Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

Nested Simulation in Portfolio Risk Measurement

Michael B. Gordy and Sandeep Juneja

2008-21

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

Nested Simulation in Portfolio Risk Measurement*

Michael B. Gordy
Federal Reserve Board

Sandeep Juneja
Tata Institute of Fundamental Research

8 April 2008

Abstract

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 re-prices 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 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 and a dynamic allocation scheme for improved efficiency.

JEL Codes: G32, C15
Key words: nested simulation, loss distribution, value-at-risk, expected shortfall, jackknife estimator, dynamic allocation

---

*We 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 our own, and do not reflect the views of the Board of Governors or its staff. Email: ⟨michael.gordy@frb.gov⟩, ⟨juneja@tifr.res.in⟩.

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. As 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 ABS), 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 re-prices each position in the portfolio at the horizon conditional on the drawn risk factors. It has been widely assumed that simulation-based pricing algorithms would be infeasible in the inner step, because the inner step must be executed once for each trial in the outer step.

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. Since 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. In addition, we introduce a dynamic allocation scheme for choosing the number of inner step trials as a function of the generated output. This technique can significantly reduce the computational effort to achieve a given level of accuracy.

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 are 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 "LSM" estimator 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.[3] 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 examine the performance of jackknife estimators and dynamic allocation schemes in a nested simulation setting.

---

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

[2] Lee (1998) is an unpublished PhD thesis, which we encountered shortly before completion of this draft. Lee's

In Section 1 we set out a very general modeling framework for a portfolio of financial instruments. We introduce the nested simulation methodology in Section 2. We characterize the bias in 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 Section 3. In the last two sections, we propose some refinements to further improve computational performance of nested simulation. Simple jackknife methods for bias reduction are developed in Section 4. Our dynamic allocation scheme is introduced and examined in Section 5.

1 Model framework

Let X_t be a vector of m state variables that govern 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 $t \leq s$ 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.[4] Position 0 represents the sub-portfolio of instruments for which there exist analytical pricing functions. Without loss of generality, we treat this as a single composite instrument. 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 cashflows up to time t are nonstochastic functions of time that depend on the contractual terms. Let $C_k(t)$ be the cumulative cashflow 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 cashflows under the risk-neutral measure Q:

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

---

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.

[3] For a very different approach to optimization of computational budget in a nested simulation, see De Prisco et al. (2007).

[4] 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.

The valuations are expressed in currency units, so there is no need for portfolio weights.[5] We set $V_k(t) = 0$ whenever $t \geq T_k$.

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 cashflows. 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 cashflows are reinvested in the money market until time H, but other conventions are easily accommodated.

2 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

1. Draw a single path X_t for $t = (0, H]$ under the physical measure. Let ξ represent the filtration $\mathcal{F}_H$ that is generated by this path.
2. Evaluate the accrued value at H of the interim cashflows.
3. Evaluate the price of each position at H.
    - Closed-form price for instrument 0.
    - Simulation with N_k "inner step" trials for remaining positions $k = 1, \ldots, K$. These paths are simulated under the risk-neutral measure.
4. Discount back to time 0 to get our loss $Y(\xi)$.

Observe that the full dependence structure across the portfolio is captured in the period up to the model horizon. Inner step simulations, in contrast, are 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. 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 very long relative to the horizon H.

---

[5] Values can be negative, so it isn't always natural to decompose portfolio value into a vector of weights and a vector of "returns".

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 (§2.1), Value-at-Risk (§2.2), and Expected Shortfall (§2.3). For simplicity, we impose a single value of N across all positions (i.e., $N_k = N$ for $k = 1, ..., K$). This restriction is relaxed in Section 2.4. In Section 2.5, we consider the asymptotic behavior of the optimal allocation of computational resources as the portfolio size grows large. Last, in Section 2.6, we elaborate on the trade-offs associated with simultaneous repricing.

2.1 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 known, the associated $Y(\xi)$ would be known and simulation would involve generating 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. However, the mark-to-market value of each position is not known and is instead estimated via the inner step simulations.

Within the inner step simulation for repricing position k, each trial gives an unbiased (but very noisy) estimate of $Y(\xi)$. 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 $\tilde{Y}(\xi) \equiv Y(\xi) + \bar{Z}^N(\xi)$ as an estimate of loss in the portfolio. By the law of large numbers,

$$\bar{Z}^N(\xi) \to 0 \quad a.s.$$

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 $(\tilde{Y}_1(\xi_1), \ldots, \tilde{Y}_L(\xi_L))$ via outer and inner step simulation and taking the average

$$\hat{\alpha}_{L,N} = \frac{1}{L}\sum_{\ell=1}^{L} 1[\tilde{Y}_\ell(\xi_\ell) > u].$$

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

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

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 sample of ξ (outer step simulation) on an average equals $\gamma_0 > 0$, and to generate an inner step simulation sample on the 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

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

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

Assumption 1 *1. The joint pdf $g_N(\cdot,\cdot)$ of Y and $\tilde{Z}_N$ and its partial derivatives $\frac{\partial}{\partial y} g_N(y,z)$ and $\frac{\partial^2}{\partial y^2} g_N(y,z)$ exist for each N and for all (y,z).*

2. For $N \geq 1$, there exist non-negative 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 $\tilde{Z}_N$ through adding to both of them mean zero, variance ϵ independent Gaussian random variables, also independent of Y and $\tilde{Z}_N$. For small ϵ this has a negligible impact on the tail measures. Then, if

$$\sup_N \int_{-\infty}^{\infty} z^4 dF_{\tilde{Z}_N}(z) < \infty, \tag{2}$$

where $F_{\tilde{Z}_N}(\cdot)$ denotes the distribution function of $\tilde{Z}_N$, Assumption 1 can be seen to hold. To see this, let Y_ϵ and $\tilde{Z}_{N,\epsilon}$ denote the random variables obtained by perturbing Y and $\tilde{Z}_N$ as described above. Then, the joint pdf of $(Y_\epsilon, \tilde{Z}_{N,\epsilon})$ 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 $(Y, \tilde{Z}_N)$. Since, $\phi_\epsilon(\cdot)$ and its first two derivatives are bounded, it is easy to see that equation (2) implies that Assumption 1 holds for $(Y_\epsilon, \tilde{Z}_{N,\epsilon})$.

Assumption 1 is sufficient to deliver a useful convergence property. Here and henceforth, let f and F denote the density and cumulative distribution function for Y, and let $\tilde{f}_N$ and $\tilde{F}_N$ denote the density and cumulative distribution function for $\tilde{Y}$. Now let y_N be some sequence of real numbers that converges to a real number y. In Appendix A.1, 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

$$\Theta(u) = \frac{1}{2} f(u)\mathrm{E}[\sigma_\xi^2 | Y = u], \tag{3}$$

where σ_ξ^2 denotes the conditional variance of Z_i (conditioned on ξ). Our approximation is given by:[6]

Proposition 1 *Under Assumption 1,*

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

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

Proof is in Appendix A.2.

For the distributions and large loss levels one might expect to appear in practice, the bias will be upwards (i.e., $\theta_u > 0$). 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 $\hat{\alpha}_{L,N} \geq \alpha$ for all $u > u^*$.

