Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Nested Simulation in Portfolio Risk Measurement

Michael B. Gordy, Sandeep Juneja,

To cite this article:
Michael B. Gordy, Sandeep Juneja, (2010) Nested Simulation in Portfolio Risk Measurement. Management Science 56(10):1833-1848. https://doi.org/10.1287/mnsc.1100.1213

MANAGEMENT SCIENCE
Vol. 56, No. 10, October 2010, pp. 1833–1848
ISSN 0025-1909 | EISSN 1526-5501 | 10 | 5610 | 1833

informs

DOI 10.1287/mnsc.1100.1213

Nested Simulation in Portfolio Risk Measurement

Michael B. Gordy
Federal Reserve Board, Washington, DC 20551, michael.gordy@frb.gov

Sandeep Juneja
School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai 400005, India, juneja@tifr.res.in

Risk measurement for derivative portfolios almost invariably calls for nested simulation. In the outer step, one draws realizations of all risk factors up to the horizon, and in the inner step, one reprices each instrument in the portfolio at the horizon conditional on the drawn risk factors. Practitioners may perceive the computational burden of such nested schemes to be unacceptable and adopt a variety of second-best pricing techniques to avoid the inner simulation. In this paper, we question whether such short cuts are necessary. We show that a relatively small number of trials in the inner step can yield accurate estimates, and we analyze how a fixed computational budget may be allocated to the inner and the outer step to minimize the mean square error of the resultant estimator. Finally, we introduce a jackknife procedure for bias reduction.

Key words: nested simulation; loss distribution; value-at-risk; expected shortfall; jackknife estimator
History: Received April 23, 2008; accepted May 8, 2010, by Gérard P. Cachon, finance. Published online in *Articles in Advance* August 20, 2010.

1. Introduction

For a wide variety of derivative instruments, computational costs may pose a binding constraint on the choice of pricing model. The more realistic and flexible the model, the less likely that there will exist an analytical pricing formula, and so the more likely that simulation-based pricing algorithms will be required. For plain-vanilla options trading in fast-moving markets, simulation is prohibitively slow. Simple models with analytical solutions are typically employed with ad hoc adjustments (such as local volatility surfaces) to obtain better fit to the cross-section of market prices. Because such models capture underlying processes in crude fashion, they tend to require frequent recalibration and perform poorly in time-series forecasting. For path-dependent options (e.g, lookback options) and complex basket derivatives (e.g., CDO of asset-backed securities), simulation is almost unavoidable, though even here computational shortcuts may be adopted at the expense of bias.[1]

Risk-management applications introduce additional challenges. Time constraints are less pressing than in trading applications, but the computational task may be more formidable. When loss is measured on a mark-to-market basis, estimation via simulation of large loss probabilities or of risk measures such as value-at-risk (VaR) calls for a nested procedure: In the outer step, one draws realizations of all risk factors up to the horizon, and in the inner step, one reprices each position in the portfolio at the horizon conditional on the drawn risk factors. Because repricing must be executed for *each* trial in the outer step, it has been widely assumed that it would be computationally prohibitive to make use of simulation in the inner step. This assumption rests on the mistaken belief that highly accurate repricing is needed for each instrument in the portfolio.

In this paper, we question whether inner step simulations must necessarily impose a large computational burden. We show that a relatively small number of trials in the inner step can yield accurate estimates for large loss probabilities and portfolio risk measures such as value-at-risk and expected shortfall, particularly when the portfolio contains a large number of positions. Because an expectation is replaced by a noisy sample mean, the estimator is biased, and we are able to characterize this bias asymptotically. We analyze how a fixed and large computational budget may be allocated to the inner and the outer step to minimize the mean square error of the resultant estimator. We show how the jackknifing technique may be applied to reduce the bias in our estimator and how this alters the optimal budget allocation.

The most studied application of nested simulation in the finance literature is the pricing of American options. An influential paper by Longstaff and Schwartz (2001) proposes a least-squares methodology in which a small number of inner step samples

[1] Many of the ideas in this paper were developed originally in Gordy and Juneja (2006) specifically for application to portfolios of CDO tranches.

is used to estimate a parametric relationship between the state vector at the horizon (in this case, the stock price) and the continuation value of the option. This methodology is applicable to a broad range of nested problems, so long as the dimension of the state vector is not too large and the relationship between state vector and continuation value is not too nonlinear. However, some care must be taken in the choice of basis functions, and in general it may be difficult to assess the associated bias (Glasserman 2004, §8.6). Our methodology, by contrast, is well-suited to portfolios of high-dimensional and highly nonlinear instruments, can be applied to a variety of derivative types without customization, and has bias of known form.

Our optimization results for large loss probabilities and value-at-risk are similar to those of Lee (1998).[2] Lee's analysis relies on a different and somewhat more intricate set of assumptions than ours, which are in the spirit of the sensitivity analysis of VaR by Gouriéroux et al. (2000) and the subsequent literature on "granularity adjustment" of credit VaR (Martin and Wilde 2002, Gordy 2004). The resulting asymptotic formulae, however, are the same. Our extension of this methodology to expected shortfall is new, as is our analysis of large portfolio asymptotics. Furthermore, so far as we are aware, we are the first to analyze the convergence properties of jackknife estimators in a nested simulation setting.

In recent related work, Lan et al. (2007a, b; 2010) also consider a two level simulation method for risk management. They focus on developing efficient methods for estimating confidence intervals for expected shortfall using ranking and selection based ideas as well as the statistical theory of empirical likelihood.

In §2, we set out a very general modeling framework for a portfolio of financial instruments. We introduce the nested simulation methodology in §3. We characterize both bias and variance of the simulation estimator and analyze the optimal allocation of computational resources between the two stages that minimizes the mean square error of the resultant estimator. Numerical illustrations of our main results are provided in §4. In §5, we develop simple jackknife methods for bias reduction in nested simulation to improve computational performance further.

[2] The work of Lee (1998) is an unpublished Ph.D. thesis, which we encountered shortly before completion of our first draft. Lee's contribution anticipates the work of Gouriéroux et al. (2000) and the literature on granularity adjustment and appears to have been overlooked in the finance literature. De Prisco et al. (2007) study a similar problem but minimize the *variance* of the estimator (rather than its mean square error) subject to the computational budget constraint.

2. Model Framework

Let X_t be the vector of m state variables that governs all prices. The vector X_t might include interest rates, commodity prices, equity prices, and other underlying prices referenced by derivatives. Let $\mathcal{F}_t$ be the filtration generated by X_t. For use in discounting future cash flows, we denote by $B_t(s)$ the value at time s of \$1 invested at time $t \le s$ in a risk free money market account, i.e.,

$$B_t(s) = \exp\left(\int_t^s r(u)\,du\right).$$

If interest rates are stochastic, then $B_t(s)$ depends on $\mathcal{F}_s$.

The portfolio consists of $K+1$ positions. The price of position k at time t depends on t, $\mathcal{F}_t$, and the contractual terms of the instrument.[3] Position 0 represents the subportfolio of instruments for which there exist analytical pricing functions. Without loss of generality, we treat this subportfolio of instruments as if it were a single position. Among the contractual terms for an instrument is its maturity. We assume maturity T_k is finite for $k = 1, \ldots, K$. As in all risk-measurement exercises, the portfolio is assumed to be held static over the model horizon.

Conditional on $\mathcal{F}_t$, the cash flows up to time t are deterministic functions of time that depend on the contractual terms. Let $C_k(t)$ be the cumulative cash flow for k on $(0, t]$. Note that increments to $C_k(t)$ can be positive or negative and can arrive at discrete time intervals or continuously. The market value of each position is the present discounted expected value of its cash flows under the risk-neutral measure Q:[4]

$$V_k(t) = E^Q\left[\int_t^{T_k} \frac{dC_k(s)}{B_t(s)} \,\middle|\, \mathcal{F}_t\right]. \tag{1}$$

The valuations are expressed in currency units, so there is no need for portfolio weights. We set $V_k(t) = 0$ whenever $t \ge T_k$.

The present time is normalized to 0 and the model horizon is H. "Loss" is defined as the difference between current value and discounted future value at the horizon, adjusting for interim cash flows. Portfolio loss is

$$Y = \sum_{k=0}^{K}\left(V_k(0) - \frac{1}{B_0(H)}\left(V_k(H) + \int_0^H B_t(H)\,dC_k(t)\right)\right).$$

The implicit assumption here is that interim cash flows are reinvested in the money market until time H, but other conventions are easily accommodated.

[3] For some exotic options, the price at t will depend on the entire path of X_s on $s = (0, t]$. This is why we need the filtration $\mathcal{F}_t$ and not just X_t.

[4] Following standard practice in the risk-management literature, we assume the existence and uniqueness of the equivalent martingale measure Q. Duffie (1996, §6.G) reviews the necessary technical assumptions.

3. Simulation Framework

We now develop notation related to the simulation process. The simulation is nested: There is an "outer step" in which we draw histories up to the horizon H. For each trial in the outer step, there is an "inner step" simulation needed for repricing at the horizon.

Let L be the number of trials in the outer step. In each of these trials, we execute the following steps:

1. Draw a single path X_t for $t = (0, H]$ under the physical measure. Let ξ denote the realization of random variables $(X_t\colon 0 < t \le H)$. Thus, ξ_l represents the generated information in the outer step of trial l.

2. Evaluate the accrued value at H of the interim cash flows.

3. Evaluate the price of each position at H.
   - Closed-form price for instrument 0.
   - Simulation with N_k "inner step" trials for time period $(H, T_k]$ for remaining positions $k = 1, \ldots, K$. These paths are simulated under the risk-neutral measure.

4. Sum the accrued value of the interim cash flows at time H and estimated price of each position at time H, and then discount back to time 0 to get our estimated loss $\widetilde{Y}(\xi)$.

Observe that the full dependence structure across the portfolio is captured in the period up to the model horizon. Inner step simulations, in contrast, can be run independently across positions. This is because the value of position k at time H is simply a conditional expectation (given $\mathcal{F}_H$ and under the risk-neutral measure) of its own subsequent cash flows and does not depend on future cash flows of other positions. An implication of running conditionally independent inner step simulations is that pricing errors are independent across positions.[5] Intuition might suggest that it would be more efficient from a simulation perspective to run inner step simulations simultaneously across all positions in order to reduce the total number of sampled paths of X_t on $(H, \max\{T_k\}]$. However, if we use the same samples of X_t across inner step simulations, pricing errors are no longer independent across the positions and so do not diversify away as effectively at the portfolio level. Furthermore, when the positions are repriced independently, to reprice position k we need only draw joint paths for the elements of X_t that influence that instrument. This may greatly reduce the memory footprint of the simulation, in particular when the number of state variables (m) is large and when some of the maturities $\{T_k\}$ are much longer than others.

We have assumed that initial prices $V_k(0)$ are already known and can be taken as constants in our algorithm. Of course, this can be relaxed.

In the following three subsections, we discuss estimation of large loss probabilities (§3.1), value-at-risk (§3.2), and expected shortfall (§3.3). For simplicity, we impose a single value of N across all positions (i.e., $N_k = N$ for $k = 1, \ldots, K$). This restriction is relaxed in §3.4. In §3.5, we consider the asymptotic behavior of the optimal allocation of computational resources as the portfolio size grows large. Last, in §3.6, we elaborate on the trade-offs associated with simultaneous repricing.

3.1. Estimating the Probability of Large Losses

We first consider the problem of efficient estimation of $\alpha = P(Y(\xi) > u)$ via simulation for a given u. If for each generated ξ, the mark-to-market values of each position were given by analytical pricing formulae, the associated $Y(\xi)$ would be known, and simulation would involve generating independent and identically distributed (i.i.d.) samples $Y(\xi_1), Y(\xi_2), \ldots, Y(\xi_L)$ and taking the average

$$\frac{1}{L}\sum_{i=1}^{L} 1[Y(\xi_i) > u]$$

as an estimator of α. We are interested in the situation in which the mark-to-market value of each position is *not* given analytically and is instead estimated via the inner step simulations.

