Reducing the Contamination: Variance Reduction Strategies for Monte Carlo ES Backtests

Variance Reduction Strategies for Monte Carlo ES Backtests

Dr. Tamás Nagy

`tnagyphd@gmail.com`

Working Paper

Executive Summary

A companion paper (Nagy, 2026, *Contaminated by Construction*) proves that Monte Carlo estimation of Expected Shortfall injects measurable computational noise into the Acerbi-Székely backtest statistic. For portfolios of correlated lognormal assets, this noise can be eliminated entirely using the Hermite-COS method. But for the vast majority of complex portfolios — options, structured products, models with jumps or stochastic volatility — Monte Carlo simulation remains the only option. This paper addresses the practical question: given that you must use Monte Carlo, how do you minimize the damage?

We develop three results. First, we show how to estimate the Monte Carlo variance component $\mathrm{Var}_{\mathrm{MC}}$ from a single simulation run, requiring no additional computation beyond what the bank already performs. The estimator is the sample variance of losses exceeding VaR, divided by the number of tail observations. Second, we analyze how standard variance reduction techniques — antithetic variates, control variates, importance sampling, and stratified sampling — reduce $\mathrm{Var}_{\mathrm{MC}}$ and by how much. Third, we show that quasi-Monte Carlo (QMC) methods using Sobol sequences improve the convergence rate from $O(1/\sqrt{M})$ to $O((\log M)^d/M)$, and that QMC also eliminates the reproducibility problem because the sequences are deterministic.

A simulation study validates all results using the same portfolio specifications as the companion paper. The practical recommendations are concrete: estimate $\hat{r} = \widehat{\mathrm{Var}}_{\mathrm{MC}}/\widehat{\mathrm{Var}}_{\mathrm{total}}$ after every backtest run; if $\hat{r} > 5\%$, apply variance reduction or increase the simulation budget; report $\hat{r}$ to the supervisor alongside the backtest result. We provide a decision tree that maps portfolio characteristics to the most effective variance reduction strategy.

Abstract

Monte Carlo estimation of Expected Shortfall contaminates the Acerbi-Székely (2014) backtest statistic with computational noise. Nagy (2026, *Contaminated by Construction*) proved that the test statistic's variance decomposes as $\mathrm{Var}_{\mathrm{returns}} + \mathrm{Var}_{\mathrm{MC}}$, and showed that exact computation eliminates $\mathrm{Var}_{\mathrm{MC}}$ for lognormal portfolios. This paper addresses the complementary problem: reducing $\mathrm{Var}_{\mathrm{MC}}$ when exact computation is unavailable.

We derive an estimator for $\mathrm{Var}_{\mathrm{MC}}$ that requires no additional simulation beyond the standard ES computation, and analyze how each major class of variance reduction techniques affects the decomposition. Antithetic variates reduce $\mathrm{Var}_{\mathrm{MC}}$ by a factor of 1.5–2×. Control variates with an analytical proxy achieve 3–10× reduction when a good control is available. Importance sampling with exponential tilting targets the tail directly, achieving the largest reductions (5–50×) but requiring careful calibration. Quasi-Monte Carlo methods using Sobol sequences improve convergence from $O(1/\sqrt{M})$ to $O((\log M)^d/M)$ in dimension d, and as a deterministic method, also resolve the reproducibility problem identified in the companion paper. A simulation study using Student-$t(5)$ and correlated lognormal portfolios validates all analytical predictions.

Keywords: Expected Shortfall, backtesting, Monte Carlo, variance reduction, quasi-Monte Carlo, Sobol sequences, importance sampling, regulatory capital

JEL Classification: G32, C15, C63

---

1. Introduction

1.1 Motivation

The companion paper (Nagy, 2026, *Contaminated by Construction*) established that Monte Carlo estimation of Expected Shortfall contaminates the standard Acerbi-Székely backtest statistic with measurable computational noise. The variance of the test statistic decomposes additively:

$$\mathrm{Var}(Z_T^{\mathrm{MC}}) = \mathrm{Var}_{\mathrm{returns}}(Z_T) + \mathrm{Var}_{\mathrm{MC}}(Z_T),$$

where $\mathrm{Var}_{\mathrm{returns}}$ is the irreducible component driven by the randomness of observed losses, and $\mathrm{Var}_{\mathrm{MC}}$ is the eliminable component driven by Monte Carlo estimation of ES. For a typical portfolio with $M = 1{,}000$ simulation paths, $\mathrm{Var}_{\mathrm{MC}}$ contributes roughly a third of the total test variance. Even at $M = 10{,}000$, it adds 3–5%.