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

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

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

$$N^* = \left(\frac{2\theta_u^2}{\alpha(1-\alpha)\gamma_1}\right)^{1/3} \Gamma^{1/3} + o_\Gamma(\Gamma^{1/3}). \tag{4}$$

Therefore optimal L has the form

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

and the mean square error at optimal N^* equals

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

---

[6] 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 $h(m)$.

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. It is easy to intuitively see the imbalance between N^* and L^*. Note that when N and L are of the same order $\sqrt{\Gamma}$, the squared bias term contributes much less to the mean square error compared to the variance term. By increasing L at the expense of N we reduce variance till it matches up in contribution to the squared bias term.

2.2 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

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

Under Assumption 1, Y is a continuous random variable so that $P(Y \geq y_\alpha) = \alpha$. As before, our nested simulation generates samples $(\tilde{Y}_1(\xi_1), \ldots, \tilde{Y}_L(\xi_L))$ where $\tilde{Y}(\xi) \equiv Y(\xi) + \bar{Z}^N(\xi)$. We sort these draws as $\tilde{Y}_{[1]} \geq \ldots \geq \tilde{Y}_{[L]}$, so that $\tilde{Y}_{\lceil \alpha L \rceil}$ 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 $\mathrm{E}[(\tilde{Y}_{\lceil \alpha L \rceil} - y_\alpha)^2]$ and then minimizing it. As before, we decompose MSE into variance and squared bias:

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

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

Proposition 2 *Under Assumption 1,*

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

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

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

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 Appendix A.3 abstract from certain mild but cumbersome regularity conditions; see the appendix for details.

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 $u = y_\alpha$. The mean square error at optimal N^* equals

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

2.3 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 may give incentives to build portfolios that are highly sensitive to extreme tail events. More formally, Value-at-Risk fails to satisfy the sub-addivity 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

$$\mathrm{ES}_\alpha[Y] = \alpha^{-1}\left(\mathrm{E}[Y \cdot 1[Y \geq y_\alpha]] + y_\alpha(1-\alpha-\Pr(Y < y_\alpha))\right), \tag{6}$$

The first term is often used as the definition of Expected Shortfall for continuous variables. It is also known as "tail conditional expectations." The second term is a correction for mass at the quantile y_α. In our setting, Y and $\tilde{Y}$ are continuous in distribution, so

$$\begin{aligned} \mathrm{ES}_\alpha[Y] &= \frac{1}{\alpha}\mathrm{E}[Y \cdot 1[Y > y_\alpha]] \\ \mathrm{ES}_\alpha[\tilde{Y}] &= \frac{1}{\alpha}\mathrm{E}[\tilde{Y} \cdot 1[\tilde{Y} > \tilde{y}_\alpha]] \end{aligned}$$

Acerbi and Tasche (2002) show that ES is coherent and equivalent to the "conditional VaR" (CVaR) measure of Rockafellar and Uryasev (2002).

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 $\mathrm{E}[Y \cdot 1[Y > y_\alpha]]$ since here u is specified while 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 $\tilde{Y} \cdot 1[\tilde{Y} > u]$. Let $\Upsilon_N(u)$ denote $\mathrm{E}[\tilde{Y} \cdot 1[\tilde{Y} > u]]$. 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 Appendix A.4.

Using similar analysis to the proof of Proposition 3, we can establish that

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

and therefore that

$$\mathrm{V}[\tilde{Y} \cdot 1[\tilde{Y} > u]] = \mathrm{V}[Y \cdot 1[Y > u]] + O_N(1/N). \tag{7}$$

Thus, analogous to Section 2.1, we consider the optimization problem

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

as $\Gamma\to\infty$.

Applying Proposition 3 and (7), the objective function reduces to finding N that minimizes

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

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

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

and optimal L has the form

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

The mean square error at optimal N^* equals

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

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

$$\begin{aligned}\mathrm{E}[\tilde{Y}|\tilde{Y}>\tilde{y}_\alpha]-\mathrm{E}[Y|Y>y_\alpha] &= \frac{1}{\alpha}(\Upsilon_N(\tilde{y}_\alpha)-\Upsilon(y_\alpha))\\ &= \frac{1}{\alpha}\left((\Upsilon_N(\tilde{y}_\alpha)-\Upsilon_N(y_\alpha))+(\Upsilon_N(y_\alpha)-\Upsilon(y_\alpha))\right)\end{aligned} \tag{10}$$

From Proposition 3, we have

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

Now from the mean value theorem

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

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

$$\Upsilon_N(\tilde{y}_\alpha)-\Upsilon_N(y_\alpha)=-(\tilde{y}_\alpha-y_\alpha)(y_\alpha f(y_\alpha)+o_N(1))=-\frac{y_\alpha\theta_\alpha}{N}+o_N(1/N). \tag{12}$$

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

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

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

2.4 Optimal allocation within the inner step

In this subsection we relax the restriction that N_k is equal across 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 $\gamma_{1,k}$. Then, total inner step simulation effort equals $N\gamma_1$ where

$$\gamma_1 = \sum_{k=1}^{K} p_k \gamma_{1,k}.$$

For a single trial of the outer step simulation, the inner step simulation generates the estimate

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

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 Section 2.1. The resultant θ_u in this setting is

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

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

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

Recall from Section 2.1 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 $\{p_k \geq 0 : k \leq K\}$. 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

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

Since 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 $(p^*_k : k \leq K)$ 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 as one expects that higher computation resources 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.

2.5 Large portfolio asymptotics

Intuition suggests that as the portfolio size increases, the optimal number of inner loops 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 increases to infinity. To avoid cumbersome notation and tedious technical arguments, we focus on the case of a portfolio of exchangeable (i.e., statistically homogeneous) positions. The arguments given are somewhat heuristic to give the flavor of analysis involved while avoiding the cumbersome and lengthy notation and assumptions needed to make it completely rigorous.

Consider an infinite sequence of exchangeable position indexed by k, and let W_k be the loss on position k. Let $\bar{Y}^K$ be the average loss per position on a portfolio consisting of the first K positions in the sequence, i.e.,

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

We assume that Assumption 1 holds for the individual W_k and their respective pricing errors ζ_{ki}.

As before, instead of observing $W_k(\xi)$, we generate N inner step samples $W_k(\xi) + \zeta_{ki}(\xi)$ for $i = 1, \ldots, N$, so that our simulation provides an unbiased estimator for the probability

$$P(\bar{Y}^K + \bar{Z}^{K,N} > u),$$

where

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

We approximate the bias $P(\bar{Y}^K + \bar{Z}^{K,N} > u) - P(\bar{Y}^K > u)$ as K and N grow large. Applying the same arguments as in proof of Proposition 1, we can show

$$P(\bar{Y}^K + \bar{Z}^{K,N} > u) = P(\bar{Y}^K > u) + \theta_u^K/(KN) + o_{KN}(1/(KN)),$$

where $\theta_u^K = -\Theta'_K(u)$ and

$$\Theta_K(u) = \frac{1}{2}\frac{d}{du}\bar{f}_K(u)\mathrm{E}[\tilde{\sigma}_\xi^2 | \bar{Y}^K = u]$$

where $\bar{f}_K(\cdot)$ denotes the pdf of $\bar{Y}^K$, and $\tilde{\sigma}_\xi^2 = \mathrm{V}[\zeta_{\cdot,\cdot}|\xi]$.

