Nested Simulation in Portfolio Risk Measurement

Michael B. Gordy\textsuperscript{1}  Sandeep Juneja\textsuperscript{2}

\textsuperscript{1}Federal Reserve Board <michael.gordy@frb.gov>

\textsuperscript{2}Tata Institute of Fundamental Research <juneja@tifr.res.in>

April 2008

The opinions expressed here are our own, and do not reflect the views of the Board of Governors or its staff.
Consider a very general derivatives portfolio: interest rate swaps, Treasury futures, equity options, default swaps, CDO tranches, etc.

In many or even most cases, preferred pricing model requires simulation.

- Models with analytical solution typically impose restrictive assumptions (Black-Scholes, most famously).
- Simulation almost unavoidable for many path-dependent and basket derivatives.

For trading applications, simulation often too slow for use in real time.

- Endless variety of short-cut approaches, but in practice many are calibrated to “deltas” from a simulation run overnight.
Talking here about risk-measurement of portfolio at some chosen horizon.

- Large loss exceedance probabilities.
- Quantiles of the loss distribution (value-at-risk).

Simulation-based algorithm is nested:

**Outer step:** Draw paths for underlying prices to horizon and calculate implied cashflows during this period.

**Inner step:** Re-price each position at horizon conditional on drawn paths.

Computational task perceived as burdensome because inner step simulation must be executed once for each outer step simulation.

Practitioners invariably use rough pricing tools in the inner step in order to avoid nested simulation.

We show the convention view is wrong – inner step simulation need not be burdensome.
The present time is normalized to 0 and the model horizon is $H$.

Let $X_t$ be a vector of $m$ state variables that govern underlying prices referenced by derivatives.

- interest rates, default intensities, commodity prices, equity prices, etc.

Let $\xi$ be the information generated by $\{X_t\}$ on $t = (0, H]$.

The portfolio consists of $K + 1$ positions.

The price of position $k$ at horizon depends on $t$, $\xi$, and the contractual terms of the instrument.

- For some exotic options, the price at $H$ will depend on the entire path of $X_t$ on $t = (0, H]$, so we need the filtration $\xi$ and not just $X_H$.

Position 0 represents the sub-portfolio of instruments for which there exist analytical pricing functions.

Positions 1 through $K$ must be priced by simulation.
Portfolio loss

- “Loss” is defined on a mark-to-market basis
  - Current value less discounted horizon value, less PDV of interim cashflows.
- Let $W_k$ be the loss on position $k$; $Y = \sum_k W_k$ is the portfolio loss.
  - Valuations are expressed in currency units, may be positive or negative.
- Conditional on $\xi$, $W_k(\xi)$ is non-stochastic.
- Except for position 0, we do not observe $W_k(\xi)$, but rather obtain noisy simulation estimates $\tilde{W}_k(\xi)$ and $\tilde{Y}(\xi)$. 
Simulation framework

Let $L$ be number of outer step trials. For each trial $\ell = 1, \ldots, L$:

1. Draw a single path $X_t$ for $t = (0, H]$ under the physical measure.
   - Let $\xi$ represent the relevant information for this path.

2. Evaluate the value of each position at horizon.
   - Accrue interim cashflows to $H$.
   - Closed-form price at $H$ for instrument 0.
   - Simulation with $N$ "inner step" trials to price each remaining positions $k = 1, \ldots, K$. Here we use the risk-neutral measure.

3. Discount back to time 0, subtract from current value, get our position losses $W_0(\xi), \tilde{W}_1(\xi), \ldots, \tilde{W}_K(\xi)$.

4. Portfolio loss $\tilde{Y}(\xi) = W_0(\xi) + \tilde{W}_1(\xi) + \ldots + \tilde{W}_K(\xi)$. 
Dependence in inner and outer steps

- Full dependence structure across the portfolio is captured in the period up to the model horizon.

- Inner step simulations are run independently across positions.
  - Value of position $k$ at time $H$ is simply a conditional expectation of its own subsequent cashflows.
  - Does not depend on future cashflows of other positions.

- Independent inner steps imply that pricing errors are independent across positions, and so tend to diversify away at portfolio level.

- Also reduces memory footprint of inner step: For position $k$, need only draw joint paths for the elements of $X_t$ upon which instrument $k$ depends.
Overview of our contribution

- Key insight of paper is that mean-zero pricing errors have minimal effect on estimation. Can set $N$ small!
- For finite $N$, estimators of exceedance probabilities, VaR and ES are biased (typically upwards).
- We obtain bias and variance of these estimators.
- Can allocate fixed computational budget between $L, N$ to minimize mean square error of estimator.
- Large portfolio asymptotics ($K \to \infty$).
- Jackknife method for bias reduction.
- Dynamic allocation scheme for greater efficiency.
Goal is efficient estimation of $\alpha = P(Y(\xi) > u)$ via simulation for a given $u$ (typically large).