Within the inner step simulation for repricing position k, each trial gives an unbiased but noisy estimate of $Y(\xi)$.[6] Let $\zeta_{ki}(\xi)$ denote the zero-mean pricing error associated with the ith such sample for position k, let $Z_i(\xi)$ denote the portfolio pricing error for the ith inner step sample, and finally let

$$\bar{Z}^N(\xi) = \frac{1}{N}\sum_{i=1}^{N} Z_i(\xi) = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} \zeta_{ki}(\xi)$$

be the zero-mean average pricing error for the portfolio as a whole. In place of $Y(\xi)$, we take as its surrogate $\widetilde{Y}(\xi) \equiv Y(\xi) + \bar{Z}^N(\xi)$ as an estimate of loss in the portfolio. By the law of large numbers, $\bar{Z}^N(\xi) \to 0$, almost surely, as $N \to \infty$ for any fixed K (assuming that $\mathrm{E}[|\zeta_{ki}|] < \infty$ for each k). The estimator for $P(Y(\xi) > u)$ then involves generating i.i.d. samples $(\widetilde{Y}(\xi_1), \ldots, \widetilde{Y}(\xi_L))$ via outer and inner step simulation

[5] Equation (1) gives the value of position k at the horizon as a conditional expectation of discounted future cash flows between times H and T_k. That is, defining $\widetilde{V}_{k,\xi}$ as the random variable $\int_H^{T_k} dC_k(s)/B_H(s)$, we have $V_k(H;\xi) = E^Q[\widetilde{V}_{k,\xi} \mid \xi]$. The inner step generates N trials of $\widetilde{V}_{k,\xi}$ under Q, for which the associated pricing error is $V_k(H;\xi) - (1/N)\sum_{i=1}^{n} \widetilde{V}_{k,\xi,i}$. Because $\widetilde{V}_{k,\xi}$ is a deterministic function of the sample paths X_t for $t = (H, T_k]$, the pricing errors are independent across positions so long as the sample paths X_t for $t = (H, T_k]$ are generated independently for each k.

[6] Recall that the value of a position is a conditional expectation of a random variable, whereas we generate a sample of that random variable under the same conditional distribution using simulation. The difference between sample and conditional expectation must be mean zero.

and taking the average

$$\hat{\alpha}_{L,N} = \frac{1}{L}\sum_{l=1}^{L} 1[\widetilde{Y}(\xi_l) > u].$$

We now examine the mean square error of $\hat{\alpha}_{L,N}$. Let α_N denote $P(\widetilde{Y}(\xi) > u)$. The mean square error of the estimator $\hat{\alpha}_{L,N}$ separates into

$$\begin{aligned}\mathrm{E}[(\hat{\alpha}_{L,N} - \alpha)^2] &= \mathrm{E}[(\hat{\alpha}_{L,N} - \alpha_N + \alpha_N - \alpha)^2] \\ &= \mathrm{E}[(\hat{\alpha}_{L,N} - \alpha_N)^2] + (\alpha_N - \alpha)^2.\end{aligned}$$

Further, note that

$$\mathrm{E}[(\hat{\alpha}_{L,N} - \alpha_N)^2] = \mathrm{V}[\hat{\alpha}_{L,N}] = \frac{\alpha_N(1-\alpha_N)}{L},$$

where for any random variable X, $\mathrm{V}[X]$ denotes its variance. Suppose that the computational effort to generate one outer step sample of ξ on average equals $\gamma_0 > 0$, and the effort to generate an inner step simulation sample on average equals γ_1. Then average effort per iteration of simulation equals $N\gamma_1 + \gamma_0$, and average effort for L iterations equals $L(N\gamma_1 + \gamma_0)$. By law of large numbers, this is close to the actual effort when L is large. We suppose that the overall computational budget is fixed at Γ. We analyze the problem of choosing (N, L) to minimize the mean square error of the estimator $\hat{\alpha}_{L,N}$ when $L(N\gamma_1 + \gamma_0) = \Gamma$ for large Γ. Thus, we consider the optimization problem

$$\begin{aligned}&\min_{N,L\geq 0} \quad \frac{\alpha_N(1-\alpha_N)}{L} + (\alpha_N - \alpha)^2 \\ &\text{subject to} \quad L(N\gamma_1 + \gamma_0) = \Gamma, \end{aligned} \tag{2}$$

as $\Gamma \to \infty$. Proposition 1 states our essential result for solving the optimization problem. We need some notation and a technical assumption for this.

Let $\widetilde{Z}_N = \bar{Z}^N\sqrt{N}$ so that $\widetilde{Z}_N$ has a nontrivial limiting distribution as $N \to \infty$. Then $\alpha_N = P(Y + \widetilde{Z}_N/\sqrt{N} > u)$. Our asymptotic analysis relies on Taylor series expansion of the joint density function $g_N(y, z)$ of $(Y, \widetilde{Z}_N)$ and its partial derivatives. Assumption 1 ensures that higher order terms in such expansions can be ignored.

Assumption 1. (i) *The joint density $g_N(\cdot,\cdot)$ of Y and $\widetilde{Z}_N$ and its partial derivatives*

$$\frac{\partial}{\partial y} g_N(y, z) \quad \textit{and} \quad \frac{\partial^2}{\partial y^2} g_N(y, z)$$

exist for each N and for all (y, z).

(ii) *For $N \geq 1$, there exist nonnegative functions $p_{0,N}(\cdot)$, $p_{1,N}(\cdot)$, and $p_{2,N}(\cdot)$ such that*

$$\begin{aligned} g_N(y, z) &\leq p_{0,N}(z), \\ \left|\frac{\partial}{\partial y} g_N(y, z)\right| &\leq p_{1,N}(z), \\ \left|\frac{\partial^2}{\partial y^2} g_N(y, z)\right| &\leq p_{2,N}(z) \end{aligned}$$

for all y, z. In addition,

$$\sup_N \int_{-\infty}^{\infty} |z|^r p_{i,N}(z)\, dz < \infty$$

for $i = 0, 1, 2$, and $0 \leq r \leq 4$.

This assumption may be expected to be true in a large portfolio where there are at least a few positions that have a sufficiently smooth payoff. Alternatively, this assumption may be satisfied by perturbing Y and $\widetilde{Z}_N$ through adding to both of them mean zero, variance ϵ independent Gaussian random variables, also independent of Y and $\widetilde{Z}_N$. For small ϵ this has a negligible impact on the tail measures. Then if

$$\sup_N \int_{-\infty}^{\infty} z^4\, dF_{\widetilde{Z}_N}(z) < \infty, \tag{3}$$

where $F_{\widetilde{Z}_N}(\cdot)$ denotes the distribution function of $\widetilde{Z}_N$, Assumption 1 can be seen to hold. To see this, let Y_ϵ and $\widetilde{Z}_{N,\epsilon}$ denote the random variables obtained by perturbing Y and $\widetilde{Z}_N$ as described above. Then the joint probability density function (pdf) of $(Y_\epsilon, \widetilde{Z}_{N,\epsilon})$ equals

$$g_{N,\epsilon}(y, z) = \int_{\Re^2} \phi_\epsilon(y - a)\phi_\epsilon(z - b)\, dG_N(a, b),$$

where $\phi_\epsilon(\cdot)$ denotes the pdf of Gaussian random variable with mean zero and variance ϵ, and $G_N(\cdot,\cdot)$ denotes the joint distribution function of $(Y, \widetilde{Z}_N)$. Because $\phi_\epsilon(\cdot)$ and its first two derivatives are bounded, it is easy to see that Equation (3) implies that Assumption 1 holds for $(Y_\epsilon, \widetilde{Z}_{N,\epsilon})$.

Assumption 1 is sufficient to deliver a useful convergence property. Here and henceforth, let f and F denote the density and cumulative distribution function (cdf) for Y, and let $\tilde{f}_N$ and $\tilde{F}_N$ denote the density and cumulative distribution function for $\widetilde{Y}$. Now let y_N be some sequence of real numbers that converges to a real number y. In §A.1 in the appendix, we prove the following lemma.

Lemma 1. *Under Assumption 1, if $y_N \to y$, then $\tilde{f}_N(y_N) \to f(y)$ and $\tilde{f}'_N(y_N) \to f'(y)$ as $N \to \infty$.*

We now approximate α_N in orders of $1/N$. We define the function

$$\Theta(u) = \tfrac{1}{2} f(u)\mathrm{E}[\sigma_\xi^2 \mid Y(\xi) = u], \tag{4}$$

where σ_ξ^2 denotes the conditional variance of Z_i (conditioned on ξ). In §A.2 in the appendix, we prove the following proposition.[7]

[7] We say that a function is $O_m(h(m))$ if for all m sufficiently large its absolute value is upper bounded by a constant times $h(m)$. We say that it is $o_m(h(m))$ if for any $\epsilon > 0$ its absolute value is upper bounded by ϵ times $h(m)$.

Proposition 1. *Under Assumption* 1,

$$\alpha_N = \alpha + \theta_u/N + O_N(1/N^{3/2})$$

where $\theta_u = -\Theta'(u)$.

By construction, the distribution of $Y(\xi) + \bar{Z}^N(\xi)$ differs from the distribution of $Y(\xi)$ by a mean-preserving spread, in the sense of Rothschild and Stiglitz (1970). Unless the two distributions have an infinite number of crossings, there will exist a u^* such that $\alpha_N \geq \alpha$ for all $u > u^*$. Therefore, setting aside the unlikely possibility of a pathological family of loss distributions, the bias will be upward (i.e., $\theta_u > 0$) for the large loss levels u that are typically of interest to risk management.

Applying Proposition 1, the objective function reduces to finding N that minimizes

$$\frac{N\gamma_1 + \gamma_0}{\Gamma}(\alpha(1-\alpha) + O_N(1/N)) + \frac{\theta_u^2}{N^2} + O_N(1/N^{5/2}).$$

It is easy to see that an optimal N for this has the form

$$N^* = \left(\frac{2\theta_u^2}{\alpha(1-\alpha)\gamma_1}\right)^{1/3}\Gamma^{1/3} + o_\Gamma(\Gamma^{1/3}). \tag{5}$$

Therefore optimal L has the form

$$L^* = \left(\frac{\alpha(1-\alpha)}{2\gamma_1^2\theta_u^2}\right)^{1/3}\Gamma^{2/3} + o_\Gamma(\Gamma^{2/3}), \tag{6}$$

and the mean square error at optimal N^* equals

$$3\left(\frac{\theta_u\alpha(1-\alpha)\gamma_1}{2\Gamma}\right)^{2/3} + o_\Gamma(\Gamma^{-2/3}).$$

For large computational budgets, we see that L^* grows with the square of N^*. Thus, marginal increments to Γ are allocated mainly to the outer step. To gain intuition for the imbalance between N^* and L^*, consider the alternative. If N and L were of the same order $\sqrt{\Gamma}$, the squared bias term would contribute much less to the mean square error than would the variance term. By increasing L at the expense of N, we reduce variance until it matches up in contribution to the squared bias term.

3.2. Estimating Value-at-Risk

We now consider the problem of efficient estimation of value-at-risk for Y. For a target insolvency probability α, VaR is the value y_α given by

$$y_\alpha = \mathrm{VaR}_\alpha[Y] = \inf\{y\colon P(Y \leq y) \geq 1-\alpha\}.$$

Under Assumption 1, Y is a continuous random variable so that $P(Y \geq y_\alpha) = \alpha$. As before, our nested simulation generates samples $(\widetilde{Y}(\xi_1), \ldots, \widetilde{Y}(\xi_L))$, where $\widetilde{Y}(\xi) \equiv Y(\xi) + \bar{Z}^N(\xi)$. We sort these draws as $\widetilde{Y}_{[1]} \geq \cdots \geq \widetilde{Y}_{[L]}$ so that $\widetilde{Y}_{\lceil \alpha L \rceil}$ provides an estimate of y_α, where $\lceil a \rceil$ denotes the integer ceiling of the real number a. Our interest is in characterizing the mean square error $\mathrm{E}[(\widetilde{Y}_{\lceil \alpha L \rceil} - y_\alpha)^2]$ and then minimizing it. As before, we decompose MSE into variance and squared bias,

$$\mathrm{E}[(\widetilde{Y}_{\lceil \alpha L \rceil} - y_\alpha)^2] = \mathrm{V}[\widetilde{Y}_{\lceil \alpha L \rceil}] + \mathrm{E}[\widetilde{Y}_{\lceil \alpha L \rceil} - y_\alpha]^2.$$

To approximate bias and variance, we use the following result.