By the law of large numbers, $|\bar{Y}^K - \mathrm{E}[W_1|\xi]| \to 0$, almost surely, so the cdf $\bar{F}_K$ converges to a non-degenerate limiting distribution $\bar{F}$, which is the distribution of $\mathrm{E}[W_1|\xi]$. Similarly, $\mathrm{E}[\tilde{\sigma}_\xi^2|\bar{Y}^K = u]$ converges to $\mathrm{E}[\tilde{\sigma}_\xi^2|\mathrm{E}[W_1|\xi] = u]$. Under suitable regularity conditions, $\theta_u^K \to \bar{\theta}_u = -\bar{\Theta}'(u)$, where

$$\bar{\Theta}(u) = \frac{1}{2}\frac{d}{du}\bar{f}(u)\mathrm{E}[\tilde{\sigma}_\xi^2|\mathrm{E}[W_1|\xi] = u].$$

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

We assume that the computational budget $\Gamma = \chi K^\beta$ for $\chi > 0$ and $\beta \geq 1$. The value of β captures the size of computational budget available relative to the time taken to generate a single inner loop sample. Note that if $\beta < 1$ then asymptotically even a single sample cannot be generated.

Recall that m denotes the number of underlying state variables that control the prices of K positions. Suppose that the computational effort to generate one sample of outer step simulation on average equals $\psi(m, K)$ for some function ψ, and to generate an inner step simulation sample on the average equals $K\gamma_1$ for a constant $\gamma_1 > 0$. Average effort per outer step trial then equals $KN\gamma_1 + \psi(m, K)$ and average effort for L such trials equals $L(KN\gamma_1 + \psi(m, K))$. We analyze the order of magnitude of N and L that minimize the resultant mean square error of the estimator.

The mean square error of the estimator equals

$$\frac{\alpha(1-\alpha)}{L} + \frac{\bar{\theta}_u^2}{K^2 N^2}$$

plus terms that are relatively negligible for large values of K. Noting that $L = \frac{\chi K^{\beta-1}}{N\gamma_1 + \psi(m,K)/K}$, it is easily seen that the value of $N \geq 1$ that minimizes the dominant terms in the mean square error equals

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

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 irrespective of the form of the fixed-cost function $\psi(m, K)$. In this case L^* is of order $\min(K^{\beta-1}, K^{\beta}/\psi(m, K))$. Only when $\beta > 3$ does N grow with K.

2.6 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. While this reduces 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

$$\beta > 1 - \frac{1}{(1 + \rho(K-1))}.$$

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.

3 Numerical examples

We illustrate our results with a parametric example. Distributions for loss Y and the pricing errors Z are specified to ensure that the bias and variance of our simulation estimators are in closed-form. While 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. We have used simulation to perform similar exercises on the somewhat more realistic example of a portfolio of equity options. All our conclusions are robust.

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 $\mathcal{N}(0, 1+\nu^2/K)$. 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 $\tilde{Y}$ 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 $\tilde{Y}$ 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.

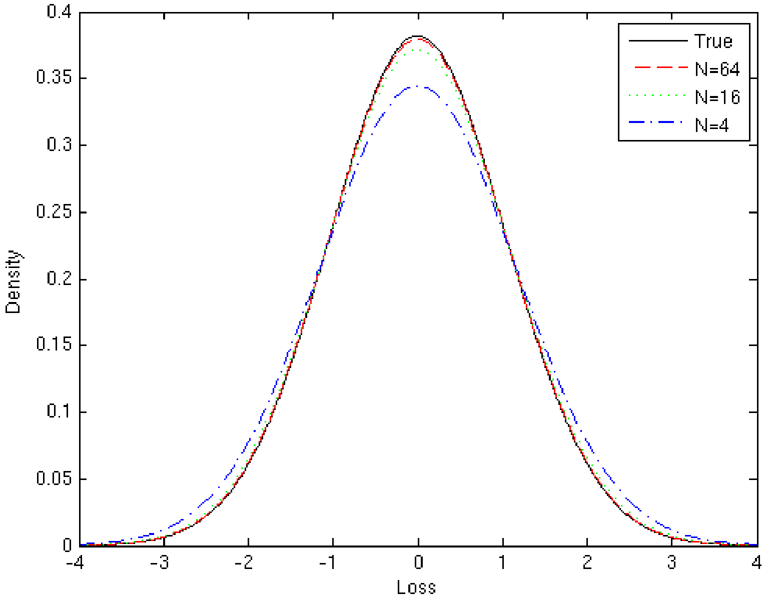
We consider our estimator $\hat{\alpha}_{L,N}$ of the large loss probability $\alpha = P(Y > u) = \Phi(-u/\sqrt{1+\nu^2/K})$ for a fixed loss level u. The expected value of $\hat{\alpha}_{L,N}$ is $\alpha_N = P(\tilde{Y} > u) = \Phi\left(-u/\sqrt{1+\nu^2/K+\sigma^2/N}\right)$. Applying Proposition 1,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

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

and where ϕ is the standard normal density. Figure 2 plots the exact bias $\alpha_N - \alpha$ and the approximation θ/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 under 17%.

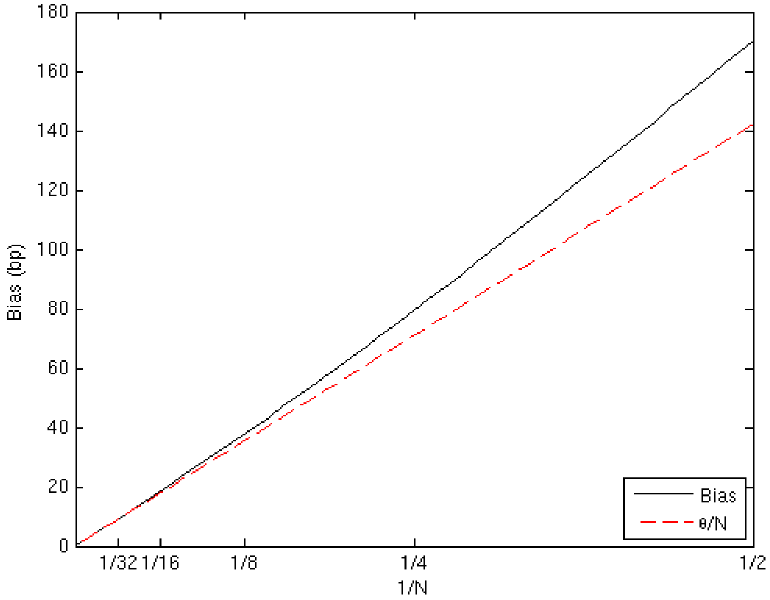
In Figure 3, we plot the exact root mean square error of $\hat{\alpha}_{L,N}$ as a function of N and with the computational budget held fixed. Here we make the assumption that the fixed cost of the outer step is negligible (i.e., $\gamma_0 \approx 0$), so that $\Gamma \propto N \cdot L$. We observe that the optimal N is increasing with the budget, but remains 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 (4). When

