Accepted by *INFORMS Journal on Computing*
manuscript JOC-2021-09-OA-261.R4

Monte Carlo Methods for Economic Capital

Yajuan Li, Zachary T. Kaplan, and Marvin K. Nakayama
Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102 {yl935@njit.edu, ztk4@njit.edu, marvin@njit.edu}

Economic capital (EC) is a risk measure used by financial firms to specify capital levels to protect (with high probability) against large unforeseen losses. Defined as the difference between an (extreme) quantile and the mean of the loss distribution, the EC is often estimated via Monte Carlo methods. While simple random sampling (SRS) may be effective in estimating the mean, it can be inefficient for the extreme quantile in the EC. Applying importance sampling (IS) may lead to an efficient quantile estimator but can do poorly for the mean. Measure-specific IS (MSIS) instead uses IS to estimate only the quantile, and the mean is independently handled via SRS. We analyze large-sample properties of EC estimators obtained via SRS only, IS only, MSIS, IS using a defensive mixture, and a double estimator using both SRS and IS to estimate both the quantile and the mean, establishing Bahadur-type representations for the EC estimators and proving they obey central limit theorems. We provide asymptotic theory comparing the estimators when the loss is the sum of a large number of independent and identically distributed random variables. Numerical and simulation results, including for a large portfolio credit risk model with dependent obligors, complement the theory.

Key words: economic capital, value-at-risk, importance sampling.
History: This paper was first submitted on Sep 24, 2021. Revisions submitted Sep 15, 2022; Jan 27, 2023; May 11, 2023.

1. Introduction

A credit portfolio comprises loans, bonds, and other financial instruments subject to default. To protect a financial firm against unforeseen large losses of such a portfolio, a risk manager may employ various *risk measures* (e.g., Section 2.3 of McNeil et al. (2015) and Hong et al. (2014)) related to the random loss Y of the portfolio over a given time period (e.g., one year) to specify capital levels. One risk measure is the p-quantile ξ for $p \approx 1$ (e.g., $p = 0.999$), also known as the *value-at-risk* (VaR) or the $100p$th *percentile*, where ξ

is a constant such that $P(Y \leq \xi) = p$. The *expected shortfall* (alternatively, *conditional tail expectation* or *conditional VaR*) is the conditional expectation of Y given that $Y > \xi$.

This paper studies the *economic capital* (EC) $\eta = \xi - \mu$, the difference between the p-quantile ξ and the mean loss μ, where $p \approx 1$; see Klaassen and van Eeghen (2009, p. 5), Lütkebohmert (2009, Section 2.4), and Scandizzo (2016, p. 194). Also called the *credit* (Jorion 2011, p. 595), *relative* (Jorion 2007, p. 108) or *mean-adjusted VaR* (McNeil et al. 2015, p. 300), the EC is used to determine capital needed to cover unexpected losses with high probability. Indeed, Deutsche Bank (2018, p. 63) appears to employ EC with $p = 0.999$: "In line with our economic capital framework, economic capital for credit risk is set at a level to absorb with a probability of 99.9% very severe aggregate unexpected losses within one year. Our economic capital for credit risk is derived from the loss distribution of a portfolio via Monte Carlo Simulation of correlated rating migrations." (The bank used $p = 0.9998$ before 2017; see Deutsche Bank (2018, p. 46).)

Monte Carlo simulation with *simple random sampling* (SRS) may produce noisy EC estimates because the rarity of extreme losses makes estimating ξ with $p \approx 1$ difficult. This motivates applying *variance-reduction techniques* (VRTs), such as *importance sampling* (IS); e.g., see Chapters V and VI of Asmussen and Glynn (2007) and Chapter 4 of Glasserman (2004) for overviews. Glasserman and Li (2005) develop IS methods for estimation of a tail probability of multifactor credit risk models using a Gaussian copula to model dependencies of default events among obligors (e.g., corporations to which the bank provided loans). Bassamboo et al. (2008) extend the IS methods to incorporate dependencies with non-Gaussian copulas.

While IS can be effective in reducing the variance of estimators of tail probabilities and extreme quantiles, it may produce worse estimators of the mean loss, the other component of the EC. An IS technique designed to work well for estimating an extreme quantile typically samples more in the tail of interest and less around the mean, degrading the mean's estimator. This motivates separately estimating the quantile and the mean via different simulation techniques. One strategy uses IS for the estimation of the quantile and independently applies SRS for estimating the mean. Goyal et al. (1992) call this approach *measure-specific importance sampling* (MSIS), which they employ to separately estimate the numerator and denominator in a ratio of means in which only one mean corresponds to a rare event. We also consider two more methods that combine IS and SRS in other ways. One applies *IS with a defensive mixture* (ISDM), as developed by Hesterberg (1995), in which the IS distribution is a mixture of a new distribution and the original one. The other approach we call a *double estimator* (DE), which estimates both ξ and μ utilizing both IS and SRS, combining all estimators with user-specified weights. We establish a *central limit theorem* (CLT) for each EC estimator (as the overall sample size $n \to \infty$).

When the loss Y is the sum of m independent and identically distributed (i.i.d.) random variables, we analytically compare the estimators of η, ξ, and μ in terms of their CLTs' asymptotic variances, in a limiting regime where $m \to \infty$ and the quantile level p simultaneously approaches 1 exponentially as $p = 1 - e^{-\beta m}$ for fixed $\beta > 0$. Originally developed by Glynn (1996) to analyze SRS and IS estimators of quantiles, this asymptotic framework has practical relevance for studying EC: bank portfolios can easily be exposed to thousands or even tens of thousands of obligors, and extreme quantiles are used in industry.

For the i.i.d. sum model, we derive asymptotic expressions (as $m \to \infty$) for the relative errors (REs) of our estimators, where an estimator's RE (e.g., L'Ecuyer et al. (2010)) is the ratio of its CLT's asymptotic standard deviation and the absolute value of the (nonzero) estimand (see (36) in Section 6.3), so an estimator with smaller RE is preferable, all other things being equal. Using the asymptotic notation (as $m \to \infty$) of exact rate $\Theta(\cdot)$ and weak and strict lower bounds $\Omega(\cdot)$ and $\omega(\cdot)$ (defined in Section 6.4), Table 1 summarizes the main findings of Theorem 5 in Section 6.5 on the limiting behavior of the RE of each EC estimator, where $\alpha_\star > 1$ is some constant, and $x \vee y = \max(x, y)$. Thus, for large m, the

Table 1 **For the i.i.d. sum model with m summands, the relative errors of the EC estimators using SRS, IS, and DE with fixed weights grow exponentially in m as $m \to \infty$, where $\alpha_\star > 1$ and $\beta > 0$ are constants. In contrast, MSIS and DE with optimal weights have vanishing RE as $m \to \infty$, and ISDM has bounded RE when $\mu \neq 0$.**

Method	RE
SRS	$\omega\left(e^{(\beta/2)m - \sqrt{m}}/\sqrt{m}\right)$
IS	$\Omega\left(\alpha_\star^{m/2}/\sqrt{m}\right)$
DE (fixed weights)	$\Omega\left([\alpha_\star \vee e^{\beta}]^{m/2} e^{-\sqrt{m}}/\sqrt{m}\right)$
DE (optimal weights)	$\Theta\left(1/\sqrt{m}\right)$
MSIS	$\Theta\left(1/\sqrt{m}\right)$
ISDM (when $\mu \neq 0$)	$\Theta(1)$

MSIS η estimator slightly outperforms ISDM (when the mean is nonzero), and both are exponentially better than SRS only, IS only, and DE with fixed weights. We further consider DE with optimally tuned weights (varying with m), which MSIS can never beat because MSIS merely specializes DE with one particular (typically suboptimal) choice of weights. However, determining the optimal DE weights encounters practical challenges as it requires specifying many parameters, whose values are unknown (although they could be estimated through a pilot simulation). But Theorem 5 additionally shows (see (57)) that the ratio of the variances of optimal DE and the much simpler MSIS converges to 1 exponentially quickly as $m \to \infty$ in the i.i.d. sum model, making MSIS a compelling alternative as its performance rapidly becomes indistinguishable from that of optimal DE. We also provide numerical (i.e., quadrature, not simulation) results for i.i.d. sums confirming the asymptotic theory; see Figures 1, A.1, and A.2 in Section 7.1 and Appendix A. (Theorem 5 additionally analyzes each estimator's work-normalized RE (WNRE), e.g., as in (37), to further account for the expected CPU time to generate a single output, which typically grows with m. Each WNRE behaves asymptotically as its corresponding RE in Table 1 multiplied by $\sqrt{m}$.)

We complement the theoretical and numerical results of the i.i.d. sum model through simulation experiments with a significantly more complex portfolio credit risk model (PCRM), further demonstrating the benefits of estimating η when $p \approx 1$ via MSIS over SRS, IS, ISDM, and DE with fixed weights. For the models in Glasserman and Li (2005) and Bassamboo et al. (2008), calculating μ may not require simulation because of their models' tractability. But more complicated stochastic models may preclude analytically evaluating the mean loss.

The main contributions of our paper are as follows. We analyze several different estimators of the EC. Although many of the techniques have been previously applied successfully to study problems arising in operations research and management science, some may not have been used before in a finance context. Our theoretical asymptotic study (see Table 1) of the i.i.d. sum model (Section 6) provides a rich body of technical analysis, yielding deep insights into the behavior of the methods observed in simulation experiments with the substantially more complicated PCRM (with 1000 dependent obligors) in Section 7.2. Another key result relates to quantile estimation. Quantile estimators often satisfy a CLT in which the asymptotic variance is a ratio, where the numerator depends on the simulation method applied and the denominator is the squared density at the quantile. The asymptotic analysis of SRS and IS quantile estimators for our i.i.d. sum model in Glynn (1996) covers only the variances' numerators, which we extend by also considering the denominators to provide a fuller understanding of the methods. Furthermore, our analysis for this model examines the other methods for estimating quantiles (Theorem 7 in Appendix F.2) and also studies all of our estimators of the EC (Theorem 5) and the mean (Theorem 6 in Appendix F.1).

The rest of the paper unfolds as follows. Section 2 gives our mathematical framework. Section 3 presents the SRS estimator of η. It further establishes a type of large-sample Bahadur (1966) representation for the estimator, and proves a CLT, which we also do for the other methods considered. Section 4 applies IS to estimate η. Section 5 describes the methods that combine IS and SRS: MSIS (Section 5.1), ISDM (Section 5.2), and DE (Section 5.3). Section 6 provides our theoretical asymptotic analyses of the η estimators when Y is the sum of m i.i.d. random variables as $m \to \infty$ with the quantile level $p = 1 - e^{-\beta m}$. Section 7 gives numerical (quadrature) and Monte Carlo results comparing the methods, with Section 7.1 considering the model from Section 6, and Section 7.2 examining a more complicated model, an extension of the PCRM from Glasserman and Li (2005). Section 8 gives concluding remarks. Appendices contain additional numerical results (Appendix A), provide all proofs (Appendices B–F), describe the simulation methodology used on the PCRM in Section 7.2 (Appendix G), and summarize the main notation and acronyms (Appendix H). Our theorems on the Bahadur representations and CLTs for the SRS, IS, and MSIS estimators previously appeared without proofs in Kaplan et al. (2018), which also describes batching and sectioning methods (Asmussen and Glynn 2007, Section V.5), briefly covered here, to construct large-sample confidence intervals (CIs) for η. Kaplan et al. (2018) do not consider any of the material in Sections 5.2, 5.3, 6, 7 and the appendices. All of the appendices are available in the online supplement. [1]

2. Mathematical Framework

Let Y be a random variable for the loss of a credit-portfolio model over a given time horizon, and let F be its *cumulative distribution function* (CDF). Assume that F is unknown or computationally intractable, but we have a simulation model that generates an observation of $Y \sim F$, where $\sim$ denotes "is distributed as". Let $\mu = E[Y]$ be the mean of $Y \sim F$, where $E[\cdot]$ is the expectation operator. For a CDF H and $0 < q < 1$, we define the *q-quantile* of

[1] https://doi.org/10.1287/ijoc.2021.0261

H as $H^{-1}(q) = \inf\{y : H(y) \geq q\}$; e.g., the median μ' is the 0.5-quantile, also known as the 50th percentile. Our goal is to use simulation to estimate the EC $\eta = \xi - \mu$, where $\xi = F^{-1}(p)$ for a given $0 < p < 1$. (Note that $\xi \equiv \xi_p$ and $\eta \equiv \eta_p$ depend on p, but we omit the subscript p to simplify notation.)

Often but not always, we assume that the loss Y has the form

$$Y = c(\mathbf{X}) \tag{1}$$

for a known function $c : \Re^d \to \Re$ with $d \geq 1$, and random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ having a specified joint CDF G, where G can allow the components of $\mathbf{X}$ to be dependent and non-identically distributed. We view the function c in (1) as a (complicated) computer code, transforming an input $\mathbf{X} \sim G$ into a loss $Y \sim F$. For example, Section 7.2 will consider a large multi-factor PCRM with dependent obligors, as in Glasserman and Li (2005), Bassamboo et al. (2008), and Lütkebohmert (2009), in which the loss Y has a form in (1), with $\mathbf{X}$ having mutually independent components.

3. Simple Random Sampling

We begin with the application of SRS to estimate η, and the results in this section do not require that the loss $Y \sim F$ has the form in (1). Let $Y_1, Y_2, \ldots, Y_n$ be a *random sample* of size n from F; i.e., $Y_1, Y_2, \ldots, Y_n$ are i.i.d. with CDF F. When Y has the form in (1), we generate $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ as i.i.d. copies of $\mathbf{X} \sim G$, and let $Y_i = c(\mathbf{X}_i)$ for each $i = 1, 2, \ldots, n$. In general, define the SRS estimator of the mean μ as the sample mean

$$\widehat{\mu}_{\mathrm{SRS},n} = \frac{1}{n} \sum_{i=1}^{n} Y_i. \tag{2}$$

We define the SRS p-quantile estimator $\widehat{\xi}_{\mathrm{SRS},n}$ by inverting the *empirical CDF* $\widehat{F}_{\mathrm{SRS},n}$:

$$\widehat{\xi}_{\mathrm{SRS},n} = \widehat{F}^{-1}_{\mathrm{SRS},n}(p), \qquad \text{where } \widehat{F}_{\mathrm{SRS},n}(y) = \frac{1}{n} \sum_{i=1}^{n} I(Y_i \leq y) = \frac{1}{n} \sum_{i=1}^{n} \left[1 - I(Y_i > y)\right] \tag{3}$$

with $I(\cdot)$ as the indicator function, equaling 1 (resp., 0) if its argument is true (resp., false). Then the SRS estimator of the EC $\eta = \xi - \mu$ is

$$\widehat{\eta}_{\mathrm{SRS},n} = \widehat{\xi}_{\mathrm{SRS},n} - \widehat{\mu}_{\mathrm{SRS},n}. \tag{4}$$

We can compute $\widehat{\xi}_{\mathrm{SRS},n}$ by order statistics. Let $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$ be the sorted values of $Y_1, Y_2, \ldots, Y_n$, and then $\widehat{\xi}_{\mathrm{SRS},n} = Y_{(\lceil np \rceil)}$, where $\lceil \cdot \rceil$ is the ceiling (i.e., round-up) function. For simplicity, we do not consider other SRS quantile estimators (Hyndman and Fan (1996)), e.g., with an interpolated CDF estimator, which typically share the same large-sample properties as $\widehat{\xi}_{\mathrm{SRS},n}$.

While the estimator $\widehat{\mu}_{\mathrm{SRS},n}$ in (2) of the mean is a sample average, the p-quantile estimator $\widehat{\xi}_{\mathrm{SRS},n} = \widehat{F}^{-1}_{\mathrm{SRS},n}(p)$ is *not*, so the EC estimator $\widehat{\eta}_{\mathrm{SRS},n}$ in (4) is also not a sample average, complicating its analysis. However, Bahadur (1966) shows that $\widehat{\xi}_{\mathrm{SRS},n}$ can be well approximated by a sample average of i.i.d. quantities when the sample size n is large, and we will do the same for $\widehat{\eta}_{\mathrm{SRS},n}$. To accomplish this, define f as the derivative (when it exists) of the CDF F. Also let $\Rightarrow$ represent convergence in distribution (e.g., Chapter 5 of Billingsley (1995)). Then if $f(\xi) > 0$, the p-quantile estimator satisfies

$$\widehat{\xi}_{\mathrm{SRS},n} = \xi - \frac{1}{f(\xi)}\left[\widehat{F}_{\mathrm{SRS},n}(\xi) - p\right] + R_n, \tag{5}$$

$$\text{with } \sqrt{n}R_n \Rightarrow 0 \quad \text{as } n \to \infty; \tag{6}$$

see Section 2.5 of Serfling (1980). If F is twice differentiable at ξ, then Kiefer (1967) proves that for either choice of sign below,

$$\limsup_{n\to\infty} \pm \frac{n^{3/4} R_n}{(\log\log n)^{3/4}} = \frac{2^{5/4}[p(1-p)]^{1/4}}{3^{3/4} f(\xi)} \quad \text{with probability 1.} \tag{7}$$

Note that (7) implies (6), and we call (5) with (6) (resp., (7)) a *weak* (resp., *strong*) *Bahadur representation* for $\widehat{\xi}_{\mathrm{SRS},n}$. The key point of (5)–(7) is that they permit analyzing the large-sample properties of $\widehat{\xi}_{\mathrm{SRS},n}$ through the simpler $\widehat{F}_{\mathrm{SRS},n}(\xi)$, which is a sample average of i.i.d. terms by (3). As next seen, the SRS EC estimator $\widehat{\eta}_{\mathrm{SRS},n}$ has similar *Bahadur-type representations* and obeys a CLT, with $N(q, s^2)$ denoting a normal random variable with mean q and variance s^2, and $\mathrm{Var}[\cdot]$ (resp., $\mathrm{Cov}[\cdot,\cdot]$) as the variance (resp., covariance) operator; see Appendix B for the proof.

THEOREM 1. *Suppose that $Y_1, Y_2, \ldots$ are i.i.d. with CDF F, and F is differentiable at ξ with $f(\xi) > 0$.*

(i) The SRS EC estimator in (4) *then satisfies*

$$\widehat{\eta}_{\mathrm{SRS},n} = \eta - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{f(\xi)}\left[[1 - I(Y_i > \xi)] - p\right] + \left[Y_i - \mu\right]\right) + R_n \tag{8}$$

with R_n from (5)*, so* (6) *holds. If also F is twice differentiable at ξ, then* (7) *further holds.*

(ii) If in addition $\sigma^2_{\mathrm{SRS}} \equiv \mathrm{Var}[Y] < \infty$*, then*

$$\sqrt{n}\,[\widehat{\eta}_{\mathrm{SRS},n} - \eta] \Rightarrow N(0, \zeta^2_{\mathrm{SRS}}) \quad \text{as } n \to \infty, \quad \text{where} \quad \zeta^2_{\mathrm{SRS}} = \frac{\chi^2_{\mathrm{SRS}}}{f^2(\xi)} + \sigma^2_{\mathrm{SRS}} - 2\frac{\gamma_{\mathrm{SRS}}}{f(\xi)}, \tag{9}$$

$$\chi^2_{\mathrm{SRS}} = p(1-p), \quad \text{and} \quad \gamma_{\mathrm{SRS}} = \mathrm{Cov}[I(Y > \xi), Y] = E[I(Y > \xi)Y] - (1-p)\mu. \tag{10}$$

The Bahadur-type representations in Theorem 1(i) give useful insight into the large-sample behavior of $\widehat{\eta}_{\mathrm{SRS},n}$, showing that approximating $\widehat{\eta}_{\mathrm{SRS},n} - \eta$ by a sample mean of the i.i.d. terms results in a remainder R_n that vanishes faster than $1/\sqrt{n}$ by (6) or (7). This then implies the CLT in (9) when $\sigma^2_{\mathrm{SRS}} < \infty$.

Under an additional assumption that F has a density f and the second derivative of F is bounded in a neighborhood of ξ, Lin et al. (1980) prove that the SRS estimators of a quantile and the mean obey a joint CLT, and Ferguson (1999) shows the same under the weaker additional assumption that the density f is continuous at ξ. While (9) follows from either result, the Bahadur-type representations in Theorem 1(i) can further be used to show the asymptotic validity of a sectioning CI for η, as in Kaplan et al. (2018).

4. Importance Sampling

When $p \approx 1$, estimators of $\xi = F^{-1}(p)$ and the corresponding EC $\eta = \xi - \mu$ may have large variance, motivating the use of a variance-reduction technique. We consider applying IS, but other VRTs are also possible. To use IS and the methods in the next section, we assume Y has the form in (1) from now on.