Proposition 2. *Under Assumption* 1,

$$\mathrm{E}[\widetilde{Y}_{\lceil \alpha L \rceil}] - y_\alpha = \frac{\theta_\alpha}{Nf(y_\alpha)} + o_N(1/N) + O_L(1/L) + o_N(1)O_L(1/L),$$

where $\theta_\alpha = -\Theta'(y_\alpha)$ *and*

$$\mathrm{V}[\widetilde{Y}_{\lceil \alpha L \rceil}] = \frac{\alpha(1-\alpha)}{(L+2)f(y_\alpha)^2} + O_L(1/L^2) + o_N(1)O_L(1/L).$$

A result parallel to the bias approximation is used in the literature on "granularity adjustment" of credit VaR to adjust asymptotic approximations of VaR for undiversified idiosyncratic risk (Martin and Wilde 2002, Gordy 2004). To avoid lengthy technical digressions, our statement of the proposition and its derivation in §A.3 in the appendix abstract from certain mild but cumbersome regularity conditions.

Our budget allocation problem reduces to minimizing the mean square error

$$\frac{\alpha(1-\alpha)}{(L+2)f(y_\alpha)^2} + \frac{\theta_\alpha^2}{N^2 f(y_\alpha)^2} + o_N(1/N^2) + o_N(1)O_L(1/L) + O_L(1/L^2),$$

subject to $L(N\gamma_1 + \gamma_0) = \Gamma$. It is easy to see that the optimal solution is

$$N^* = \left(\frac{2\theta_\alpha^2}{\alpha(1-\alpha)\gamma_1}\right)^{1/3}\Gamma^{1/3} + o_\Gamma(\Gamma^{1/3}).$$

and

$$L^* = \left(\frac{\alpha(1-\alpha)}{2\gamma_1^2\theta_\alpha^2}\right)^{1/3}\Gamma^{2/3} + o_\Gamma(\Gamma^{2/3}).$$

These values are identical up to terms of size $o_\Gamma(\Gamma^{2/3})$ to the optimal values for estimating $P(Y > u)$ derived in the previous section when $u = y_\alpha$.

3.3. Estimating Expected Shortfall

Although value-at-risk is ubiquitous in industry practice, it is well understood that it has significant theoretical and practical shortcomings. It ignores the distribution of losses beyond the target quantile, so it may give incentives to build portfolios that are highly

sensitive to extreme tail events. More formally, value-at-risk fails to satisfy the subadditivity property, so a merger of two portfolios can yield VaR greater than the sum of the two stand-alone VaRs. For this reason, value-at-risk is not a *coherent* risk measure, in the sense of Artzner et al. (1999).

As an alternative to VaR, Acerbi and Tasche (2002) propose using generalized *expected shortfall* (ES), defined by

$$\mathrm{ES}_\alpha[Y] = \alpha^{-1}(\mathrm{E}[Y\cdot 1[Y \ge y_\alpha]] + y_\alpha(1-\alpha-\Pr(Y<y_\alpha))). \quad (7)$$

The first term is often used as the definition of expected shortfall for continuous variables.[8] The second term is a correction for mass at the quantile y_α. Acerbi and Tasche (2002) show that ES is coherent. In our setting, Y and $\widetilde{Y}$ are continuous in distribution, so

$$\mathrm{ES}_\alpha[Y] = \frac{1}{\alpha}\mathrm{E}[Y\cdot 1[Y>y_\alpha]],$$
$$\mathrm{ES}_\alpha[\widetilde{Y}] = \frac{1}{\alpha}\mathrm{E}[\widetilde{Y}\cdot 1[\widetilde{Y}>\tilde{y}_\alpha]].$$

We begin with the more general problem of optimally allocating a computational budget to efficiently estimate $\Upsilon(u)=\mathrm{E}[Y\cdot 1[Y>u]]$ for arbitrary u. This is easier than the problem of estimating $\mathrm{E}[Y\cdot 1[Y>y_\alpha]]$ because here u is specified whereas in the latter case y_α is estimated. We return later to analyze the bias associated with the estimate of $\mathrm{E}[Y\cdot 1[Y>y_\alpha]]$.

Again, our sample output from the simulation to estimate $Y\cdot 1[Y>u]$ equals $\widetilde{Y}\cdot 1[\widetilde{Y}>u]$. Let $\Upsilon_N(u)$ denote $\mathrm{E}[\widetilde{Y}\cdot 1[\widetilde{Y}>u]]$. The following proposition evaluates the bias associated with this term.

Proposition 3. *Under Assumption* 1,

$$\Upsilon_N(u) = \Upsilon(u) + (\Theta(u)-u\Theta'(u))/N + O_N(1/N^{3/2}).$$

The proof is given in §A.4 in the appendix.

As before, the allocation problem is formalized as the constrained optimization

$$\min_{N,L\ge 0}\ \frac{\mathrm{V}[\widetilde{Y}\cdot 1[\widetilde{Y}>u]]}{L} + (\Upsilon_N(u)-\Upsilon(u))^2$$
$$\text{subject to } L(N\gamma_1+\gamma_0)=\Gamma,$$

as $\Gamma\to\infty$. Using similar analysis to the proof of Proposition 3, we can establish that

$$\mathrm{E}[\widetilde{Y}^2\cdot 1[\widetilde{Y}>u]] - \mathrm{E}[Y^2\cdot 1[Y>u]] = O_N(1/N)$$

and therefore that

$$\mathrm{V}[\widetilde{Y}\cdot 1[\widetilde{Y}>u]] = \mathrm{V}[Y\cdot 1[Y>u]] + O_N(1/N).$$

Applying Proposition 3, the objective function then reduces to finding N that minimizes

$$\frac{\mathrm{V}[Y\cdot 1[Y>u]]+O_N(1/N)}{L} + \frac{(\Theta(u)-u\Theta'(u))^2}{N^2} + O_N(1/N^{5/2}).$$

It is easy to see that an optimal N for this has the form

$$N^* = \left(\frac{2(\Theta(u)-u\Theta'(u))^2}{\mathrm{V}[Y\cdot 1[Y>u]]\gamma_1}\right)^{1/3}\Gamma^{1/3} + o_\Gamma(\Gamma^{1/3}), \quad (8)$$

and optimal L has the form

$$L^* = \left(\frac{\mathrm{V}[Y\cdot 1[Y>u]]}{2\gamma_1^2(\Theta(u)-u\Theta'(u))^2}\right)^{1/3}\Gamma^{2/3} + o_\Gamma(\Gamma^{2/3}). \quad (9)$$

We return now to the problem of the bias of $\mathrm{ES}_\alpha[\widetilde{Y}]$. We can write the difference between the expected shortfall of random variables $\widetilde{Y}$ and Y as

$$\begin{aligned}&\mathrm{E}[\widetilde{Y}\mid\widetilde{Y}>\tilde{y}_\alpha]-\mathrm{E}[Y\mid Y>y_\alpha]\\ &=\frac{1}{\alpha}(\Upsilon_N(\tilde{y}_\alpha)-\Upsilon(y_\alpha))\\ &=\frac{1}{\alpha}((\Upsilon_N(\tilde{y}_\alpha)-\Upsilon_N(y_\alpha))+(\Upsilon_N(y_\alpha)-\Upsilon(y_\alpha))). \quad (10)\end{aligned}$$

From Proposition 3, we have

$$\Upsilon_N(y_\alpha)-\Upsilon(y_\alpha) = \frac{1}{N}(\Theta(y_\alpha)-y_\alpha\Theta'(y_\alpha)) + O_N(1/N^{3/2}). \quad (11)$$

Now from the mean value theorem

$$\Upsilon_N(\tilde{y}_\alpha)-\Upsilon_N(y_\alpha) = (\tilde{y}_\alpha-y_\alpha)\Upsilon_N'(\check{y}),$$

where $\check{y}$ lies between y_α and $\tilde{y}_\alpha$. Note that $\Upsilon_N'(y) = -y\tilde{f}_N(y)$ and that $\check{y}\to y_\alpha$ as $N\to\infty$. From Lemma 1 it follows that $\tilde{f}_N(\check{y})\to f(y)$. Therefore,

$$\begin{aligned}\Upsilon_N(\tilde{y}_\alpha)-\Upsilon_N(y_\alpha) &= -(\tilde{y}_\alpha-y_\alpha)(y_\alpha f(y_\alpha)+o_N(1))\\ &= -\frac{y_\alpha\theta_\alpha}{N}+o_N(1/N), \quad (12)\end{aligned}$$

where the last equality follows from Proposition 2. By substituting Equations (11) and (12) into (10), we arrive at

$$\mathrm{ES}_\alpha[\widetilde{Y}]-\mathrm{ES}_\alpha[Y] = \Theta(y_\alpha)/N + o_N(1/N). \quad (13)$$

A similar result is noted by Martin and Tasche (2007) and Gordy (2004).

[8] It is also known in the literature as tail conditional expectation (TCE), conditional value-at-risk (CVaR), and expected tail loss (ETL).

3.4. Optimal Allocation Within the Inner Step

In this subsection, we relax the restriction that N_k is equal for $k = 1, 2, \ldots, K$. We focus on estimation of large loss probabilities. Similar analysis would allow us to vary N_k across positions in estimating VaR and ES.

We redefine N as the *total* number of inner step simulations. This aggregate N is to be divided up among the positions $k = 1, \ldots, K$ by allocating $p_k N$ simulations for position k, where each $p_k \geq 0$ and $\sum_{k \leq K} p_k = 1$. Suppose that the average effort to generate a single such inner step simulation for k is $\gamma_{1,k}$ so that total inner step simulation effort equals $N\gamma_1$, where $\gamma_1 = \sum_k p_k \gamma_{1,k}$. For a single trial of the outer step simulation, the inner step simulation generates the estimate

$$\widetilde{Y}(\xi) = Y(\xi) + \sum_{k=1}^{K} \frac{1}{p_k N} \sum_{i=1}^{p_k N} \zeta_{ki}(\xi).$$

Here we ignore minor technicalities associated with $p_k N$ not being an integer.

The analysis to compute the mean square error proceeds exactly as in §3.1. The resultant θ_u in this setting is

$$\theta_u = -\frac{1}{2}\frac{d}{du} f(u) \sum_{k=1}^{K} \frac{1}{p_k} \mathrm{E}[\sigma^2_{k,\xi} \mid Y = u] = \sum_{k=1}^{K} \frac{\theta_{u,k}}{p_k}, \quad (14)$$

where $\sigma^2_{k,\xi}$ denotes the variance of $\zeta_{k,\cdot}(\xi)$ conditioned on ξ, and where we define

$$\theta_{u,k} = -\frac{1}{2}\frac{d}{du} f(u) \mathrm{E}[\sigma^2_{k,\xi} \mid Y = u].$$

Recall from §3.1 that the mean square error at optimal N^* equals

$$3\left(\frac{\theta_u \alpha(1-\alpha)\gamma_1}{2\Gamma}\right)^{2/3} + o_\Gamma(\Gamma^{-2/3}).$$

Holding N^* fixed, we now consider the problem of determining an approximation to optimal $\{p_k \geq 0\colon k \leq K\}$. As u and hence α are fixed, it is reasonable to ignore the $o_\Gamma(\Gamma^{-2/3})$ residual and simply minimize the product $\theta_u \gamma_1$, which we can write as

$$\theta_u \gamma_1 = \left(\sum_{k=1}^{K} \frac{\theta_{u,k}}{p_k}\right) \times \left(\sum_{j=1}^{K} p_j \gamma_{1,j}\right).$$

Because the terms p_k and p_j appear as ratios in the objective, the constraint $\sum_{k \leq K} p_k = 1$ simply involves normalizing any solution of the unconstrained problem. From the first-order conditions, we can easily verify that the solution $(p^*_k\colon k \leq K)$ is

$$p^*_k = \frac{\sqrt{\theta_{u,k}/\gamma_{1,k}}}{\sum_{j=1}^{K} \sqrt{\theta_{u,j}/\gamma_{1,j}}}.$$

This is intuitive because one expects that a larger number of samples should be allocated to a position with higher contribution to bias and lower computational effort. This is captured by $\theta_{u,k}$ in the numerator and $\gamma_{1,k}$ in the denominator.

3.5. Large Portfolio Asymptotics

Intuition suggests that as the portfolio size increases, the optimal number of inner steps needed becomes small, even falling to $N^* = 1$ for a sufficiently large portfolio. We formalize this intuition by considering an asymptotic framework where both the portfolio size and the computational budget increase to infinity. The arguments that we provide in this section are somewhat heuristic, to give the flavor of the analysis involved while avoiding the cumbersome and lengthy notation and assumptions needed to make the analysis completely rigorous.