Figure 1: Density of the loss distribution



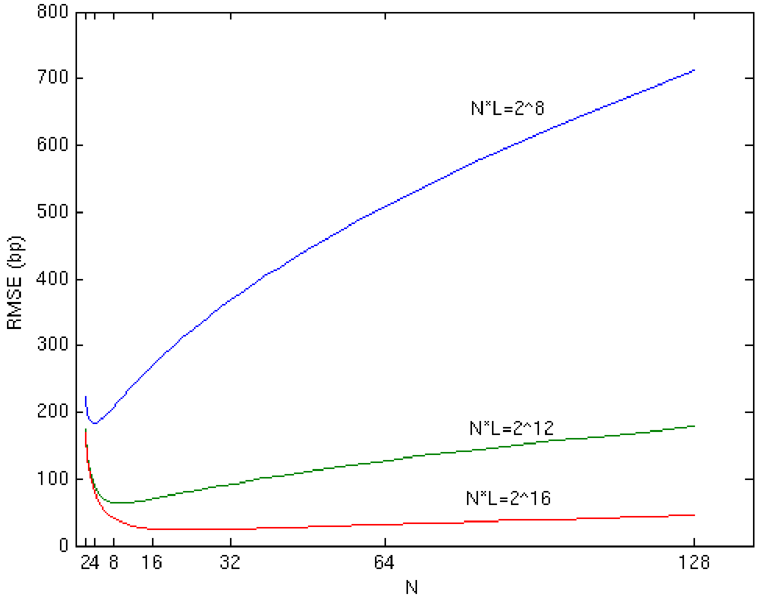
Gaussian example with parameters: $\nu = 3$, $\eta = 10$, $K = 100$.

Figure 2: Exact and approximated bias



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

Figure 3: Root Mean Square Error



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 $u = \mathrm{VaR}_{0.01}[Y]$.

$\Gamma = 2^{10}$, the relative error of the approximation is 8.3%. Increasing the budget to $\Gamma = 2^{16}$ shrinks the relative error to under 2.5%.

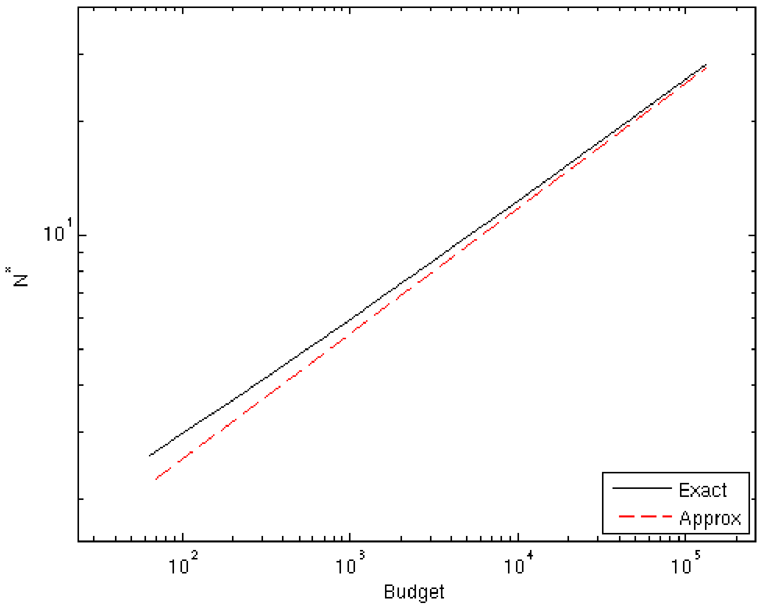
The optimal N^* and the accuracy of its approximation may depend on the exceedance threshold u, and not necessarily in a monotonic fashion. This is demonstrated in the top panel of Figure 5. The budget here is large, and we see that the approximation to N^* is accurate over the entire range of interest (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 under six. If we increase the portfolio size by a factor of 10, we find that N^* is under two. These results suggest that the large portfolio asymptotics of section 2.5 may pertain to portfolios of realistic size.

As Value-at-Risk is the risk-measure most commonly used in practice, we turn briefly to the estimation of VaR. The results of Section 2.2 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., $\mathrm{VaR}_\alpha[Y]$. The distance on the y-axis between $\mathrm{VaR}_\alpha[\tilde{Y}]$ (on the solid line) 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

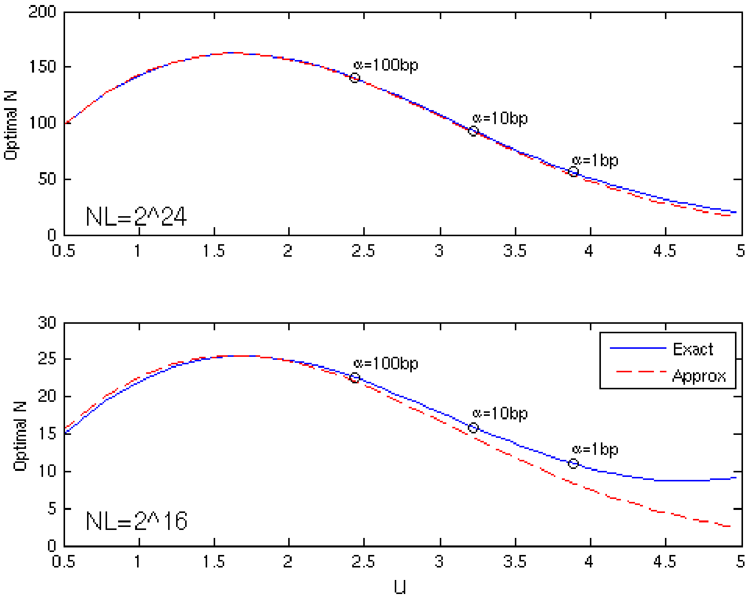
$$\mathrm{E}[\tilde{Y}_{\lceil \alpha L \rceil}] \approx y_\alpha + \frac{1}{N} \frac{\theta_\alpha}{f(y_\alpha)}.$$

Figure 4: Optimal N grows with computational budget



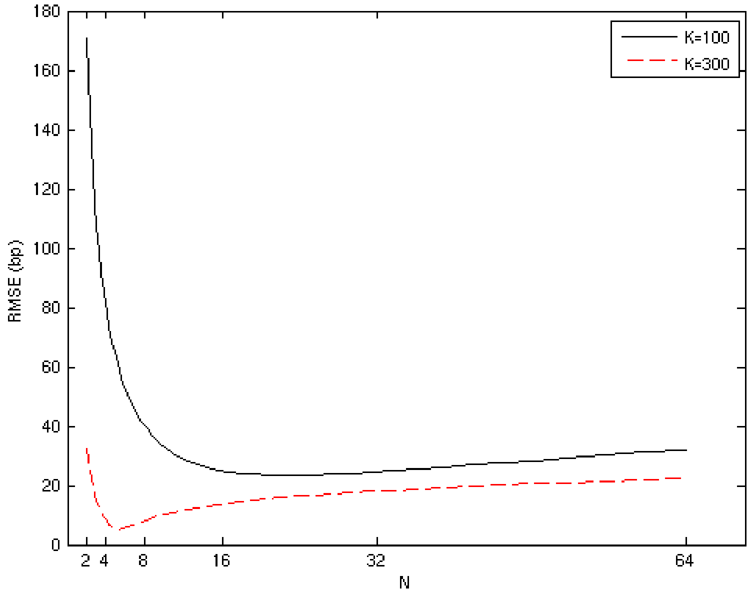
Budget is $\Gamma = N \cdot L$. Parameters: $\nu = 3$, $\eta = 10$, $K = 100$ and $u = \mathrm{VaR}_{0.01}[Y]$.

