.7	0.01075	24.863	20.3	0.1933
(c.2) EO, ±	setting 1	2.5	0.009956	27.3565	32.3998	-0.1634
	setting 2	2.5	0.009951	23.6959	21.2771	0.08941
(c.3) EO, C & P,+	setting 1	2.7	0.01105	22.8095	24.8185	-0.07761
	setting 2	2.7	0.01105	18.8686	19.9206	-0.04591
(c.4) EO, C & P, –	setting 1	2.45	0.009501	29.6147	34.2297	-0.1011
	setting 2	2.45	0.009504	20.6845	19.7364	0.0354
(c.5) EO, C & P, ±	setting 1	2.65	0.01075	23.2812	25.6462	-0.09771
	setting 2	2.65	0.01074	18.5544	19.5501	-0.04434

Table d2: Comparison of the numerical results of the estimator with importance samping alone (IS) and the estimaotr with importance smapling and antithetic variate. $\hat{\rho}$ is the sample correlation between $H(Z) = I(L > x)R(Z)$ and the antithetic variable $H(Z^*)$ under the importance sampling change of measure, where $Z^* = 2\mu(\theta_x) - Z$.

E. Full Algorithm

Given $\boldsymbol{x}, \boldsymbol{\Sigma}, \boldsymbol{a_0}, \ \boldsymbol{a}, \boldsymbol{and}\ \boldsymbol{A}\ (\boldsymbol{Note}: \boldsymbol{L} \approx \boldsymbol{a_0} + \boldsymbol{a^T}\Delta\boldsymbol{S} + \Delta\boldsymbol{S^T}\boldsymbol{A}\Delta\boldsymbol{S} \equiv \boldsymbol{a_0} + \boldsymbol{Q})$.

Assume $\Delta\boldsymbol{S} \sim \boldsymbol{N_m}(\boldsymbol{0}, \boldsymbol{\Sigma})$, estimate $\boldsymbol{P}\{\boldsymbol{L} > \boldsymbol{x}\}$.

1. Express $Q = b^T Z + Z^T \Lambda Z$, where $Z \sim \mathrm{N_m}(0, \mathrm{I})$

- Find $\tilde{C}\tilde{C}^T = \Sigma$
- Solve the Eigen-Decomposition $\left(\tilde{C}^T A \tilde{C}\right) = U \Lambda U^T$
- Set $C = \tilde{C}U, \ b^T = a^T C$

2. Identify the IS distribution $Z \sim \mathrm{N_m}\big(\mu(\theta), \mathrm{B}(\theta)\big)$, with $\mathrm{B}(\theta) = (1 - 2\theta\Lambda)^{-1}, \ \mu(\theta) = \theta \mathrm{B}(\theta) b$

- Set $\theta = \theta_x$, where $\psi'(\theta_x) = (x - a_0)$

3. Define k strata,

- Given p_j, find s_j such that $P\{Q \le s_j\} = \sum_{i=1}^{j} p_j$, $\mathrm{j} = 1, \ldots, \mathrm{k} - 1$

4. Perform the simulation,

- Generate $Z^{(ij)}, Q^{(ij)}$ in stratum j and set $\Delta S^{(ij)} = CZ^{(ij)}$
- Use $Z^{(ij)}$ to evaluate $L^{(ij)}$ and $R\left(Z^{(ij)}\right)$

5. Estimate $P\{L > x\}$ by $\hat{P}$

- $\hat{P} = \sum_{j=1}^{k} \left[\sum_{i=1}^{n_j} I\left(L^{(ji)} > x\right) R\left(Z^{(ji)}\right) \left(\frac{1}{n_j}\right)\right] (p_j)$

$= \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left[I\left(L^{(ji)} > x\right) R\left(Z^{(ji)}\right) \left(\frac{1}{n}\right)\right]$, if we employ equiprobable strata and let $n_j = np_j$

Reference

Glasserman, P., P. Heidelberger, P. Shahabuddin. 1999a. Variance reduction techniques for estimating Value-at-Risk. Management Science 2000 INFORMS Vol.46, No 10, October pp. 1349-1364

Glasserman, P., P. Heidelberger, P. Shahabuddin. 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 Internet. Conf. Comput. Finance)* MIT Press, Cambridge, MA, 7-24.

Glasserman, P., P. Heidelberger, P. Shahabuddin. 1999c. Stratification issues in estimating Value-at-Risk. *Proc. 1999 Winter Simulation Conf.*, IEEE Computer Society Press, Phoenix, AZ, 351-358.

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

Hall, J. C. 2012. Options, Futures and other derivatives, 8th ed. Prentice Hall.

Peter G. Z. 1998. Exotic Options, A Guide to Second Generation Options, 2nd ed. World Scientific.