Let $W_k(\xi)$ denote the loss from position k for $k = 1, \ldots, K$. Let $\bar{Y}^K(\xi)$ be the average loss per position, i.e.,

$$\bar{Y}^K(\xi) = \frac{1}{K} \sum_{k=1}^{K} W_k(\xi).$$

Let $\alpha_K \equiv P(\bar{Y}^K(\xi) > u)$ be the desired exceedance probability for fixed K. As before, we do not observe $W_k(\xi)$ directly but rather must generate N inner step samples $W_k(\xi) + \zeta_{ki}(\xi)$ for $i = 1, \ldots, N$, where $\zeta_{ki}(\xi)$ have mean zero. Our simulation then provides an unbiased estimator for the probability $\alpha_{K,N} \equiv P(\bar{Y}^K(\xi) + \bar{Z}^{K,N}(\xi) > u)$, where

$$\bar{Z}^{K,N}(\xi) = \frac{1}{KN} \sum_{i=1}^{N} \sum_{k=1}^{K} \zeta_{ki}(\xi).$$

We approximate the bias $\alpha_{K,N} - \alpha_K$ as K and N grow large. Let $\bar{f}_K(\cdot)$ denote the pdf of $\bar{Y}^K(\xi)$, and let $\tilde{\sigma}^2_{K,\xi}$ denote the average conditional variance of the pricing errors, i.e.,

$$\tilde{\sigma}^2_{K,\xi} = \frac{1}{K} \sum_{k=1}^{K} \mathrm{V}[\zeta_{k,i}(\xi) | \xi]$$

for any trial $i = 1, \ldots, N$. Applying the same arguments as in the proof of Proposition 1, under assumptions analogous to Assumption 1, we can show

$$\alpha_{K,N} = \alpha_K + \theta^K_u/(KN) + o_{KN}(1/(KN)), \quad (15)$$

where $\theta^K_u = -\Theta'_K(u)$ and

$$\Theta_K(u) = \tfrac{1}{2} \bar{f}_K(u) \mathrm{E}[\tilde{\sigma}^2_{K,\xi} \mid \bar{Y}^K(\xi) = u].$$

We assume that $\bar{Y}^K(\xi)$ converges almost surely to a random variable $Y(\xi)$ as $K \to \infty$. This implies that the cdf $\bar{F}_K$ converges to a limiting distribution $\bar{F}$ of Y. We further assume that $\tilde{\sigma}^2_{K,\xi}$ converges to a limit $\tilde{\sigma}^2_\xi$ as $K \to \infty$. Under suitable regularity conditions, we

can show that $\mathrm{E}[\tilde{\sigma}^2_{K,\xi} \mid \bar{Y}^K(\xi) = u]$ converges to $\mathrm{E}[\tilde{\sigma}^2_\xi \mid Y(\xi) = u]$ and finally that $\theta^K_u \to \bar{\theta}_u = -\bar{\Theta}'(u)$, where

$$\bar{\Theta}(u) = \tfrac{1}{2}\bar{f}(u)\mathrm{E}[\tilde{\sigma}^2_\xi \mid Y(\xi) = u].$$

Therefore, the bias has the form $\bar{\theta}_u/(KN) + o_{KN}(1/KN)$.

Now consider an asymptotic regime where the available computational budget increases as a function of the portfolio size K. Specifically, assume that the computational budget $\Gamma = \chi K^\beta$ for constants $\chi > 0$ and β. Observe that we must have $\beta \geq 1$, because otherwise the available computational budget per position would shrink to zero as K grew toward infinity, so that asymptotically even a single sample could not be generated.

Suppose that the computational effort to generate one sample of outer step simulation on average equals $\psi(m, K)$, where ψ is some nondecreasing function and the number of underlying state variables (m) is held fixed. Assume that $\psi(m, K)$ is $O_K(K)$, and that to generate an inner step simulation sample on average equals $K\gamma_1$ for a constant $\gamma_1 > 0$. For a nested simulation of L outer step and N inner step trials, total computational effort is $L(KN\gamma_1 + \psi(m, K))$. We analyze the orders of magnitude of N and L that minimize the resultant mean square error of the estimator.

The mean square error of the estimator equals $\alpha_K(1 - \alpha_K)/L + (\bar{\theta}_u/(KN))^2$ plus terms that are relatively negligible for large values of K. By a relabeling (i.e., from N to K) and slight generalization of the arguments in the proof of Proposition 1, we have $\alpha_K = \alpha + O_K(1/K)$ for $\alpha \equiv P(Y(\xi) > u)$.[9] Thus, the dominant terms in the mean square error are

$$\alpha(1 - \alpha)/L + (\bar{\theta}_u/(KN))^2.$$

From the budget constraint, we can substitute $L = \chi K^{\beta-1}/(N\gamma_1 + \psi(m, K)/K)$. It is then easily verified that the value of $N \geq 1$ that minimizes the dominant terms in the mean square error equals

$$N^* = \max\left(1, \left(\frac{2\bar{\theta}^2_u \chi}{\alpha(1-\alpha)\gamma_1}\right)^{1/3} K^{\beta/3-1}\right).$$

In particular, if $\beta < 3$ then $N = 1$ for K sufficiently large. Intuitively, this means that if the portfolio has a large number of positions and the computational budget is limited, then N may be kept equal to 1. In this case L^* is of order $K^{\beta-1}$. Only when $\beta > 3$ does N grow with K.

[9] Indeed, the rate of convergence of α_K to α is the central theme in the granularity adjustment literature that inspired this study.

3.6. Simultaneous Repricing

Up to this point, we have stipulated that the inner step samples for each position in the portfolio are generated independently (conditional on ξ) across positions. In application to derivative portfolios, there may be factors common to many positions (e.g., the prices of underlying securities), and it may be computationally efficient to generate these factors once for all positions rather than generating them independently for each position. Although this may reduce the computational effort required to generate a single sample of each position, it induces dependence across positions in the generated samples. If the dependence is such that the sum of the resultant noise from each position has lesser variance than if these samples were generated independently, as might be the case when there are many offsetting positions, then the former is a preferred method. However, typically the noises generated may have positive dependence, and that may enhance the variance of the resulting samples, thereby increasing the total number of samples required to achieve specified accuracy.

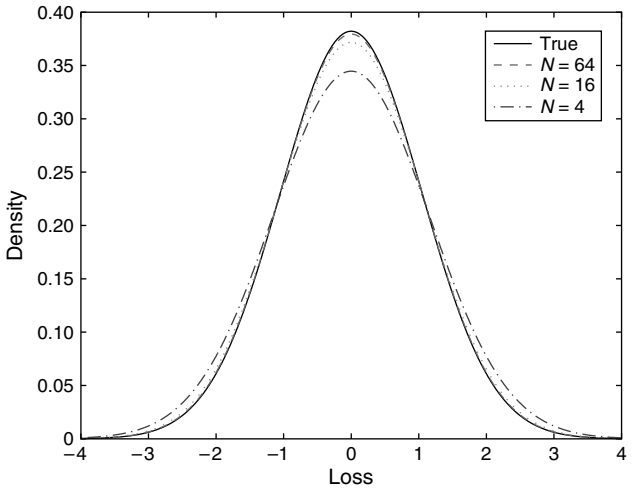
We now make this idea precise in a very simple setting. Consider the case where we want to find the expectation of $\sum_{i=1}^K X_i$ via simulation. Suppose that average computational effort needed to generate a sample of $\sum_{i=1}^K X_i$ by generating independent samples of $(X_1, \ldots, X_K)$ equals $K\gamma$ for some constant $\gamma > 0$. Let $\mathrm{V}[X]$ denote the variance of these X_i's (to keep the discussion simple, we assume that all rv have the same variance). Then the computational effort required to get a specified accuracy is proportional to the variance of the sample $K \cdot \mathrm{V}[X]$ times the expected effort required to generate a single sample $K\gamma$ (see Glynn and Whitt 1992), i.e., $K^2\gamma\mathrm{V}[X]$. We refer to this measure as the simulation efficiency.

Now consider the case where we generate $\sum_{i=1}^K X_i$ by generating dependent samples of $(X_1, \ldots, X_K)$. Suppose that the computational effort to generate these samples on average equals $K\gamma(1 - \beta)$ for some $\beta < 1$. Further suppose that the correlation between any two random variables X_i and X_j for $i \neq j$ is $\rho \in (0, 1)$. Then the variance of $\sum_{i=1}^K X_i$ equals $K(1 + \rho(K - 1))\mathrm{V}[X]$. So the simulation efficiency equals $K^2\gamma(1-\beta)(1+\rho(K-1))\mathrm{V}[X]$. We therefore prefer to draw dependent samples whenever $\beta > 1 - 1/(1 + \rho(K - 1))$. Unless $1 - \beta = o_K(1/K)$, we will prefer to draw independent samples for any K sufficiently large. This broadly indicates the benefits of independent samples in many finance settings. For the remainder of this paper, we reinstate the stipulation of independent repricing.

4. Numerical Examples

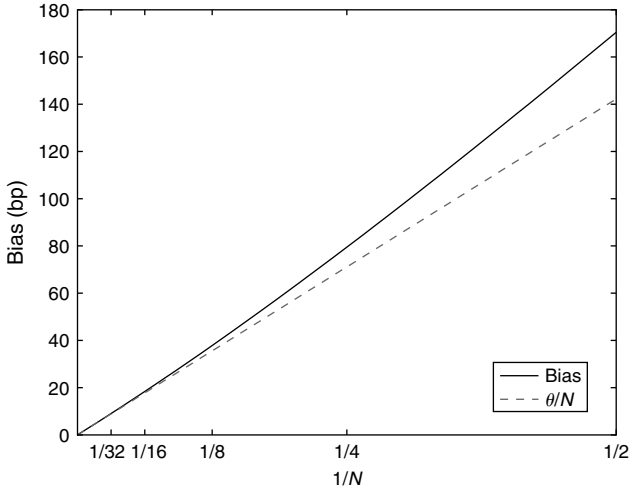
We illustrate our results with a parametric example. Distributions for loss Y and the inner step pricing

Figure 1 Density of the Loss Distribution



Note. Gaussian example with parameters: $\nu = 3$, $\eta = 10$, and $K = 100$.

Figure 2 Exact and Approximated Bias



Notes. The solid line plots exact bias for Gaussian example, and the dashed line plots first-order approximation of Proposition 1. Parameters: $\nu = 3$, $\eta = 10$, $K = 100$, and $u = \mathrm{VaR}_{0.01}[Y]$.

errors are specified to ensure that the bias and variance of our simulation estimators are in closed-form. Although the example is highly stylized, it allows us to compare our asymptotically optimal (N^*, L^*) to the exact optimal solution under a finite computational budget.

Consider a homogeneous portfolio of K positions. Let the state variable, X_t, represent a single-dimensional market risk factor and assume $X_H \sim \mathcal{N}(0,1)$. Let ϵ_k be the idiosyncratic component to the return on position k at the horizon so that the loss on position k is $(X_H + \epsilon_k)$ per unit of exposure. We assume that the ϵ_k are i.i.d. $\mathcal{N}(0, \nu^2)$. To facilitate comparative statics on K, we scale exposure sizes by $1/K$. The exact distribution for portfolio loss Y is then $\mathcal{N}(0, 1 + \nu^2/K)$.

We assume that the position-level inner step pricing errors $\zeta_{k,\cdot}$ are i.i.d. $\mathcal{N}(0, \eta^2)$ per unit of exposure so that the portfolio pricing error has variance $\sigma^2 = \eta^2/K$ across inner step trials. This implies that the simulated loss variable $\widetilde{Y}$ is distributed $\mathcal{N}(0, 1 + \nu^2/K + \sigma^2/N)$. Figure 1 shows how the density of the simulated loss distribution varies with the choice of N. For the baseline parameter values $\nu = 3$, $\eta = 10$, and $K = 100$, we observe that the density of $\widetilde{Y}$ for $N = 64$ is a close approximation to the "true" density for Y. Even for $N = 16$, the error due to inner step simulation appears modest.