Figure 5: Optimal N depends on exceedance threshold



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

Figure 6: Optimal N as portfolio size varies



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

For the distributional assumptions of our example, we have

$$\frac{\theta_\alpha}{f(y_\alpha)} = \frac{\sigma^2}{2} \frac{\Phi^{-1}(1-\alpha)}{\sqrt{1+\nu^2/K}}.$$

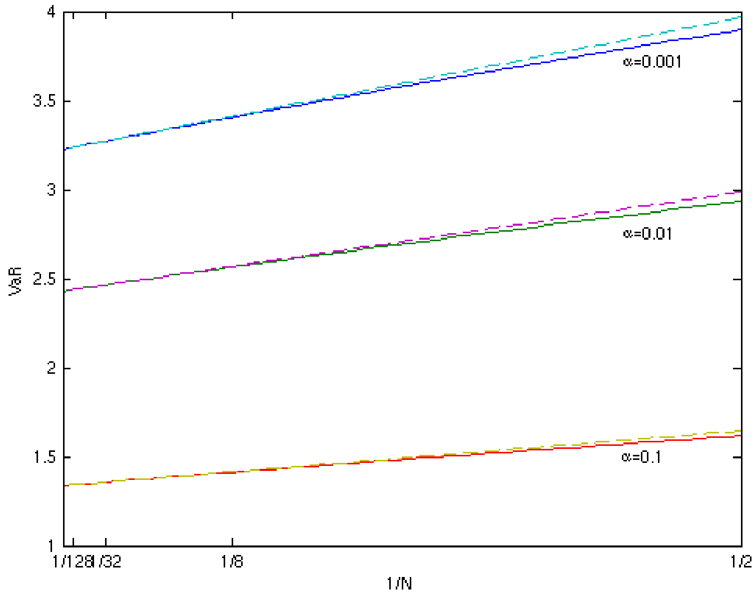
Clearly, the linearized approximation to the bias is highly accurate in this example.

4 Jackknife estimation

Jackknife estimators, introduced over 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 loop sample of N draws into I non-overlapping *sections* (N and I are selected so that N/I is an integer). Section i covers the N/I draws $N\frac{(i-1)}{I} + 1, \ldots, N\frac{i}{I}$. For a single outer step trial, represented by ξ, let a_N denote the sample output from the inner step $1[\tilde{Y}_\xi > u]$ as proposed in earlier sections. Let $\tilde{Y}_\xi(-i)$ be the estimate of Y_ξ that is obtained when section i is omitted, and define $a_N(-i)$ similarly, e.g., $a_N(-I) = 1[\tilde{Y}_\xi(-I) > u]$. 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.

Figure 7: Bias of estimated Value-at-Risk



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

We now propose the jackknife inner step simulation output

$$a^\dagger = I a_N - (I-1)\frac{1}{I}\sum_{i=1}^{I} a_N(-i) = a_N + \frac{I-1}{I}\sum_{i=1}^{I}(a_N - a_N(-i)). \tag{15}$$

The bias in $a^\dagger$ is

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

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 (15), we see that the variance of the estimator $a^\dagger$ is

$$\mathrm{V}[a^\dagger] = \mathrm{V}[a_N] + \frac{(I-1)^2}{I}\mathrm{V}[b_N(-I)] + 2(I-1)\mathrm{Cov}[a_N, b_N(-I)] + \frac{(I-1)^3}{I}\mathrm{Cov}[b_N(-1), b_N(-I)], \tag{16}$$

where $\mathrm{Cov}[X_1, X_2]$ denotes the covariance between X_1 and X_2. We now evaluate the dominant term in this 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. In Appendix A.5, we demonstrate that

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

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

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

Similarly, $\mathrm{Cov}[b_N(-1), b_N(-I)]$ is $O_N(1/N^{1/2})$. Therefore, from equation (16) 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., $\hat{\alpha}^\dagger = (1/L)\sum_{\ell=1}^{L} a^\dagger{}_\ell(\xi_\ell)$. 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 $\frac{\pi}{N^3}$ (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 section 2.1, 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

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

so even fewer inner step samples are needed for the jackknife estimator. 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 $\Gamma^{-3/4}$.

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 Section 3. For any parametric example, it is useful to re-write the terms in equation (16) for the variance of $a^\dagger$ as:

$$\begin{aligned} \mathrm{V}[b_N(-I)] &= \mathrm{V}[a_N] + \mathrm{V}[a_N(-I)] - 2\mathrm{Cov}[a_N, a_N(-I)] \\ \mathrm{Cov}[a_N, b_N(-I)] &= \mathrm{V}[a_N] - \mathrm{Cov}[a_N, a_N(-I)] \\ \mathrm{Cov}[b_N(-1), b_N(-I)] &= \mathrm{V}[a_N] - \mathrm{Cov}[a_N, a_N(-I)] - \mathrm{Cov}[a_N, a_N(-1)] + \mathrm{Cov}[a_N(-1), a_N(-I)]. \end{aligned}$$

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

$$\mathrm{Cov}[a_N, a_N(-1)] = \mathrm{Cov}[a_N, a_N(-I)] = P(\tilde{Y} > u, \tilde{Y}(-I) > u) - \alpha_N \alpha_{N(I-1)/I}. \tag{17}$$

Similarly, we have

$$\mathrm{Cov}[a_N(-1), a_N(-I)] = P(\tilde{Y}(-1) > u, \tilde{Y}(-I) > u) - \alpha^2_{N(I-1)/I}. \tag{18}$$

For the Gaussian example, let

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

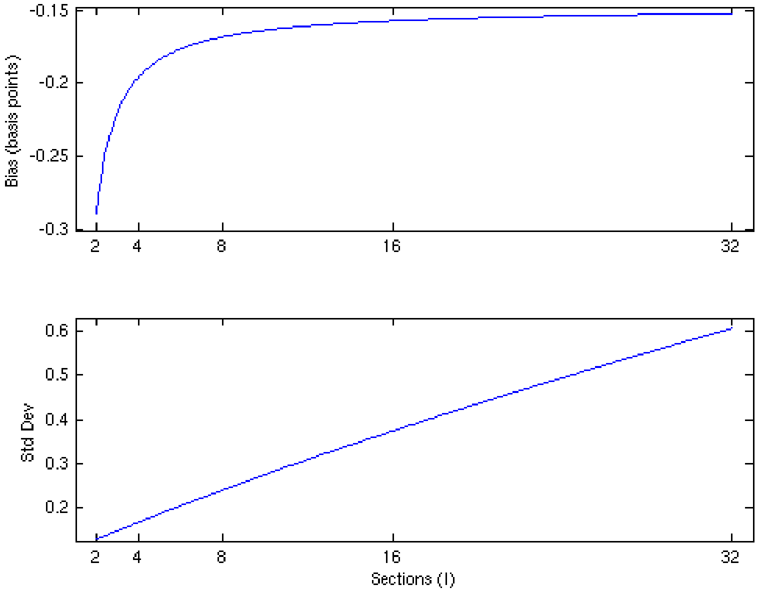
so that $\alpha = \Phi(U_\infty)$ and $\alpha_N = \Phi(U_N)$. The bivariate probabilities in Equations (17) and (18) are

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

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 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 $a_{L,N}$ 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 $\sqrt{\mathrm{V}[a^\dagger]} = 0.128$ for $I = 2$ and $\sqrt{\mathrm{V}[a^\dagger]} = 0.606$ for $I = N = 32$.