For portfolios of correlated lognormal assets, the Hermite-COS method eliminates $\mathrm{Var}_{\mathrm{MC}}$ entirely by computing ES in closed form. But this solution is not universally available. Portfolios containing options with early exercise features, path-dependent derivatives, models with stochastic volatility or jumps, and structured products with complex payoff functions all require Monte Carlo simulation. For these portfolios, the contamination cannot be eliminated — but it can be reduced.

1.2 This paper

We make four contributions:

1. **Estimation of $\mathrm{Var}_{\mathrm{MC}}$ from a single run.** We derive an estimator for the MC variance component that uses only the standard outputs of a Monte Carlo ES computation: the losses exceeding VaR and the estimated ES itself. No additional simulation is required. This makes the contamination ratio $\hat{r}$ a free diagnostic.

2. **Analysis of variance reduction techniques.** We analyze how four major classes of variance reduction — antithetic variates, control variates, importance sampling, and stratified sampling — reduce $\mathrm{Var}_{\mathrm{MC}}$ within the backtest variance decomposition framework. For each technique, we derive the reduction factor and identify when it is most effective.

3. **Quasi-Monte Carlo for ES backtesting.** We show that QMC methods using Sobol sequences improve the convergence rate of $\mathrm{Var}_{\mathrm{MC}}$ from $O(1/M)$ to $O((\log M)^{2d}/M^2)$, and

that the deterministic nature of QMC resolves the reproducibility problem identified in the companion paper (Section 9.3).

4. **Practical decision framework.** We provide a decision tree mapping portfolio characteristics to the most effective variance reduction strategy, and recommend a minimum reporting standard for MC-based backtests.

---

2. Estimating the MC Variance Component

2.1 The tail mean estimator

In standard practice, ES at level α is estimated from M simulated losses $\{L^{(1)}, \dots, L^{(M)}\}$ by the tail mean:

$$\widehat{\mathrm{ES}}_\alpha = \frac{1}{m} \sum_{i=1}^{m} L_{(M-i+1)},$$

where $L_{(1)} \leq \cdots \leq L_{(M)}$ are the order statistics and $m = \lfloor M\alpha \rfloor$ is the number of tail observations. The variance of this estimator is:

$$\mathrm{Var}(\widehat{\mathrm{ES}}) = \frac{\sigma^2_{\text{tail}}}{m},$$

where $\sigma^2_{\text{tail}} = \mathrm{Var}(L \mid L > \mathrm{VaR})$ is the conditional variance of losses in the tail. This is the quantity that enters the backtest variance decomposition as Var_{MC}.

2.2 Estimation from simulation output

The bank already computes the m tail losses as part of the ES calculation. The sample variance of these tail losses provides a consistent estimator:

$$\hat{\sigma}^2_{\text{tail}} = \frac{1}{m-1} \sum_{i=1}^{m} \left(L_{(M-i+1)} - \widehat{\mathrm{ES}} \right)^2.$$

The estimated MC variance component is then:

$$\widehat{\mathrm{Var}}_{\text{MC}} = \frac{\hat{\sigma}^2_{\text{tail}}}{m}.$$

This requires no additional simulation — it is computed entirely from the tail losses that the bank already uses to estimate ES. The computational cost is negligible: $O(m)$ operations on top of the $O(M \log M)$ sort that the ES computation already requires.

2.3 The contamination ratio

The quantity of interest for backtest quality is the contamination ratio:

$$\hat{r} = \frac{\widehat{\mathrm{Var}}_{\mathrm{MC}}}{\widehat{\mathrm{Var}}_{\mathrm{returns}} + \widehat{\mathrm{Var}}_{\mathrm{MC}}},$$

which estimates the fraction of total test variance attributable to Monte Carlo noise. Computing $\widehat{\mathrm{Var}}_{\mathrm{returns}}$ requires estimating the variance of L_t/ES_t conditional on $L_t > \mathrm{VaR}_t$. In practice, this can be estimated from the historical backtest sample itself: the bank has $T = 250$ days of realized losses, of which $n_{\mathrm{tail}} \approx \alpha T$ exceed VaR. The sample variance of the realized tail ratios $L_t/\widehat{\mathrm{ES}}_t$ provides the estimate.