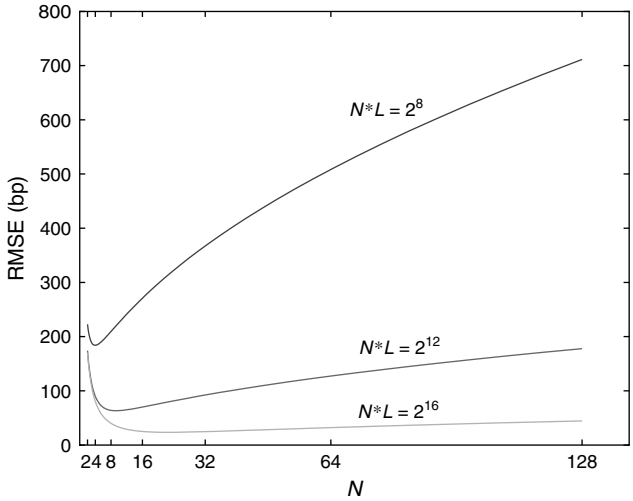
We consider our estimator $\hat{\alpha}_{L,N}$ of the large loss probability $\alpha = P(Y > u) = \Phi(-u/\sqrt{1 + \nu^2/K})$ for a fixed loss level u. The expected value of $\hat{\alpha}_{L,N}$ is $\alpha_N = P(\widetilde{Y} > u) = \Phi(-u/\sqrt{1 + \nu^2/K + \sigma^2/N})$. Applying Proposition 1, bias in $\hat{\alpha}_{L,N}$ expands as $\alpha_N - \alpha \approx \theta_u/N$, where

$$\theta_u = \frac{\sigma^2}{2} u (1 + \nu^2/K)^{-3/2} \phi(-u/\sqrt{1 + \nu^2/K})$$

and where ϕ is the standard normal density. Figure 2 plots the exact bias $\alpha_N - \alpha$ and the approximation θ_u/N as a function of $1/N$. The first-order approximation to the bias is quite accurate at modest values of N. At $N = 8$, the relative error of the approximation is roughly 6%, and even at $N = 2$, the relative error is less than 17%.

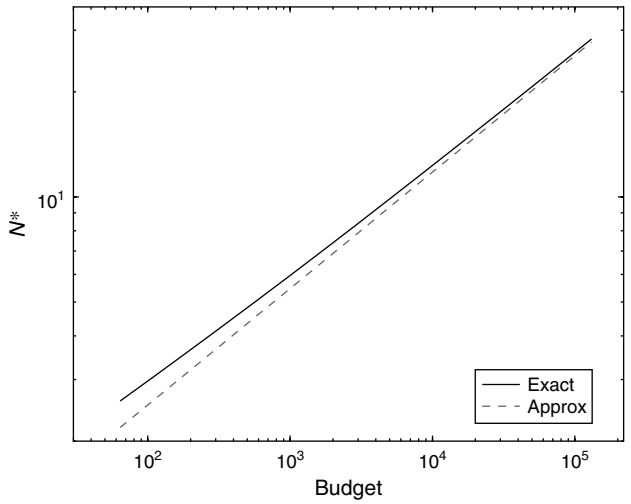
In Figure 3, we plot the exact root mean square error of $\hat{\alpha}_{L,N}$ as a function of N and with the computational budget held fixed. Here we make the assumption that the fixed cost of the outer step is negligible (i.e., $\gamma_0 \approx 0$) so that $\Gamma \propto N \cdot L$. We observe that the optimal N is increasing with the budget but remains

Figure 3 Root Mean Square Error



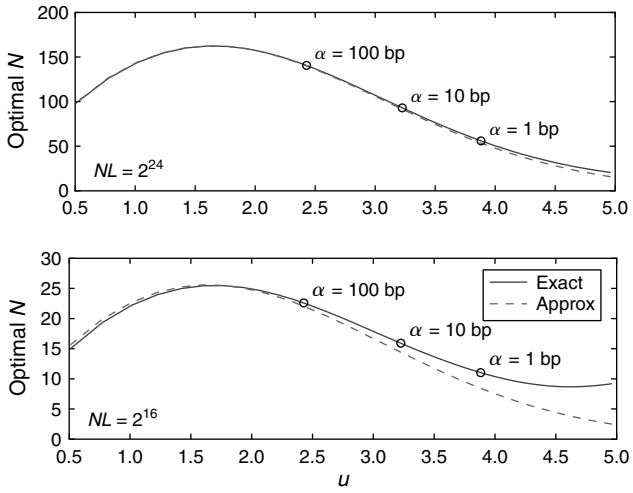
Notes. RMSE in Gaussian example. Each line depicts the relationship between N and RMSE for a fixed computational budget. Parameters: $\nu = 3$, $\eta = 10$, $K = 100$, and $u = \mathrm{VaR}_{0.01}[Y]$.

Figure 4 Optimal N Grows with Computational Budget



Notes. Budget is $\Gamma = N \cdot L$. Parameters: $\nu = 3$, $\eta = 10$, $K = 100$, and $u = \mathrm{VaR}_{0.01}[Y]$.

Figure 5 Optimal N Depends on Exceedance Threshold



Notes. Quantiles of the distribution of Y marked in basis points. Budget is $\Gamma = N \cdot L$. Parameters: $\nu = 3$, $\eta = 10$, and $K = 100$.

quite modest even for the largest budget depicted ($N \cdot L = 2^{16}$). The relationship between the computational budget and optimal N is explored further in Figure 4. We solve for the N that minimizes the (exact) mean square error and plot N^* as a function of the budget $\Gamma = N \cdot L$. Again, we see that N^* is much smaller than L^* and grows at a slower rate with Γ. For example, when $N \cdot L = 2^{10}$, we find N^* is under 6 and L^* is over 165. Increasing the budget by a factor of 64, we find N^* roughly quadruples while L^* increases by a factor of roughly 16. The figure demonstrates the accuracy of the approximation to N^* given by Equation (5). When $\Gamma = 2^{10}$, the relative error of the approximation is 8.3%. Increasing the budget to $\Gamma = 2^{16}$ shrinks the relative error to less than 2.5%.

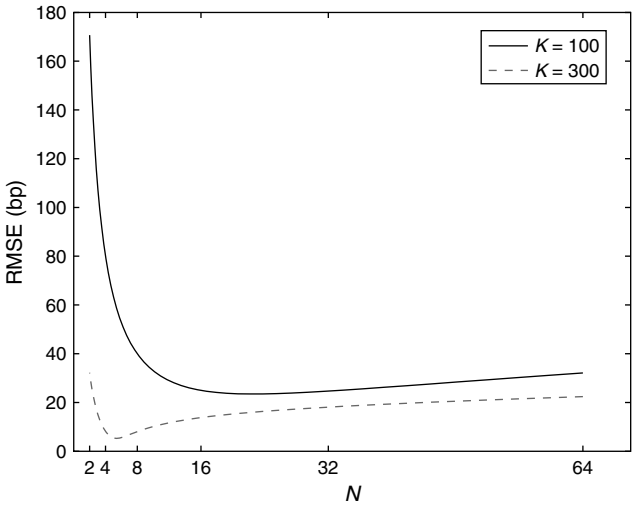
The optimal N^* and the accuracy of its approximation may depend on the exceedance threshold u, and not necessarily in a monotonic fashion. This is demonstrated in the top panel of Figure 5. The budget here is large, and we see that the approximation to N^* is accurate over the entire range of interest (say, for $u < \mathrm{VaR}_{0.0001}[Y]$). When the budget is smaller (bottom panel), the accuracy of the approximation is most severely degraded in the tail.

In Figure 6, we explore the relationship between portfolio size K and optimal N. For simplicity, we assume here that the budget grows linearly with K. In the baseline case of $K = 100$, we find that N^* is roughly 23. When we triple the portfolio size (and budget), we find that N^* falls to less than six. If we increase the portfolio size by a factor of 10, we find that N^* is less than two. These results suggest that the large portfolio asymptotics of §3.5 may pertain to portfolios of realistic size.

Because value-at-risk is the risk measure most commonly used in practice, we turn briefly to the estimation of VaR. The results of §3.2 show that the bias in $\hat{y}_\alpha$ vanishes with $1/N$. This is demonstrated in Figure 7 for three values of α. For each line, the y-axis intercept at $1/N = 0$ is the unbiased benchmark, i.e., $\mathrm{VaR}_\alpha[Y]$. Along the solid line, the distance on the y-axis between $\mathrm{VaR}_\alpha[\widetilde{Y}]$ for any finite N and the corresponding unbiased benchmark $\mathrm{VaR}_\alpha[Y]$ is the exact bias attributable to errors in pricing at the horizon. The dashed lines show the approximated VaR based on

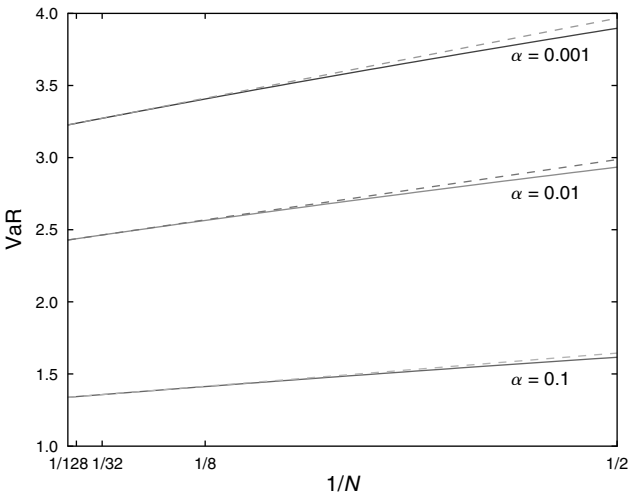
$$\mathrm{E}[\widetilde{Y}_{\lceil \alpha L \rceil}] \approx y_\alpha + \frac{1}{N}\frac{\theta_\alpha}{f(y_\alpha)} = y_\alpha + \frac{1}{N}\frac{\sigma^2}{2}\frac{\Phi^{-1}(1-\alpha)}{\sqrt{1+\nu^2/K}}.$$

Figure 6 Optimal N as Portfolio Size Varies



Notes. Budget is $\Gamma = N \cdot L = 2^{16}$ for $K = 100$ and grows linearly with K. Parameters: $\nu = 3$, $\eta = 10$, and $u = \mathrm{VaR}_{0.01}[Y]$.

Figure 7 Bias of Estimated Value-at-Risk



Notes. The solid lines are exact $\mathrm{VaR}_\alpha[\widetilde{Y}]$, and the dashed lines are approximations based on Proposition 2. Intercept on y-axis is $\mathrm{VaR}_\alpha[Y]$. Parameters: $\nu = 3$, $\eta = 10$, and $K = 100$.

Clearly, the linearized approximation to the bias is highly accurate in this example.

As a complement to the stylized Gaussian example, we have studied the somewhat more realistic example of a portfolio of equity options in a Black-Scholes setting. In this example, the loss density f does not have tractable form, so we cannot examine directly the performance of the approximations in Propositions 1 and 2. Nonetheless, the existence of closed-form pricing formulae as a benchmark does allow us to measure precisely the bias due to simulation noise in the inner step. We can also look at how mean square error varies with N for a fixed computational budget (as in Figure 3) and at how the resulting optimal N^* varies with the budget (as in Figure 4). All our conclusions in these exercises appear robust.

5. Jackknife Estimation

Jackknife estimators, introduced more than 50 years ago by Quenouille (1956), are commonly applied when bias reduction is desired. We show how jackknife methods can be applied in our setting to the estimation of large loss probabilities. Parallel methods would apply to VaR and ES.

We divide the inner step sample of N draws into I nonoverlapping *sections* (N and I are selected so that N/I is an integer). Section i covers the N/I draws $N((i-1)/I)+1, \ldots, N(i/I)$. For a single outer step trial, represented by ξ, let a_N denote the sample output from the inner step $1[\widetilde{Y}_\xi > u]$ as proposed in earlier sections. Let $\widetilde{Y}_\xi(-i)$ be the inner step estimate of Y_ξ that is obtained when section i is omitted, and define $a_N(-i)$ similarly, e.g., $a_N(-I) = 1[\widetilde{Y}_\xi(-I) > u]$. Observe that the bias in $a_N(-i)$ is $\theta/(N(I-1)/I)$ plus $O_N(1/N^{3/2})$ terms and furthermore that we can construct I different sample outputs this way.

We now propose the bias-corrected jackknife inner step simulation output

$$a^\dagger = Ia_N - (I-1)\frac{1}{I}\sum_{i=1}^{I} a_N(-i)$$
$$= a_N + \frac{I-1}{I}\sum_{i=1}^{I}(a_N - a_N(-i)). \qquad (16)$$

The bias in $a^\dagger$ is

$$\begin{aligned}
&\mathrm{E}[a^\dagger] - \alpha \\
&= I\alpha_N - (I-1)\alpha_{N(I-1)/I} - \alpha \\
&= I(\alpha + \theta/N + O_N(1/N^{3/2})) \\
&\quad - (I-1)(\alpha + \theta/(N(I-1)/I) + O_N(1/N^{3/2})) - \alpha \\
&= \theta\left(\frac{I}{N} - \frac{I-1}{(N(I-1)/I)}\right) + O_N(1/N^{3/2}) \\
&= O_N(1/N^{3/2}).
\end{aligned}$$

Thus, the first-order term in the bias is eliminated.