Figure 8: Bias and variance of jackknife estimators



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

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

5 Dynamic allocation

Through dynamic allocation of workload in the inner step we can further reduce the computational effort in the inner step while increasing the bias by a negligible controlled amount or even (in many cases) decreasing the bias. Consider the estimation of large loss probabilities $P(Y > u)$ for a given u. Bias in our estimated $\hat{\alpha}$ is increasing with the likelihood that the indicator $1[\tilde{Y}(\xi) > u]$ obtained from the inner step simulation is not equal to the true $1[Y(\xi) > u]$ for a given realization ξ of the outer step. Say we form a preliminary estimate $\tilde{Y}^n$ based on the average of a small number n of inner step trials. If this estimate is much smaller or much larger than u, it may be a waste of effort to generate many more samples in the inner simulation step. However, if this average is close to u, it makes sense to generate many more inner step samples in order to increase the probability that the estimated $1[\tilde{Y}(\xi) > u]$ is equal to the true value $1[Y(\xi) > u]$. The variance of $\hat{\alpha}$ is dominated by $\alpha(1-\alpha)/L$, so dynamic allocation in the inner step will have little effect on the variance. As in Section 4, we develop our dynamic allocation methods for the estimation of large loss probabilities. Similar methods would apply to VaR and ES.

Our proposed dynamic allocation (DA) scheme is very simple. For each trial ξ of the outer step, we generate δN inner step trials (selecting δ so that δN is an integer), and let the resulting preliminary loss estimate be denoted $\tilde{Y}^\delta$. If $\tilde{Y}^\delta < u - \epsilon$ for some well-chosen $\epsilon > 0$ then we terminate the inner step and our sample output is zero.[7] Otherwise, we generate a second estimate $\tilde{Y}^{1-\delta}$ based on an additional $(1-\delta)N$ conditionally independent samples. Defining

$$\tilde{Y}^* = \delta \tilde{Y}^\delta + (1-\delta)\tilde{Y}^{1-\delta},$$

our final estimate $\tilde{Y}_\xi^{DA}$ is taken as

$$\tilde{Y}_\xi^{DA} = \begin{cases} \tilde{Y}^\delta, & \text{if } \tilde{Y}^\delta < u - \epsilon \\ \tilde{Y}^*, & \text{otherwise.} \end{cases}$$

The cost of computation per inner loop is proportional to the average number of inner step trials executed, which is

$$\bar{N}^{DA} \equiv N \cdot (\delta + (1-\delta)P(\tilde{Y}^\delta > u - \epsilon)) < N,$$

so the DA estimator always has lower computational effort than the static estimator (for any N held fixed).

With dynamic allocation, the bias in the estimated exceedance probability is

$$\begin{aligned} P(\tilde{Y}^{DA} > u, \tilde{Y}^\delta > u - \epsilon) - P(Y > u) &= P(\tilde{Y}^* > u) - P(Y > u) - P(\tilde{Y}^* > u, \tilde{Y}^\delta \leq u - \epsilon) \\ &= (\alpha_N - \alpha) - P(\tilde{Y}^* > u, \tilde{Y}^\delta \leq u - \epsilon). \quad (19) \end{aligned}$$

---

[7] We could improve the scheme slightly by incorporating an upper bound stopping rule as well, setting sample output to 1 if, say, $tY^\delta > u + \epsilon$. As this event occurs only rarely, the computational savings are generally modest so we do not develop this modification here.

Thus, dynamic allocation introduces a negative increment to bias, relative to the static estimator. As we have seen earlier, in typical application the static estimator has a positive bias. The joint probability $P(\tilde{Y}^* > u, \tilde{Y}^\delta \leq u - \epsilon)$ can be made arbitrarily small by choosing δ large (near one) and/or choosing a large ϵ. Therefore, when $\alpha_N - \alpha > 0$, there must exist pairs (δ, ϵ) such that the DA estimator has a lower bias than the static estimator. In this case, dynamic allocation reduces both computational effort and mean square error.

If $\alpha_N \leq \alpha$ then dynamic allocation will increase absolute bias. Even in this general case, we can derive an upper bound on the increase in absolute bias. As shown in Appendix B, for every δ one can find an ϵ so that the resultant increase in bias is within a specified tolerance. The issue then is to select a δ that most reduces the computational effort.

For numerical illustration, we return to our Gaussian example of Section 3. The exact expected value of $\hat{\alpha}^{DA}$ in this example is

$$\begin{aligned} \mathrm{E}[\hat{\alpha}^{DA}] &= P(\tilde{Y}^{DA} > u) \\ &= \Phi_2\left(\frac{-u}{\sqrt{1+\nu^2/K+\sigma^2/N}}, \frac{\epsilon - u}{\sqrt{1+\nu^2/K+\sigma^2/(\delta N)}}, \sqrt{\frac{1+\nu^2/K+\sigma^2/N}{1+\nu^2/K+\sigma^2/(\delta N)}}\right). \end{aligned} \tag{20}$$

As in our baseline examples, we fix $\nu = 3$, $\eta = 10$, $K = 100$ and $u = \mathrm{VaR}_{0.01}[Y]$. For the static scheme with $N = 32$, we find a bias of 9.0 basis points. For the dynamic allocation scheme, say we choose $\delta = 1/32$ and $\epsilon = \sqrt{\mathrm{V}[Y]} = 1.044$. Observe that the preliminary sample in the inner step will consist of a *single* trial. We find the average number of inner step trials executed is $\bar{N}^{DA} = 6.24$, so computational effort is reduced by over 80%. Bias is -0.4 basis points, so also a large improvement over the static scheme.

More complex dynamic schemes are possible. For example, instead of a single stopping test at δN, one can have a series of j stopping tests after $\delta_1 N, \delta_2 N, \ldots, \delta_j N$ inner step trials. The bandwidths would narrow successively, i.e., $\epsilon_1 > \epsilon_2 > \ldots \epsilon_j$. It may also be possible to design adaptive strategies for choosing δ and ϵ that will converge to the optimal values as the outer step simulation progresses.

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 mainly 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. Simple extensions of the basic nested algorithm can eliminate much of the bias at modest cost.

Our results suggest that current practice is misguided. In order 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,[8] 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.