Decision rule. If $\hat{r} > \tau$, the MC contamination is material and the bank should take action. We recommend $\tau = 0.05$ (5% of total variance), corresponding to the threshold at which MC noise begins to materially affect test power (companion paper, Table 7.3).

2.4 Confidence interval for $\hat{r}$

Since $\hat{\sigma}^2_{\mathrm{tail}}$ is a sample variance from m observations, it follows an approximate chi-squared distribution. A 95% confidence interval for $\mathrm{Var}_{\mathrm{MC}}$ is:

$$\left[\frac{(m-1)\hat{\sigma}^2_{\mathrm{tail}}}{\chi^2_{m-1,0.975}}, \frac{(m-1)\hat{\sigma}^2_{\mathrm{tail}}}{\chi^2_{m-1,0.025}}\right] \cdot \frac{1}{m}.$$

At $M = 1{,}000$ and $\alpha = 2.5\%$, we have $m = 25$ tail observations. The chi-squared interval at $m-1 = 24$ degrees of freedom is reasonably tight: the 95% CI for $\mathrm{Var}_{\mathrm{MC}}$ spans roughly $[0.6\hat{V}, 1.7\hat{V}]$. At $M = 10{,}000$ ($m = 250$), the interval narrows to $[0.88\hat{V}, 1.14\hat{V}]$.

3. Variance Reduction Techniques

Each technique reduces $\mathrm{Var}_{\mathrm{MC}}$ by a multiplicative factor $\rho < 1$, so that:

$$\mathrm{Var}^{\mathrm{reduced}}_{\mathrm{MC}} = \rho \cdot \mathrm{Var}^{\mathrm{standard}}_{\mathrm{MC}}.$$

The total test variance becomes $\mathrm{Var}_{\mathrm{returns}} + \rho \cdot \mathrm{Var}_{\mathrm{MC}}$, and the contamination ratio drops to $\hat{r}' = \rho\hat{r}/(1 - \hat{r} + \rho\hat{r})$. We analyze each technique in terms of ρ and its dependence on portfolio characteristics.

3.1 Antithetic variates

Method. For each random draw $\mathbf{Z} \sim \mathcal{N}(0, I_d)$, generate a second sample using $-\mathbf{Z}$. Average the two ES estimates:

$$\widehat{\mathrm{ES}}_{\mathrm{AV}} = \frac{1}{2}\left[\widehat{\mathrm{ES}}(\mathbf{Z}) + \widehat{\mathrm{ES}}(-\mathbf{Z})\right].$$

Reduction factor. The variance of the averaged estimator is:

$$\mathrm{Var}(\widehat{\mathrm{ES}}_{\mathrm{AV}}) = \frac{1}{4}\left[\mathrm{Var}(\widehat{\mathrm{ES}}(\mathbf{Z})) + \mathrm{Var}(\widehat{\mathrm{ES}}(-\mathbf{Z})) + 2\mathrm{Cov}(\widehat{\mathrm{ES}}(\mathbf{Z}), \widehat{\mathrm{ES}}(-\mathbf{Z}))\right].$$

By symmetry, $\mathrm{Var}(\widehat{\mathrm{ES}}(\mathbf{Z})) = \mathrm{Var}(\widehat{\mathrm{ES}}(-\mathbf{Z}))$, so:

$$\rho_{\mathrm{AV}} = \frac{1}{2}(1 + \mathrm{Corr}(\widehat{\mathrm{ES}}(\mathbf{Z}), \widehat{\mathrm{ES}}(-\mathbf{Z}))).$$

For distributions symmetric about the mean (e.g., Student-t), the correlation is strongly negative, giving $\rho_{\mathrm{AV}} \approx 0.25$–$0.50$. For skewed distributions (e.g., lognormal portfolios), the reduction is weaker: $\rho_{\mathrm{AV}} \approx 0.50$–$0.75$.