The mean of Y is then $\mu = E_G[c(\mathbf{X})]$, where, for any CDF $G^\dagger$ on $\Re^d$, $E_{G^\dagger}$ (resp., $P_{G^\dagger}$, $\mathrm{Var}_{G^\dagger}$, $\mathrm{Cov}_{G^\dagger}$) is the expectation (resp., probability, variance, covariance) operator when $\mathbf{X} \sim G^\dagger$. Let $\widetilde{G}$ be a CDF on $\Re^d$ such that (the measure of) G is absolutely continuous (Billingsley 1995, p. 422) with respect to $\widetilde{G}$. A *change of measure* ensures

$$\mu = E_G[c(\mathbf{X})] = \int_{\Re^d} c(\mathbf{x})\mathrm{d}G(\mathbf{x}) = \int_{\Re^d} c(\mathbf{x})\frac{\mathrm{d}G(\mathbf{x})}{\mathrm{d}\widetilde{G}(\mathbf{x})}\mathrm{d}\widetilde{G}(\mathbf{x}) = E_{\widetilde{G}}[c(\mathbf{X})L(\mathbf{X})], \text{ for } L(\mathbf{x}) = \frac{\mathrm{d}G(\mathbf{x})}{\mathrm{d}\widetilde{G}(\mathbf{x})} \tag{11}$$

as the *likelihood ratio* (LR), for $\mathbf{x} \in \Re^d$. To estimate μ via IS, we sample i.i.d. $\mathbf{X}_i \sim \widetilde{G}$, $i = 1, 2, \ldots, n$, and

$$\widehat{\mu}_{\mathrm{IS},n} = \frac{1}{n}\sum_{i=1}^{n} c(\mathbf{X}_i)L(\mathbf{X}_i) \tag{12}$$

is an unbiased estimator of μ by (11). (IS reduces to SRS when $\widetilde{G} = G$ as then $L(\mathbf{x}) \equiv 1$.)

Suppose that under both G and $\widetilde{G}$, the components of $\mathbf{X} = (X_1, \ldots, X_d)$ are mutually independent. Then for G_j (resp., $\widetilde{G}_j$) denoting the marginal CDF of X_j under G (resp., $\widetilde{G}$), we have $G(\mathbf{x}) = \prod_{j=1}^d G_j(x_j)$ and $\widetilde{G}(\mathbf{x}) = \prod_{j=1}^d \widetilde{G}_j(x_j)$ for $\mathbf{x} = (x_1, \ldots, x_d)$. If we further suppose that each G_j (resp., $\widetilde{G}_j$) has a density or probability mass function g_j (resp., $\tilde{g}_j$), the likelihood ratio in (11) becomes $L(\mathbf{x}) = \prod_{j=1}^d \frac{g_j(x_j)}{\tilde{g}_j(x_j)}$.

To estimate the p-quantile ξ by IS, we use an approach of Glynn (1996): first apply IS to estimate the CDF F, and then invert the estimated CDF to obtain the IS quantile estimator. Specifically, write

$$1 - F(y) = E_G[I(c(\mathbf{X}) > y)] = E_{\widetilde{G}}[I(c(\mathbf{X}) > y)L(\mathbf{X})] \tag{13}$$

through a change of measure. By (13), an unbiased estimator of $F(y)$ is $\widehat{F}_{\mathrm{IS},n}(y)$, with

$$\widehat{F}_{\mathrm{IS},n}(y) = 1 - \frac{1}{n}\sum_{i=1}^n I(c(\mathbf{X}_i) > y)L(\mathbf{X}_i), \quad \text{and} \quad \widehat{\xi}_{\mathrm{IS},n} = \widehat{F}^{-1}_{\mathrm{IS},n}(p), \tag{14}$$

where $\mathbf{X}_i \sim \widetilde{G}$, $i = 1, 2, \ldots, n$, are the same as in (12). We call $\widehat{F}_{\mathrm{IS},n}(y)$ and $\widehat{\xi}_{\mathrm{IS},n}$ the *IS estimators* of $F(y)$ and ξ, respectively. To compute $\widehat{\xi}_{\mathrm{IS},n}$, let $Y_i = c(\mathbf{X}_i)$, and let $Y_{1:n} \le Y_{2:n} \le \cdots \le Y_{n:n}$ be the sorted values of $Y_1, Y_2, \ldots, Y_n$. Defining $\mathbf{X}_{i::n}$ as the $\mathbf{X}_j$ corresponding to $Y_{i:n}$ results in $\widehat{\xi}_{\mathrm{IS},n} = Y_{i_p:n}$, with i_p the greatest integer for which $\sum_{\ell=i_p}^n L(\mathbf{X}_{\ell::n}) \ge n(1-p)$. Chu and Nakayama (2012) establish that the quantile estimator obtained via a combination of IS and stratified sampling obeys a weak Bahadur representation, with $\widehat{\xi}_{\mathrm{IS},n}$ in (14) being a special case of IS only; i.e., their Theorem 4.2 shows that if

$$\text{there exist constants } \epsilon > 0 \text{ and } \lambda > 0 \text{ such that } E_{\widetilde{G}}[I(c(\mathbf{X}) > \xi - \lambda)L^{2+\epsilon}(\mathbf{X})] < \infty, \tag{15}$$

then

$$\widehat{\xi}_{\mathrm{IS},n} = \xi - \frac{1}{f(\xi)}[\widehat{F}_{\mathrm{IS},n}(\xi) - p] + \tilde{R}_n, \quad \text{with} \quad \sqrt{n}\tilde{R}_n \Rightarrow 0 \quad \text{as } n \to \infty. \tag{16}$$

The fact that $F(y) = E_G[I(c(\mathbf{X}) \le y)] = E_{\widetilde{G}}[I(c(\mathbf{X}) \le y)L(\mathbf{X})]$ suggests another CDF estimator, $\widehat{F}'_{\mathrm{IS},n}(y) = \frac{1}{n}\sum_{i=1}^n I(c(\mathbf{X}_i) \le y)L(\mathbf{X}_i)$, with each $\mathbf{X}_i \sim \widetilde{G}$, which leads to another p-quantile estimator $\widehat{\xi}'_{\mathrm{IS},n} = \widehat{F}'^{-1}_{\mathrm{IS},n}(p)$. Theorem 4.1 of Chu and Nakayama (2012) (resp., Sun and Hong (2010)) establishes a weak (resp., strong) Bahadur representation for $\widehat{\xi}'_{\mathrm{IS},n}$. When estimating the p-quantile with $p \approx 1$ using IS, Glynn (1996) shows that for a simple example, the p-quantile estimator $\widehat{\xi}_{\mathrm{IS},n}$ in (14) has smaller asymptotic variance than the estimator $\widehat{\xi}'_{\mathrm{IS},n}$. (But $\widehat{\xi}'_{\mathrm{IS},n}$ can have smaller asymptotic variance than $\widehat{\xi}_{\mathrm{IS},n}$ when $p \approx 0$.)

The IS estimator of the EC is then

$$\widehat{\eta}_{\mathrm{IS},n} = \widehat{\xi}_{\mathrm{IS},n} - \widehat{\mu}_{\mathrm{IS},n}, \tag{17}$$

with both $\widehat{\xi}_{\mathrm{IS},n}$ and $\widehat{\mu}_{\mathrm{IS},n}$ computed from the same i.i.d. sample $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, with each $\mathbf{X}_i \sim \widetilde{G}$. The following result, proven in Appendix C, shows that $\widehat{\eta}_{\mathrm{IS},n}$ has a Bahadur-type representation and obeys a CLT.

Theorem 2. *Suppose that $Y \sim F$ has the form in* (1)*, and $f(\xi) > 0$. Suppose that $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are i.i.d. with CDF $\widetilde{G}$, where (the measure induced by) G is absolutely continuous with respect to $\widetilde{G}$. Also suppose that* (15) *holds for $L(\mathbf{x})$ in* (11) *and* (13)*. Then the following hold.*

(i) The IS EC estimator in (17) *satisfies*

$$\widehat{\eta}_{\mathrm{IS},n} = \eta - \frac{1}{n}\sum_{i=1}^{n}\left[\frac{[1 - I(c(\mathbf{X}_i) > \xi)L(\mathbf{X}_i)] - p}{f(\xi)} + c(\mathbf{X}_i)L(\mathbf{X}_i) - \mu\right] + \tilde{R}_n \tag{18}$$

with $\sqrt{n}\tilde{R}_n \Rightarrow 0$ *as* $n \to \infty$.

(ii) If in addition $\sigma^2_{\mathrm{IS}} \equiv \mathrm{Var}_{\widetilde{G}}[c(\mathbf{X})L(\mathbf{X})] < \infty$, then $\sqrt{n}\,[\widehat{\eta}_{\mathrm{IS},n} - \eta] \Rightarrow N(0, \zeta^2_{\mathrm{IS}})$ as $n \to \infty$, where

$$\zeta^2_{\mathrm{IS}} = \frac{\chi^2_{\mathrm{IS}}}{f^2(\xi)} + \sigma^2_{\mathrm{IS}} - 2\frac{\gamma_{\mathrm{IS}}}{f(\xi)}, \quad \textit{with} \ \ \chi^2_{\mathrm{IS}} \equiv \mathrm{Var}_{\widetilde{G}}[I(c(\mathbf{X}) > \xi)L(\mathbf{X})], \quad \textit{and} \tag{19}$$

$$\gamma_{\mathrm{IS}} \equiv \mathrm{Cov}_{\widetilde{G}}[I(c(\mathbf{X}) > \xi)L(\mathbf{X}), c(\mathbf{X})L(\mathbf{X})] = E_G[I(c(\mathbf{X}) > \xi)c(\mathbf{X})L(\mathbf{X})] - (1-p)\mu. \tag{20}$$

5. Methods that Combine SRS and IS

Section 4 estimates ξ and μ from the same data generated from IS distribution $\widetilde{G}$, but the resulting estimator of $\eta = \xi - \mu$ can have large variance. When $p \approx 1$, ξ is a property of the right tail of F, whereas μ typically measures the distribution's central tendency. Thus, while SRS can often effectively estimate the mean μ of F, a VRT designed to analyze only the tail of F may fare poorly in estimating μ.

When $p \approx 1$, the heuristic reason that an IS CDF $\widetilde{G}$ for $\mathbf{X}$ designed to estimate only ξ can do badly for μ arises from the LR in (11) often being immense. To see why, first express the second moment of the IS estimator of μ as $m_2 \equiv E_{\widetilde{G}}[c^2(\mathbf{X})L^2(\mathbf{X})] = E_G[c^2(\mathbf{X})L(\mathbf{X})]$

by a change of measure. The original CDF G usually assigns much of its probability to points $\mathbf{x}$ with $c(\mathbf{x})$ near the mean μ. But $\widetilde{G}$ shifts most of that mass to values $\mathbf{x}'$ with $c(\mathbf{x}') \approx \xi$, making points $\mathbf{x}$ with $c(\mathbf{x}) \approx \mu$ rare under $\widetilde{G}$. Thus, the LR $L(\mathbf{x}) = \mathrm{d}G(\mathbf{x})/\mathrm{d}\widetilde{G}(\mathbf{x})$ is enormous for these common $\mathbf{x}$ under G, leading to $m_2 = E_G[c^2(\mathbf{X})L(\mathbf{X})]$ and the variance $\sigma^2_{\mathrm{IS}} = \mathrm{Var}_{\widetilde{G}}[c(\mathbf{X})L(\mathbf{X})]$ of the IS estimator of μ being large (μ is unchanged by (11)).

5.1. Measure-Specific Importance Sampling (MSIS)

To address these issues, MSIS estimates only ξ by IS and independently estimates μ using SRS. Goyal et al. (1992) apply MSIS to estimate a ratio of means, in which only one corresponds to a rare event and is thus handled via IS, and the other (non-rare) mean is simulated independently without IS. More generally, we can use one VRT to estimate ξ and another to estimate μ, where VRTs other than IS may instead be employed.

We next give the details of MSIS. For an overall sample size n, we specify a fraction $0 < \delta < 1$ of the sample size to estimate ξ by IS, and we use SRS to estimate μ with the rest of the sample size. Let δn be the sample size estimating ξ via IS, and $(1-\delta)n$ be the sample size estimating μ by SRS, both assumed to be integer-valued; if not, replace δn and $(1-\delta)n$ by $\lfloor \delta n \rfloor$ and $\lfloor (1-\delta)n \rfloor$, respectively, where $\lfloor \cdot \rfloor$ is the floor function. Let $\widehat{F}_{\mathrm{IS},\delta n}$ be the IS CDF estimator in (14) but with sample size δn instead of n, and $\widehat{\xi}_{\mathrm{IS},\delta n} = \widehat{F}^{-1}_{\mathrm{IS},\delta n}(p)$ is the resulting p-quantile estimator. Also let $\widehat{\mu}_{\mathrm{SRS},(1-\delta)n}$ be the SRS estimator of μ in (2) with sample size $(1-\delta)n$ instead of n. Then the MSIS estimator of η is

$$\widehat{\eta}_{\mathrm{MSIS},n} = \widehat{\xi}_{\mathrm{IS},\delta n} - \widehat{\mu}_{\mathrm{SRS},(1-\delta)n}. \tag{21}$$

The next result, proven in Appendix D, gives a weak Bahadur-type representation and CLT for $\widehat{\eta}_{\mathrm{MSIS},n}$.

THEOREM 3. *Suppose that $Y \sim F$ has the form in (1), $f(\xi) > 0$, (15) holds, and (the measure induced by) G is absolutely continuous with respect to $\widetilde{G}$. Then the following hold for any fixed $0 < \delta < 1$.*

(i) The MSIS EC estimator in (21) satisfies

$$\widehat{\eta}_{\mathrm{MSIS},n} = \eta - \frac{1}{f(\xi)}[\widehat{F}_{\mathrm{IS},\delta n}(\xi) - p] - (\widehat{\mu}_{\mathrm{SRS},(1-\delta)n} - \mu) + \tilde{R}_{n,\delta}, \text{ with } \sqrt{n}\tilde{R}_{n,\delta} \Rightarrow 0 \text{ as } n \to \infty. \tag{22}$$

(ii) If in addition $\sigma^2_{\mathrm{SRS}} < \infty$, then for χ^2_{IS} from (19),

$$\sqrt{n}\,[\widehat{\eta}_{\mathrm{MSIS},n} - \eta] \Rightarrow N(0, \zeta^2_{\mathrm{MSIS}}) \quad \text{as} \quad n \to \infty, \qquad \text{where} \qquad \zeta^2_{\mathrm{MSIS}} = \frac{\chi^2_{\mathrm{IS}}}{\delta f^2(\xi)} + \frac{\sigma^2_{\mathrm{SRS}}}{1-\delta}. \tag{23}$$

In contrast to (9) and (19), (23) has no covariance term as MSIS estimates ξ and μ independently. Also, the value of $\zeta^2_{\rm MSIS}$ depends on δ, with $\delta^* = [\chi_{\rm IS}/f(\xi)]/[\sigma_{\rm SRS} + (\chi_{\rm IS}/f(\xi))]$ minimizing $\zeta^2_{\rm MSIS}$. But the values of $\sigma^2_{\rm SRS}$, $\chi^2_{\rm IS}$ and $f(\xi)$ are unknown. However we can employ a two-stage procedure, with a pilot run to roughly estimate the unknown parameters, which are used in the second stage with the resulting estimated δ^*.

5.2. Importance Sampling with a Defensive Mixture Distribution (ISDM)

To estimate simultaneously multiple metrics (including the mean and a tail probability), Hesterberg (1995) develops ISDM, which applies IS, as in Section 4, with $\mathbf{X} \sim \widetilde{G}_{\rm ISDM} \equiv \delta G^* + (1-\delta)G$, where G^* (resp., G) is a new (resp., original) joint CDF for $\mathbf{X}$, and $0 \le \delta \le 1$ is a user-specified constant. We can sample $\mathbf{X}$ from the mixture $\widetilde{G}_{\rm ISDM}$ by generating $\mathbf{X}$ from G^* (resp., G) with probability δ (resp., $1-\delta$). The ISDM EC estimator has the form (17) based on (12) and (14), with the LR in (11) and (13) as

$$L_{\rm ISDM}(\mathbf{x}) = \frac{\mathrm{d}G(\mathbf{x})}{\mathrm{d}\widetilde{G}_{\rm ISDM}(\mathbf{x})} = \frac{\mathrm{d}G(\mathbf{x})}{\delta \mathrm{d}G^*(\mathbf{x}) + (1-\delta)\mathrm{d}G(\mathbf{x})}, \quad \text{so} \quad L_{\rm ISDM}(\mathbf{x}) \le \frac{1}{1-\delta} \quad \text{for all } \mathbf{x}. \tag{24}$$

Thus, for $\delta \in (0,1)$, ISDM prevents the LR from being too big (Section 5), making $\widetilde{G}_{\rm ISDM}$ a defensive mixture. A special case of IS, the ISDM EC estimator obeys Theorem 2, where $\delta \in (0,1)$ ensures the assumed absolute continuity and (15) hold. When $\delta = 0$ (resp., $\delta = 1$), ISDM reduces to SRS (resp., IS with $\mathbf{X} \sim G^*$).

CDF G^* itself can be a mixture of r CDFs, so then $\widetilde{G}_{\rm ISDM}$ mixes $r+1$ CDFs. Other works using a mixture for IS include Owen and Zhou (2000) and Glasserman and Juneja (2008).

5.3. Double Estimator

A *double estimator* (DE) provides another way of combining IS and SRS to estimate the EC. As with MSIS, DE generates an IS (resp., SRS) sample of size δn (resp., $(1-\delta)n$), with the two samples independent. But in contrast to MSIS, DE employs both the IS and SRS samples to estimate both ξ and μ. More specifically, we use the IS sample of size δn to construct estimators $\widehat{\xi}_{{\rm IS},\delta n}$ and $\widehat{\mu}_{{\rm IS},\delta n}$ in (14) and (12), respectively. Also, we form estimators $\widehat{\xi}_{{\rm SRS},(1-\delta)n}$ and $\widehat{\mu}_{{\rm SRS},(1-\delta)n}$ in (3) and (2), respectively, from the SRS sample of size $(1-\delta)n$. For user-specified constants $\upsilon_1, \upsilon_2 \in [0,1]$, we define the DE EC estimator as

$$\widehat{\eta}_{{\rm DE},n} = \left[\upsilon_1 \widehat{\xi}_{{\rm IS},\delta n} + \upsilon_1' \widehat{\xi}_{{\rm SRS},(1-\delta)n}\right] - \left[\upsilon_2 \widehat{\mu}_{{\rm IS},\delta n} + \upsilon_2' \widehat{\mu}_{{\rm SRS},(1-\delta)n}\right] \equiv \widehat{\xi}_{{\rm DE},n} - \widehat{\mu}_{{\rm DE},n}, \tag{25}$$

where $\upsilon_1' = 1 - \upsilon_1$ and $\upsilon_2' = 1 - \upsilon_2$. When $\upsilon_1 = \upsilon_2$, $\widehat{\eta}_{\mathrm{DE},n}$ is a weighted sum of two independent EC estimators: $\widehat{\eta}_{\mathrm{IS},\delta n}$ of (17) with an IS sample of size δn and weight $\upsilon_1 = \upsilon_2$, and $\widehat{\eta}_{\mathrm{SRS},(1-\delta)n}$ of (4) with an SRS sample of size $(1-\delta)n$ and weight $\upsilon_1' = \upsilon_2'$. But (25) also allows $\upsilon_1 \neq \upsilon_2$, and DE becomes MSIS when $\upsilon_1 = 1$ and $\upsilon_2 = 0$. Also, DE reduces to SRS (resp., IS) when $\upsilon_1 = \upsilon_2 = \delta = 0$ (resp., $\upsilon_1 = \upsilon_2 = \delta = 1$). The following, whose proof appears in Appendix E, gives a Bahadur-type representation and CLT for $\widehat{\eta}_{\mathrm{DE},n}$.

THEOREM 4. *Suppose that $Y \sim F$ has the form in* (1) *and $f(\xi) > 0$. Suppose that $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{(1-\delta)n}$ are i.i.d. with CDF G, and $\mathbf{X}_1', \mathbf{X}_2', \ldots, \mathbf{X}_{\delta n}'$ are i.i.d. with CDF $\widetilde{G}$ and independent of $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{(1-\delta)n}$, where (the measure induced by) G is absolutely continuous with respect to $\widetilde{G}$, which satisfies* (15). *Then the following hold for any $\delta, \upsilon_1, \upsilon_2 \in [0,1]$.*

(i) The DE estimator $\widehat{\eta}_{\mathrm{DE},n}$ in (25), *constructed from $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{(1-\delta)n}$ and $\mathbf{X}_1', \mathbf{X}_2', \ldots, \mathbf{X}_{\delta n}'$, satisfies*

$$\begin{aligned}
\widehat{\eta}_{\mathrm{DE},n} = \eta - \Bigg[& \left(\frac{\upsilon_1}{\delta n} \sum_{i=1}^{\delta n} \frac{[1 - I(c(\mathbf{X}_i') > \xi) L(\mathbf{X}_i')] - p}{f(\xi)} \right) + \left(\frac{\upsilon_1'}{(1-\delta)n} \sum_{i=1}^{(1-\delta)n} \frac{[1 - I(c(\mathbf{X}_i) > \xi)] - p}{f(\xi)} \right) \\
& - \left(\frac{\upsilon_2}{\delta n} \sum_{i=1}^{\delta n} [c(\mathbf{X}_i') L(\mathbf{X}_i') - \mu] \right) + \left(\frac{\upsilon_2'}{(1-\delta)n} \sum_{i=1}^{(1-\delta)n} [c(\mathbf{X}_i) - \mu] \right) \Bigg] + (\upsilon_1 \tilde{R}_{\delta n} + \upsilon_1' R_{(1-\delta)n}), \\
& \textit{with } \sqrt{n}(\upsilon_1 \tilde{R}_{\delta n} + \upsilon_1' R_{(1-\delta)n}) \Rightarrow 0 \quad \textit{as } n \to \infty, \qquad (26)
\end{aligned}$$

where $\tilde{R}_{\delta n}$ is from (16) *and $R_{(1-\delta)n}$ from* (5).