A Proofs

In this section we prove Lemma 1 and Proposition 1. We then derive the expression for the mean square error of Value-at-Risk discussed in Section 2.2. We then provide a proof of Proposition 3. In our notation, we suppress the dependence on ξ and N wherever it is visually convenient.

A.1 Proof of Lemma 1

Note that

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

Differentiating both sides by y, noting that in the 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,

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

This equals

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

where $\breve{y}_N$ 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 $N \to \infty$.

---

[8] Closed-form solutions to even the simplest pricing models may require special functions, e.g., the confluent hypergeometric function.

A.2 Proof of Proposition 1

Note that

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

so that

$$\alpha_N - \alpha = \int_{\Re} \int_{u-z/\sqrt{N}}^{u} g_N(y,z)\,dydz. \tag{21}$$

Consider the Taylor's series expansion of the pdf in (21),

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

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

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

Note that the first term above equals

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

This term equals zero since

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

The second term can be re-written as:

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

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 $\mathrm{E}[\tilde{Y}_{\lceil \alpha L \rceil}] - y_\alpha$. To evaluate $\mathrm{E}[\tilde{Y}_{\lceil \alpha L \rceil}]$, it is useful to consider the order statistics $U_{[1]} \geq \ldots \geq U_{[L]}$ of L random variables uniformly distributed over the unit interval. Within this appendix only, let $\tilde{G}_N$ denote the inverse of the cdf of $\tilde{Y}$. Observe that $\tilde{Y}_{\lceil \alpha L \rceil}$ has the same distribution as $\tilde{G}_N(U_{\lceil \alpha L \rceil})$.

Consider the Taylor's series expansion

$$\tilde{G}_N(U_{\lceil\alpha L\rceil}) = \tilde{G}_N(\mathrm{E}[U_{\lceil\alpha L\rceil}]) + (U_{\lceil\alpha L\rceil} - \mathrm{E}[U_{\lceil\alpha L\rceil}])\tilde{G}'_N(\mathrm{E}[U_{\lceil\alpha L\rceil}])$$
$$+ \frac{1}{2}(U_{\lceil\alpha L\rceil} - \mathrm{E}[U_{\lceil\alpha L\rceil}])^2\tilde{G}''_N(\mathrm{E}[U_{\lceil\alpha L\rceil}]) + \text{ remainder terms.} \quad (22)$$

Taking expectations

$$\mathrm{E}[\tilde{Y}_{\lceil\alpha L\rceil}] = \mathrm{E}[\tilde{G}_N(U_{\lceil\alpha L\rceil})]$$
$$= \tilde{G}_N(\mathrm{E}[U_{\lceil\alpha L\rceil}]) + \frac{1}{2}\mathrm{V}[U_{\lceil\alpha L\rceil}]\tilde{G}''_N(\mathrm{E}[U_{\lceil\alpha L\rceil}]) + \mathrm{E}[\text{remainder terms}]. \quad (23)$$

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

$$\mathrm{E}[U_{\lceil\alpha L\rceil}] = 1 - \lceil\alpha L\rceil/(L+1) = 1 - \alpha + O_L(1/L) \quad (24)$$
$$\mathrm{V}[U_{\lceil\alpha L\rceil}] = \frac{\alpha(1-\alpha)}{L+2} + O_L(1/L^2). \quad (25)$$

Through differentiation of the relation $\tilde{F}_N(\tilde{G}_N(u)) = u$, it can be seen that

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

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

$$\mathrm{E}[\tilde{Y}_{\lceil\alpha L\rceil}] = \hat{y}_\alpha - \frac{\alpha(1-\alpha)}{2(L+2)}\frac{\tilde{f}'_N(\hat{y}_\alpha)}{\tilde{f}_N(\hat{y}_\alpha)^3} + O_L(1/L^2)(\text{constant} + o_N(1)). \quad (26)$$

Here we omit the lengthy discussion on the technical assumptions needed to ensure that the expectation of the remainder terms in (23) has the form $O_L(1/L^2)(\text{constant} + o_N(1))$.

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

$$\hat{y}_\alpha = \tilde{y}_\alpha + O_L(1/L) \quad (27)$$

and that

$$\tilde{y}_\alpha - y_\alpha = \frac{\theta_\alpha}{Nf(y_\alpha)} + o_N(1/N), \quad (28)$$

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

$$\hat{y}_\alpha = y_\alpha + \frac{\theta_\alpha}{Nf(y_\alpha)} + O_L(1/L) + o_N(1/N). \quad (29)$$

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

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

Similarly,

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

In particular, from this, (26) and (29), it follows that the bias

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

Equation (27) follows since equation (24) implies that

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

for N sufficiently large.

We now show (28). Using the Taylor's series expansion,

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

where $\breve{y}_\alpha$ 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

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

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

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

Since, $\tilde{f}_N(y_\alpha) \to f(y_\alpha)$, (28) follows.

The expression for variance can be determined by subtracting (26) from (22) and squaring the difference to get

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

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

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

A.4 Proof of Proposition 3

The $\Upsilon(\cdot)$ and $\Upsilon_N(\cdot)$ can be written as

$$\begin{aligned} \Upsilon_N(u) &= \mathrm{E}[\tilde{Y} \cdot 1[\tilde{Y} > u]] = \int_{\Re} \int_{u - z/\sqrt{N}}^{\infty} (y + z/\sqrt{N}) g_N(y,z)\,dy\,dz & (30) \\ \Upsilon(u) &= \mathrm{E}[Y \cdot 1[Y > u]] = \int_{\Re} \int_{u}^{\infty} y g_N(y,z)\,dy\,dz. & (31) \end{aligned}$$

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

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

Substituting into (30), we have

$$\Upsilon_N(u) - \Upsilon(u) = \int_{\Re}\int_{u-z/\sqrt{N}}^{u} y g_N(y,z)dydz + \int_{\Re} z/\sqrt{N}\int_{u-z/\sqrt{N}}^{u} g_N(y,z)dydz. \tag{32}$$

We apply a Taylor series expansion to the first term in (32):

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

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 (32),

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

Proposition 3 follows by noting that

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

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

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

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

$$\mathrm{E}[b_N(-I)^2] = P(Y + \bar{Z}^N > u, Y + \bar{Z}^{N(I-1)/I} \le u) + P(Y + \bar{Z}^N \le u, Y + \bar{Z}^{N(I-1)/I} > u). \tag{33}$$

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

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

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

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

Let $g_N^{\dagger}$ denote the joint pdf of (Y, A_N, B_N). Then the first term in (33) can be written as

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

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

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

Since 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

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

We proceed in exactly the same fashion for the second term in (33), and find that this term is also $O_N(1/N^{1/2})$.