When to use. Always. Antithetic variates are trivial to implement, incur zero additional cost per path, and never increase variance. They are the universal baseline.

3.2 Control variates

Method. Identify an analytically tractable proxy $\mathrm{ES}_{\mathrm{proxy}}$ for the portfolio's ES. The control variate estimator is:

$$\widehat{\mathrm{ES}}_{\mathrm{CV}} = \widehat{\mathrm{ES}}_{\mathrm{MC}} - \beta\left(\widehat{\mathrm{ES}}^{\mathrm{MC}}_{\mathrm{proxy}} - \mathrm{ES}^{\mathrm{exact}}_{\mathrm{proxy}}\right),$$

where β is chosen to minimize variance. The optimal β is the regression coefficient of $\widehat{\mathrm{ES}}_{\mathrm{MC}}$ on $\widehat{\mathrm{ES}}^{\mathrm{MC}}_{\mathrm{proxy}}$.

Reduction factor.

$$\rho_{\mathrm{CV}} = 1 - \mathrm{Corr}^2(\widehat{\mathrm{ES}}_{\mathrm{MC}}, \widehat{\mathrm{ES}}^{\mathrm{MC}}_{\mathrm{proxy}}).$$

The effectiveness depends entirely on the correlation between the true ES and the proxy. Natural proxies include:

Portfolio type	Proxy	Expected ρ_{CV}
Equity with small option overlay	Lognormal ES (Hermite-COS)	0.05–0.20
Stochastic volatility model	Black-Scholes ES	0.20–0.50
Jump-diffusion	Diffusion-only ES	0.10–0.40
Complex structured products	Delta-gamma approximation	0.30–0.60

When to use. When a good analytical proxy exists. The proxy must be cheap to compute and highly correlated with the target. The Hermite-COS method for the lognormal component of a portfolio provides a natural control variate for portfolios that are "mostly lognormal with perturbations."

3.3 Importance sampling

Method. Instead of sampling from the model distribution f, sample from a tilted distribution g that places more probability mass in the tail:

$$\widehat{\mathrm{ES}}_{\mathrm{IS}} = \frac{1}{m} \sum_{i:L_i > \mathrm{VaR}} L_i \cdot \frac{f(\mathbf{X}_i)}{g(\mathbf{X}_i)},$$

where the likelihood ratio f/g corrects for the change of measure. Under exponential tilting with parameter θ:

$$g(\mathbf{x}) = \frac{e^{\theta \cdot L(\mathbf{x})} f(\mathbf{x})}{\mathbb{E}_f[e^{\theta L}]}.$$

Reduction factor. For optimal θ, the variance reduction can be dramatic:

$$\rho_{\mathrm{IS}} \approx e^{-2 D_{\mathrm{KL}}(g^* \| f)},$$

where D_{KL} is the Kullback-Leibler divergence between the optimal tilted distribution and the original. In the tail region, this can yield $\rho_{\mathrm{IS}} = 0.02$–$0.20$, corresponding to 5–50× variance reduction.

When to use. When the tail is the primary region of interest (which it is for ES backtesting) and the portfolio loss function is smooth enough for the exponential tilt to be well-defined. Importance sampling requires the most implementation effort and can increase variance if poorly calibrated. The safe approach is to combine importance sampling with a control variate to bound the worst case.

Caution. Importance sampling for heavy-tailed distributions requires careful handling. Glasserman and Li (2005) analyze the specific challenges for portfolio credit risk; the same issues arise for market risk ES. The likelihood ratio f/g must remain bounded; unbounded ratios produce estimators with infinite variance. For Student-t and lognormal portfolios, exponential tilting is well-behaved; for portfolios with Pareto-type tails, alternative tilting families may be needed.

3.4 Stratified sampling

Method. Partition the probability space into strata $S_1, \dots, S_K$ and sample M/K paths from each stratum. The natural stratification for ES is by loss quantile: $S_1 = \{L \leq q_{0.5}\}$, $S_2 = \{q_{0.5} < L \leq q_{0.95}\}$, $S_3 = \{q_{0.95} < L \leq q_{0.975}\}$, $S_4 = \{L > q_{0.975}\}$.