(ii) If also $\sigma_{\mathrm{IS}}^2 < \infty$ and $\sigma_{\mathrm{SRS}}^2 < \infty$, then $\sqrt{n}\,[\widehat{\eta}_{\mathrm{DE},n} - \eta] \Rightarrow N(0, \zeta_{\mathrm{DE}}^2)$ as $n \to \infty$, where

$$\zeta_{\mathrm{DE}}^2 = \left[\frac{\upsilon_1^2}{\delta} \frac{\chi_{\mathrm{IS}}^2}{f^2(\xi)} + \frac{\upsilon_1'^2}{1-\delta} \frac{\chi_{\mathrm{SRS}}^2}{f^2(\xi)} \right] + \left[\frac{\upsilon_2^2}{\delta} \sigma_{\mathrm{IS}}^2 + \frac{\upsilon_2'^2}{1-\delta} \sigma_{\mathrm{SRS}}^2 \right] - 2 \left[\frac{\upsilon_1 \upsilon_2}{\delta} \frac{\gamma_{\mathrm{IS}}}{f(\xi)} + \frac{\upsilon_1' \upsilon_2'}{1-\delta} \frac{\gamma_{\mathrm{SRS}}}{f(\xi)} \right] \qquad (27)$$

with χ_{IS}^2, σ_{IS}^2, and γ_{IS} from (19)–(20), *and χ_{SRS}^2, σ_{SRS}^2 and γ_{SRS} from* (9)–(10). *For a fixed $\delta \in (0,1)$, the optimal choice of (υ_1, υ_2) to minimize ζ_{DE}^2 is*

$$\begin{aligned}
(\upsilon_1^*, \upsilon_2^*) &= \left(\frac{a_1}{a_0}, \frac{a_2}{a_0} \right), \quad \textit{where} \qquad (28) \\
a_0 &= \mathrm{V}_{\mathrm{SRS}}^{(\xi)} \mathrm{V}_{\mathrm{IS}}^{(\mu)} - \mathrm{C}_{\mathrm{IS}}^2 - 2\mathrm{C}_{\mathrm{IS}} \mathrm{C}_{\mathrm{SRS}} - \mathrm{C}_{\mathrm{SRS}}^2 + \mathrm{V}_{\mathrm{IS}}^{(\xi)} \mathrm{V}_{\mathrm{IS}}^{(\mu)} + \mathrm{V}_{\mathrm{IS}}^{(\xi)} \mathrm{V}_{\mathrm{SRS}}^{(\mu)} + \mathrm{V}_{\mathrm{SRS}}^{(\xi)} \mathrm{V}_{\mathrm{SRS}}^{(\mu)}, \\
a_1 &= \mathrm{V}_{\mathrm{SRS}}^{(\xi)} \mathrm{V}_{\mathrm{IS}}^{(\mu)} + \mathrm{V}_{\mathrm{SRS}}^{(\xi)} \mathrm{V}_{\mathrm{SRS}}^{(\mu)} - \mathrm{V}_{\mathrm{IS}}^{(\mu)} \mathrm{C}_{\mathrm{SRS}} + \mathrm{V}_{\mathrm{SRS}}^{(\mu)} \mathrm{C}_{\mathrm{IS}} - \mathrm{C}_{\mathrm{IS}} \mathrm{C}_{\mathrm{SRS}} - \mathrm{C}_{\mathrm{SRS}}^2, \quad \textit{and} \\
a_2 &= \mathrm{V}_{\mathrm{IS}}^{(\xi)} \mathrm{V}_{\mathrm{SRS}}^{(\mu)} + \mathrm{V}_{\mathrm{SRS}}^{(\xi)} \mathrm{V}_{\mathrm{SRS}}^{(\mu)} - \mathrm{V}_{\mathrm{IS}}^{(\xi)} \mathrm{C}_{\mathrm{SRS}} + \mathrm{V}_{\mathrm{SRS}}^{(\xi)} \mathrm{C}_{\mathrm{IS}} - \mathrm{C}_{\mathrm{IS}} \mathrm{C}_{\mathrm{SRS}} - \mathrm{C}_{\mathrm{SRS}}^2,
\end{aligned}$$

with $\mathrm{V}_{\mathrm{IS}}^{(\xi)} = \frac{\chi_{\mathrm{IS}}^2}{\delta f^2(\xi)}$, $\mathrm{V}_{\mathrm{SRS}}^{(\xi)} = \frac{\chi_{\mathrm{SRS}}^2}{(1-\delta) f^2(\xi)}$, $\mathrm{V}_{\mathrm{IS}}^{(\mu)} = \frac{\sigma_{\mathrm{IS}}^2}{\delta}$, $\mathrm{V}_{\mathrm{SRS}}^{(\mu)} = \frac{\sigma_{\mathrm{SRS}}^2}{1-\delta}$, $\mathrm{C}_{\mathrm{IS}} = \frac{\gamma_{\mathrm{IS}}}{\delta f(\xi)}$, *and* $\mathrm{C}_{\mathrm{SRS}} = \frac{\gamma_{\mathrm{SRS}}}{(1-\delta) f(\xi)}$.

6. Asymptotic Analysis of i.i.d. Sum

We now provide a theoretical comparison of the EC estimators from Sections 3–5, showing MSIS (Section 5.1) is a compellingly simple and effective approach. Our study considers the loss Y as a sum of m i.i.d. random variables (i.e., a random walk, often a building block in more complex models) in an asymptotic regime of Glynn (1996), where $m \to \infty$ with the quantile level p simultaneously approaching 1 exponentially fast in m, i.e.,

$$p \equiv p_m = 1 - e^{-\beta m}, \text{ for some constant } \beta > 0. \tag{29}$$

In addition to its theoretical convenience, the framework also has practical relevance: bank portfolios are commonly exposed to thousands of obligors (i.e., large m), and as noted in Section 1, Deutsche Bank (2018, p. 46), e.g., has used $p = 0.999$ and $p = 0.9998$ in its EC computations. Although the analysis in this section is for an i.i.d. sum model, the dependent sum in the more complicated PCRM (Section 7.2) can be reduced to an independent (but not necessarily identically distributed) sum via conditioning arguments (Glasserman and Li (2005), Bassamboo et al. (2008)). Thus, the i.i.d. sum asymptotics provide insights about how exponential twisting may behave for factor models with dependence.

6.1. Model and Assumptions

Throughout the rest of Section 6, the loss Y in (1) has $d = m \geq 1$ with

$$Y = c(\mathbf{X}) = \sum_{j=1}^{m} X_j, \qquad \text{where } \mathbf{X} = (X_1, X_2, \ldots, X_m) \sim G \text{ has } m \text{ i.i.d. components.} \tag{30}$$

Each X_j has marginal CDF G_0, where G_0 does not depend on m and has mean $\mu_0 \equiv E_0[X_j] = \int x \, \mathrm{d}G_0(x)$, and variance $\sigma_0^2 \equiv \mathrm{Var}_0[X_j] = E_0[(X_j - \mu_0)^2]$, for E_0 and Var_0 as the expectation and variance operators, respectively, when $X_j \sim G_0$.

Assumption 1. *Each i.i.d. summand $X_j \sim G_0$ has $\sigma_0^2 > 0$, and its moment generating function (MGF) $M_0(\theta) = E_0[e^{\theta X_j}] = \int e^{\theta x} \, \mathrm{d}G_0(x)$, $\theta \in \Re$, has domain $\Delta = \{\theta \in \Re : M_0(\theta) < \infty\}$ with interior Δ° containing* 0.

Since $M_0(0) = 1$, the domain Δ of M_0 always contains 0, but its interior Δ° may not, as for heavy-tailed distributions, such as the lognormal or Pareto (Asmussen and Glynn 2007, Section VI.3), both having $\Delta = (-\infty, 0]$. Thus, Assumption 1's stipulation that $0 \in \Delta^\circ$, which often appears in the large-deviations literature (e.g., Section 2.2.1 of Dembo and Zeitouni (1998)), restricts us to light-tailed summands (Asmussen and Glynn 2007, Section VI.2), e.g., normal or gamma. In this case (Billingsley 1995, p. 278), all moments of $X_j \sim G_0$ are finite, and $M_0(\theta)$ has derivatives of all orders for $\theta \in \Delta^\circ$; let $M_0'(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta} M_0(\theta)$ and $M_0''(\theta) = \frac{\mathrm{d}^2}{\mathrm{d}\theta^2} M_0(\theta)$.

Define $Q_0(\theta) = \ln M_0(\theta)$ as the *cumulant generating function* (CGF) of $X_j \sim G_0$, with $Q_0'(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta} Q_0(\theta)$ and $Q_0''(\theta) = \frac{\mathrm{d}^2}{\mathrm{d}\theta^2} Q_0(\theta)$ as its first two derivatives. For $F = F_m$ as the CDF of the i.i.d. sum Y, let $f = f_m$ be its derivative (when it exists).

Assumption 2. *The characteristic function $C_0(\theta) \equiv M_0(\theta\sqrt{-1})$, $\theta \in \Re$, of $X_j \sim G_0$ satisfies*

$$\int_{\Re} |C_0(\theta)|^{q_0}\,\mathrm{d}\theta < \infty \ \textit{ for some } \ q_0 \geq 1. \tag{31}$$

For $\beta > 0$ in (29),

$$\textit{there exists } \theta = \theta_\star \in \Delta^\circ \textit{ with } \theta_\star > 0 \textit{ and } -\theta_\star Q_0'(\theta_\star) + Q_0(\theta_\star) = -\beta. \tag{32}$$

Also, f_m exists at the p_m-quantile $\xi = \xi_m = F_m^{-1}(p_m)$, with $f_m(\xi_m) > 0$.

In contrast to Assumption 1's requirement that $0 \in \Delta^\circ$, which constrains G_0 to be light-tailed, Assumption 2's condition (31) on the characteristic function $C_0(\theta)$ of each i.i.d. summand relates instead to the smoothness of G_0. As $|C_0(\theta)| \leq 1$ for all $\theta \in \Re$ (Durrett 1996, p. 92), the class of distributions G_0 with $\int_{\Re} |C_0(\theta)|^{q_0}\,\mathrm{d}\theta < \infty$ shrinks as q_0 decreases. For example, if G_0 is a normal distribution or a gamma with shape parameter $\alpha_0 > 1$, its (Lebesgue) density is continuous on all of $\Re$, and (31) holds for all $q_0 \geq 1$. But for a gamma distribution with $\alpha_0 \in (0, 1]$ and unit scale parameter, whose density g_{α_0} has a discontinuity at the origin (since $g_{\alpha_0}(x) = 0$ for all $x < 0$, so at $x = 0$, g_{α_0} has a jump, infinitely large when $\alpha_0 < 1$), (31) requires $q_0 > 1/\alpha_0 \geq 1$ as $C_0(\theta) = (1 - \theta\sqrt{-1})^{-\alpha_0}$. If G_0 has a (Lebesgue) density g_0 with $\int [g_0(x)]^{r_0}\,\mathrm{d}x < \infty$ for some $r_0 \in (1, 2]$, then (31) holds with $q_0 = r_0/(r_0 - 1) \geq 2$ by the Hausdorff-Young inequality (Theorem 1.2.3 of Jensen (1995) or Theorem IX.8 of Reed and Simon (1975)). While the marginal CDF G_0 is guaranteed to have a (Lebesgue) density when (31) is true for some $q_0 \in [1, 2]$, G_0 can be singular (i.e., G_0 is continuous but has no Lebesgue density) when (31) holds for only some $q_0 > 2$ but not for any $q_0 \in [1, 2]$ (Kawata 1972, p. 438).

But for the CDF F_m of the i.i.d. sum $Y = \sum_{j=1}^m X_j$, (31) secures the existence of its (Lebesgue) density f_m for all $m \geq q_0$; see (91) in Appendix F.3. The asymptotic variance of each EC estimator has at least one term that includes $f_m(\xi_m)$ in the denominator, as seen in (9), (19), (23), and (27). Thus, Assumption 2 further imposes $f_m(\xi_m) > 0$, as required by the CLTs in Theorems 1–4, and this holds if, e.g., G_0 has a density and the support of G_0 is a single (possibly infinitely long) interval, as for a normal or gamma. Our analysis handles f_m through a *saddlepoint approximation* (Jensen 1995, Chapter 2), given in (105) of Appendix F.3, which will be justified under Assumptions 1 and 2; see Lemma 2 in Appendix F.3.

Glynn (1996) shows that if there is a $\theta_\star > 0$ solving (32), it is unique. When the CGF Q_0 is "steep" (i.e., $\lim_{\theta \to \partial\Delta : \theta \in \Delta^\circ} |Q_0'(\theta)| = \infty$, where $\partial\Delta$ is the boundary of Δ (Dembo and Zeitouni 1998, p. 44)), a unique $\theta_\star > 0$ satisfies (32). For example, Q_0 is steep for the normal or gamma, but when Q_0 is not steep, then (32) can have no root for some values of $\beta > 0$. For example, if G_0 is the "perverted exponential" (Durrett 1996, p. 74), which has density $g_0(x) = c_0 x^{-3} e^{-x} I(x \geq 1)$ and $\Delta = \{\theta \leq 1\}$, where $c_0 \equiv 1/\int_1^\infty x^{-3} e^{-x}\,\mathrm{d}x \doteq 9.116$, then it can be shown that there exists $\theta_\star > 0$ with $\theta_\star \in \Delta^\circ$ solving (32) if and only if $0 < \beta < 2 - \ln(c_0/2) \doteq 0.483$. The last paragraph of Section 6.2 below will provide further discussions on (32).

We sometimes (but not always) emphasize the dimension m of $\mathbf{X}$ in (30) by writing, e.g., $\mathbf{X} = \mathbf{X}_m$, $\eta = \eta_m$ for the EC, $F = F_m$ as the CDF of Y, $\xi = \xi_m = F_m^{-1}(p_m)$ as the p_m-quantile of Y, and $\mu = \mu_m = E_G[Y] = E_G[c(\mathbf{X})]$. As in Section 3, SRS samples i.i.d. copies of $c(\mathbf{X})$ with $\mathbf{X} \sim G$. Some of our asymptotic analysis will account for the computational effort to construct an EC estimator, and we assume that the computation (CPU) time to generate a single $c(\mathbf{X})$ with $\mathbf{X} \sim G$ is a random variable (a constant being a special case) with expectation $m\tau_{\mathrm{SRS}}$ for some constant $\tau_{\mathrm{SRS}} \in (0, \infty)$.

6.2. Importance Sampling via Exponential Twisting

A common IS approach applies *exponential twisting*, also called *exponential tilting* or an *exponential change of measure* (Asmussen and Glynn 2007, Section VI.2). The exponential twist $\widetilde{G}_{0,\theta}$ of the marginal CDF G_0 of X_j is given by

$$\mathrm{d}\widetilde{G}_{0,\theta}(x) = \frac{e^{\theta x}\,\mathrm{d}G_0(x)}{M_0(\theta)} = e^{\theta x - Q_0(\theta)}\,\mathrm{d}G_0(x), \quad x \in \Re, \quad \theta \in \Delta^\circ, \tag{33}$$

and setting $\theta = 0$ reduces $\widetilde{G}_{0,\theta}$ to G_0. We next describe methods IS(θ), MSIS(θ), ISDM(θ), and DE(θ), which correspond to Sections 4, 5.1, 5.2, and 5.3, respectively, and utilize twisting with parameter θ in various ways, where we often (but not always) will choose $\theta = \theta_\star$ from (32).

- IS(θ) applies exponential twisting with parameter $\theta \in \Delta^\circ$. Specifically, under IS(θ), random vector $\mathbf{X}$ has joint CDF $\widetilde{G}_\theta$ such that the components $X_1, \dots, X_m$ of $\mathbf{X}$ are i.i.d., where the marginal CDF of each X_j is the exponential twist $\widetilde{G}_{0,\theta}$ of G_0 in (33). For the resulting LR

$$L_\theta(\mathbf{x}) \equiv \frac{\mathrm{d}G(\mathbf{x})}{\mathrm{d}\widetilde{G}_\theta(\mathbf{x})} = \prod_{j=1}^{m} \frac{\mathrm{d}G_0(x_j)}{\mathrm{d}\widetilde{G}_{0,\theta}(x_j)} = \exp\Big(mQ_0(\theta) - \theta \sum_{j=1}^{m} x_j\Big) = [M_0(\theta)]^m e^{-\theta c(\mathbf{x})} \tag{34}$$

in (11), we assume that the expected time to generate $(c(\mathbf{X}), L_\theta(\mathbf{X}))$, $\mathbf{X} \sim \widetilde{G}_\theta$, is $m\tau_{\mathrm{IS}(\theta)}$ for a constant $\tau_{\mathrm{IS}(\theta)} \in (0, \infty)$.

- ISDM(θ) $\equiv$ ISDM(θ, δ) corresponds to IS that samples

$$\mathbf{X} \sim \widetilde{G}_{\mathrm{ISDM}(\theta)} \equiv \delta \widetilde{G}_\theta + (1-\delta) G \tag{35}$$

for $\theta \in \Delta^\circ$ and fixed $\delta \in (0,1)$. For LR $L_{\mathrm{ISDM}(\theta)}(\mathbf{x}) \equiv \mathrm{d}G(\mathbf{x}) / \mathrm{d}\widetilde{G}_{\mathrm{ISDM}(\theta)}(\mathbf{x})$ in (24), we assume that the expected time to generate $(c(\mathbf{X}), L_{\mathrm{ISDM}(\theta)}(\mathbf{X}))$, $\mathbf{X} \sim \widetilde{G}_{\mathrm{ISDM}(\theta)}$, is $m\tau_{\mathrm{ISDM}(\theta)}$ for a constant $\tau_{\mathrm{ISDM}(\theta)} \in (0, \infty)$.

- MSIS(θ) $\equiv$ MSIS(θ, δ) and DE(θ) $\equiv$ DE($\theta, \delta, \upsilon_1, \upsilon_2$) use IS($\theta$) as their IS, with fixed $\delta \in (0,1)$ and fixed $\upsilon_1, \upsilon_2 \in [0,1]$.

In our asymptotic regime with $m \to \infty$ and p_m as in (29), Glynn (1996) employs IS($\theta_\star$) with $\theta_\star$ from (32) to estimate the p_m-quantile ξ_m, motivated by the following heuristic argument. Large-deviations analysis (Durrett 1996, Section 1.9) suggests that when $\mathbf{X} \sim G$ and x is large (i.e., $x \geq m(\mu_0 + \epsilon)$ for any constant $\epsilon > 0$), the tail probability $P(c(\mathbf{X}) > x) \approx \exp[-m(\theta_x Q_0'(\theta_x) - Q_0(\theta_x))]$ for θ_x as the root of the equation $mQ_0'(\theta_x) = x$. When $\mathbf{X} \sim \widetilde{G}_\theta$ with parameter $\theta \in \Delta^\circ$, each X_j has mean $Q_0'(\theta)$ (see p. 72 of Durrett (1996) or (96) of Appendix F.3), so the sum $c(\mathbf{X})$ has mean $mQ_0'(\theta)$. The p_m-quantile ξ_m of $c(\mathbf{X})$ satisfies $P(c(\mathbf{X}) > \xi_m) = 1 - p_m = e^{-\beta m}$ by (29), so equating the two tail probabilities results in the twisting parameter $\theta_\star$ in (32) shifting the mean of $c(\mathbf{X})$ under $\widetilde{G}_{\theta_\star}$ to about ξ_m. Hence, IS($\theta_\star$) often samples $\mathbf{X}$ so that $c(\mathbf{X})$ is around ξ_m, leading to more efficient estimation of ξ_m.

6.3. Relative Error and Work-Normalized Relative Error

We will compare our EC estimators in terms of their relative errors (e.g., L'Ecuyer et al. (2010)). We explain this idea in a general context of a Monte Carlo method $\mathfrak{M}$ (e.g., SRS, IS(θ), MSIS(θ), ISDM(θ), or DE(θ)) for an estimand $\varphi \equiv \varphi_m$ (e.g., η, ξ, or μ) of a sequence of stochastic models indexed by a parameter m (e.g., dimension of $\mathbf{X}$). Let $\widehat{\varphi}_{\mathfrak{M},n} \equiv \widehat{\varphi}_{\mathfrak{M},n,m}$ be the $\mathfrak{M}$ estimator of φ based on a total sample size n. For each fixed m, assume the estimator obeys a CLT $\sqrt{n}[\widehat{\varphi}_{\mathfrak{M},n} - \varphi] \Rightarrow N(0, \varsigma^2_{\mathfrak{M}})$ as $n \to \infty$, where $\varsigma^2_{\mathfrak{M}} \equiv \varsigma^2_{\mathfrak{M},m} < \infty$ is the asymptotic variance. When $\varphi \neq 0$, the *relative error* (RE) of the $\mathfrak{M}$ estimator of φ is

$$\mathrm{RE}_{\mathfrak{M},m}[\varphi] = \frac{\varsigma_{\mathfrak{M}}}{|\varphi|} \equiv \frac{\varsigma_{\mathfrak{M},m}}{|\varphi_m|}, \tag{36}$$

which we will study as $m \to \infty$ and fixed (large) n. (Our definition of RE ignores that $\widehat{\varphi}_{\mathfrak{M},n}$ may be biased, as is often the case when $\varphi = \xi$ or $\varphi = \eta$. But when applying SRS with fixed dimension m, the simplification is reasonable because as $n \to \infty$, the SRS quantile estimator's mean-squared error is determined primarily by its asymptotic variance, with negligible contribution from the bias (Avramidis and Wilson 1998, Theorem 2).)