If analytical pricing formulae were available, then for each generated $\xi$, $Y(\xi)$ would be observable.

In this case, outer step simulation would generate iid samples $Y_1(\xi_1), Y_2(\xi_2), \ldots, Y_L(\xi_L)$, and we would take average

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

as an estimator of $\alpha$. 
Pricing errors in inner step

- When analytical pricing formulae unavailable, we estimate $Y(\xi)$ via inner step simulation.
- Let $\zeta_{ki}(\xi)$ be zero-mean pricing error associated with $i^{th}$ “inner step” trial for position $k$.
- Let $Z_i(\xi)$ be the zero-mean portfolio pricing error associated with this inner step trial, i.e., $Z_i(\xi) = \sum_{k=1}^{K} \zeta_{ki}(\xi)$.
- Average portfolio error across trials is $\bar{Z}^N(\xi) = \frac{1}{N} \sum_{i=1}^{N} Z_i(\xi)$.
- Instead of $Y(\xi)$, we take as surrogate $\tilde{Y}(\xi) \equiv Y(\xi) + \bar{Z}^N(\xi)$.
- By the law of large numbers,

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

i.e., pricing error vanishes as $N$ grows large.
Mean square error in nested simulation

- We generate iid samples \((\tilde{Y}_1(\xi_1), \ldots, \tilde{Y}_L(\xi_L))\) via outer and inner step simulation, and take average

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

- Let \(\alpha_N \equiv P(\tilde{Y}(\xi) > u) = E[\hat{\alpha}_{LN}]\).

- Mean square error decomposes as

\[
E[\hat{\alpha}_{LN} - \alpha]^2 = E[\hat{\alpha}_{LN} - \alpha_N + \alpha_N - \alpha]^2 = E[\hat{\alpha}_{LN} - \alpha_N]^2 + (\alpha_N - \alpha)^2.
\]

- \(\hat{\alpha}_{LN}\) has binomial distribution, so variance term is

\[
E[\hat{\alpha}_{LN} - \alpha_N]^2 = \frac{\alpha_N(1 - \alpha_N)}{L}.
\]
Approximation for bias

Proposition:

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

where

$$\theta = \frac{-1}{2} \frac{d}{du} f(u) E[\sigma_\xi^2 | Y = u],$$

and where $$\sigma_\xi^2 = V[Z_1 | \xi]$$ is the conditional variance of the portfolio pricing error, and $$f(u)$$ is density of $$Y$$.

- Our approach follows Gouriéroux, Laurent and Scaillet (JEF, 2000) and Martin and Wilde (Risk, 2002) on sensitivity of VaR to portfolio allocation.
- $$\tilde{Y}$$ is mean-preserving spread of $$Y$$. Bias is upwards for large enough $$u$$, except under pathological cases.
- Similar approximations for bias in VaR and ES.
Example: Gaussian loss and pricing errors

- Highly stylized example for which RMSE has analytical expression.
- Homogeneous portfolio of $K$ positions.
- Let $X \sim \mathcal{N}(0, 1)$ be a market risk factor.
- Loss on position $k$ is $W_k = (X + \epsilon_k)/K$ per unit exposure where the $\epsilon_k$ are iid $\mathcal{N}(0, \nu^2)$.
  - Scale exposures by $1/K$ to ensure that portfolio loss distribution converges to $\mathcal{N}(0, 1)$ as $K \to \infty$.
- Implies portfolio loss $Y \sim \mathcal{N}(0, 1 + \nu^2/K)$.
- Assume pricing errors $\zeta_k$ iid $\mathcal{N}(0, \eta^2)$, so portfolio pricing error has variance $\sigma^2 = \eta^2/K$ for each inner step trial.
- Implies $\tilde{Y} = Y + \tilde{Z}^N \sim \mathcal{N}(0, 1 + \nu^2/K + \sigma^2/N)$. 

Gordy/Juneja (FRB/TIFR) Nested Simulation April 2008 13 / 28
Density of the loss distribution

Parameters: \( \nu = 3, \eta = 10, K = 100 \).
Exact and approximate bias in Gaussian example

- Variance of $Y$ is $s^2 = 1 + \nu^2/K$, variance of $\tilde{Y}$ is $\tilde{s}^2 = s^2 + \sigma^2/N$.
- Exact bias is
  \[ \alpha_N - \alpha = \Phi \left( -\frac{u}{\tilde{s}} \right) - \Phi \left( -\frac{u}{s} \right) \]
- Apply Proposition to approximate $\alpha_N - \alpha \approx \theta / N$ where
  \[ \theta = \phi \left( -\frac{u}{s} \right) \frac{u\sigma^2}{2s^3}. \]
Bias in Gaussian example

Parameters: $\nu = 3$, $\eta = 10$, $K = 100$, $u = F^{-1}(0.99)$. 
Optimal allocation of workload

- Total computational effort is $L(N\gamma_1 + \gamma_0)$ where
  - $\gamma_0$ is average cost to sample $\xi$ (outer step).
  - $\gamma_1$ is average cost per inner step sample.