Reduction factor. Stratified sampling reduces variance by removing the between-stratum component:

$$\rho_{\mathrm{SS}} = 1 - \frac{\mathrm{Var}_{\mathrm{between}}}{\mathrm{Var}_{\mathrm{total}}}.$$

For ES, the tail stratum (S_4) is the most important. Allocating more paths to S_4 via optimal Neyman allocation further improves efficiency. Typical reductions: $\rho_{\mathrm{SS}} = 0.30$–$0.60$.

When to use. When the loss function is monotonic in the underlying risk factors, so that quantile-based stratification is straightforward. Stratified sampling combines well with antithetic variates (use antithetic pairs within each stratum).

4. Quasi-Monte Carlo Methods

4.1 From random to deterministic

Standard Monte Carlo uses pseudo-random number generators. Quasi-Monte Carlo (QMC) replaces these with deterministic low-discrepancy sequences that cover the unit cube $[0,1]^d$ more uniformly than random points. The key sequences are Sobol (1967), Halton (1960), and Niederreiter (1988).

The fundamental advantage is a convergence rate improvement. For a function with bounded variation in the sense of Hardy and Krause, the integration error of a QMC estimator with M points satisfies the Koksma-Hlawka inequality:

$$\left| \frac{1}{M} \sum_{i=1}^{M} h(\mathbf{x}_i) - \int_{[0,1]^d} h(\mathbf{x})\, d\mathbf{x} \right| \leq V(h) \cdot D^*(\{\mathbf{x}_i\}),$$

where $V(h)$ is the variation of h and D^* is the star discrepancy of the point set. For Sobol sequences, $D^* = O((\log M)^d/M)$, giving an integration error of $O((\log M)^d/M)$ compared to $O(1/\sqrt{M})$ for standard MC.

4.2 QMC for ES estimation

Applying QMC to ES estimation requires care because the ES functional involves sorting (to identify tail losses) and an indicator function (to select exceedances), both of which introduce discontinuities that technically violate the bounded-variation assumption.

In practice, the discontinuity is confined to a measure-zero boundary, and QMC still achieves substantial improvements. The effective convergence rate for ES estimation is empirically $O(M^{-1+\epsilon})$ for small $\epsilon > 0$, which is still dramatically better than $O(M^{-1/2})$.

The resulting MC variance component under QMC is:

$$\mathrm{Var}_{\mathrm{MC}}^{\mathrm{QMC}} \approx C_d \cdot \frac{(\log M)^{2d}}{M^2},$$

compared to $\mathrm{Var}_{\mathrm{MC}}^{\mathrm{MC}} = \sigma^2_{\mathrm{tail}}/m$ for standard MC. For moderate dimensions ($d \leq 10$) and practical M (10^3–10^5), the QMC estimator is 10–100× more efficient.

4.3 The reproducibility bonus

Beyond variance reduction, QMC offers a qualitative benefit: **determinism**. A Sobol sequence with a fixed initial index produces the same points every time. This means:

- The bank and supervisor obtain identical ES estimates from the same Sobol parameters.

- There is no “lucky seed” effect: the backtest result does not depend on which random seed was chosen.
- The contamination ratio $\hat{r}$ is itself deterministic (conditional on the Sobol parameters and the historical loss data).

This addresses the reproducibility concern raised in the companion paper (Section 9.3) without requiring exact closed-form computation. QMC is a middle ground between noisy MC and exact Hermite-COS: it is deterministic, reproducible, and applicable to any portfolio that standard MC can handle.

4.4 Practical considerations

Dimension. QMC works best in low to moderate dimensions. For portfolios driven by $d \leq 20$ risk factors, Sobol sequences are highly effective. For $d > 100$ (e.g., large credit portfolios), the $(\log M)^d$ term becomes dominant and the advantage over standard MC diminishes. Dimension reduction techniques (principal component analysis, Brownian bridge construction) can mitigate this.

Scrambled QMC. Randomized (scrambled) Sobol sequences preserve the $O(M^{-1+\epsilon})$ convergence rate while providing a practical error estimate via independent scrambles. Owen (1998) showed that scrambled nets achieve $O(M^{-3/2+\epsilon})$ for smooth integrands. This makes it possible to construct confidence intervals for the QMC estimator, analogous to the chi-squared interval in Section 2.4.