To motivate the study of RE, consider a 95% confidence interval $(\widehat{\varphi}_{\mathfrak{M},n} \pm 1.96\varsigma_{\mathfrak{M}}/\sqrt{n})$ for φ based on the CLT for $\widehat{\varphi}_{\mathfrak{M},n}$. (In practice, $\varsigma_{\mathfrak{M}}$ is typically unknown, and the CI replaces it with a consistent estimator.) Suppose that we want to determine a sample size n so that the CI is roughly $(\widehat{\varphi}_{\mathfrak{M},n} \pm \varepsilon|\widehat{\varphi}_{\mathfrak{M},n}|)$ for a specified desired relative precision $\varepsilon > 0$; e.g., if $\varepsilon = 0.1$, then the desired CI has 10% relative half-width. Thus, we seek n so that $1.96\varsigma_{\mathfrak{M}}/\sqrt{n} \approx \varepsilon|\varphi|$, or equivalently, $n \approx (1.96\,\mathrm{RE}_{\mathfrak{M},m}[\varphi]/\varepsilon)^2$. If $\mathrm{RE}_{\mathfrak{M},m}[\varphi]$ is bounded (resp., grows to ∞) as $m \to \infty$, then the sample size n needed to achieve a fixed relative precision ε remains bounded (resp., blows up) as m increases. L'Ecuyer et al. (2010) and Asmussen and Glynn (2007, Chapter VI) review a variety of simulation methods $\mathfrak{M}$ that achieve the desirable property of bounded or even vanishing RE when estimating some parameter φ for various stochastic models and asymptotic regimes.

As m grows, the computation (CPU) time to generate one output for method $\mathfrak{M}$ often increases with m. For example, the end of Section 6.1 specifies $m\tau_{\mathrm{SRS}}$ as the expected CPU time to generate an SRS output $c(\mathbf{X})$ for $\mathbf{X} \in \Re^m$ with $\mathbf{X} \sim G$, and Section 6.2 imposes similar structure for IS(θ) and ISDM(θ). For a method $\mathfrak{M}$ that estimates φ through a *single* i.i.d. sample of size n (as for SRS, IS(θ), and ISDM(θ)), let $m\tau_{\mathfrak{M}}$ be the expected CPU time to generate one output, with $\tau_{\mathfrak{M}} \in (0, \infty)$ a constant. To account for the CPU time for such a method $\mathfrak{M}$, we define the *work-normalized RE* (WNRE) of the $\mathfrak{M}$ estimator of $\varphi \neq 0$ as

$$\mathrm{WNRE}_{\mathfrak{M},m}[\varphi] = \frac{\sqrt{m\tau_{\mathfrak{M}}}\,\varsigma_{\mathfrak{M}}}{|\varphi|} \equiv \frac{\sqrt{m\tau_{\mathfrak{M}}}\,\varsigma_{\mathfrak{M},m}}{|\varphi_m|} = \sqrt{m\tau_{\mathfrak{M}}}\,\mathrm{RE}_{\mathfrak{M},m}[\varphi]; \tag{37}$$

see L'Ecuyer et al. (2010). To motivate the WNRE, suppose that we have a (large) CPU budget $b_0 > 0$. Within our budget b_0, method $\mathfrak{M}$ obtains a sample of size approximately $n_{\mathfrak{M},b_0} \equiv \lfloor b_0/(m\tau_{\mathfrak{M}})\rfloor$. When $n_{\mathfrak{M},b_0} \geq 1$, the resulting $\mathfrak{M}$ estimator of φ based on budget b_0 is then roughly $\widehat{\varphi}_{\mathfrak{M},n_{\mathfrak{M},b_0}}$, whose variance is approximately $\varsigma^2_{\mathfrak{M}}/n_{\mathfrak{M},b_0} \approx m\tau_{\mathfrak{M}}\varsigma^2_{\mathfrak{M}}/b_0$, which we can express through the CLT $\sqrt{b_0}[\widehat{\varphi}_{\mathfrak{M},n_{\mathfrak{M},b_0}} - \varphi] \Rightarrow N(0, m\tau_{\mathfrak{M}}\varsigma^2_{\mathfrak{M}})$ as $b_0 \to \infty$. (We can formalize this argument through a random-time-change CLT, e.g., see (Chung 2001, Theorem 7.3.2).) Thus, the budget-constrained estimator's standard deviation is roughly scaled by the square root of the expected time to generate one output, which leads to the definition of the WNRE in (37).

While (37) is appropriate when $\mathfrak{M}$ utilizes only a *single* i.i.d. sample, MSIS(θ) and DE(θ) instead collect *multiple* samples, and we will define their WNRE by slightly adjusting how these estimators are constructed. Consider estimating $\varphi = \mu$ or $\varphi = \xi$ via DE(θ), which takes two independent samples: one with IS(θ) and the other with SRS. Rather than dividing the total *sample size* n between IS(θ) and SRS using allocation parameter $\delta \in (0,1)$, as in the DE(θ) estimator in (25), we instead split the *CPU budget* b_0 when considering WNRE, where δb_0 (resp., $(1-\delta)b_0$) of the budget is for IS(θ) (resp., SRS). Then the IS(θ) and SRS samples have approximately sizes $n'_{b_0,1} \equiv \lfloor \delta b_0/(m\tau_{\mathrm{IS}(\theta)})\rfloor$ and $n'_{b_0,2} \equiv \lfloor (1-\delta)b_0/(m\tau_{\mathrm{SRS}})\rfloor$, respectively, so the variances of the budget-constrained IS(θ) and SRS estimators of φ are roughly $\varsigma^2_{\mathrm{IS}(\theta)}/n'_{b_0,1}$ and $\varsigma^2_{\mathrm{SRS}}/n'_{b_0,2}$. We form the budget-constrained DE(θ) estimator of φ as a weighted average of the budget-constrained IS(θ) and SRS estimators of φ using respective weights υ and $\upsilon' = 1-\upsilon$, where $\upsilon = \upsilon_1$ when $\varphi = \xi$, and $\upsilon = \upsilon_2$ when $\varphi = \mu$, with υ_1 and υ_2 as in (25). As DE(θ) applies IS(θ) and SRS independently, the variance of the budget-constrained DE(θ) estimator of φ is roughly $\frac{\upsilon^2\varsigma^2_{\mathrm{IS}(\theta)}}{n'_{b_0,1}} + \frac{\upsilon'^2\varsigma^2_{\mathrm{SRS}}}{n'_{b_0,2}} \approx \frac{m}{b_0}\left[\frac{\tau_{\mathrm{IS}(\theta)}\upsilon^2\varsigma^2_{\mathrm{IS}(\theta)}}{\delta} + \frac{\tau_{\mathrm{SRS}}\upsilon'^2\varsigma^2_{\mathrm{SRS}}}{1-\delta}\right]$. This motivates defining the WNRE for the DE(θ) estimator of $\varphi \neq 0$ for $\varphi = \xi$ or μ as

$$\mathrm{WNRE}_{\mathrm{DE}(\theta),m}[\varphi] = \frac{1}{|\varphi|}\left[m\left(\frac{\tau_{\mathrm{IS}(\theta)}\upsilon^2\varsigma^2_{\mathrm{IS}(\theta)}}{\delta} + \frac{\tau_{\mathrm{SRS}}\upsilon'^2\varsigma^2_{\mathrm{SRS}}}{1-\delta}\right)\right]^{1/2}. \tag{38}$$

Appendix F.6 will similarly define $\mathrm{WNRE}_{\mathrm{DE}(\theta_\star),m}[\eta]$ and $\mathrm{WNRE}_{\mathrm{MSIS}(\theta_\star),m}[\eta]$ in (134) and (135), respectively.

6.4. Asymptotic Notation and Properties

For our i.i.d. sum model (30), Theorem 5 in Section 6.5 below will derive asymptotic expressions for the RE and WNRE of estimators of the EC $\eta_m = \xi_m - \mu_m$ as $m \to \infty$, building on analogous results (Appendix F) for the mean μ_m and p_m-quantile ξ_m. The limiting results will adopt the following asymptotic notation. For functions $r_1(m)$ and $r_2(m)$, we write $r_1(m) = O(r_2(m))$ (resp., $r_1(m) = \Omega(r_2(m))$) as $m \to \infty$ if there are constants d_0 and $m_0 > 0$ such that $|r_1(m)| \le d_0 |r_2(m)|$ (resp., $|r_1(m)| \ge d_0 |r_2(m)|$) for all $m \ge m_0$, so $d_0 |r_2(m)|$ provides an asymptotic *upper* (resp., *lower*) bound for $|r_1(m)|$. Also, $r_1(m) = \Theta(r_2(m))$ if both $r_1(m) = O(r_2(m))$ and $r_1(m) = \Omega(r_2(m))$. Moreover, $r_1(m) = o(r_2(m))$ means $r_1(m)/r_2(m) \to 0$ as $m \to \infty$, and $r_1(m) = \omega(r_2(m))$ denotes that $r_2(m) = o(r_1(m))$.

We next review some (mainly asymptotic) properties that will arise in our analysis. Let $b_1, b_2, c_1, c_2 \in \Re$ be constants, and let $m \ge 1$. For $x, y \in \Re$, define $x \vee y = \max(x, y)$ and $x \wedge y = \min(x, y)$. If $b_1, b_2 > 1$, then $b_1^{-m} \vee b_2^{-m} = [b_1 \wedge b_2]^{-m}$. If $c_1 > 0$, then for any c_2 and for each $c_3 \in (0, c_1)$, we have that $e^{c_1 m + c_2 \sqrt{m}} = \omega(e^{c_3 m})$ as $m \to \infty$ because $e^{c_1 m + c_2 \sqrt{m}} / e^{c_3 m} = e^{c_1 m} e^{c_2 \sqrt{m}} / [e^{c_1 m} e^{(c_3 - c_1) m}] = e^{(c_1 - c_3) m + c_2 \sqrt{m}} \to \infty$ as $m \to \infty$ since $c_1 - c_3 > 0$; thus, $e^{c_1 m + c_2 \sqrt{m}}$ grows exponentially quickly in m. Similarly, $c_1 > 0$ implies for any c_2 and each $c_3 \in (0, c_1)$ that $e^{-c_1 m + c_2 \sqrt{m}} = o(e^{-c_3 m})$ as $m \to \infty$, so $e^{-c_1 m + c_2 \sqrt{m}}$ shrinks exponentially fast in m. If a function $r_1(m)$ satisfies $r_1(m) = e^{o(\sqrt{m})}$ as $m \to \infty$, then there exists some function $r_2(m)$ such that $r_1(m) = e^{r_2(m)}$ with $r_2(m)/\sqrt{m} \to 0$ as $m \to \infty$; for each $t \in \Re$, this then implies, as $m \to \infty$, that $r_1^t(m) = e^{o(\sqrt{m})}$, $r_1^t(m) = o(e^{c_1 \sqrt{m}})$, and $r_1^t(m) = \omega(e^{-c_1 \sqrt{m}})$ for each $c_1 > 0$.

6.5. Estimating EC

For method $\mathfrak{M}$ equaling SRS, IS($\theta_\star$), MSIS($\theta_\star$), or ISDM($\theta_\star$), the asymptotic variance $\zeta_{\mathfrak{M}}^2 \equiv \zeta_{\mathfrak{M},m}^2$ in (9), (19), and (23) of the resulting η_m estimator has the form

$$\zeta_{\mathfrak{M}}^2 = \Lambda_{\mathfrak{M}} \kappa_{\mathfrak{M}}^2 + \Lambda_{\mathfrak{M}}^\dagger \sigma_{\mathfrak{M}}^2 - 2 \frac{\gamma_{\mathfrak{M}}}{f_m(\xi_m)} \tag{39}$$

for constants $\Lambda_{\mathfrak{M}}$ and $\Lambda_{\mathfrak{M}}^\dagger$ depending on the method $\mathfrak{M}$ but not on m. In (39), $\kappa_{\mathfrak{M}}^2 \equiv \kappa_{\mathfrak{M},m}^2$ denotes the asymptotic variance of the method-$\mathfrak{M}$ estimator of the p_m-quantile ξ_m, where

$$\kappa_{\mathfrak{M}}^2 = \frac{\chi_{\mathfrak{M}}^2}{f_m^2(\xi_m)}, \tag{40}$$

as seen through Theorems 1, 2, and 3, and will be more fully developed for our i.i.d. sum model in Theorem 7 of Appendix F.2. Also, $\sigma_{\mathfrak{M}}^2 \equiv \sigma_{\mathfrak{M},m}^2$ in (39) represents the asymptotic variance of method $\mathfrak{M}$'s estimator of the mean μ_m, which Theorem 6 of Appendix F.1 will analyze, and $\gamma_{\mathfrak{M}}/f_m(\xi_m)$ with $\gamma_{\mathfrak{M}} \equiv \gamma_{\mathfrak{M},m}$ is the asymptotic covariance of the estimators of ξ_m and μ_m. For example, (39) and (40) for $\mathfrak{M} = \mathrm{IS}(\theta_\star)$ have $\Lambda_{\mathrm{IS}(\theta_\star)} = \Lambda^\dagger_{\mathrm{IS}(\theta_\star)} = 1$ and $\chi^2_{\mathrm{IS}(\theta_\star)}$, $\sigma^2_{\mathrm{IS}(\theta_\star)}$, and $\gamma_{\mathrm{IS}(\theta_\star)}$ as in (19) and (20) with LR in (34) and twisting parameter $\theta = \theta_\star$ from (32). For the other methods $\mathfrak{M}$, the specific forms of the terms in (39) and (40) will be given explicitly in the proof of Theorem 5 in Appendix F.6, but they can be inferred from (9) and (10) for $\mathfrak{M} = \mathrm{SRS}$, from (23) for $\mathfrak{M} = \mathrm{MSIS}(\theta_\star)$ (where $\gamma_{\mathrm{MSIS}(\theta_\star)} = 0$, as ξ and μ are estimated independently), and from (19) and (20) for $\mathfrak{M} = \mathrm{ISDM}(\theta_\star)$ (since $\mathrm{ISDM}(\theta_\star)$ is a special case of IS, sampling $\mathbf{X}$ as in (35) with $\theta = \theta_\star$). In the right side of (40), the numerator $\chi^2_{\mathfrak{M}} \equiv \chi^2_{\mathfrak{M},m}$ depends on $\mathfrak{M}$, but the denominator does not; both depend on m for our model (30), as do $\zeta^2_{\mathfrak{M}}$, $\kappa^2_{\mathfrak{M}}$, $\sigma^2_{\mathfrak{M}}$ and $\gamma_{\mathfrak{M}}$ in (39).

For $\mathfrak{M} = \mathrm{DE}(\theta_\star)$, we can also write its asymptotic variance from (27) to fit into (39), but to handle its WNRE, defined in (134) of Appendix F.6, it is more convenient to treat it differently as

$$\zeta^2_{\mathrm{DE}(\theta_\star)} = \frac{1}{\delta}\left(\upsilon_1^2 \kappa^2_{\mathrm{IS}(\theta_\star)} + \upsilon_2^2 \sigma^2_{\mathrm{IS}(\theta_\star)} - 2\upsilon_1\upsilon_2 \frac{\gamma_{\mathrm{IS}(\theta_\star)}}{f_m(\xi_m)}\right) + \frac{1}{1-\delta}\left(\upsilon_1'^2 \kappa^2_{\mathrm{SRS}} + \upsilon_2'^2 \sigma^2_{\mathrm{SRS}} - 2\upsilon_1'\upsilon_2' \frac{\gamma_{\mathrm{SRS}}}{f_m(\xi_m)}\right). \tag{41}$$

Studying the asymptotic properties of the RE and WNRE of the estimators of $\eta_m = \xi_m - \mu_m$ entails examining the two components of η_m and each term in (39), but several challenges arise with ξ_m and the denominator $f_m(\xi_m)$ in (39) and (40). As the CDF F_m of the sum $c(\mathbf{X})$ is a convolution, an explicit expression for F_m is generally analytically intractable for large m. This complicates deriving the exact values of ξ_m and $f_m(\xi_m)$, but we manage to analyze their asymptotic behaviors via the following ideas. Lemma 1 of Appendix F.3 will show that the p_m-quantile $\xi_m = F_m^{-1}(p_m)$ satisfies $\xi_m = mQ_0'(\theta_\star) + o(\sqrt{m})$ as $m \to \infty$, for Q_0' as the derivative of the CGF of G_0 and $\theta_\star$ in (32). Lemma 2 of Appendix F.3 handles f_m through a saddlepoint approximation (Jensen 1995, Chapter 2), which when approximating $f_m(x)$ can be viewed as first exponentially twisting (Section 6.2) the distribution of $c(\mathbf{X})$ so its mean is x, and then applying an Edgeworth expansion (Jensen 1995, Section 1.5).

Recall that $(\upsilon_1^*, \upsilon_2^*) = (\frac{a_1}{a_0}, \frac{a_2}{a_0})$ in (28) minimizes ζ^2_{DE} in (27) for fixed $\delta \in (0,1)$. Letting $(a_0, a_1, a_2) = (a_{0,m}, a_{1,m}, a_{2,m})$ now depend on m leads to $(\upsilon^*_{1,m}, \upsilon^*_{2,m}) = (\frac{a_{1,m}}{a_{0,m}}, \frac{a_{2,m}}{a_{0,m}})$. Since the value of η_m does not depend on υ_1 and υ_2, minimizing the asymptotic variance $\zeta^2_{\mathrm{DE}(\theta_\star)}$ or

$\mathrm{RE}_{\mathrm{DE}(\theta_\star),m}[\eta]$ results in the same optimal value $(\upsilon^*_{1,m}, \upsilon^*_{2,m})$. Define $\mathrm{DE}_*(\theta_\star)$ as the method $\mathrm{DE}(\theta_\star)$ with optimally varying weights $(\upsilon^*_{1,m}, \upsilon^*_{2,m})$, and let $\zeta^2_{\mathrm{DE}_*(\theta_\star)}$ be the asymptotic variance of the corresponding $\mathrm{DE}_*(\theta_\star)$ estimator $\widehat{\eta}_{\mathrm{DE}_*(\theta_\star),n}$ of η_m, as in (25). We also define the optimal value $(\upsilon^{**}_{1,m}, \upsilon^{**}_{2,m})$ of (υ_1, υ_2) that minimizes $\mathrm{WNRE}_{\mathrm{DE}(\theta_\star),m}[\eta]$ defined in (134) of Appendix F.6, and let $\mathrm{DE}_{**}(\theta_\star)$ be the method $\mathrm{DE}(\theta_\star)$ using weights $(\upsilon^{**}_{1,m}, \upsilon^{**}_{2,m})$, with $\zeta^2_{\mathrm{DE}_{**}(\theta_\star)}$ as the asymptotic variance of the $\mathrm{DE}_{**}(\theta_\star)$ estimator of η_m.

As Table 1 (Section 1) previously summarized, Theorem 5 below shows that when estimating the EC η_m, the methods SRS, $\mathrm{IS}(\theta_\star)$, and $\mathrm{DE}(\theta_\star)$ with fixed weights behave poorly as $m \to \infty$, with exponentially increasing RE and WNRE; see parts (i), (ii), and (v). In contrast, $\mathrm{ISDM}(\theta_\star)$ performs well, yielding bounded (resp., vanishing) RE when $\mu_0 \neq 0$ (resp., $\mu_0 = 0$) (part (iv)). $\mathrm{MSIS}(\theta_\star)$ and optimal $\mathrm{DE}_*(\theta_\star)$ can do even better, producing vanishing RE (parts (iii) and (vi)) for all μ_0, the latter when (56) holds.