Define $b_N(-i) = a_N - a_N(-i)$ for $i \le I$. From the second representation of $a^\dagger$ in (16), we see that the variance of the estimator $a^\dagger$ is

$$\begin{aligned}
\mathrm{V}[a^\dagger] = \mathrm{V}[a_N] &+ \frac{(I-1)^2}{I}\mathrm{V}[b_N(-I)] \\
&+ 2(I-1)\mathrm{Cov}[a_N, b_N(-I)] \\
&+ \frac{(I-1)^3}{I}\mathrm{Cov}[b_N(-1), b_N(-I)], \qquad (17)
\end{aligned}$$

where $\mathrm{Cov}[X_1, X_2]$ denotes the covariance between X_1 and X_2. We now evaluate the dominant term in Equation (17) as $N \to \infty$. To keep the analysis simple, we assume that I is fixed even as N increases. As we note later, keeping I small has computational benefits. The following proposition is proved in §A.5 in the appendix.

Proposition 4. $\mathrm{V}[b_N(-i)] = O_N(1/N^{1/2})$.

From this result and the Cauchy-Schwartz inequality, we have

$$|\mathrm{Cov}[a_N, b_N(-I)]| \le \sqrt{\mathrm{V}[a_N]\cdot\mathrm{V}[b_N(-I)]} = O_N(1/N^{1/4}).$$

Similarly, $\mathrm{Cov}[b_N(-1), b_N(-I)]$ is $O_N(1/N^{1/2})$. Therefore, from Equation (17) for fixed I independent of N, $\mathrm{V}[a^\dagger] = \alpha(1-\alpha) + O_N(1/N^{1/4})$.

Our jackknife estimator is the average of L samples of $a^\dagger$, i.e., $\hat{\alpha}^\dagger = (1/L)\sum_{l=1}^{L} a^\dagger(\xi_l)$. The contribution of variance to its mean square error is

$$\mathrm{V}[\hat{\alpha}^\dagger] = \frac{\alpha(1-\alpha) + O_N(1/N^{1/4})}{L}.$$

The contribution from the squared bias term is $O_N(1/N^3)$. Suppose that the dominant term in the squared bias has the form π/N^3 (an expression for π could be obtained by taking one more term in the Taylor series expansion of α_N in the proof of Proposition 1). Proceeding as in §3.1, we assume that the computation effort takes the form $\Gamma = L(N\gamma_1 + \gamma_0)$ and choose N^* and L^* to minimize the mean square error subject to the budget constraint. We find

$$N^* = \left(\frac{3\pi}{\alpha(1-\alpha)\gamma_1}\right)^{1/4}\Gamma^{1/4} + o_\Gamma(\Gamma^{1/4}),$$

$$L^* = \left(\frac{\alpha(1-\alpha)}{3\pi\gamma_1^3}\right)^{1/4}\Gamma^{3/4} + o_\Gamma(\Gamma^{3/4}),$$

so even fewer inner step samples are needed for the jackknife estimator than for the estimator of §3.1. The mean square error at optimal N^* equals

$$4\pi^{1/4}\left(\frac{\alpha(1-\alpha)\gamma_1}{3\Gamma}\right)^{3/4} + o_\Gamma(\Gamma^{-3/4}).$$

Thus, the rate of convergence of the mean square error to zero reduces from order $\Gamma^{-2/3}$ to a faster $\Gamma^{-3/4}$.

Under some circumstances, the MSE will converge even more quickly. If we assume that the normalized sum of pricing errors

$$\widetilde{Z}_N = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{k=1}^{K}\zeta_{ki}(\xi)$$

has a Gaussian distribution with mean zero, then the third moment $\mathrm{E}[\widetilde{Z}_N^3 \mid Y = u]$ is zero. The $O_N(1/N^{3/2})$ term drops out in the Taylor expansion underpinning Proposition 1, so that

$$\alpha_N = \alpha + \theta_u/N + c/N^2 + O_N(1/N^{5/2})$$

for an appropriate constant c. This is also observed by Lee (1998), who further generalized the result for non-Gaussian $\widetilde{Z}_N$ under certain restrictions on the distribution of each $\zeta_{ki}(\xi)$. When this expansion for the bias term holds, it follows that the bias for the jackknife bias-corrected estimator $a^\dagger$ is dominated by $\tilde{c}/N^2$ for some constant $\tilde{c}$. Therefore, the optimal N^* and L^* that minimize the asymptotic mean square error are proportional to $\Gamma^{1/5}$ and $\Gamma^{4/5}$, respectively. Furthermore, under this allocation, the rate of convergence of the mean square error to zero is $\Gamma^{-4/5}$.

In most applications of jackknife methods, bias reduction comes at the price of increased variance. This trade-off applies here as well. To illustrate, we return to the Gaussian example of §4. For any parametric example, it is useful to rewrite the terms in Equation (17) for the variance of $a^\dagger$ as

$$\mathrm{V}[b_N(-I)] = \mathrm{V}[a_N] + \mathrm{V}[a_N(-I)] - 2\mathrm{Cov}[a_N, a_N(-I)],$$

$$\mathrm{Cov}[a_N, b_N(-I)] = \mathrm{V}[a_N] - \mathrm{Cov}[a_N, a_N(-I)],$$

$$\begin{aligned}\mathrm{Cov}[b_N(-1), b_N(-I)] = {}& \mathrm{V}[a_N] - \mathrm{Cov}[a_N, a_N(-I)] \\ & - \mathrm{Cov}[a_N, a_N(-1)] \\ & + \mathrm{Cov}[a_N(-1), a_N(-I)].\end{aligned}$$

By the symmetry of the $a_N(-i)$, we have

$$\begin{aligned}&\mathrm{Cov}[a_N, a_N(-1)] \\ &\quad = \mathrm{Cov}[a_N, a_N(-I)] \\ &\quad = P(\widetilde{Y} > u, \widetilde{Y}(-I) > u) - \alpha_N\alpha_{N(I-1)/I}. \qquad (18)\end{aligned}$$

Similarly, we have

$$\begin{aligned}&\mathrm{Cov}[a_N(-1), a_N(-I)] \\ &\quad = P(\widetilde{Y}(-1) > u, \widetilde{Y}(-I) > u) - \alpha^2_{N(I-1)/I}. \qquad (19)\end{aligned}$$

For the Gaussian example, let

$$U_n = \frac{-u}{\sqrt{1 + \nu^2/K + \sigma^2/n}}$$

so that $\alpha = \Phi(U_\infty)$ and $\alpha_N = \Phi(U_N)$. The bivariate probabilities in Equations (18) and (19) are

$$\begin{aligned}&P(\widetilde{Y} > u, \widetilde{Y}(-I) > u) \\ &\quad = \Phi_2\left(U_N, U_{N(I-1)/I}, \sqrt{\frac{1+\nu^2/K+\sigma^2/N}{1+\nu^2/K+\sigma^2/(N(I-1)/I)}}\right),\end{aligned}$$

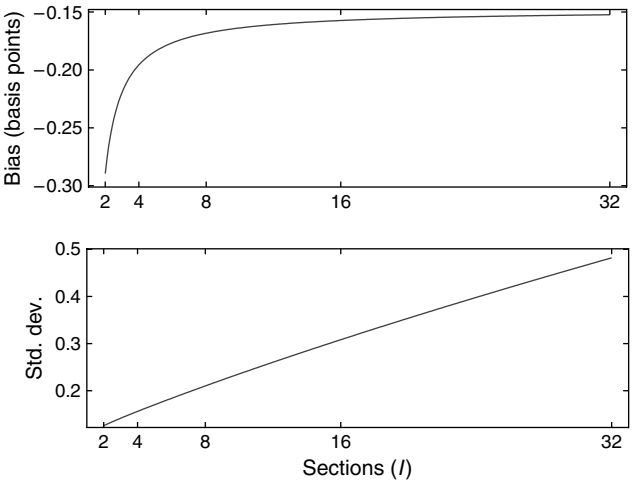
$$\begin{aligned}&P(\widetilde{Y}(-1) > u, \widetilde{Y}(-I) > u) \\ &= \Phi_2\Bigg(U_{N(I-1)/I}, U_{N(I-1)/I}, \\ &\qquad \frac{1+\nu^2/K+(\sigma^2/(N(I-1)/I))\cdot((I-2)/(I-1))}{1+\nu^2/K+\sigma^2/(N(I-1)/I)}\Bigg),\end{aligned}$$

where $\Phi_2(z_1, z_2, \rho)$ is the bivariate normal cdf for standard normal marginals and correlation ρ.

Holding fixed $N = 32$, we plot the bias of the jackknife estimator as a function of the number of sections (I) in the top panel of Figure 8. The magnitude of the bias is decreasing in I, but the sensitivity to I is modest in both absolute terms and relative to the bias of the uncorrected estimator. The bias of the uncorrected a_N is 9.04 basis points, whereas the bias of $a^\dagger$ is -0.29 basis points when $I = 2$ and -0.15 basis points when $I = N = 32$. The standard deviation of $a^\dagger$ is plotted as a function of I in the bottom panel of the figure. We find the standard deviation increases in roughly linear fashion with I. For the uncorrected estimator, we find $\sqrt{\mathrm{V}[a_N]} = 0.104$, whereas $\sqrt{\mathrm{V}[a^\dagger]} = 0.126$ for $I = 2$ and $\sqrt{\mathrm{V}[a^\dagger]} = 0.482$ for $I = N = 32$.

The optimal choice of I for minimizing mean square error will depend on L. The larger is L, the smaller the contribution of variance to MSE, so the larger the optimal I. As a practical matter, we advocate setting $I = 2$, which eliminates nearly all the bias at little cost

Figure 8 Bias and Variance of Jackknife Estimators



Note. Parameters: $N = 32$, $\nu = 3$, $\eta = 10$, $K = 100$, and $u = \text{VaR}_{0.01}[Y]$.

to variance. Setting $I = 2$ has the further advantage of minimal memory overhead, in that one can estimate $1[\widetilde{Y}_\xi(-1) > u]$, $1[\widetilde{Y}_\xi(-2) > u]$, and $1[\widetilde{Y}_\xi > u]$ for each outer step ξ in a single pass through the inner step. By contrast, to implement the jackknife with $I = N$ sections, we need to save each of the N inner step draws in order to calculate each of the N estimates $1[\widetilde{Y}_\xi(-1) > u], \ldots, 1[\widetilde{Y}_\xi(-N) > u]$.

In §§3.2 and 3.3, we observed that sampling error in the inner step is not the only source of bias in estimation of VaR and expected shortfall. Taking the case of VaR, the bias $\text{E}[\widetilde{Y}_{\lceil \alpha L \rceil}] - y_\alpha$ can be decomposed as $\text{E}[\widetilde{Y}_{\lceil \alpha L \rceil}] - \text{E}[Y_{\lceil \alpha L \rceil}]$ and $\text{E}[Y_{\lceil \alpha L \rceil}] - y_\alpha$. The first term is due to sampling error in the inner step, which is $O(1/N)$ when the jackknife method is not used. The second term is the familiar bias associated with quantile estimation based on empirical samples. This "outer step" bias is $O(1/L)$, so it is dominated by the sampling error term in optimal allocation of computational budget. In this paper, we do not address the bias in term $\text{E}[Y_{\lceil \alpha L \rceil}] - y_\alpha$. Kim and Hardy (2007) develop a bootstrapping technique for reducing outer step bias in estimation of VaR and ES. They achieve better success in reducing bias for ES compared to VaR.

6. Conclusion

We have shown that nested simulation of loss distributions poses a much less formidable computational obstacle than it might initially appear. The essential intuition is similar to the intuition for diversification in the long-established literature on portfolio choice. In the context of a large, well-diversified portfolio, risk-averse investors need not avoid high-variance investments so long as most of the variance is idiosyncratic to the position. In our risk-measurement problem, we see that large errors in pricing at the model horizon can be tolerated so long as the errors are zero mean and idiosyncratic. In the aggregate, such errors have modest impact on our estimated loss distribution. More formally, we are able to quantify that impact in terms of bias and variance of the resulting estimator and allocate workload in the simulation algorithm to minimize mean square error.