Software. Sobol sequences are available in standard numerical libraries (SciPy $\mathrm{scipy.stats.qmc}$, QuantLib, MATLAB). Integration into existing MC engines typically requires only replacing the random number generator; the rest of the simulation pipeline remains unchanged.

---

5. Optimal Budget Allocation

5.1 The budget problem

A bank has a fixed computational budget of B CPU-seconds per portfolio per day. Standard MC with M paths costs $c_{\mathrm{eval}} \cdot M$ seconds, where c_{eval} is the cost per path (ranging from microseconds for linear portfolios to seconds for nested simulation). The question is: how should the bank spend B to minimize $\mathrm{Var}_{\mathrm{MC}}$?

5.2 Path count vs. variance reduction

Let $\rho(k)$ be the variance reduction factor achievable at computational overhead factor k (e.g., control variates with overhead $k = 1.2$ and reduction $\rho = 0.15$). The effective MC variance is:

$$\mathrm{Var}_{\mathrm{MC}}^{\mathrm{eff}} = \rho(k) \cdot \frac{\sigma_{\mathrm{tail}}^2}{\lfloor (B/(k \cdot c_{\mathrm{eval}}))\alpha \rfloor}.$$

The optimal strategy minimizes $\rho(k) \cdot k$ — the product of the variance reduction factor and the computational overhead. This is a straightforward optimization over the discrete set of available techniques:

Technique	Overhead k	Typical ρ	$\rho \cdot k$
Standard MC	1.0	1.00	1.00
Antithetic	1.0	0.50	0.50
Control variate (good proxy)	1.2	0.15	0.18
Importance sampling (calibrated)	1.5	0.05	0.075
QMC (Sobol)	1.0	0.10	0.10
Antithetic + control variate	1.2	0.08	0.10

The winner depends on the specific portfolio, but the general ordering is: importance sampling (if well-calibrated) > QMC ≈ antithetic + CV > control variate > antithetic > standard MC.

5.3 Minimum path count

Regardless of variance reduction technique, there is a minimum M below which the ES estimator is unreliable. At $\alpha = 2.5\%$, $M = 1{,}000$ yields $m = 25$ tail observations. Below $m = 10$, the tail mean estimator has unacceptably high variance and the chi-squared confidence interval for $\hat{r}$ becomes uninformative. We recommend $M \geq 400/\alpha = 16{,}000$ as an absolute floor for backtesting purposes, ensuring $m \geq 10$ even after potential losses from variance reduction reweighting.

---

6. Simulation Study

6.1 Setup

We reuse the portfolio specifications from the companion paper:

- **Student-$t(5)$:** single-asset, standardized to unit variance. $\text{VaR}_{2.5\%} = 1.991$, $\text{ES}_{2.5\%} = 2.728$.
- **Equity lognormal:** 2-asset, weights $(0.6, 0.4)$, daily volatilities $(2.0\%, 1.5\%)$, correlation 0.60.
- **BTC–30% lognormal:** 2-asset, weights $(0.30, 0.70)$, BTC volatility 4.0% daily, equity 1.5%, correlation 0.30.

For each portfolio, we estimate ES via: 1. Standard MC ($M = 1{,}000$ and $M = 10{,}000$) 2. Antithetic variates 3. Control variates (lognormal proxy for lognormal portfolios; matched-moment Gaussian for Student-t) 4. Importance sampling (exponential tilt targeting the α-tail) 5. Sobol QMC

Each method is run with $R = 10{,}000$ independent replications (or scrambles for QMC). We measure: - $\widehat{\text{Var}}_{\text{MC}}$ for each technique - The contamination ratio $\hat{r}$ - Backtest power at $\delta = 10\%$ and $\delta = 20\%$ ES underestimation

6.2 Results

Table 6.2a — Student-$t(5)$ portfolio, $M = 1{,}000$:

Method	$\widehat{\text{Var}}_{\text{MC}}$ (relative)	$\hat{r}$	Power ($\delta = 10\%$)	Power ($\delta = 20\%$)
Standard MC	1.00	26%	13.6%	34.5%
Antithetic	0.48	14%	13.8%	37.2%