THEOREM 5. *For the i.i.d. sum model* (30) *with* $m \geq 1$ *summands, suppose that Assumptions 1 and 2 hold. Also, assume that* $\theta_\star \in \Delta^\circ$ *in* (32) *further satisfies* $-\theta_\star \in \Delta^\circ$ *for methods* $\mathfrak{M} = \mathrm{IS}(\theta_\star)$, $\mathrm{DE}(\theta_\star)$, $\mathrm{DE}_*(\theta_\star)$, *and* $\mathrm{DE}_{**}(\theta_\star)$ *(but not necessarily for* SRS, $\mathrm{ISDM}(\theta_\star)$, *and* $\mathrm{MSIS}(\theta_\star)$*). Then for the EC* $\eta \equiv \eta_m = \xi_m - \mu_m$, *the method-*$\mathfrak{M}$ *estimators* $\widehat{\eta}_{\mathfrak{M},n}$ *with asymptotic variance* $\zeta^2_{\mathfrak{M}} \equiv \zeta^2_{\mathfrak{M},m}$ *satisfy the following as* $m \to \infty$, *with* $\beta > 0$ *from* (29), $\alpha_\star \equiv M_0(\theta_\star) M_0(-\theta_\star) > 1$ *for* M_0 *as the MGF of* G_0, *and* $\Upsilon_m = e^{o(\sqrt{m})}$ *(defined in* (110) *of Appendix F.3).*

(i) The $\mathfrak{M} = \mathrm{SRS}$ *estimator* $\widehat{\eta}_{\mathrm{SRS},n}$ *in* (4) *satisfies* $\lim_{m\to\infty} \frac{1}{m} \ln \zeta^2_{\mathrm{SRS}} = \beta$,

$$\zeta^2_{\mathrm{SRS}} = \frac{\Theta(m e^{\beta m})}{\Upsilon^2_m} = \omega(m e^{\beta m - \sqrt{m}}), \tag{42}$$

$$\mathrm{RE}_{\mathrm{SRS},m}[\eta] = \omega\left(\frac{e^{(\beta/2)m - \sqrt{m}}}{\sqrt{m}}\right) \to \infty, \ \textit{and} \ \mathrm{WNRE}_{\mathrm{SRS},m}[\eta] = \omega(e^{(\beta/2)m - \sqrt{m}}) \to \infty. \tag{43}$$

(ii) The $\mathfrak{M} = \mathrm{IS}(\theta_\star)$ *estimator* $\widehat{\eta}_{\mathrm{IS}(\theta_\star),n}$ *in* (17) *has*

$$\zeta^2_{\mathrm{IS}(\theta_\star)} = \Omega(m \alpha^m_\star), \tag{44}$$

$$\mathrm{RE}_{\mathrm{IS}(\theta_\star),m}[\eta] = \Omega(\alpha^{m/2}_\star / \sqrt{m}) \to \infty, \ \textit{and} \ \mathrm{WNRE}_{\mathrm{IS}(\theta_\star),m}[\eta] = \Omega(\alpha^{m/2}_\star) \to \infty. \tag{45}$$

(iii) The $\mathfrak{M} = \mathrm{MSIS}(\theta_\star)$ *estimator* $\widehat{\eta}_{\mathrm{MSIS}(\theta_\star),n}$ *in* (21) *has*

$$\zeta^2_{\mathrm{MSIS}(\theta_\star)} = \Theta(m), \tag{46}$$

$$\mathrm{RE}_{\mathrm{MSIS}(\theta_\star),m}[\eta] = \Theta(1/\sqrt{m}) \to 0, \ \textit{and} \ \mathrm{WNRE}_{\mathrm{MSIS}(\theta_\star),m}[\eta] = \Theta(1). \tag{47}$$

(iv) The $\mathfrak{M} = \mathrm{ISDM}(\theta_\star)$ *estimator* $\widehat{\eta}_{\mathrm{ISDM}(\theta_\star),n}$ *in* (17) *has*

$$\zeta^2_{\mathrm{ISDM}(\theta_\star)} = O(m^2), \tag{48}$$

$$\mathrm{RE}_{\mathrm{ISDM}(\theta_\star),m}[\eta] = O(1), \text{ and } \mathrm{WNRE}_{\mathrm{ISDM}(\theta_\star),m}[\eta] = O(\sqrt{m}). \tag{49}$$

If $\mu_0 \neq 0$*, then*

$$\zeta^2_{\mathrm{ISDM}(\theta_\star)} = \Theta(m^2), \tag{50}$$

$$\mathrm{RE}_{\mathrm{ISDM}(\theta_\star),m}[\eta] = \Theta(1), \text{ and } \mathrm{WNRE}_{\mathrm{ISDM}(\theta_\star),m}[\eta] = \Theta(\sqrt{m}). \tag{51}$$

If $\mu_0 = 0$*, then*

$$\zeta^2_{\mathrm{ISDM}(\theta_\star)} = O(m), \ \mathrm{RE}_{\mathrm{ISDM}(\theta_\star),m}[\eta] = O(1/\sqrt{m}) \to 0, \text{ and } \mathrm{WNRE}_{\mathrm{ISDM}(\theta_\star),m}[\eta] = O(1). \tag{52}$$

(v) The $\mathfrak{M} = \mathrm{DE}(\theta_\star)$ *estimator* $\widehat{\eta}_{\mathrm{DE}(\theta_\star),n}$ *in* (25) *with any fixed* $\delta, \upsilon_1, \upsilon_2 \in (0,1)$ *satisfies* $\liminf_{m\to\infty} \frac{1}{m} \ln \zeta^2_{\mathrm{SRS}} \geq s_0$*, where* $s_0 \equiv s_0(\theta_\star, \beta) = \alpha_\star \vee e^\beta > 1$*. Also, as* $m \to \infty$*,*

$$\zeta^2_{\mathrm{DE}(\theta_\star)} = \Omega(m s_0^m e^{-\sqrt{m}}), \tag{53}$$

$$\mathrm{RE}_{\mathrm{DE}(\theta_\star),m}[\eta] = \Omega\left(\frac{[s_0^{1/2}]^m e^{-\sqrt{m}}}{\sqrt{m}}\right) \to \infty, \text{ and } \mathrm{WNRE}_{\mathrm{DE}(\theta_\star),m}[\eta] = \Omega([s_0^{1/2}]^m e^{-\sqrt{m}}) \to \infty. \tag{54}$$

(vi) For the $\mathfrak{M} = \mathrm{DE}_*(\theta_\star)$ *estimator* $\widehat{\eta}_{\mathrm{DE}_*(\theta_\star),n}$ *in* (25) *with optimal weights* $(\upsilon_1^*, \upsilon_2^*) = (\upsilon_{1,m}^*, \upsilon_{2,m}^*)$ *in* (28) *that vary with* m *but with the sampling-allocation parameter* $\delta \in (0,1)$ *still fixed,*

$$(\upsilon_{1,m}^*, \upsilon_{2,m}^*) \to (1,0) \quad \text{exponentially fast as } m \to \infty. \tag{55}$$

If in addition

$$e^\beta < \alpha_\star^4, \tag{56}$$

then for $\mathrm{DE}_*(\theta_\star)$ *and* $\mathrm{MSIS}(\theta_\star)$ *with the same fixed* $\delta \in (0,1)$*,*

$$\frac{\zeta^2_{\mathrm{DE}_*(\theta_\star)}}{\zeta^2_{\mathrm{MSIS}(\theta_\star)}} \to 1 \quad \text{exponentially fast as } \ m \to \infty, \tag{57}$$

so $\mathrm{RE}_{\mathrm{DE}_*(\theta_\star),m}[\eta] = \Theta(1/\sqrt{m})$ *as* $m \to \infty$ *by* (47)*. Moreover, if we replace* $(\upsilon_{1,m}^*, \upsilon_{2,m}^*)$ *and* $\zeta^2_{\mathrm{DE}_*(\theta_\star)}$ *with* $(\upsilon_{1,m}^{**}, \upsilon_{2,m}^{**})$ *and* $\zeta^2_{\mathrm{DE}_{**}(\theta_\star)}$*, respectively, for minimizing* $\mathrm{WNRE}_{\mathrm{DE}(\theta_\star),m}[\eta]$*, then* (55) *and* (57) *still hold, the latter under* (56)*, so* $\mathrm{WNRE}_{\mathrm{DE}_{**}(\theta_\star),m}[\eta] = \Theta(1)$ *as* $m \to \infty$ *by* (47)*.*

The assumption (56) can restrict the choices of $\beta > 0$ in (29) for certain G_0. While always true when G_0 is normal, (56) holds for G_0 as Erlang if and only if $\beta < \bar{\beta}$ for some $\bar{\beta} > 0$. For example, an Erlang with $s \geq 1$ stages, each with mean 1, has $\bar{\beta} \doteq 6.6029s$.

In the case that $\delta, \upsilon_1, \upsilon_2 \in (0,1)$ are fixed, we now sketch Theorem 5's proof, which is in Appendix F.6. For a method-$\mathfrak{M}$ estimator of $\eta_m = \xi_m - \mu_m$, the growth rate (as $m \to \infty$) of its asymptotic variance $\zeta^2_{\mathfrak{M}}$ in (39)–(41) is governed by the largest growth rate of the variances of its constituent estimators of μ_m and ξ_m, analyzed in Theorem 6 (Appendix F.1) and Theorem 7 (Appendix F.2). (Covariance terms in $\zeta^2_{\mathfrak{M}}$ are nondominant, by the Cauchy-Schwarz inequality.) Also, η_m grows linearly in m (see (93) in Appendix F.3).

Applying these insights to SRS shows that its exponential growth in (42) and (43) is due to that same behavior of the SRS quantile estimator, by Theorem 7(i). For IS($\theta_\star$), the exponential behavior in (44) and (45) arises from that of the mean estimator by Theorem 6(ii). As DE($\theta_\star$) uses both SRS and IS($\theta_\star$) to estimate both ξ_m and μ_m, its η_m estimator's behavior with any fixed weights $\upsilon_1, \upsilon_2 \in (0,1)$ is determined by the *worst* of those estimators; the base s_0 of the dominant exponential term s_0^m in (53)–(54) is the *larger* of the SRS base $e^\beta > 1$ of $(e^\beta)^m$ in (42), which comes from Theorem 7(i), and the IS($\theta_\star$) base $\alpha_\star > 1$ in (44), resulting from Theorem 6(ii).

In contrast, the MSIS($\theta_\star$) and ISDM($\theta_\star$) estimators of $\eta_m = \xi_m - \mu_m$ behave polynomially in m because the same holds for its constituent estimators of ξ_m and μ_m, where Theorem 7(ii)–(iii) cover the ξ_m estimators, and parts (i) and (iii) of Theorem 6 analyze the μ_m estimators. Specifically, Theorem 5(iii)–(iv) establish that as $m \to \infty$, MSIS($\theta_\star$) has vanishing RE and bounded WNRE by (47), and ISDM($\theta_\star$) does the same when $\mu_m = 0$ by (52). But when $\mu_m \neq 0$, ISDM($\theta_\star$) has only bounded (but not vanishing) RE and unbounded WNRE (but growing only as $\Theta(\sqrt{m})$) by (51), strictly worse than MSIS($\theta_\star$).

When the optimal DE weights $(\upsilon_1, \upsilon_2) = (\upsilon^*_{1,m}, \upsilon^*_{2,m})$ to minimize ζ^2_{DE} and $\mathrm{RE}_{\mathrm{DE}(\theta_\star),m}[\eta]$ vary with m (but δ is still fixed), (55) establishes that $(\upsilon^*_{1,m}, \upsilon^*_{2,m})$ converges exponentially quickly to $(1,0)$ as $m \to \infty$. Appendix F.6.8 also shows the same when instead minimizing $\mathrm{WNRE}_{\mathrm{DE}(\theta_\star),m}[\eta]$. Moreover, (57) analyzes the DE variance with weights $(\upsilon^*_{1,m}, \upsilon^*_{2,m})$, showing that the ratio of the variances of optimal DE$_*(\theta_\star)$ and MSIS($\theta_\star$) converges to 1 exponentially quickly as $m \to \infty$. Thus, even though MSIS($\theta_\star$) can never beat optimal DE$_*(\theta_\star)$ as MSIS($\theta_\star$) is a special case of DE($\theta_\star$) that uses just a particular (typically suboptimal) choice of DE weights, MSIS($\theta_\star$) provides a compelling alternative because it is much simpler to implement but does virtually the same as optimal $\mathrm{DE}_*(\theta_\star)$ for large m by (57).

7. Numerical and Simulation Results

We next present results for two models of a loss $Y = c(\mathbf{X})$ as in (1), but with different definitions for c and $\mathbf{X} = (X_1, \ldots, X_d)$. Section 7.1 studies the random-walk model (30) of Section 6, so $c(\mathbf{X}) = \sum_{j=1}^{d} X_j$ has i.i.d. summands with $d = m$; we take the summand CDF G_0 as exponential, which permits numerical computation using exact analytics and quadrature (not simulation). Section 7.2 examines a more complicated portfolio credit risk model, which we instead simulate. We compare EC estimators for SRS (Section 3), IS (Section 4), MSIS (Section 5.1), ISDM (Section 5.2), and DE (Section 5.3). For MSIS, ISDM, and DE (with fixed weights), we let $\delta = \upsilon_1 = \upsilon_2 = 1/2$. Section 7.1 also considers DE with optimal weights $(\upsilon^*_{1,m}, \upsilon^*_{2,m})$ in (28) and (55) that vary with m but with $\delta = 1/2$ still fixed. For each model, we specify below the joint CDF $\widetilde{G}$ of $\mathbf{X}$ for IS, MSIS, and DE, and we use this same CDF as G^* in ISDM; see (24). Although the PCRM is much more complex with dependent obligors, the results will show that the methods behave similarly on the two models. Thus, our theory in Section 6.4 for the i.i.d. sum provide considerable insight into the methods. All of the codes that produced the numerical and simulation results are available at a GitHub repository (Li et al. (2023)).

7.1. Exact Relative Error for i.i.d. Sum

As in Section 6, we define here the loss as $Y = \sum_{j=1}^{m} X_j$, with the X_j as i.i.d. with marginal CDF G_0, and the quantile level $p \equiv p_m$ satisfies (29). Theorem 5 establishes that as $m \to \infty$, the relative errors of the estimators of the EC $\eta \equiv \eta_m$ using SRS, $\mathrm{IS}(\theta_\star)$, and $\mathrm{DE}(\theta_\star)$ (with fixed weights) grow exponentially, but $\mathrm{MSIS}(\theta_\star)$ and $\mathrm{DE}(\theta_\star)$ with optimally varying weights (resp., $\mathrm{ISDM}(\theta_\star)$) has RE that shrinks to 0 (resp., is $\Theta(1)$ for $\mu_0 \neq 0$). For each EC estimator, we want to investigate numerically (no simulation) the behavior (as m increases) of the *exact* RE in (36) (based on (9), (19), (23), or (27)) to show that our asymptotic theory in Theorems 5–7 (the latter two in Appendix F) accurately captures the behavior of the exact values as m grows. We present here results for G_0 as exponential with mean $\mu_0 = 1$. Appendix A provides other results when G_0 is $N(1, 1)$ and Erlang ($s = 8$ stages).

Figure 1 gives log-log plots of the exact RE of our estimators of η, ξ, and μ as functions of the dimension m. Lemma 1 (Appendix F.3) will establish that η_m, ξ_m, and μ_m share the same growth rate: linear in m. Thus, the RE and WNRE of η, ξ, and μ are directly comparable. As m grows, the top left panel shows that the estimators of η using SRS, IS$(\theta_\star)$, and DE$(\theta_\star)$ (with *fixed* weights (υ_1, υ_2) that do not vary with m) have exponentially increasing RE, in line with our asymptotic theory in (43), (45), and (54). For SRS (resp., IS$(\theta_\star)$) the RE of the η estimator grows exponentially because the same holds for ξ (resp., μ) by (78) (resp., (71)) of Appendix F; see bottom panels. Also, as explained after Theorem 5, when DE$(\theta_\star)$ uses fixed weights, the RE of the DE$(\theta_\star)$ estimator of η is governed by the worst of the SRS and IS$(\theta_\star)$ estimators of ξ and μ, which in this case is the SRS estimator of ξ, as seen in the bottom panels of Figure 1. (For other G_0 in Appendix A, the IS$(\theta_\star)$ estimator of μ is worst.)

In contrast, Figure 1 also shows that the MSIS$(\theta_\star)$ and ISDM$(\theta_\star)$ estimators of η have decreasing RE as m grows; see (47) and (51) for RE. As m gets large, MSIS$(\theta_\star)$ is a bit better than ISDM$(\theta_\star)$ when estimating η, with MSIS$(\theta_\star)$ continually decreasing, but ISDM$(\theta_\star)$ flattening out. The reason becomes apparent from the bottom right panel: the estimator of μ using SRS (which is how MSIS$(\theta_\star)$ estimates μ) has shrinking RE as m grows by (67) of Theorem 6 in Appendix F.1, while the ISDM$(\theta_\star)$ estimator of μ does not; (73) of Theorem 6 applies here because $\mu_0 \neq 0$ and $\theta_\star > 0$, as explained in the penultimate paragraph of Section 6.5.

Figure 1 also plots the RE of the DE$_*(\theta_\star)$ estimator of η using the optimal weights $(\upsilon^*_{1,m}, \upsilon^*_{2,m})$ in (28) that vary with m. Now the DE$_*(\theta_\star)$ estimator has vanishing RE. Zooming in on the top left panel reveals that the optimal DE$_*(\theta_\star)$ does slightly better than MSIS$(\theta_\star)$ when m is small, but the difference rapidly vanishes as m increases. This agrees with (57), which shows that the ratio of their variances converges to 1 exponentially quickly as $m \to \infty$. Additional results for G_0 as exponential (not presented) appear to indicate that (57) remains valid even when condition (56) does not hold, so it may be possible to weaken (56).

7.2. Portfolio Credit Risk Model

We next present Monte Carlo results from estimating EC for a large credit portfolio with dependent obligors. We consider a multi-factor portfolio-credit-risk model as in Glasserman and Li (2005), Bassamboo et al. (2008), and Lütkebohmert (2009), in which the loss Y has a form in (1), with mutually independent components in $\mathbf{X}$, defined as follows. The portfolio has $m \geq 1$ obligors, and dependence among the default events across obligors is induced

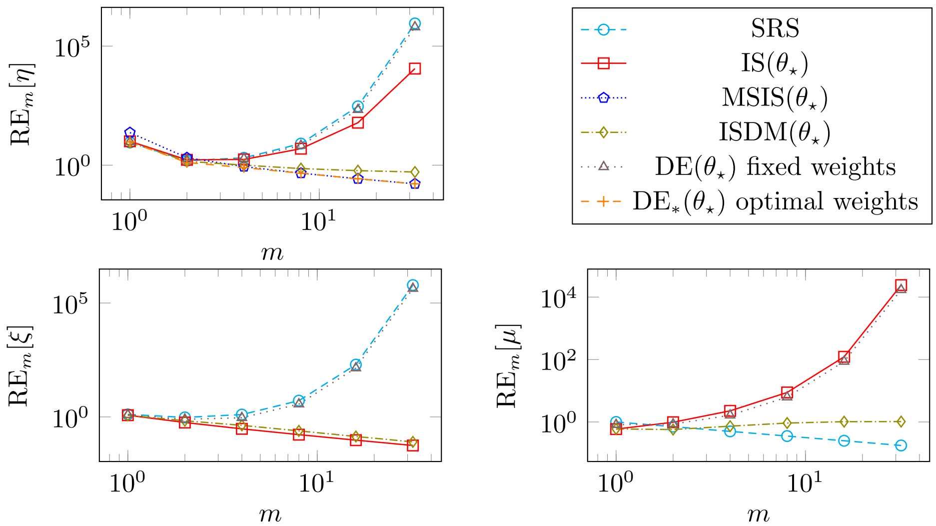

Figure 1 **For G_0 as exponential (mean 1) and $\beta = 1.1$ in (29) the log-log plots show the exact RE computed numerically (i.e., not estimated via simulation), as functions of the dimension m, of estimators of the EC η (top left panel), the p-quantile ξ (bottom left panel), and the mean μ (bottom right). The bottom panels do not give results for MSIS($\theta_\star$), which uses IS($\theta_\star$) (resp., SRS) to estimate ξ (resp., μ).**

through common factors. Let $\mathbf{Z} = (Z_1, \ldots, Z_r)$ be a column vector of $r \geq 1$ *systematic risk factors*, which are i.i.d. $N(0,1)$ random variables, modeling global, country, and sector factors that impact all obligors. For each $k = 1, 2, \ldots, m$, let ϵ_k be another independent $N(0,1)$ random variable denoting the *idiosyncratic risk* associated with obligor k. The *loading factors* are specified constant row vectors $\mathbf{a}_k = (a_{k,j} : j = 1, 2, \ldots, r)$, $k = 1, 2, \ldots, m$, satisfying $\mathbf{a}_k \mathbf{a}_k^\top \leq 1$ for each k, where $\top$ denotes transpose. Let $b_k = (1 - \mathbf{a}_k \mathbf{a}_k^\top)^{1/2}$, so $\mathbf{a}_k \mathbf{Z} + b_k \epsilon_k \sim N(0,1)$ for each k. Let $S > 0$ be another independent random variable denoting a *common shock* affecting all obligors. For each $k = 1, 2, \ldots, m$, obligor k defaults if and only if $(\mathbf{a}_k \mathbf{Z} + b_k \epsilon_k)/S > w_k$ for a constant w_k chosen so that obligor k has a specified marginal default probability $\dot{p}_k$. Glasserman and Li (2005) and Bassamboo et al. (2008) assume that the *loss given default* (LGD) of obligor k is a constant c_k, but they state their methods also allow LGD to be stochastic, which we need to ensure F is differentiable at ξ and $f(\xi) > 0$, as required by our theorems. For obligor k, let J_k be another independent random variable, and define the LGD for obligor k as $v_k(\mathbf{Z}, S, \epsilon_1, \ldots, \epsilon_m, J_k)$ for a given function $v_k : \Re^{r+m+2} \to \Re_+$. Therefore, the LGD may depend on J_k, as well as the systematic and idiosyncratic risk factors and common shock, as in Andersen and Sidenius (2005) and Farinelli and Shkolnikov (2012). Finally, let $\mathbf{X} = (\mathbf{Z}, S, \epsilon_1, \ldots, \epsilon_m, J_1, \ldots, J_m)$, which has

$d = r + 2m + 1$ independent components, and the function c in (1) for the total loss is

$$c(\mathbf{X}) = \sum_{k=1}^{m} v_k(\mathbf{Z}, S, \epsilon_1, \ldots, \epsilon_m, J_k)\, I\left(\frac{\mathbf{a}_k\mathbf{Z} + b_k\epsilon_k}{S} > w_k\right). \tag{58}$$

Our experiments have $m = 1000$ obligors and $r = 10$ factors. As in Glasserman and Li (2005), we take the common shock to be $S \equiv 1$. Let $D_k = I(\mathbf{a}_k\mathbf{Z} + b_k\epsilon_k > w_k)$ be the indicator function in (58) that obligor k defaults. Because $\mathbf{a}_k\mathbf{Z} + b_k\epsilon_k \sim N(0,1)$, if we set $w_k = \Phi^{-1}(1 - \dot{p}_k)$ for some constant $0 < \dot{p}_k < 1$, where $\Phi(\cdot)$ is the $N(0,1)$ CDF, then obligor k has marginal default probability $P(D_k = 1) = \dot{p}_k$. Our experiments used $\dot{p}_k = 0.01 \cdot (1 + \sin(16\pi k/m))$, $k = 1, \ldots, m$, as in Glasserman and Li (2005). For each obligor $k = 1, 2, \ldots, m$, the constant LGD in Glasserman and Li (2005) is modified to $C_k = v_k(\mathbf{Z}, S, \epsilon_1, \ldots, \epsilon_m, J_k) = J_k \sim \mathrm{Unif}(0, \beta_k)$, where $\beta_k = 2 \cdot (\lceil 5k/m \rceil)^2$ and $\mathrm{Unif}(c_0, c_1)$ denotes a continuous uniform distribution on (c_0, c_1). As in Glasserman and Li (2005), we randomly generated the loading factors $a_{k,j}$ in (58) once as independent $\mathrm{Unif}(0, 1/\sqrt{r})$, and used these values in all experiments.