- Fix overall computational budget $\Gamma$.
- Minimize mean square error subject to $\Gamma = L(N\gamma_1 + \gamma_0)$.
- For $\Gamma$ large, get

  \[
  N^* \approx \left( \frac{2\theta^2}{\alpha(1-\alpha)\gamma_1} \right)^{1/3} \Gamma^{1/3}
  \]

  \[
  L^* \approx \left( \frac{\alpha(1-\alpha)}{2\gamma_1^2\theta^2} \right)^{1/3} \Gamma^{2/3}
  \]

- Similar results in Lee (1998).
- Analysis for VaR and ES proceeds similarly, also find $N^* \propto \Gamma^{1/3}$. 
Approximate $\Gamma \propto N \cdot L$. Parameters: $\nu = 3$, $\eta = 10$, $K = 100$, $u = F^{-1}(0.99)$. 
Optimal $N$ in Gaussian example

Approximate $\Gamma \propto N \cdot L$. Parameters: $\nu = 3$, $\eta = 10$, $K = 100$, $u = F^{-1}(0.99)$. 
Consider an infinite sequence of exchangeable positions.

Let $\bar{Y}^K$ be average loss per position on a portfolio consisting of the first $K$ positions, i.e.,

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

Assume budget is $\chi K^\beta$ for $\chi > 0$ and $\beta \geq 1$.

Assume fixed cost per outer step is $\psi(m, K)$, so budget constraint is

$$L(KN_{\gamma_1} + \psi(m, K)) \leq \chi K^\beta$$ 

Proposition: For $\beta \leq 3$, $N^* \to 1$ as $K \to \infty.$
Optimal allocation as portfolio size varies

Budget is $\Gamma \propto N \cdot L$ for $K = 100$ and grows linearly with $K$. Parameters: $\nu = 3$, $\eta = 10$, $\Gamma = 2^{14}$, $u = F^{-1}(0.99)$. 
Jackknife estimators for bias correction

- In simplest version, divide inner step sample into two subsamples of \( N/2 \) each.
- Let \( \hat{\alpha}_j \) be the estimator of \( \alpha \) based on subsample \( j \).
- Observe that the bias in \( \hat{\alpha}_j \) is \( \theta/(N/2) \) plus terms of order \( O(1/N^{3/2}) \).
- We define the jackknife estimator \( a_{LN} \) as

\[
a_{LN} = 2\hat{\alpha}_{LN} - \frac{1}{2}(\hat{\alpha}_1 + \hat{\alpha}_2)
\]

- Jackknife estimator requires no additional simulation work.
- Can generalize by dividing the inner step sample into \( I \) overlapping subsamples of \( N - N/I \) trials each.
Bias reduction

The bias in $a_{LN}$ is

$$E[a_{LN}] - \alpha = 2\alpha_N - \alpha_{N/2} - \alpha$$

$$= 2(\alpha + \theta/N + O(1/N^{3/2})) - (\alpha + \theta/(N/2) + O(1/N^{3/2})) - \alpha$$

$$= \theta \left( \frac{2}{N} - \frac{1}{N/2} \right) + O(1/N^{3/2}) = O(1/N^{3/2}).$$

- First-order term in the bias is eliminated.
- Variance of $a_{LN}$ depends on covariances among $\hat{\alpha}_{LN}, \hat{\alpha}_1, \hat{\alpha}_2$. Tedious but tractable. Find $\text{Var}[a_{LN}] > \text{Var}[\hat{\alpha}_{LN}]$.
- Optimal choice of $N^*$ and $L^*$ changes because bias is a lesser consideration and variance a greater consideration.
  - Find $N^* \propto \Gamma^{1/4}$ (versus $1/3$ for uncorrected estimator) and $L^* \propto \Gamma^{3/4}$ (versus $2/3$).
Jackknife estimator for Gaussian example

- Both bias and variance have analytical expressions in this example.
  - Variance involves bivariate normal cdfs.
- Example with $N = 8$, $\nu = 3$, $\eta = 10$, $K = 100$, $u = F^{-1}(0.99)$:

<table>
<thead>
<tr>
<th></th>
<th>Bias (bp)</th>
<th>Std Dev (pct)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncorrected $\hat{\alpha}_{LN}$</td>
<td>37.8</td>
<td>$11.7/\sqrt{L}$</td>
</tr>
<tr>
<td>Jackknife $a_{LN}$</td>
<td>-3.8</td>
<td>$14.5/\sqrt{L}$</td>
</tr>
</tbody>
</table>

- Optimizing for fixed budget $N \cdot L = 2^{16}$:

<table>
<thead>
<tr>
<th></th>
<th>$N^*$</th>
<th>Bias (bp)</th>
<th>RMSE (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncorrected $\hat{\alpha}_{LN}$</td>
<td>22.6</td>
<td>12.9</td>
<td>23.5</td>
</tr>
<tr>
<td>Jackknife $a_{LN}$</td>
<td>6.0</td>
<td>-6.2</td>
<td>17.7</td>
</tr>
</tbody>
</table>
Dynamic allocation

- For given $\xi$, say we estimate $Y(\xi)$ with a small number $n_1$ of inner step trials.
- If $|\tilde{Y}^{n_1}(\xi) - u| \gg 0$, then $1[\tilde{Y}^{n_1}(\xi) > u]$ is a good estimator of $1[Y(\xi) > u]$, even though $\tilde{Y}^{n_1}(\xi)$ not a good estimator of $Y(\xi)$. ⇒ No need to do more inner step trials for this $\xi$!
- To implement this intuition in algorithm, fix $n_1, n_2$ and bandwidth $\epsilon$. For each outer step draw $\xi$:
  1. Simulate $n_1$ inner step trials to get $\tilde{Y}^{n_1}(\xi)$.
  2. If $\tilde{Y}^{n_1}(\xi) > u - \epsilon$, generate another $n_2$ inner step trials, set $\tilde{Y}^{DA}(\xi) = \tilde{Y}^{n_1+n_2}(\xi)$.
  3. Otherwise, we stop and set $\tilde{Y}^{DA}(\xi) = \tilde{Y}^{n_1}(\xi)$.
- Dynamic allocation estimator is

$$\hat{\alpha}^{DA} = \frac{1}{L} \sum_{\ell=1}^{L} 1[\tilde{Y}^{DA}(\xi_{\ell}) > u].$$
Lower bias, lower effort

- Average effort proportional to $n_1 + n_2 \cdot P(\tilde{Y}^{n_1}(\xi) > u - \epsilon) < n_1 + n_2$, so reduced relative to static estimator with $N = n_1 + n_2$.
- Bias under DA is

\[
P(\tilde{Y}^{DA} > u, \tilde{Y}^{n_1} > u - \epsilon) - P(Y > u) \\
= P(\tilde{Y}^{n_1+n_2} > u) - P(Y > u) - P(\tilde{Y}^{n_1+n_2} > u, \tilde{Y}^{n_1} \leq u - \epsilon) \\
= (\alpha_N - \alpha) - P(\tilde{Y}^{n_1+n_2} > u, \tilde{Y}^{n_1} \leq u - \epsilon) < \alpha_N - \alpha
\]

so DA introduces negative increment to bias, relative to static estimator.
- In typical application, $\alpha_N - \alpha > 0$. In this case, by choosing large enough $\epsilon$ can always reduce absolute bias relative to static estimator with $N = n_1 + n_2$.
  - Even when $\alpha_N - \alpha$ cannot be signed, we can bound the increase in bias relative to static scheme, so can trade off increase in bias vs reduction in effort.
- Variance is dominated by $\alpha(1 - \alpha)/L$, so insensitive to DA.
Dynamic allocation in Gaussian example

- $\mathbb{E}[\hat{\alpha}^{DA}]$ has analytical expression as a bivariate normal cdf.
- Fix $\nu = 3$, $\eta = 10$, $K = 100$, $u = F^{-1}(0.99)$ as in baseline examples.
- Static scheme with $N = 32$ has bias of 9.0 bp.
- DA with $n_1 = 1$, $n_2 = 31$, $\epsilon = \sqrt{\text{Var}[Y]}$ has bias of -0.4 bp and $\bar{N}^{DA} = 6.24$.
- Effort reduced by 80%, absolute bias by 95%.
Conclusion

- Large errors in pricing individual position can be tolerated so long as they can be diversified away.
  - Inner step gives errors that are zero mean and independent. Ideal for diversification!
  - In practice, large banks have many thousands of positions, so might have $N^* \approx 1$.

- Results suggest current practice is misguided.
  - Use of short-cut pricing methods introduces model misspecification.
  - Errors hard to bound and do not diversify away at portfolio level.
  - Practitioners should retain best pricing models that are available, run inner step with few trials.

- Dynamic allocation is robust and easily implemented in a setting with many state prices and both long and short exposures.
  - Stands in contrast to importance sampling, control variates, and other variance reduction methods.