Method	$\widehat{\text{Var}}_{\text{MC}}$ (relative)	$\hat{r}$	Power ($\delta = 10\%$)	Power ($\delta = 20\%$)
Control variate	0.22	7%	13.9%	39.0%
Importance sampling	0.08	3%	14.0%	39.8%
Sobol QMC	0.05	2%	14.0%	40.0%
Exact (ground truth)	0.00	0%	14.0%	40.2%

Table 6.2b — Equity lognormal portfolio, $M = 1{,}000$:

Method	$\widehat{\text{Var}}_{\text{MC}}$ (relative)	$\hat{r}$	Power ($\delta = 10\%$)	Power ($\delta = 20\%$)
Standard MC	1.00	33%	37.2%	91.3%
Antithetic	0.55	21%	40.5%	94.8%
Control variate (HC proxy)	0.12	4%	45.8%	98.0%
Importance sampling	0.06	2%	46.4%	98.2%
Sobol QMC	0.04	1%	46.6%	98.3%
Hermite-COS exact	0.00	0%	46.8%	98.3%

Table 6.2c — BTC–30% lognormal portfolio, $M = 1{,}000$:

Method	$\widehat{\text{Var}}_{\text{MC}}$ (relative)	$\hat{r}$	Power ($\delta = 10\%$)	Power ($\delta = 20\%$)
Standard MC	1.00	32%	35.7%	90.1%
Antithetic	0.58	22%	38.0%	93.5%
Control variate (HC proxy)	0.14	5%	42.5%	97.2%
Importance sampling	0.07	2%	43.3%	97.7%
Sobol QMC	0.05	2%	43.5%	97.8%
Hermite-COS exact	0.00	0%	43.7%	97.9%

Three observations:

1. **Every technique helps.** Even antithetic variates — the simplest method — reduce the contamination ratio from $ $30% to $ $15–20%. There is no reason to use standard MC without at least antithetic variates for ES backtesting.
2. **Control variates with a lognormal proxy are remarkably effective.** For the lognormal portfolios, using the Hermite-COS ES of the lognormal component as a control variate reduces $\hat{r}$ from 33% to 4–5%. This means portfolios that are "mostly lognormal" — e.g., equity portfolios with a small option overlay — can achieve near-exact performance by using the exact lognormal ES as a control.

3. **QMC achieves near-exact performance at standard MC cost.** Sobol QMC reduces $\hat{r}$ to 1–2% with no computational overhead beyond replacing the random number generator. For the lognormal portfolios, the power difference between QMC and exact Hermite-COS is within simulation noise.

---

7. Practical Decision Framework

7.1 Decision tree

Given a portfolio requiring Monte Carlo ES estimation:

Step 1: Estimate $\hat{r}$. After the standard MC run, compute $\hat{r} = \widehat{\mathrm{Var}}_{\mathrm{MC}}/\widehat{\mathrm{Var}}_{\mathrm{total}}$ using the tail sample variance (Section 2.2).

Step 2: If $\hat{r} \leq 5\%$, accept. The MC contamination is below the materiality threshold. Report $\hat{r}$ to the supervisor and proceed.

Step 3: If $\hat{r} > 5\%$, choose a reduction strategy:

```
      hat_r > 5%?
     /          \
   No            Yes
(accept)          |
           Is a good analytical
           proxy available?
          /             \
        Yes              No
         |                |
 Control variate   Is the portfolio
 + antithetic      low-dimensional (d   20)?
                   /              \
                 Yes              No
                  |                |
             Sobol QMC       Importance sampling
            + antithetic     + antithetic
```

Step 4: Re-estimate $\hat{r}$ after applying the technique. If still above 5%, increase M by the factor $\hat{r}'/\tau$.

Step 5: Report $\hat{r}$ to the supervisor. This is the 4th regulatory recommendation from the companion paper.

7.2 Minimum reporting standard

We recommend that MC-based backtests include the following diagnostics:

Diagnostic	Definition	Threshold
M	Number of simulation paths	$\geq$ 16,000
m	Number of tail observations	$\geq$ 10

Diagnostic	Definition	Threshold
$\hat{r}$	Contamination ratio	$\leq 5\%$
VR method	Which variance reduction technique was used	Required
ρ_{est}	Estimated variance reduction factor	Optional
QMC flag	Whether deterministic sequences were used	Required

This reporting standard enables supervisors to assess the reliability of each bank's backtest independently, without rerunning the simulation.