We ran simulation experiments to estimate this model's EC η for $p = 0.999$. For this model we can compute analytically the mean as $\mu = 104.02$, but this may not be possible for more complicated models, and our simulation experiments treat μ as unknown, requiring estimation. The value of ξ is not analytically tractable, and we obtained its "true" value as $\xi = 1885.9$ from an SRS simulation with sample size 10^7, giving the "true" value for EC as $\eta = 1781.9$.

We construct nominal 95% confidence intervals for η using two approaches: batching and sectioning. For an estimation method $\mathfrak{M}$ and total overall sample size n, we first construct $b \geq 2$ i.i.d. estimators $\widehat{\eta}^{(j)}_{\mathfrak{M},n/b}$, $j = 1, 2, \ldots, b$, of η, each based on a sample size n/b. Batching uses their sample average $\bar{\eta}_{\mathfrak{M},b,n} = (1/n)\sum_{j=1}^{b} \widehat{\eta}^{(j)}_{\mathfrak{M},n/b}$ and sample variance $S^2_{\mathfrak{M},b,n} = (1/(b-1))\sum_{j=1}^{b}[\widehat{\eta}^{(j)}_{\mathfrak{M},n/b} - \bar{\eta}_{\mathfrak{M},b,n}]^2$ to build an approximate $\alpha = 0.95$-level CI $I_{\mathfrak{M},b,n} = (\bar{\eta}_{\mathfrak{M},b,n} \pm t_{b-1,0.95} S_{\mathfrak{M},b,n}/\sqrt{b})$, where $t_{b-1,\alpha} = H^{-1}_{b-1}(1 - \alpha/2)$, and H_{b-1} the Student-t CDF with $b-1$ degrees of freedom. Sectioning (Asmussen and Glynn 2007, Section V.5) replaces $\bar{\eta}_{\mathfrak{M},b,n}$ in $S^2_{\mathfrak{M},b,n}$ and $I_{\mathfrak{M},b,n}$ with the overall point estimator $\widehat{\eta}_{\mathfrak{M},n}$ to get a CI $J_{\mathfrak{M},b,n}$, centered at $\widehat{\eta}_{\mathfrak{M},n}$. Because η estimators are biased, with the bias shrinking (nonmonotonically) as the sample size increases, the sectioning CI can have better coverage (but not always; see He and Lam (2021)) than $I_{\mathfrak{M},b,n}$ as $J_{\mathfrak{M},b,n}$ is better centered on average (Kaplan et al. (2018)).

Table 2 gives results of coverage experiments to construct batching and sectioning CIs for η using SRS, IS_{φ} (explained below), MSIS, ISDM, and DE, each with overall sample size $n = 2000$. We take $b = 10$, as suggested by Nakayama (2014). From 10^3 independent replications, we estimated the batching and sectioning CIs' coverage and *average relative half width* (ARHW), and the point estimators' *root-mean-squared relative error* (RMSRE), defined as $\sqrt{E[(\hat{\eta}-\eta)^2]}/\eta$ for a generic estimator $\widehat{\eta}$ of η. When the coverage is low, the ARHW and RMSRE results may not be reliable.

For SRS, the batching CI has poor coverage, while the coverage for sectioning is reasonably close to nominal, for the reasons explained before. Also, for sectioning, the ARHW (resp., RMSRE) for SRS is about 7 (resp., 13) times larger than for MSIS.

IS_φ is a modification of a method of Glasserman and Li (2005) for estimating a tail probability $\lambda_x \equiv P(Y > x)$ for a given large threshold x to estimate λ_φ, where φ is either η or ξ, and then use the generated IS data to compute an estimator of η. But as these choices for φ are unknown, we cannot directly apply the Glasserman and Li (2005) IS algorithm to estimate λ_φ. Rather, when $\varphi = \xi$, we first run $j_0 = 5$ pilot IS simulations, each with small sample size $n_0 = 100$, to estimate λ_x at j_0 different thresholds x, and interpolate to obtain a crude approximation $\mathring{\xi}$ to ξ. Then IS_ξ runs another IS simulation with sample size $n - j_0 n_0$ to estimate $\lambda_{\mathring{\xi}}$, finally employing the generated IS data to estimate both ξ and μ to obtain an estimator of η. Each independent replication repeated these steps. Appendix G gives the approach's full details for $\varphi = \xi$. For ISDM, the only difference from IS_ξ is that we sample $\mathbf{X} \sim \widetilde{G}_{\mathrm{ISDM}} = \delta G^* + (1-\delta)G$ in Section 5.2, where G^* corresponds to IS_ξ.

For IS_φ with $\varphi = \eta$, we execute an additional pilot SRS simulation with sample size n_0 to produce an approximation $\mathring{\mu}$ to μ, and compute $\mathring{\eta} = \mathring{\xi} - \mathring{\mu}$ as an approximation to η. Then IS_φ for $\varphi = \eta$ runs an IS simulation with sample size $n - (j_0+1)n_0$ to estimate λ_φ for $\varphi = \mathring{\eta}$, and employs the resulting IS data to compute estimators of both ξ and μ, resulting in our final estimator of η.

Table 2 shows that for each choice of φ, the IS_φ CI does poorly, with coverage near 0. This occurs because in IS_φ, we apply the same IS data from estimating ξ to also estimate μ, leading to the problems discussed in Section 5 and the poor coverage for our CIs. In particular, the average across 10^3 replications of the IS_η estimator of μ is about 11.1, quite far from the true value 104.02. As noted on pp. 134–135 of Asmussen and Glynn (2007), these types of discrepancies can occur with IS when the sample size is not sufficiently large, especially when an IS approach is applied inappropriately for the estimand. (To investigate this further, we ran additional simulations (not reported) verifying that the

IS_φ CIs approach nominal coverage with larger sample size n for $p = 0.95$. For larger p, rather than conducting converge experiments, which would require a massive sample size and extremely long CPU time (several months), we did experiments showing that the IS_φ estimators of η appear to converge to its "true" values as n gets larger, indicating that the CIs should approach nominal coverage with a large enough sample size.) Also, the ARHW and RMSRE results for IS_φ may not be reliable because of the poor coverages.

For MSIS, we use a total sample of size $j_0 n_0$ for computing the crude quantile approximation $\mathring{\xi}$, as is done with IS_ξ; then generate an IS sample of size $\delta(n - j_0 n_0)$ to estimate $\lambda_{\mathring{\xi}}$, and use the resulting IS data to estimate ξ; and finally employ an SRS sample of size $(1-\delta)(n - j_0 n_0)$ for the estimation of μ. Table 2 shows that MSIS sectioning and batching CIs achieve nominal coverage, with about the same ARHW, but with the sectioning point estimator having roughly 10% smaller RMSRE. MSIS outperforms SRS for both batching and sectioning, with the mean-squared error (MSE) for sectioning being reduced by a factor of $(2.276\text{e–}01/1.801\text{e–}02)^2 \approx 160$. In our python implementations, the IS code, including the pilot runs to obtain the crude quantile approximation $\mathring{\xi}$, requires about thrice the CPU time as SRS to execute. Taking this into account, MSIS improves work-normalized MSE by about a factor of 50 compared to SRS. DE and MSIS differ only in computing their estimator of η from the generated data; see (25) and (21). For simplicity, we consider DE with only fixed weights for the PCRM.

We next compare the methods (MSIS, ISDM, DE of Section 5) that combine SRS and IS. For the i.i.d. sum model in Section 6, recall that Theorem 5 and Figure 1 (Section 7.1) established the following properties for the methods' RE of η:

- MSIS does better than ISDM (but not by a lot);
- Both MSIS and ISDM greatly outperform DE; and
- DE (with fixed weights) behaves about the same as the worse of IS_ξ and SRS.

For the more complicated PCRM, Table 2 shows that the methods perform similarly in terms of ARHW and RMSRE. First, comparing MSIS and ISDM shows that MSIS has about 30% smaller ARHW and RMSRE than ISDM for sectioning; while MSIS and ISDM produce CIs achieving close to nominal coverage, MSIS perhaps does a bit better. Second, relative to MSIS, DE has about 5 (resp., 10) times larger ARHW (resp., RMSE) for sectioning. While DE has sectioning coverage for η reasonably close to nominal, it is not as good as MSIS. We expect DE to do about the same as the worse of SRS and IS_ξ, and the ARHW and RMSRE of DE are reasonably close to those of SRS for sectioning, but IS_ξ has very poor coverages so its ARHW and RMSRE results may not be reliable. The coverages of batching and sectioning for DE differ substantially, which is similar to what we see for SRS for the same reasons. Both ISDM and DE incur about the same CPU time as MSIS. Thus, the methods exhibit the same behavior for the PCRM as we saw for the i.i.d. sum model. Also compared to SRS, MSIS improved precision by reducing the ARHW of the sectioning CI from roughly 0.3 to only 0.04.

Table 2 **We ran 10^3 independent replications of the PCRM to estimate the coverage and average relative half width (ARHW) of sectioning and batching CIs with nominal 95% confidence level for the EC η for $p = 0.999$ estimated with sample size $n = 2000$. We also give the root-mean-squared relative error (RMSRE). Numbers marked with $*$ may not reliable due to very low coverage.**

	Batching			Sectioning		
Method	Coverage	ARHW	RMSRE	Coverage	ARHW	RMSRE
MSIS	0.921	0.038	2.016e–02	0.956	0.041	1.801e–02
ISDM	0.867	0.060	5.199e–02	0.915	0.060	2.574e–02
DE	0.076	0.136*	2.274e–01*	0.884	0.220	1.803e–01
SRS	0.365	0.273*	2.633e–01*	0.892	0.292	2.276e–01
IS_η	0.096	0.024*	7.253e–02*	0.087	0.024*	7.370e–02*
IS_ξ	0.074	0.027*	5.212e–02*	0.047	0.028*	5.356e–02*

8. Concluding Remarks

The economic capital is a risk measure, which is used to determine capital levels (e.g., Deutsche Bank (2018)). Defined as the difference between the p-quantile ξ and the mean μ of the loss distribution, the EC in practice takes $p \approx 1$, in which case SRS, which typically estimates μ well, is ineffective for ξ. Applying IS to estimate both ξ and μ can be detrimental for μ, leading to a poor estimator of η. We thus also considered methods that combine SRS and IS to estimate η in various ways: MSIS, which applies IS to estimate ξ only, and independently employs SRS to estimate μ only; ISDM, which samples from a mixture of IS and SRS; and DE, which estimates both ξ and μ using both IS and SRS, and takes a weighted linear combination of all the estimators. For DE, we considered both fixed weights and optimal weights to minimize the η estimator's asymptotic variance. Our asymptotic theory (Theorem 5 of Section 6.5) for the i.i.d. sum model with m summands as

$m \to \infty$ shows that SRS alone, IS alone, and DE with fixed weights do poorly in estimating η. In contrast, ISDM, MSIS, and DE with optimal weights perform well, with ISDM being slightly outperformed when $\mu \neq 0$ by the latter two, whose asymptotic variances quickly become indistinguishable as $m \to \infty$. Hence, MSIS provides a compelling simple alternative to DE with optimal weights, which is more complicated to implement. Our numerical and simulation studies (Section 7) provide results agreeing with the theory, even for a more complicated portfolio credit risk model, with dependent obligors.

As a measure of central tendency, the median $\mu' \equiv F^{-1}(1/2)$ is sometimes adopted instead of μ as a location parameter. This motivates a modified EC, denoted EC$'$, defined as the difference $\xi - \mu'$. Typically analytically intractable for nontrivial models, μ' often lies below μ for positively skewed distributions, as can be the case for a portfolio loss. Thus, EC$'$ can be a more conservative risk measure than EC, which may be of interest to regulators. We can also apply the methods in our paper to construct EC$'$ estimators, which for SRS are special cases of L-estimates or L-statistics (Chapter 8 of Serfling (1980)).

While our theoretical results for EC estimators for the i.i.d. sum model (Section 6 and Appendix F) provide deep insights for problems in rare-event simulation and financial risk management, they also have implications for techniques that reuse simulation data (Liu and Zhou (2020), Dong et al. (2018)), which is also called "green simulation" (Feng and Staum (2017)). To estimate mean performances when parameters of underlying distributions in the same simulation model differ across experiments, green simulation reuses outputs from previous experiments by weighting them with likelihood ratios. In estimating a mean, as has been the focus of green simulation, our Theorem 6 in Appendix F.1 shows that IS can result in estimators with an extremely large variance when a single simulation run requires generating many independent random variables, so resuing simulation data through likelihood ratios may be less effective in such contexts. Feng and Staum (2017) and Dong et al. (2018) further apply (a slight variation of) ISDM in green simulations when estimating mean performances, and our Theorem 6 also reveals that ISDM can be quite effective to control the variance but slightly worse than SRS when the mean is nonzero.

Acknowledgments

This work has been supported in part by the National Science Foundation under Grant No. CMMI-1537322. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors thank

Bruno Tuffin for discussions related to the optimal weights for DE in (25). The authors are further deeply indebted to the (anonymous) reviewers for their numerous insightful comments that led to a more complete analysis of the asymptotic behavior of the estimators for Section 6's i.i.d. sum model as the number m of summands grows to infinity. In our original submission, rather than proving limiting properties of the true p_m-quantile ξ_m, the EC η_m, and their estimators' asymptotic variances, which include $f_m(\xi_m)$, as in (40) and (39), where f_m is the sum's density, Theorems 5 and 7 (in Appendix F.2) instead previously provided limiting analyses based on only the approximations $\breve{\xi}_m$ in (85) of ξ_m and $\breve{f}_m(\breve{\xi}_m)$, where $\breve{f}_m$ in (105) of Appendix F.3 is the saddlepoint approximation of f_m. An anonymous referee graciously gave detailed suggestions on how to extend our original proofs of Theorems 5 and 7 to instead work for the true quantities η_m, ξ_m, and $f_m(\xi_m)$. Also, Theorem 6 (Appendix F.1) in our original submission presented an analysis for ISDM based on only the inequality (72), which led to (48) in Theorem 5 merely suggesting (but not conclusively establishing) that MSIS (see (46)) is strictly better than ISDM to estimate η_m; the same referee helpfully provided detailed suggestions to strengthen (72) to instead show (73) and (74) when $\mu_0 \neq 0$, allowing us to definitively prove in this case the asymptotic superiority of MSIS over ISDM through (50) in Theorem 5. Finally, Theorem 5 in the original submission examined the η_m estimator for DE with mainly fixed weights $(\upsilon_1, \upsilon_2) \in (0,1)^2$, as in part (v) of Theorem 5, with only a basic analysis of the optimal DE weights that vary with m, showing that they converge to $(1,0)$ as $m \to \infty$; two referees suggested delving more deeply into the latter case, which eventually led to part (vi) of Theorem 5.

References

Andersen L, Sidenius J (2005) Extensions to the Gaussian copula: Random recovery and random factor loadings. *Journal of Credit Risk* 1(1):29–70.

Asmussen S, Glynn P (2007) *Stochastic Simulation: Algorithms and Analysis* (New York: Springer).

Avramidis AN, Wilson JR (1998) Correlation-induction techniques for estimating quantiles in simulation. *Operations Research* 46:574–591.

Bahadur RR (1966) A note on quantiles in large samples. *Annals of Mathematical Statistics* 37(3):577–580.

Bassamboo A, Juneja S, Zeevi A (2008) Portfolio credit risk with extremal dependence: Asymptotic analysis and efficient simulation. *Operations Research* 56(3):593–606.

Billingsley P (1995) *Probability and Measure* (New York: John Wiley and Sons), 3rd edition.

Chu F, Nakayama MK (2012) Confidence intervals for quantiles when applying variance-reduction techniques. *ACM Transactions On Modeling and Computer Simulation* 22(2):10:1–10:25.

Chung KL (2001) *A Course in Probability Theory* (San Diego: Academic Press), third edition.

Dembo A, Zeitouni O (1998) *Large Deviations Techniques and Applications* (New York: Springer), second edition.

Deutsche Bank (2018) Annual report 2018. Frankfurt am Main, Germany.

Dong J, Feng MB, Nelson BL (2018) Unbiased metamodeling via likelihood ratios. Rabe M, Juan AA, Mustafee N, Skoogh A, Jain S, Johansson B, eds., *Proceedings of the 2018 Winter Simulation Conference*, 1778–1789 (Piscataway, New Jersey: Institute of Electrical and Electronics Engineers).

Durrett R (1996) *Probability: Theory and Examples* (Belmont, California: Duxbury Press), 2nd edition.

Farinelli S, Shkolnikov M (2012) Two models of stochastic loss given default. *Journal of Credit Risk* 8(2):3–20.

Feng M, Staum J (2017) Green simulation: Reusing the output of repeated experiments. *ACM Transactions on Modeling and Computer Simulation* 27(4):28, ISSN 1049-3301.

Ferguson T (1999) Asymptotic joint distribution of sample mean and a sample quantile. `https://www.math.ucla.edu/~tom/papers/unpublished/meanmed.pdf`.

Ghosh JK (1971) A new proof of the Bahadur representation of quantiles and an application. *Annals of Mathematical Statistics* 42:1957–1961.

Glasserman P (2004) *Monte Carlo Methods in Financial Engineering* (New York: Springer).

Glasserman P, Juneja S (2008) Uniformly efficient importance sampling for the tail distribution of sums of random variables. *Mathematics of Operations Research* 33(1):36–50.

Glasserman P, Li J (2005) Importance sampling for portfolio credit risk. *Management Science* 51(11):1643–1656.

Glynn PW (1996) Importance sampling for Monte Carlo estimation of quantiles. Ermakov SM, Melas VB, eds., *Mathematical Methods in Stochastic Simulation and Experimental Design: Proceedings of the 2nd St. Petersburg Workshop on Simulation*, 180–185 (St. Petersburg, Russia: Publishing House of St. Petersburg Univ.).

Goyal A, Shahabuddin P, Heidelberger P, Nicola V, Glynn PW (1992) A unified framework for simulating Markovian models of highly dependable systems. *IEEE Transactions on Computers* C-41(1):36–51.

He S, Lam H (2021) Higher-order coverage error analysis for batching and sectioning. Kim S, Feng B, Smith K, Masoud S, Zheng Z, Szabo C, Loper M, eds., *2021 Winter Simulation Conference (WSC)*, 12 pages (Piscataway, New Jersey: IEEE), URL `http://dx.doi.org/10.1109/wsc52266.2021.9715418`.

Hesterberg T (1995) Weighted average importance sampling and defensive mixture distributions. *Technometrics* 37(2):185–194.

Hong LJ, Hu Z, Liu G (2014) Monte Carlo methods for value-at-risk and conditional value-at-risk: A review. *ACM Transactions on Modeling and Computer Simulataion* 24(4):22:1–22:37.

Hyndman RJ, Fan Y (1996) Sample quantiles in statistical packages. *American Statistician* 50(4):361–365.

Jensen JL (1995) *Saddlepoint Approximations* (New York: Oxford University Press).