Extensions on our basic nested algorithm can improve performance. We show that a simple jackknife extension can eliminate much of the bias at modest cost to variance. In more sophisticated schemes, such as dynamic allocation (Gordy and Juneja 2008, §5) and restarting (Boesel et al. 2003, Lesnevski et al. 2007), one chooses the number of inner step trials as a function of the generated output. The idea is to redirect computational effort toward trials that appear likely to generate large losses. Such schemes can significantly reduce the computational effort to achieve a given level of accuracy.

Our results suggest that current practice is misguided. To avoid the "inner step" simulation for repricing at the model horizon, practitioners typically rely on overly stylized pricing models with closed-form solution. Unlike the simulation pricing error in our nested approach, the pricing errors that arise due to model misspecification are difficult to quantify, need not be zero mean, and are likely to be correlated across positions in the portfolio. At the portfolio level, therefore, the error in estimates of value-at-risk (or other quantities of interest) cannot readily be bounded and does not vanish asymptotically. Our results imply that practitioners should retain the best pricing models that are available, regardless of their computational tractability. A single trial of a simulation algorithm for the preferred model will often be less costly than a single call to the stylized pricing function, so running a nested simulation with a small number of trials in the inner step may be comparable in computational burden to the traditional approach. Despite the high likelihood of grotesque pricing errors at the instrument level, the impact on estimated VaR is small. In the limit of an asymptotically large, fine-grained portfolio, even a *single* inner step trial per instrument is sufficient to obtain exact pricing at the portfolio level.

Our methods have application to other problems in finance. Nested simulation may arise in pricing options on complex derivatives (e.g., a European call option on a CDO tranche). When parameters in an option pricing model are estimated with uncertainty, nested simulation may also be needed to determine confidence intervals on model prices for thinly traded complex derivatives. Similar problems arise in the rating of CDOs and other structured debt instruments when model parameters are subject to uncertainty. These applications will be developed in future work.

Acknowledgments

The authors thank Aaron Brown and Darrell Duffie for helpful suggestions. Much of this work was conducted while both authors were visiting the Indian School of Business. The opinions expressed here are the authors' and do not reflect the views of the Board of Governors or its staff.

Appendix

A.1. Proof of Lemma 1

Our asymptotic analysis relies on Taylor series expansion of the joint density function $g_N(y,z)$ of $(Y,\widetilde{Z}_N)$ and its partial derivatives. Note that

$$P(\widetilde{Y}\le y)=\int_{\Re}\int_{-\infty}^{y-z/\sqrt{N}} g_N(v,z)\,dv\,dz.$$

Differentiating both sides by y, noting that in the right-hand side (RHS) the interchange of integral and the derivative follows from Assumption 1 (Lang 1968, p. 249), we get

$$\tilde{f}_N(y)=\int_{\Re} g_N(y-z/\sqrt{N},z)\,dz.$$

Therefore,

$$\tilde{f}_N(y_N)-f(y)=\int_{\Re}(g_N(y_N-z/\sqrt{N},z)-g_N(y,z))\,dz.$$

This equals

$$(y_N-y)\int_{\Re}\frac{\partial}{\partial y}g_N(\check{y}_N,z)\,dz-\frac{1}{\sqrt{N}}\int_{\Re} z\frac{\partial}{\partial y}g_N(\check{y}_N,z)\,dz,$$

where $\check{y}_N$ lies between $y_N-z/\sqrt{N}$ and y. Clearly, due to Assumption 1, both these terms converge to zero. Similarly, we see that $\tilde{f}_N'(y_N)\to f'(y)$ as $N\to\infty$.

A.2. Proof of Proposition 1

Note that

$$\alpha_N=\int_{\Re}\int_{u-z/\sqrt{N}}^{\infty} g_N(y,z)\,dy\,dz,$$

so that

$$\alpha_N-\alpha=\int_{\Re}\int_{u-z/\sqrt{N}}^{u} g_N(y,z)\,dy\,dz. \tag{20}$$

Consider the Taylor series expansion of the pdf in (20),

$$g_N(y,z)=g_N(u,z)+(y-u)\frac{\partial}{\partial y}g_N(u,z)+\frac{(y-u)^2}{2}\frac{\partial^2}{\partial y^2}g_N(u_y,z),$$

where u_y denotes an appropriate number between u and y. From this and Assumption 1, it follows that

$$\alpha_N-\alpha=\int_{\Re}\frac{z}{\sqrt{N}}g_N(u,z)\,dz-\int_{\Re}\frac{z^2}{2N}\frac{\partial}{\partial y}g_N(u,z)\,dz+O_N(1/N^{3/2}).$$

The first term on the RHS equals

$$\frac{f(u)}{\sqrt{N}}\mathrm{E}[\widetilde{Z}_N\mid Y=u].$$

This term equals zero because

$$\mathrm{E}[\widetilde{Z}_N\mid Y=u]=\mathrm{E}[\mathrm{E}[\widetilde{Z}_N(\xi)\mid Y(\xi)=u,\xi]]=0.$$

The second term can be rewritten as

$$\begin{aligned}\int_{\Re} z^2\frac{\partial}{\partial y}g_N(u,z)\,dz&=\frac{d}{du}\int_{\Re} z^2 g_N(u,z)\,dz\\&=\frac{d}{du}f(u)\mathrm{E}[\widetilde{Z}_N^2\mid Y=u]\\&=\frac{d}{du}f(u)\mathrm{E}[\mathrm{E}[\widetilde{Z}_N^2\mid\xi]\mid Y=u]\\&=\frac{d}{du}f(u)\mathrm{E}[\sigma_\xi^2\mid Y=u],\end{aligned}$$

which completes the proof.

A.3. Mean Square Error for Value-at-Risk

We derive the bias and variance approximations of Proposition 2. Our treatment in places is heuristical to avoid lengthy technical issues. We first develop an asymptotic expression for the bias $\mathrm{E}[\widetilde{Y}_{\lceil\alpha L\rceil}]-y_\alpha$. To evaluate $\mathrm{E}[\widetilde{Y}_{\lceil\alpha L\rceil}]$, it is useful to consider the order statistics $U_{[1]}\ge\cdots\ge U_{[L]}$ of L random variables uniformly distributed over the unit interval. Within this section only, let $\widetilde{G}_N$ denote the inverse of the cdf of $\widetilde{Y}$. Observe that $\widetilde{Y}_{\lceil\alpha L\rceil}$ has the same distribution as $\widetilde{G}_N(U_{\lceil\alpha L\rceil})$.

Consider the Taylor series expansion

$$\begin{aligned}&\widetilde{G}_N(U_{\lceil\alpha L\rceil})\\&\quad=\widetilde{G}_N(\mathrm{E}[U_{\lceil\alpha L\rceil}])+(U_{\lceil\alpha L\rceil}-\mathrm{E}[U_{\lceil\alpha L\rceil}])\widetilde{G}_N'(\mathrm{E}[U_{\lceil\alpha L\rceil}])\\&\qquad+\tfrac{1}{2}(U_{\lceil\alpha L\rceil}-\mathrm{E}[U_{\lceil\alpha L\rceil}])^2\widetilde{G}_N''(\mathrm{E}[U_{\lceil\alpha L\rceil}])\\&\qquad+\text{ remainder terms.}\end{aligned}\tag{21}$$

Taking expectations

$$\begin{aligned}\mathrm{E}[\widetilde{Y}_{\lceil\alpha L\rceil}]&=\mathrm{E}[\widetilde{G}_N(U_{\lceil\alpha L\rceil})]\\&=\widetilde{G}_N(\mathrm{E}[U_{\lceil\alpha L\rceil}])+\tfrac{1}{2}\mathrm{V}[U_{\lceil\alpha L\rceil}]\widetilde{G}_N''(\mathrm{E}[U_{\lceil\alpha L\rceil}])\\&\qquad+\mathrm{E}[\text{remainder terms}].\end{aligned}\tag{22}$$

It is well known (see, e.g., David 1981) that

$$\mathrm{E}[U_{\lceil\alpha L\rceil}]=1-\lceil\alpha L\rceil/(L+1)=1-\alpha+O_L(1/L),\tag{23}$$

$$\mathrm{V}[U_{\lceil\alpha L\rceil}]=\frac{\alpha(1-\alpha)}{L+2}+O_L(1/L^2).\tag{24}$$

Through differentiation of the relation $\widetilde{F}_N(\widetilde{G}_N(u))=u$, we obtain

$$\widetilde{G}_N''(u)=-\frac{\tilde{f}_N'(\widetilde{G}_N(u))}{\tilde{f}_N(\widetilde{G}_N(u))^3}.$$

Letting $\hat{y}_\alpha=\widetilde{G}_N(\mathrm{E}[U_{\lceil\alpha L\rceil}])$, it can be seen that

$$\mathrm{E}[\widetilde{Y}_{\lceil\alpha L\rceil}]=\hat{y}_\alpha-\frac{\alpha(1-\alpha)}{2(L+2)}\frac{\tilde{f}_N'(\hat{y}_\alpha)}{\tilde{f}_N(\hat{y}_\alpha)^3}+O_L(1/L^2)(\text{constant}+o_N(1)).\tag{25}$$

Here we omit the lengthy discussion on the technical assumptions needed to ensure that the expectation of the remainder terms in (22) has the form $O_L(1/L^2)(\text{constant}+o_N(1))$.

Let $\tilde{y}_\alpha = \widetilde{G}_N(1-\alpha)$ denote the α quantile corresponding to the random variable $\widetilde{Y}$. If we can show that

$$\hat{y}_\alpha = \tilde{y}_\alpha + O_L(1/L) \tag{26}$$

and that

$$\tilde{y}_\alpha - y_\alpha = \frac{\theta_\alpha}{Nf(y_\alpha)} + o_N(1/N), \tag{27}$$

where $\theta_\alpha = -\Theta(y_\alpha) = (-1/2)(d/du)f(u)\mathrm{E}[\sigma_\xi^2 \mid Y = u]|_{u=y_\alpha}$, then

$$\hat{y}_\alpha = y_\alpha + \frac{\theta_\alpha}{Nf(y_\alpha)} + O_L(1/L) + o_N(1/N). \tag{28}$$

To see that

$$\frac{\alpha(1-\alpha)}{2(L+2)}\frac{\tilde{f}_N'(\hat{y}_\alpha)}{\tilde{f}_N(\hat{y}_\alpha)^3}$$

is $O_L(1/L)(1+o_N(1))$, note that due to Assumption 1, $\tilde{f}_N(\cdot)$ and $\tilde{f}_N'(\cdot)$ have bounded derivatives, which implies

$$\begin{aligned}\tilde{f}_N(\hat{y}_\alpha) &= \tilde{f}_N(y_\alpha) + O_N(1/N) + O_L(1/L)\\ &= f(y_\alpha) + o_N(1) + O_L(1/L).\end{aligned}$$

Similarly,

$$\tilde{f}_N'(\hat{y}_\alpha) = f'(y_\alpha) + o_N(1) + O_L(1/L).$$

In particular, from this, (25), and (28), it follows that

$$\mathrm{E}[\widetilde{Y}_{\lceil \alpha L\rceil}] = y_\alpha + \frac{\theta_\alpha}{Nf(y_\alpha)} + O_L(1/L) + O_L(1/L)o_N(1) + o_N(1/N).$$

Equation (26) follows because Equation (23) implies that

$$\hat{y}_\alpha = \widetilde{G}_N(1-\alpha) + O_L(1/L) = \tilde{y}_\alpha + O_L(1/L)$$

for N sufficiently large.

We now show (27). Using the Taylor series expansion,

$$\alpha = P(\widetilde{Y} > \tilde{y}_\alpha) = P(\widetilde{Y} > y_\alpha) - (\tilde{y}_\alpha - y_\alpha)\tilde{f}_N(y_\alpha) + (\tilde{y}_\alpha - y_\alpha)^2\tilde{f}_N'(\check{y}_\alpha)$$

where $\check{y}_\alpha$ lies between $\tilde{y}_\alpha$ and y_α. From Assumption 1 it follows that $\tilde{f}_N'(y)$ is uniformly bounded for all y. By Proposition 1, we have

$$P(\widetilde{Y} > y_\alpha) = P(Y > y_\alpha) + \theta_\alpha/N + O_N(1/N^{3/2}).$$

Because $P(Y > y_\alpha) = \alpha$, it follows that

$$\tilde{y}_\alpha - y_\alpha = \frac{\theta_\alpha}{N\tilde{f}_N(y_\alpha)} + o_N(1/N).$$

Because $\tilde{f}_N(y_\alpha) \to f(y_\alpha)$, (27) follows.