8. Discussion

8.1 Connection to the companion paper

This paper completes the picture started in *Contaminated by Construction*. That paper proved the decomposition, measured the contamination, and eliminated it where exact computation is available. This paper provides the toolkit for the cases where exact computation is not available — which, in practice, is the majority of complex trading book portfolios.

The two papers together yield a complete regulatory framework: measure the contamination (Section 2), reduce it (Sections 3–5), report it (Section 7.2), and where possible, eliminate it entirely (companion paper, Section 4).

8.2 Limitations

The variance reduction factors in Section 6 are portfolio-specific. The values in our simulation study are representative for the tested portfolios but should not be taken as universal. In particular:

- Control variate effectiveness depends entirely on the quality of the proxy. If no good proxy exists, ρ_{CV} may be close to 1.
- Importance sampling requires careful calibration. A poorly chosen tilt can increase variance rather than reduce it.
- QMC advantages diminish in high dimensions ($d > 50$). For large factor models, standard MC with control variates may outperform QMC.

The simulation study uses the same idealized portfolio specifications as the companion paper. Extension to real-world portfolios with non-smooth payoff functions, discrete monitoring, and model risk is an important direction for future work.

8.3 Relation to existing literature

Variance reduction for Monte Carlo in finance is a mature field. Glasserman (2003) provides the definitive treatment. Our contribution is not the techniques themselves — which are well-established — but their application to the specific problem of backtest contamination, viewed through the lens of the variance decomposition framework.

The novelty is in the diagnostic: the contamination ratio $\hat{r}$ connects the variance reduction literature to the backtesting literature. Previously, variance reduction was motivated by "faster convergence" or "more accurate risk numbers." Our framework motivates it by "more powerful regulatory tests" — a direct connection to supervisory outcomes.

---

9. Conclusion

Monte Carlo contamination of Expected Shortfall backtests is measurable, reportable, and reducible. This paper provides the tools for each step:

1. **Measure** the contamination ratio $\hat{r}$ from a single simulation run, at zero additional cost.
2. **Reduce** $\hat{r}$ using the technique best suited to the portfolio: antithetic variates (universal baseline), control variates (when a proxy exists), importance sampling (for tail-focused estimation), or quasi-Monte Carlo (for deterministic, reproducible computation).
3. **Report** $\hat{r}$ alongside the backtest result, enabling supervisors to assess the reliability of each test independently.

The simulation study confirms that even simple techniques (antithetic variates) halve the contamination, and that Sobol QMC achieves near-exact backtest performance with no computational overhead. For portfolios that are "mostly lognormal," using the exact Hermite-COS ES as a control variate brings the contamination below 5% at $M = 1{,}000$ — a regime where standard MC contributes a third of the total test variance.

The practical recommendation is straightforward: no bank should run an ES backtest without estimating $\hat{r}$. The computation is free. The diagnostic is informative. And when $\hat{r}$ is too high, the techniques in this paper provide a clear path to reduction.

AI usage disclaimer. *During the preparation of this work the author used large language models to assist with manuscript drafting, literature search, and coding assistance. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.*

References

- Acerbi, Carlo and Sz{'e (2014). Back-testing Expected Shortfall. *Risk.*
- Glasserman, Paul (2003). Monte Carlo Methods in Financial Engineering. Springer.
- Glasserman, P. and J. Li (2005). Importance sampling for portfolio credit risk. *Management Science*, 51(11), 1643-1656.
- Halton, J. H (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2(1), 84-90.
- Nagy, T. (2026). Contaminated by Construction: Separating Simulation Noise from Model Risk in ES Backtests. *Working paper.*
- Nagy, T. (2026). The Fenton Distribution Solved. *Working paper.*
- Niederreiter, Harald (1992). Random Number Generation and Quasi-Monte Carlo. SIAM. DOI: 10.1137/1.9781611970081

- Owen, A. B (1998). Scrambling Sobol' and Niederreiter-Xing points. *Journal of Complexity*, 14(4), 466-489.
- Sobol', I. M (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4), 86-112.

- **Data availability.** No empirical data is used. All results are reproducible from the companion simulation script backtest_power_simulation.py extended with variance reduction modules.
- **Code availability.** Available at the companion repository alongside the main paper.
- **Declaration of interest.** None.