Jorion P (2007) *Value at Risk* (New York: McGraw-Hill), 3rd edition.

Jorion P (2011) *Financial Risk Manager Handbook* (Hoboken, New Jersey: John Wiley and Sons, Inc.), 6th edition.

Kaplan ZT, Li Y, Nakayama MK (2018) Monte Carlo estimation of economic capital. Rabe M, Juan AA, Mustafee N, Skoogh A, Jain S, Johansson B, eds., *Proceedings of the 2018 Winter Simulation Conference*, 1754–1765 (Piscataway, New Jersey: Institute of Electrical and Electronics Engineers).

Kawata T (1972) *Fourier Analysis in Probability Theory* (New York: Academic Press).

Kiefer J (1967) On Bahadur's representation of sample quantiles. *Annals of Mathemetical Statistics* 38:1350–1353.

Klaassen P, van Eeghen I (2009) *Economic Capital: How It Works, and What Every Manager Needs to Know* (Burlington, MA: Elsevier).

L'Ecuyer P, Blanchet JH, Tuffin B, Glynn PW (2010) Asymptotic robustness of estimators in rare-event simulation. *ACM Transactions on Modeling and Computer Simulation* 20(1):Article 6.

Li Y, Kaplan ZT, Nakayama MK (2023) Github repository: Monte Carlo methods for economic capital. URL `http://dx.doi.org/10.1287/ijoc.2021.0261.cd`, available for download at https://github.com/INFORMSJoC/2021.0261.

Lin PE, Wu KT, Ahmad IA (1980) Asymptotic joint distribution of sample quantiles and sample mean with applications. *Communications in Statistics - Theory and Methods* A9(1):51–60.

Liu T, Zhou E (2020) Simulation optimization by reusing past replications: don't be afraid of dependence. *Winter Simulation Conference* .

Lütkebohmert E (2009) *Concentration Risk in Credit Portfolios* (Berlin: Springer).

McNeil AJ, Frey R, Embrechts P (2015) *Quantitative Risk Management: Concepts, Techniques, Tools* (Princeton, New Jersey: Princeton University Press), revised edition.

Nakayama MK (2014) Confidence intervals using sectioning for quantiles when applying variance-reduction techniques. *ACM Transactions on Modeling and Computer Simulation* 24(4):19:1–19:21.

Owen A, Zhou Y (2000) Safe and effective importance sampling. *Journal of the American Statistical Association* 95(449):135–143.

Reed M, Simon B (1975) *Methods of Modern Mathematical Physics*, volume II: Fourier Analysis, Self-Adjointness (San Diego, California: Academic Press), ISBN 9780125850025.

Scandizzo S (2016) *The Validation of Risk Models: A Handbook for Practitioners* (New York: Palgrave Macmillan).

Serfling RJ (1980) *Approximation Theorems of Mathematical Statistics* (New York: John Wiley and Sons).

Sun L, Hong LJ (2010) Asymptotic representations for importance-sampling estimators of value-at-risk and conditional value-at-risk. *Operations Research Letters* 38(4):246–251.

Whitt W (2002) *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues* (New York: Springer-Verlag).

Online Supplement: Appendices of "Monte Carlo Methods for Economic Capital"

Yajuan Li, Zachary T. Kaplan, and Marvin K. Nakayama
Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102 {yl935@njit.edu, ztk4@njit.edu, marvin@njit.edu}

A: Further Numerical Study of Relative Error for i.i.d. Sum Model (30)

Recall that Section 7.1 presented numerical results for the RE for the model in Section 6 when the i.i.d. summands have marginal distribution G_0 that is exponential. We now present some additional results for G_0 as normal $N(1,1)$ and Erlang ($s=8$ stages, scale parameter 1).

For G_0 as $N(1,1)$ (resp., Erlang), Figure A.1 (resp., Figure A.2) provides log-log plots of the exact RE for estimators of η, ξ, and μ. These two figures mostly exhibit the same basic trends that we saw in Figure 1 when G_0 is exponential: the η estimators have exponentially increasing RE as m grows for SRS, IS($\theta_\star$), and DE($\theta_\star$) with fixed weights; and decreasing RE for MSIS($\theta_\star$), ISDM($\theta_\star$), and DE$_*$($\theta_\star$) with optimal weights, where MSIS($\theta_\star$) and DE$_*$($\theta_\star$) rapidly become indistinguishable as m grows. But one difference is that for RE[η] for large m, SRS is the worst for exponential G_0, whereas IS($\theta_\star$) is the worst for G_0 as $N(1,1)$ and Erlang. Also, as noted in the two paragraphs after Theorem 5, the behavior of RE[η] for DE($\theta_\star$) with fixed weights is governed by the *worst* of the SRS and IS($\theta_\star$) estimators of ξ and μ, which, comparing the scales in the bottom panels of the figures shows, is the SRS estimator of ξ in Figure 1 and the IS($\theta_\star$) estimator of μ in Figures A.1 and A.2.

B: Proof of Theorem 1

Ghosh (1971) proves (5) and (6) hold when $f(\xi)>0$. Further assuming that F is twice differentiable at ξ, Kiefer (1967) (see also p. 100 of Serfling (1980)) derives the exact rate of convergence of R_n, given in (7), improving on the original result of Bahadur (1966) (also see Theorem 2.5.1 of Serfling (1980)). In all cases, putting (5), (3), and (2) into (4) then leads to

$$\widehat{\eta}_{\mathrm{SRS},n} = \xi - \frac{\widehat{F}_{\mathrm{SRS},n}(\xi)-p}{f(\xi)} + R_n - \widehat{\mu}_{\mathrm{SRS},n} = \xi - \mu - \frac{1}{f(\xi)}\left[\frac{1}{n}\sum_{i=1}^{n}\left[1-I(Y_i>\xi)\right]-p\right] + R_n - \frac{1}{n}\sum_{i=1}^{n} Y_i + \mu,$$

which equals (8), establishing part (i).

We next prove part (ii). Rearranging (8) and scaling by $\sqrt{n}$ leads to

$$\sqrt{n}\,[\widehat{\eta}_{\mathrm{SRS},n}-\eta] = -\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\left[\frac{1-I(Y_i>\xi)}{f(\xi)}+Y_i\right]-\left[\frac{p}{f(\xi)}+\mu\right]\right)\right]+\sqrt{n}R_n. \tag{59}$$

Let $A=[(1-I(Y>\xi))/f(\xi)]+Y$ and $A_i=[(1-I(Y_i>\xi))/f(\xi)]+Y_i$. Now A_i, $i=1,2,\ldots,n$, are i.i.d. copies of A, where $\phi\equiv E[A]=[p/f(\xi)]+\mu$ as $f(\xi)>0$ ensures $E[1-I(Y>\xi)]=E[I(Y\le\xi)]=F(\xi)=p$. Hence, the right side of (59) equals $-\sqrt{n}[\frac{1}{n}\sum_{i=1}^{n}A_i-\phi]+\sqrt{n}R_n$. Also, $\mathrm{Var}[1-I(Y>\xi)]=p(1-p)=\chi^2_{\mathrm{SRS}}$ implies

$$\mathrm{Var}[A]=\mathrm{Var}[(1-I(Y>\xi))/f(\xi)]+\mathrm{Var}[Y]+2\mathrm{Cov}[(1-I(Y>\xi))/f(\xi),Y]=\zeta^2_{\mathrm{SRS}}$$

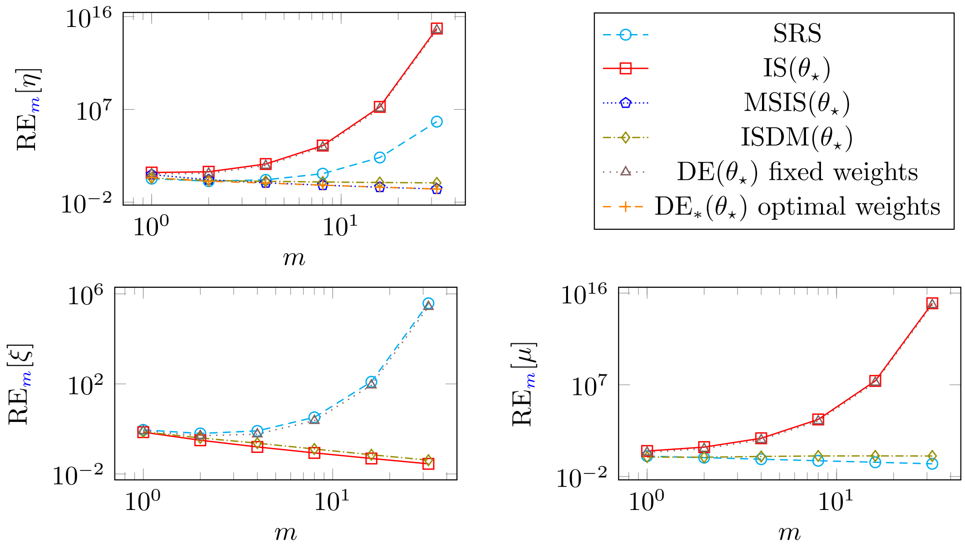

Figure A.1 **For G_0 as $N(1,1)$ and $\beta = 1.1$ in (29), the log-log plots show the RE, computed numerically (i.e., not estimated via simulation), as functions of the dimension m. The plots display the exact RE of estimators of the EC η (top left panel), the p-quantile ξ (bottom left panel), and the mean μ (bottom right panel). The bottom panels do not give results for MSIS($\theta_\star$), which uses IS($\theta_\star$) (resp., SRS) to estimate ξ (resp., μ).**

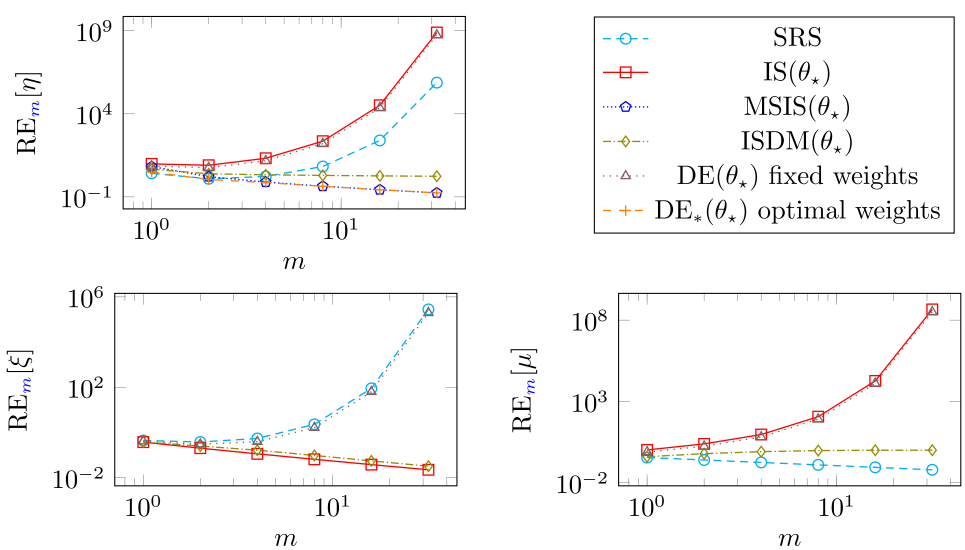

Figure A.2 **For G_0 as Erlang ($s = 8$ stages, scale parameter 1) and $\beta = 1.1$ in (29), the log-log plots show RE, computed numerically (i.e., not simulation), as functions of dimension m. The plots display the exact RE of estimators of the EC η (top left panel), the p-quantile ξ (bottom left panel), and the mean μ (bottom right panel). The bottom panels do not give results for MSIS($\theta_\star$), which uses IS($\theta_\star$) (resp., SRS) to estimate ξ (resp., μ).**

by (9) as $\mathrm{Cov}[(1 - I(Y > \xi))/f(\xi), Y] = [-E[I(Y > \xi)Y] + \mu(1-p)]/f(\xi) = -\gamma_{\mathrm{SRS}}$. We assumed that $\sigma^2_{\mathrm{SRS}} < \infty$ and $f(\xi) > 0$, so $\zeta^2_{\mathrm{SRS}} < \infty$ by the Cauchy-Schwarz inequality. Thus, the ordinary CLT ensures that

$$-\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} A_i - \phi\right] \Rightarrow N(0, \zeta^2_{\mathrm{SRS}}) \qquad \text{as } n \to \infty. \tag{60}$$

As the limit in (6) is deterministic, the left sides of (60) and (6) jointly converge to their respective limits by Theorem 11.4.5 of Whitt (2002). Hence, applying the continuous-mapping theorem (e.g., Theorem 3.4.3 of Whitt (2002)) establishes (9). $\square$

C: Proof of Theorem 2

Put the first part of (16) and (12) into (17) and use the second part of (16). This establishes part (i).

We next prove part (ii). The sum in (18) has i.i.d. summands, where each summand has mean 0 and variance ζ_{IS}^2. From the first part of each summand, we have that $\mathrm{Var}_{\widetilde{G}}[1-I(c(\mathbf{X})>\xi)L(\mathbf{X})] = \mathrm{Var}_{\widetilde{G}}[I(c(\mathbf{X})>\xi)L(\mathbf{X})] = E_{\widetilde{G}}[(I(c(\mathbf{X})>\xi)L(\mathbf{X}))^2] - (1-p)^2$, and

$$E_{\widetilde{G}}[I(c(\mathbf{X})>\xi)L^2(\mathbf{X})] \le E_{\widetilde{G}}[I(c(\mathbf{X})>\xi-\lambda)(L^{2+\epsilon}(\mathbf{X})+1)] < \infty \tag{61}$$

by (15), so $\chi_{\mathrm{IS}}^2 < \infty$. Also, we assumed that $\sigma_{\mathrm{IS}}^2 < \infty$, so the Cauchy-Schwarz inequality ensures that γ_{IS} is finite, implying the same is true for ζ_{IS}^2 because $f(\xi) > 0$. Thus, rearranging (18) and scaling it by $\sqrt{n}$ leads to $\sqrt{n}[\widehat{\eta}_{\mathrm{IS},n} - \eta]$ satisfying part (ii) by (16) and Slutsky's theorem (e.g., p. 19 of Serfling (1980)). $\square$

D: Proof of Theorem 3

Put (16) with $n_1 = \delta n$ replacing n into (21) to get

$$\widehat{\eta}_{\mathrm{MSIS},n} = \xi - \mu - \frac{\widehat{F}_{\mathrm{IS},\delta n}(\xi) - p}{f(\xi)} + \tilde{R}_{n,\delta} - \widehat{\mu}_{\mathrm{SRS},(1-\delta)n} + \mu,$$

so (22) follows. This establishes part (i).

We next prove part (ii). Rearrange (22) and scale by $\sqrt{n}$ to get

$$\sqrt{n}\,[\widehat{\eta}_{\mathrm{MSIS},n} - \eta] = -\frac{\sqrt{n}}{f(\xi)}\left[\widehat{F}_{\mathrm{IS},\delta n}(\xi) - p\right] - \sqrt{n}\left[\widehat{\mu}_{\mathrm{SRS},(1-\delta)n} - \mu\right] + \sqrt{n}\tilde{R}_{n,\delta}.$$

As we showed in (61), (15) implies that $\chi_{\mathrm{IS}}^2 < \infty$, so $f(\xi) > 0$ ensures that

$$\sqrt{n}[\widehat{F}_{\mathrm{IS},\delta n}(\xi) - p] \Rightarrow N_1' \sim N\left(0, \frac{\chi_{\mathrm{IS}}^2}{\delta f^2(\xi)}\right), \tag{62}$$

where the δ appears in the denominator of the asymptotic variance because the left side of (62) scales by $\sqrt{n}$ rather than $\sqrt{n_1}$, and the sample size used to construct $\widehat{F}_{\mathrm{IS},\delta n}$ is $n_1 = \delta n$. Also,

$$\sqrt{n}[\widehat{\mu}_{\mathrm{SRS},(1-\delta)n} - \mu] \Rightarrow N_2' \sim N\left(0, \frac{\sigma_{\mathrm{SRS}}^2}{1-\delta}\right) \quad \text{as } n \to \infty \tag{63}$$

since $\sigma_{\mathrm{SRS}}^2 < \infty$, where the $1-\delta$ appears in the denominator of the asymptotic variance because the scaling in (63) is $\sqrt{n}$ rather than $\sqrt{n_2}$. Under MSIS, $\widehat{\mu}_{\mathrm{SRS},(1-\delta)n}$ is independent of $\widehat{\xi}_{\mathrm{IS},\delta n}$ and $\widehat{F}_{\mathrm{IS},\delta n}$, guaranteeing the joint convergence of (62) and (63) as $n \to \infty$ by Theorem 11.4.4 of Whitt (2002). Moreover, because the limit in (22) is deterministic, it follows that

$$(\sqrt{n}\left[\widehat{F}_{\mathrm{IS},\delta n}(\xi) - p\right], \sqrt{n}\left[\widehat{\mu}_{\mathrm{SRS},(1-\delta)n} - \mu\right], \sqrt{n}\tilde{R}_{n,\delta}) \Rightarrow (N_1', N_2', 0) \quad \text{as } n \to \infty$$

by Theorem 11.4.5 of Whitt (2002), where N_1' and N_2' are independent. Finally, applying the continuous-mapping theorem completes the proof. $\square$

E: Proof of Theorem 4

By (25), we have $\widehat{\eta}_{\mathrm{DE},n} = \upsilon_1 \widehat{\xi}_{\mathrm{IS},\delta n} + \upsilon_1' \widehat{\xi}_{\mathrm{SRS},(1-\delta)n} - \upsilon_2 \widehat{\mu}_{\mathrm{IS},\delta n} - \upsilon_2' \widehat{\mu}_{\mathrm{SRS},(1-\delta)n}$, where $\widehat{\xi}_{\mathrm{IS},\delta n}$ and $\widehat{\mu}_{\mathrm{IS},\delta n}$ are from (17), and $\widehat{\xi}_{\mathrm{SRS},(1-\delta)n}$ and $\widehat{\mu}_{\mathrm{SRS},(1-\delta)n}$ are from (4). Use the corresponding expressions for $\widehat{\xi}_{\mathrm{IS},\delta n}$ and $\widehat{\mu}_{\mathrm{IS},\delta n}$ from (18), and expressions for $\widehat{\xi}_{\mathrm{SRS},(1-\delta)n}$ and $\widehat{\mu}_{\mathrm{SRS},(1-\delta)n}$ from (8), and then rearrange these to get (26).

To prove part (ii), employ (25) to split $\sqrt{n}\,[\widehat{\eta}_{\mathrm{DE},n} - \eta]$ into terms based on IS and terms based on SRS, which will be analyzed separately. For the IS estimators, modify the proof of Theorem 2(ii) to get the CLT

$$\sqrt{n}\left(\left[\upsilon_1 \widehat{\xi}_{\mathrm{IS},\delta n} - \upsilon_2 \widehat{\mu}_{\mathrm{IS},\delta n}\right] - [\upsilon_1 \xi - \upsilon_2 \mu]\right) \Rightarrow N_1 \tag{64}$$

as $n \to \infty$, where $N_1 \sim N(0, \psi^2_{\mathrm{IS},\delta,\upsilon_1,\upsilon_2})$ and

$$\psi^2_{\mathrm{IS},\delta,\upsilon_1,\upsilon_2} = \frac{1}{\delta}\left[\upsilon_1^2 \frac{\chi^2_{\mathrm{IS}}}{f^2(\xi)} + \upsilon_2^2 \sigma^2_{\mathrm{IS}} - 2\upsilon_1\upsilon_2 \frac{\gamma_{\mathrm{IS}}}{f(\xi)}\right].$$

Similarly, for the SRS estimators, we modify the proof of Theorem 1(ii) to get the CLT

$$\sqrt{n}\left(\left[\upsilon_1' \widehat{\xi}_{\mathrm{SRS},(1-\delta)n} - \upsilon_2' \widehat{\mu}_{\mathrm{SRS},(1-\delta)n}\right] - [\upsilon_1' \xi - \upsilon_2' \mu]\right) \Rightarrow N_2 \tag{65}$$

as $n \to \infty$, where $N_2 \sim N(0, \psi^2_{\mathrm{SRS},\delta,\upsilon_1,\upsilon_2})$ and

$$\psi^2_{\mathrm{SRS},\delta,\upsilon_1,\upsilon_2} = \frac{1}{1-\delta}\left[\upsilon_1'^2 \frac{\chi^2_{\mathrm{SRS}}}{f^2(\xi)} + \upsilon_2'^2 \sigma^2_{\mathrm{SRS}} - 2\upsilon_1'\upsilon_2' \frac{\gamma_{\mathrm{SRS}}}{f(\xi)}\right].$$

Hence, $\sqrt{n}\,[\widehat{\eta}_{\mathrm{DE},n} - \eta]$ equals the sum of the left sides of (64) and (65) by (25). The estimators in (64) and (65) are independent, so the CLTs in (64) and (65) hold jointly with N_1 and N_2 independent (Whitt 2002, Theorem 11.4.5). Thus, the continuous-mapping theorem (Whitt 2002, Theorem 3.4.3) implies $\sqrt{n}\,[\widehat{\eta}_{\mathrm{DE},n} - \eta] \Rightarrow N_1 + N_2 \sim N(0, \psi^2_{\mathrm{IS},\delta,\upsilon_1,\upsilon_2} + \psi^2_{\mathrm{SRS},\delta,\upsilon_1,\upsilon_2})$ as $n \to \infty$, proving (27).