The expression for variance can be determined by subtracting (25) from (21), squaring the difference, and substituting $\widetilde{G}_N'(u) = 1/\tilde{f}_N(\widetilde{G}_N(u))$ to get

$$\mathrm{V}[\widetilde{Y}_{\lceil \alpha L\rceil}] = \frac{\alpha(1-\alpha)}{(L+2)\tilde{f}_N(\tilde{y}_\alpha)^2} + O_L(1/L^2) + O_L(1/L)\cdot o_N(1).$$

Again from Lemma 1, $\tilde{f}_N(\tilde{y}_\alpha) \to f(y_\alpha)$ as $N \to \infty$. Therefore,

$$\mathrm{V}[\widetilde{Y}_{\lceil \alpha L\rceil}] = \frac{\alpha(1-\alpha)}{(L+2)f(y_\alpha)^2} + O_L(1/L^2) + O_L(1/L)\cdot o_N(1).$$

A.4. Proof of Proposition 3

The conditional expected loss functions $\Upsilon(\cdot)$ and $\Upsilon_N(\cdot)$ can be written as

$$\begin{aligned}\Upsilon_N(u) &= \mathrm{E}[\widetilde{Y}\cdot 1[\widetilde{Y} > u]]\\ &= \int_{\Re}\int_{u-z/\sqrt{N}}^{\infty}(y + z/\sqrt{N})g_N(y,z)\,dy\,dz,\end{aligned} \tag{29}$$

$$\Upsilon(u) = \mathrm{E}[Y\cdot 1[Y > u]] = \int_{\Re}\int_{u}^{\infty} y g_N(y,z)\,dy\,dz. \tag{30}$$

As $\mathrm{E}[\widetilde{Z}_N \mid Y = y] = 0$ for each y, we have

$$\int_{\Re}\int_{u}^{\infty} z g_N(y,z)\,dy\,dz = \int_{u}^{\infty}\mathrm{E}[\widetilde{Z}_N \mid Y = y]f(y)\,dy = 0.$$

Substituting into (29), we have

$$\begin{aligned}\Upsilon_N(u) - \Upsilon(u) &= \int_{\Re}\int_{u-z/\sqrt{N}}^{u} y g_N(y,z)\,dy\,dz\\ &\quad + \int_{\Re} z/\sqrt{N}\int_{u-z/\sqrt{N}}^{u} g_N(y,z)\,dy\,dz.\end{aligned} \tag{31}$$

We apply a Taylor series expansion to the first term on the RHS:

$$\begin{aligned}&\int_{\Re}\int_{u-z/\sqrt{N}}^{u} y g_N(y,z)\,dy\,dz\\ &\quad= \int_{\Re}\int_{u-z/\sqrt{N}}^{u} y\Bigg(g_N(u,z) + (y-u)\frac{\partial}{\partial}g_N(u,z)\\ &\qquad\qquad + \frac{(y-u)^2}{2}\frac{\partial^2}{\partial y^2}g_N(u_y,z)\Bigg)\,dy\,dz\\ &\quad= \int_{\Re}\left(\frac{uz}{N^{1/2}} - \frac{z^2}{2N}\right)g_N(u,z)\,dz\\ &\qquad - \int_{\Re}\left(\frac{uz^2}{2N} - \frac{z^3}{3N^{3/2}}\right)\frac{\partial}{\partial y}g_N(u,z)\,dz + o_N(1/N^{3/2})\\ &\quad= -\frac{1}{2N}f(u)\mathrm{E}[\widetilde{Z}_N^2 \mid Y = u]\\ &\qquad - \frac{u}{2N}\frac{d}{du}(f(u)\mathrm{E}[\widetilde{Z}_N^2 \mid Y = u]) + O_N(1/N^{3/2}).\end{aligned}$$

Here in the first equality, u_y denotes an appropriate number between u and y; in the third equality we use the fact that $\int_{\Re} z g_N(u,z)\,dz = 0$.

For the second term in (31),

$$\begin{aligned}&\int_{\Re} z/\sqrt{N}\int_{u-z/\sqrt{N}}^{u} g_N(y,z)\,dy\,dz\\ &\quad= \int_{\Re} z/\sqrt{N}\int_{u-z/\sqrt{N}}^{u}\Bigg(g_N(u,z) + (y-u)\frac{\partial}{\partial y}g_N(u,z)\\ &\qquad\qquad + \frac{(y-u)^2}{2}\frac{\partial^2}{\partial y^2}g_N(u_y,z)\Bigg)\,dy\,dz\\ &\quad= \int_{\Re}\frac{z^2}{N}g_N(u,z)\,dz - \int_{\Re}\frac{z^3}{2N^{3/2}}\frac{\partial}{\partial y}g_N(u,z)\,dz + o_N(1/N^{3/2})\\ &\quad= \frac{1}{N}f(u)\mathrm{E}[\widetilde{Z}_N^2 \mid Y = u] + O_N(1/N^{3/2}).\end{aligned}$$

Proposition 3 follows by noting that

$$\mathrm{E}[\widetilde{Z}_N^2 \mid Y = u] = \mathrm{E}[\mathrm{E}[\widetilde{Z}_N^2 \mid Y, \xi] \mid Y = u] = \mathrm{E}[\sigma_\xi^2 \mid Y = u].$$

A.5. Proof of Proposition 4

We decompose the variance of $b_N(-i)$ as $\mathrm{E}[b_N(-i)^2] - \mathrm{E}[b_N(-i)]^2$. The expectation of $b_N(-i)$ is

$$\begin{aligned}\mathrm{E}[b_N(-i)] &= \mathrm{E}[a_N] - \mathrm{E}[a_N(-i)] \\ &= \alpha + \frac{\theta_u}{N} + O_N(1/N^{3/2}) \\ &\quad - \left(\alpha + \frac{\theta_u}{N(I-1)/I} + O_N(1/N^{3/2})\right) \\ &= \frac{-\theta_u}{N(I-1)} + O_N(1/N^{3/2}),\end{aligned}$$

which implies that $\mathrm{E}[b_N(-i)]^2 = O_N(1/N^2)$.

We now argue that $\mathrm{E}[b_N(-I)^2]$ is $O_N(1/N^{1/2})$, ignoring some minor technical issues. From expanding the square and taking expectations, observe that

$$\begin{aligned}\mathrm{E}[b_N(-I)^2] = {} & P(Y + \bar{Z}^N > u,\, Y + \bar{Z}^{N(I-1)/I} \le u) \\ & + P(Y + \bar{Z}^N \le u,\, Y + \bar{Z}^{N(I-1)/I} > u). \quad (32)\end{aligned}$$

We now show that these two probabilities are equal in their dominant term and are $O(1/N^{1/2})$. Some notation is needed for this purpose. Let

$$A_N \equiv \left(\frac{I}{N}\right)^{1/2} \sum_{i=N(I-1)/I}^{N} Z_i,$$

$$B_N \equiv \left(\frac{I}{N(I-1)}\right)^{1/2} \sum_{i=1}^{N(I-1)/I} Z_i.$$

We assume that both A_N and B_N have limiting distributions with finite expectations. Then

$$\bar{Z}^N = \frac{1}{N^{1/2}}((1/I)^{1/2} A_N - ((I-1)/I)^{1/2} B_N).$$

Let $g_N^\dagger$ denote the joint pdf of (Y, A_N, B_N). Then the first term in (32) can be written as

$$\begin{aligned}&P(Y + \bar{Z}^N > u,\, Y + \bar{Z}^{N(I-1)/I} \le u) \\ &= \int_{a \ge b/(I-1)^{1/2}} \int_{u - a/(NI)^{1/2} - b/(NI/(I-1))^{1/2}}^{u - b/(N(I-1)/I)^{1/2}} g_N^\dagger(y, a, b)\, dy\, da\, db.\end{aligned}$$

Taking a Taylor series expansion of $g_N^\dagger(y, a, b)$ at the first argument set to u and under assumptions similar to Assumption 1, this can be seen to equal

$$\begin{aligned}&\frac{1}{(NI)^{1/2}} f(u) \mathrm{E}[(A_N - B_N/(I-1)^{1/2}) \\ &\quad \cdot 1[A_N - B_N/(I-1)^{1/2} > 0] \mid Y = u] + O_N(1/N).\end{aligned}$$

Because A_N and B_N have limiting distributions with finite expectations, the expectation term converges to a constant as $N \to \infty$, which implies that

$$\begin{aligned}&\frac{1}{(NI)^{1/2}} f(u) \mathrm{E}[(A_N - B_N/(I-1)^{1/2}) \\ &\quad \cdot 1[A_N - B_N/(I-1)^{1/2} > 0] \mid Y = u] = O_N(1/N^{1/2}).\end{aligned}$$

We proceed in exactly the same fashion for the second term in (32) and find that this term is also $O_N(1/N^{1/2})$.

References

Acerbi, C., D. Tasche. 2002. On the coherence of expected shortfall. *J. Banking Finance* **26**(7) 1487–1503.

Artzner, P., F. Delbaen, J.-M. Eber, D. Heath. 1999. Coherent measures of risk. *Math. Finance* **9**(3) 203–228.

Boesel, J., B. L. Nelson, S.-H. Kim. 2003. Using ranking and selection to "clean up" after simulation optimization. *Oper. Res.* **51**(5) 814–825.

David, H. A. 1981. *Order Statistics*, 2nd ed. John Wiley & Sons, New York.

De Prisco, B., I. Iscoe, Y. Jiang, H. Mausser. 2007. Compound scenarios: An efficient framework for integrated market-credit risk. *J. Risk* **11**(2) 1–36.

Duffie, D. 1996. *Dynamic Asset Pricing Theory*, 2nd ed. Princeton University Press, Princeton, NJ.

Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York.

Glynn, P. W., W. Whitt. 1992. The asymptotic efficiency of simulation estimators. *Oper. Res.* **40**(3) 505–520.

Gordy, M. B. 2004. Granularity adjustment in portfolio credit risk measurement. G. P. Szegö, ed. *Risk Measures for the 21st Century*. John Wiley & Sons, Chichester, UK.

Gordy, M. B., S. Juneja. 2006. Efficient simulation for risk measurement in a portfolio of CDOs. L. F. Perrone, B. G. Lawson, J. Liu, F. P. Wieland, eds. *Proc. 2006 Winter Simulation Conf.*, IEEE Press, Piscataway, NJ, 749–756.

Gordy, M. B., S. Juneja. 2008. Nested simulation in portfolio risk measurement. FEDS 2008-21, Federal Reserve Board, Washington, DC.

Gouriéroux, C., J. P. Laurent, O. Scaillet. 2000. Sensitivity analysis of values at risk. *J. Empirical Finance* **7**(3–4) 225–245.

Kim, J. H. T., M. R. Hardy. 2007. Quantifying and correcting the bias in estimated risk measures. *ASTIN Bull.* **37**(2) 365–386.

Lan, H., B. L. Nelson, J. Staum. 2007a. A confidence interval for tail conditional expection via two-level simulation. S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, R. R. Barton, eds. *Proc. 2007 Winter Simulation Conf.*, IEEE Press, Piscataway, NJ, 949–957.

Lan, H., B. L. Nelson, J. Staum. 2007b. Two-level simulations for risk management. S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, R. R. Barton, eds. *Proc. 2007 Winter Simulation Conf.*, IEEE Press, Piscataway, NJ, 102–107.

Lan, H., B. L. Nelson, J. Staum. 2010. A confidence interval procedure for expected shortfall risk measurement via two-level simulation. *Oper. Res.* **58**(5) 1481–1490.

Lang, S. 1968. *Analysis I*. Addison-Wesley, Reading, MA.

Lee, S.-H. 1998. Monte Carlo computation of conditional expectation quantiles. Ph.D. thesis, Stanford University, Stanford, CA.

Lesnevski, V., B. L. Nelson, J. Staum. 2007. Simulation of coherent risk measures based on generalized scenarious. *Management Sci.* **53**(11) 1756–1769.

Longstaff, F. A., E. S. Schwartz. 2001. Valuing American options by simulation: A simple least-squares approach. *Rev. Financial Stud.* **14**(1) 113–147.

Martin, R., D. Tasche. 2007. Shortfall: A tail of two parts. *Risk* **20**(2) 84–89.

Martin, R., T. Wilde. 2002. Unsystematic credit risk. *Risk* **15**(11) 123–128.

Quenouille, M. H. 1956. Notes on bias in estimation. *Biometrika* **43**(3–4) 353–360.

Rothschild, M., J. E. Stiglitz. 1970. Increasing risk I: A definition. *J. Econom. Theory* **2**(3) 225–243.