B Bounding the bias under dynamic allocation

We can bound the absolute bias under dynamic allocation via the triangle inequality:

$$\begin{aligned} |P(\tilde{Y}^* > u, \tilde{Y}^{\delta} > u - \epsilon) - P(Y > u)| &= |P(\tilde{Y}^* > u) - P(Y > u) - P(\tilde{Y}^* > u, \tilde{Y}^{\delta} \leq u - \epsilon)| \\ &\leq |P(\tilde{Y}^* > u) - P(Y > u)| + P(\tilde{Y}^* > u, \tilde{Y}^{\delta} \leq u - \epsilon) \end{aligned}$$

so the increase in the absolute bias is no greater than $P(\tilde{Y}^* > u, \tilde{Y}^{\delta} \leq u - \epsilon)$. Decomposing this as

$$P(\tilde{Y}^* > u, \tilde{Y}^{\delta} \leq u - \epsilon) = P(\tilde{Y}^* > u, \tilde{Y}^{\delta} \leq u - \epsilon, Y > u) + P(\tilde{Y}^* > u, \tilde{Y}^{\delta} \leq u - \epsilon, Y \leq u), \quad (34)$$

we can upper bound each of these terms. Let $\bar{Z}^{\delta}$ and $\bar{Z}^{1-\delta}$ be the average portfolio pricing errors for the preliminary sample of δN and the continuation sample of $(1-\delta)N$ inner step trials. For the first term in equation (34), we have the inequality

$$P(\tilde{Y}^{*} > u, \tilde{Y}^{\delta} \leq u-\epsilon, Y > u) \leq P(\bar{Z}^{\delta} \leq -\epsilon).$$

For the second term, we have

$$P(\tilde{Y}^{*} > u, \tilde{Y}^{\delta} \leq u-\epsilon, Y \leq u) \leq P\left(\tilde{Y}^{1-\delta} > u + \frac{\delta}{1-\delta}\epsilon, Y \leq u\right) \leq P\left(\bar{Z}^{1-\delta} > \frac{\delta}{1-\delta}\epsilon\right).$$

Hence, an upper bound on the additional bias is

$$P(\bar{Z}^{\delta} \leq -\epsilon) + P\left(\bar{Z}^{1-\delta} > \frac{\delta}{1-\delta}\epsilon\right). \tag{35}$$

From here we can take two different approaches:

1. Hoeffding's inequality (Hoeffding, 1963) can be used to develop exact bounds if the zero mean noise associated with each position is upper as well as lower bounded. This is generally true of put options, credit derivatives, and many other types of derivatives. Let $a_k < \zeta_{k,\cdot} < b_k$ denote the bounds on position k. Then,

$$\begin{aligned} P(\bar{Z}^{\delta} \leq -\epsilon) &\leq \exp\left(-\frac{2\delta N\epsilon^2}{\sum_{i=1}^{K}(b_i-a_i)^2}\right) \\ P\left(\bar{Z}^{1-\delta} > \frac{\delta}{1-\delta}\epsilon\right) &\leq \exp\left(-\frac{2\delta^2 N\epsilon^2}{(1-\delta)\sum_{i=1}^{K}(b_i-a_i)^2}\right). \end{aligned}$$

2. As Z_i is the sum of zero mean errors from K positions, we can appeal to the central limit theorem and take each $Z_i(\xi)$ as approximately normal in distribution with mean zero and variance σ_{ξ}^2.[9] Then,

$$\begin{aligned} P(\bar{Z}^{\delta} \leq -\epsilon) &= \Phi\left(\frac{-\epsilon}{\sigma_{\xi}/\sqrt{\delta N}}\right) \\ P\left(\bar{Z}^{1-\delta} > \frac{\delta}{1-\delta}\epsilon\right) &= \Phi\left(\frac{-\epsilon\delta/(1-\delta)}{\sigma_{\xi}/\sqrt{(1-\delta)N}}\right). \end{aligned}$$

Thus, given an acceptable threshold on the increase in bias, for every δ one can find an ϵ so that the resultant increase in bias is controlled. The issue then, is to select a δ that most reduces the computational effort.

We illustrate this computation using our simple Gaussian example with baseline parameter values $K = 100$, $\nu = 3$, and $\eta = 10$. This gives us $\sigma = \eta/\sqrt{K} = 1$ as the standard deviation of the Z_i. Let $u = \mathrm{VaR}_{0.01}[Y] = \sqrt{1+\nu^2/K}\,\Phi^{-1}(0.99)$, so that $\alpha = 0.01$. Say we fix $N = 30$, $\delta = 1/3$ and

---

[9] One can, if necessary, batch the sum of a few Z_i's to improve the approximation to the normal distribution. If σ_{ξ}^2 is unknown, one can estimate it from the sample variance within the inner step. In this case, the standardized Z_i have Student t distribution.

$\epsilon = 2$. The upper bound in equation (35) for the increase in bias for $\hat{\alpha}^{DA}$ over the static estimator $\hat{\alpha}_{L,N}$ is under 4×10^{-6}. The probability of stopping with the preliminary sample is

$$P(\tilde{Y}^{\delta} \leq u - \epsilon) = \Phi\left(\frac{u - \epsilon}{\sqrt{1 + \nu^2/K + \sigma^2/(\delta N)}}\right) \approx 0.653.$$

The average number of inner step trials executed works out to $\bar{N}^{DA} \approx 0.565N$. For a negligible upper bound on the increase in bias, we cut computational effort nearly in half.

References

Carlo Acerbi and Dirk Tasche. On the coherence of expected shortfall. *Journal of Banking and Finance*, 26(7):1487–1503, 2002.

Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

Herbert Aron David. *Order Statistics.* John Wiley & Sons, second edition, 1981.

Ben De Prisco, Ian Iscoe, Yijun Jiang, and Helmut Mausser. Compound scenarios: An efficient framework for integrated market-credit risk. Research paper series, Algorithmics, May 2007.

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

Peter W. Glynn and Ward Whitt. The asymptotic efficiency of simulation estimators. *Operations Research*, 40:505–520, 1992.

Michael Gordy and Sandeep Juneja. Efficient simulation for risk measurement in a portfolio of cdos. In L. Felipe Perrone, Barry G. Lawson, Jason Liu, and Frederick P. Wieland, editors, *Proceedings of the 2006 Winter Simulation Conference*, Piscataway, NJ, 2006. IEEE Press.

Michael B. Gordy. Granularity adjustment in portfolio credit risk measurement. In Giorgio P. Szegö, editor, *Risk Measures for the 21st Century.* John Wiley & Sons, 2004.

C. Gouriéroux, J.P. Laurent, and O. Scaillet. Sensitivity analysis of values at risk. *Journal of Empirical Finance*, 7:225–245, 2000.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, March 1963.

Serge Lang. *Analysis I.* Addison-Wesley, 1968.

Shing-Hoi Lee. *Monte Carlo Computation of Conditional Expectation Quantiles.* PhD thesis, Stanford University, October 1998.

Francis A. Longstaff and Eduardo S. Schwartz. Valuing American options using simulation: A simple least-squares approach. *Review of Financial Studies*, 14:113–147, 2001.

Richard Martin and Dirk Tasche. Shortfall: a tail of two parts. *Risk*, 20(2):84–89, February 2007.

Richard Martin and Tom Wilde. Unsystematic credit risk. *Risk*, 15(11):123–128, November 2002.

Maurice Quenouille. Notes on bias and estimation. *Biometrika*, 43:353–360, 1956.

R. Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26(7):1443–1471, 2002.

Michael Rothschild and Joseph E. Stiglitz. Increasing risk I: A definition. *Journal of Economic Theory*, 2(3):225–243, 1970.