Next we prove for fixed $\delta \in (0,1)$, the optimal value of υ_1 and υ_2 is as in (28). The partial derivative of ζ^2_{DE} with respect to υ_1 is $\frac{\partial \zeta^2_{\mathrm{DE}}}{\partial \upsilon_1} = 2\mathrm{V}^{(\xi)}_{\mathrm{IS}}\upsilon_1 - 2\mathrm{V}^{(\xi)}_{\mathrm{SRS}}(1-\upsilon_1) + 2[-\mathrm{C}_{\mathrm{IS}}\upsilon_2 + \mathrm{C}_{\mathrm{SRS}}(1-\upsilon_2)]$. The partial derivative of ζ^2_{DE} with respect to υ_2 is $\frac{\partial \zeta^2_{\mathrm{DE}}}{\partial \upsilon_2} = 2\mathrm{V}^{(\mu)}_{\mathrm{IS}}\upsilon_2 - 2\mathrm{V}^{(\mu)}_{\mathrm{SRS}}(1-\upsilon_2) + 2[-\mathrm{C}_{\mathrm{IS}}\upsilon_1 + \mathrm{C}_{\mathrm{SRS}}(1-\upsilon_1)]$. By setting these two partial derivatives equal to 0 and solving, we get $(\upsilon_1^*, \upsilon_2^*)$ in (28). □

F: Proofs and Additional Results for i.i.d. Sum Model of Section 6

Theorem 5 in Section 6.5 specifies the asymptotic behavior of the estimators of the EC $\eta_m = \xi_m - \mu_m$. Its proof in Appendix F.6 builds on analogous properties for the mean μ_m and the p_m-quantile ξ_m, which we will establish in this appendix. Appendices F.1 and F.2 will first state the results about μ_m and ξ_m in Theorems 6 and 7, which are later proved in Appendices F.4 and F.5 after first securing several lemmas in Appendix F.3.

F.1. Estimating μ_m

We want to analyze the asymptotic behavior (as $m \to \infty$) of estimators of the mean $\mu \equiv \mu_m = E_G[c(\mathbf{X})] = E_{\widetilde{G}_\theta}[c(\mathbf{X})L_\theta(\mathbf{X})] = m\mu_0$ for methods $\mathfrak{M} = \mathrm{SRS}$ (also used by MSIS(θ)), IS(θ), ISDM(θ), and DE(θ). The next result, proven in Appendix F.4, provides expressions and bounds for the estimators' exact variances, the RE in (36) with $\varphi = \mu_m$, and the WNRE in (37)–(38), where we recall $Q_0(\theta) = \ln M_0(\theta)$ is the CGF of $X_j \sim G_0$, with $Q_0'(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}Q_0(\theta)$ and $Q_0''(\theta) = \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}Q_0(\theta)$.

Theorem 6. *Under Assumption 1 for the i.i.d. sum model* (30) *with* $m \geq 1$ *summands, the following hold for method-*$\mathfrak{M}$ *estimators of* $\mu \equiv \mu_m = E_G[c(\mathbf{X})]$.

(i) The $\mathfrak{M} = \mathrm{SRS}$ *estimator* $\widehat{\mu}_{\mathrm{SRS},n}$ *in* (2) *has variance* $\sigma^2_{\mathrm{SRS}}/n$*, with*

$$\sigma^2_{\mathrm{SRS}} = \mathrm{Var}_G[c(\mathbf{X})] = \sigma_0^2 m \quad \textit{and} \quad \sigma_0^2 \in (0,\infty), \tag{66}$$

so when $\mu_0 \neq 0$*,*

$$\mathrm{RE}_{\mathrm{SRS},m}[\mu] = \left(\frac{\sigma_0}{|\mu_0|}\right)\frac{1}{\sqrt{m}} \to 0 \ \textit{as}\ m\to\infty, \ \textit{and}\ \mathrm{WNRE}_{\mathrm{SRS},m}[\mu] = \frac{\sqrt{\tau_{\mathrm{SRS}}}\,\sigma_0}{|\mu_0|} \ \textit{for all}\ m. \tag{67}$$

All of the remaining parts consider any $\theta \neq 0$ *with* $\theta \in \Delta^\circ$*, where methods* $\mathfrak{M} = \mathrm{IS}(\theta)$ *and* $\mathrm{DE}(\theta)$ *(but not* $\mathrm{ISDM}(\theta)$*) also require that* $-\theta \in \Delta^\circ$*. In particular, the following hold for* $\theta = \theta_\star > 0$ *in* (32) *of Assumption 2 (when also* $-\theta_\star \in \Delta^\circ$ *for* $\mathrm{IS}(\theta_\star)$ *and* $\mathrm{DE}(\theta_\star)$*), although this choice of* θ *is not required.*

(ii) The $\mathfrak{M} = \mathrm{IS}(\theta)$ *estimator* $\widehat{\mu}_{\mathrm{IS}(\theta),n}$ *in* (12) *has variance* $\sigma^2_{\mathrm{IS}(\theta)}/n$*, with*

$$\sigma^2_{\mathrm{IS}(\theta)} \equiv \mathrm{Var}_{\widetilde{G}_\theta}[c(\mathbf{X})L_\theta(\mathbf{X})] = m[\alpha(\theta)]^m\Big(m[Q_0'(-\theta)]^2 + Q_0''(-\theta)\Big) - (m\mu_0)^2, \tag{68}$$

$$\textit{where} \quad \alpha(\theta) \equiv M_0(\theta)M_0(-\theta) \in (1,\infty) \quad \textit{and} \quad Q_0''(-\theta) > 0, \tag{69}$$

so

$$\sigma^2_{\mathrm{IS}(\theta)} = O(m^2[\alpha(\theta)]^m) \quad \textit{and} \quad \sigma^2_{\mathrm{IS}(\theta)} = \Omega(m[\alpha(\theta)]^m) \quad \textit{as} \quad m\to\infty. \tag{70}$$

If in addition $\mu_0 \neq 0$*, then as* $m\to\infty$*,*

$$\mathrm{RE}_{\mathrm{IS}(\theta),m}[\mu] = \Omega([\alpha(\theta)]^{m/2}/\sqrt{m}) \to \infty, \ \textit{and}\ \mathrm{WNRE}_{\mathrm{IS}(\theta),m}[\mu] = \Omega([\alpha(\theta)]^{m/2}) \to \infty. \tag{71}$$

(iii) The $\mathfrak{M} = \mathrm{ISDM}(\theta)$ *estimator* $\widehat{\mu}_{\mathrm{ISDM}(\theta),n}$ *in* (12) *has variance* $\sigma^2_{\mathrm{ISDM}(\theta)}/n$*, with*

$$\sigma^2_{\mathrm{ISDM}(\theta)} \equiv \mathrm{Var}_{\widetilde{G}_{\mathrm{ISDM}(\theta)}}[c(\mathbf{X})L_{\mathrm{ISDM}(\theta)}(\mathbf{X})] \le \frac{\delta\mu_0^2}{1-\delta}m^2 + \frac{\sigma_0^2}{1-\delta}m. \tag{72}$$

If $\mu_0 \neq 0$*, then as* $m\to\infty$*,*

$$\sigma^2_{\mathrm{ISDM}(\theta)} = \frac{\delta\mu_0^2}{1-\delta}\big[1+o(1)\big]m^2 = \Theta(m^2), \tag{73}$$

$$\mathrm{RE}_{\mathrm{ISDM}(\theta),m}[\mu] = \Theta(1), \ \textit{and}\ \mathrm{WNRE}_{\mathrm{ISDM}(\theta),m}[\mu] = \Theta(\sqrt{m}). \tag{74}$$

If $\mu_0 = 0$*, then* $\sigma^2_{\mathrm{ISDM}(\theta)} \le \frac{\sigma_0^2}{1-\delta}m$ *by* (72)*.*

(iv) For any fixed weight $\upsilon_2 \in (0,1)$*, the* $\mathfrak{M} = \mathrm{DE}(\theta)$ *estimator* $\widehat{\mu}_{\mathrm{DE}(\theta),n}$ *in* (25) *has variance* $\sigma^2_{\mathrm{DE}(\theta)}/n$*, with*

$$\sigma^2_{\mathrm{DE}(\theta)} \equiv \frac{\upsilon_2^2}{\delta}\sigma^2_{\mathrm{IS}(\theta)} + \frac{\upsilon_2'^2}{1-\delta}\sigma^2_{\mathrm{SRS}} = \Omega(m[\alpha(\theta)]^m) \quad \textit{as} \quad m\to\infty, \tag{75}$$

$$\textit{so}\ \mathrm{RE}_{\mathrm{DE}(\theta),m}[\mu] = \Omega([\alpha(\theta)]^{m/2}/\sqrt{m}) \to \infty \ \textit{and}\ \mathrm{WNRE}_{\mathrm{DE}(\theta),m}[\mu] = \Omega([\alpha(\theta)]^{m/2}) \to \infty. \tag{76}$$

Theorem 6 shows that when we estimate μ_m via SRS or $\mathrm{ISDM}(\theta)$, the variance, RE, and WNRE behave polynomially in m as $m\to\infty$, by (66), (67), (72), (73), and (74). But $\mathrm{IS}(\theta)$ with *any* fixed $\theta \neq 0$ results in exponential growth, by (70) and (71). As seen in (75), $\mathrm{DE}(\theta)$ with fixed weight in $\upsilon_2 \in (0,1)$ takes on the asymptotic characteristics of the *worse* of SRS and $\mathrm{IS}(\theta)$. For some stochastic models of fixed dimension m, Hesterberg (1995), who proves a generalization of (72) (not just for an i.i.d. sum), provides numerical/simulation results showing that an IS method designed to estimate only a tail probability leads to poor mean estimators compared to SRS. Our Theorem 6(ii) provides supporting theory for the setting of a sum of m i.i.d. random variables as $m\to\infty$. While $\mathrm{ISDM}(\theta)$ leads to a much smaller asymptotic variance than $\mathrm{IS}(\theta)$ (compare the polynomial behavior of (72) and the exponential behavior in (70)), $\mathrm{ISDM}(\theta)$ does worse than SRS by a factor of m when $\mu_0 \neq 0$ and $\theta \neq 0$ (compare (73) and (66)). Thus, compared with SRS, $\mathrm{ISDM}(\theta)$ with $\theta \neq 0$ incurs some penalty in estimating $\mu_m \neq 0$.

F.2. Estimating ξ_m

Next we analyze the asymptotic behavior (as $m \to \infty$) of the p_m-quantile $\xi_m = F_m^{-1}(p_m)$ for p_m satisfying (29). For the methods $\mathfrak{M}$ described in Sections 6.1 and 6.2, when $f_m(\xi_m) > 0$ (Assumption 2), the asymptotic variance $\varsigma_{\mathfrak{M}}^2 = \kappa_{\mathfrak{M}}^2 \equiv \kappa_{\mathfrak{M},m}^2$ of the resulting estimator $\widehat{\xi}_{\mathfrak{M},n}$ has the form $\kappa_{\mathfrak{M}}^2 = \chi_{\mathfrak{M}}^2 / f_m^2(\xi_m)$ in (40), as can be seen through Theorems 1–4, and will be more explicitly explained in Theorem 7 below. While the numerator $\chi_{\mathfrak{M}}^2 \equiv \chi_{\mathfrak{M},m}^2$ of (40) depends on $\mathfrak{M}$, the denominator $f_m^2(\xi_m)$ does not, but both depend on m.

We now want to study the RE and WNRE of the SRS, IS($\theta_\star$) (also used by MSIS($\theta_\star$)), ISDM($\theta_\star$), and DE($\theta_\star$) estimators of $\xi \equiv \xi_m = F_m^{-1}(p_m)$ when $F \equiv F_m$ as $m \to \infty$ for quantile level $p \equiv p_m$ as in (29). Glynn (1996) analyzes (only) the numerator $\chi_{\mathfrak{M}}^2$ in (40) for $\mathfrak{M} = \mathrm{SRS}$ and IS($\theta_\star$) in this asymptotic regime, proving that $\lim_{m\to\infty} (1/m)\ln(\chi_{\mathrm{SRS}}^2) = -\beta$ and $\limsup_{m\to\infty}(1/m)\ln(\chi_{\mathrm{IS}(\theta_\star)}^2) \le -2\beta$. This indicates that IS($\theta_\star$) can produce substantial variance reductions.

Further analyzing the quantile estimators' RE and WNRE in (36)–(38) for $\varphi = \xi_m$ requires understanding the asymptotic properties of ξ_m and the denominator $f_m^2(\xi_m)$ in (40). Section 6.5 noted several technical challenges in studying these terms (e.g., F_m is generally an intractable convolution), which are resolved in the following result, whose proof appears in Appendix F.5.

THEOREM 7. *Under Assumptions 1 and 2 for the i.i.d. sum model* (30) *with* $m \ge 1$ *summands, the following hold for method-*$\mathfrak{M}$ *estimators of* $\xi \equiv \xi_m = F_m^{-1}(p_m)$ *for all sufficiently large* m*, where* $\beta > 0$ *is from* (29)*,* $Q_0''(\theta_\star) > 0$ *for* $\theta_\star$ *in* (32)*, and* $\Upsilon_m > 0$ *is defined in* (110) *of Appendix F.3 with* $\Upsilon_m = e^{o(\sqrt{m})}$ *as* $m \to \infty$ *(so* $\Upsilon_m^t = \omega(e^{-\sqrt{m}})$ *for each* $t \in \Re$*).*

(i) For the $\mathfrak{M} = \mathrm{SRS}$ *estimator* $\widehat{\xi}_{\mathrm{SRS},n}$ *in* (3)*, the asymptotic variance* $\kappa_{\mathrm{SRS}}^2 = \chi_{\mathrm{SRS}}^2 / f_m^2(\xi_m)$ *in* (40)*, with* $\chi_{\mathrm{SRS}}^2 = p_m(1-p_m)$ *from* (10)*, satisfies* $\lim_{m\to\infty} \frac{1}{m} \ln \kappa_{\mathrm{SRS}}^2 = \beta$*, and*

$$\kappa_{\mathrm{SRS}}^2 = \frac{2\pi Q_0''(\theta_\star)\left[1 + o\left(\frac{1}{\sqrt{m}}\right)\right] m e^{\beta m}}{\Upsilon_m^2} = \frac{\Theta(m e^{\beta m})}{\Upsilon_m^2} = \omega(m e^{\beta m - \sqrt{m}}) \tag{77}$$

as $m \to \infty$*. If* $Q_0'(\theta_\star) \ne 0$*, then as* $m \to \infty$*,*

$$\mathrm{RE}_{\mathrm{SRS},m}[\xi] = \omega\left(\frac{e^{(\beta/2)m - \sqrt{m}}}{\sqrt{m}}\right) \to \infty \text{ and } \mathrm{WNRE}_{\mathrm{SRS},m}[\xi] = \omega\left(e^{(\beta/2)m - \sqrt{m}}\right) \to \infty. \tag{78}$$

(ii) For the $\mathfrak{M} = \mathrm{IS}(\theta_\star)$ *estimator* $\widehat{\xi}_{\mathrm{IS}(\theta_\star),n}$ *from* (14)*, the asymptotic variance* $\kappa_{\mathrm{IS}(\theta_\star)}^2 = \chi_{\mathrm{IS}(\theta_\star)}^2 / f_m^2(\xi_m)$ *in* (40)*, with* $\chi_{\mathrm{IS}(\theta_\star)}^2 = \mathrm{Var}_{\widetilde{G}_{\theta_\star}}[L_{\theta_\star}(\mathbf{X}) I(c(\mathbf{X}) > \xi_m)]$ *as in* (19)*, satisfies*

$$\kappa_{\mathrm{IS}(\theta_\star)}^2 \le \left[2\pi Q_0''(\theta_\star)\right]\left[1 + o(1)\right] m = O(m) \tag{79}$$

as $m \to \infty$*. If* $Q_0'(\theta_\star) \ne 0$*, then*

$$\mathrm{RE}_{\mathrm{IS}(\theta_\star),m}[\xi] = O(1/\sqrt{m}) \to 0 \text{ and } \mathrm{WNRE}_{\mathrm{IS}(\theta_\star),m}[\xi] = O(1) \text{ as } m \to \infty. \tag{80}$$

(iii) For the $\mathfrak{M} = \mathrm{ISDM}(\theta_\star)$ *estimator* $\widehat{\xi}_{\mathrm{ISDM}(\theta_\star),n}$ *from Section 5.2 the asymptotic variance* $\kappa_{\mathrm{ISDM}(\theta_\star)}^2 = \chi_{\mathrm{ISDM}(\theta_\star)}^2 / f_m^2(\xi_m)$ *in* (40)*, with* $\chi_{\mathrm{ISDM}(\theta_\star)}^2 = \mathrm{Var}_{\widetilde{G}_{\mathrm{ISDM}(\theta_\star)}}[L_{\mathrm{ISDM}(\theta_\star)}(\mathbf{X}) I(c(\mathbf{X}) > \xi_m)]$ *as in* (19)*, satisfies*

$$\kappa_{\mathrm{ISDM}(\theta_\star)}^2 \le \left[\frac{2\pi}{\delta^2} Q_0''(\theta_\star)\right]\left[1 + o(1)\right] m = O(m), \tag{81}$$

as $m \to \infty$*. If* $Q_0'(\theta_\star) \ne 0$*, then*

$$\mathrm{RE}_{\mathrm{ISDM}(\theta_\star),m}[\xi] = O(1/\sqrt{m}) \to 0 \text{ and } \mathrm{WNRE}_{\mathrm{ISDM}(\theta_\star),m}[\xi] = O(1) \text{ as } m \to \infty. \tag{82}$$

(iv) For the $\mathfrak{M} = \mathrm{DE}(\theta_\star)$ *estimator* $\widehat{\xi}_{\mathrm{DE}(\theta_\star),n}$ *from* (25) *with fixed weight* $\upsilon_1 \in (0,1)$*, the asymptotic variance* $\kappa^2_{\mathrm{DE}(\theta_\star)} = \chi^2_{\mathrm{DE}(\theta_\star)}/f_m^2(\xi_m)$ *in* (40)*, with* $\chi^2_{\mathrm{DE}(\theta_\star)} = \frac{\upsilon_1^2}{\delta}\chi^2_{\mathrm{IS}(\theta_\star)} + \frac{\upsilon_1'^2}{1-\delta}\chi^2_{\mathrm{SRS}}$ *as in* (27)*, satisfies* $\lim_{m\to\infty}\frac{1}{m}\ln\kappa^2_{\mathrm{DE}(\theta_\star)} = \beta$*. Also, as* $m \to \infty$*,*

$$\kappa^2_{\mathrm{DE}(\theta_\star)} = \left[\frac{\upsilon_1'^2}{1-\delta}\right]\frac{2\pi Q_0''(\theta_\star)\left[1+o\left(\frac{1}{\sqrt{m}}\right)\right]me^{\beta m}}{\Upsilon_m^2} = \frac{\Theta(me^{\beta m})}{\Upsilon_m^2} = \omega(me^{\beta m-\sqrt{m}}). \tag{83}$$

If $Q_0'(\theta_\star) \neq 0$*, then as* $m \to \infty$*,*

$$\mathrm{RE}_{\mathrm{DE}(\theta_\star),m}[\xi] = \omega\left(\frac{e^{(\beta/2)m-\sqrt{m}}}{\sqrt{m}}\right) \to \infty, \text{ and } \mathrm{WNRE}_{\mathrm{DE}(\theta_\star),m}[\xi] = \omega(e^{(\beta/2)m-\sqrt{m}}) \to \infty. \tag{84}$$

Theorems 6 and 7 show that SRS and IS($\theta_\star$) have opposite effects when estimating μ_m and ξ_m. SRS (resp., IS($\theta_\star$)) leads to polynomial (resp., exponential) behavior (in m, as $m \to \infty$) for the (asymptotic) variance, RE, and WNRE when estimating μ_m, by Theorem 6(i)–(ii), but the estimator of ξ_m behaves exponentially (resp., polynomially), by Theorem 7(i)–(ii). For estimating ξ_m, ISDM($\theta_\star$) inflates the upper bound of the asymptotic variance of IS($\theta_\star$) by a factor of $1/\delta^2$ (compare (79) and (81)), but as the parameter $\delta \in (0,1)$ is fixed, ISDM($\theta_\star$) still has polynomial behavior. By (83), DE($\theta_\star$) with fixed weight $\upsilon_1 \in (0,1)$ adopts the limiting characteristics of the *worse* of SRS and IS($\theta_\star$) in (77) and (79).

F.3. Lemmas for the i.i.d. Sum Model of Section 6

The proofs of Theorems 5 and 7 require getting a handle on the true p_m-quantile $\xi_m = F_m^{-1}(p_m)$, which we will approximate by

$$\breve{\xi}_m = mQ_0'(\theta_\star), \text{ for } \theta_\star \in \Delta^\circ \text{ in (32)}. \tag{85}$$

Glynn (1996, Theorem 2) shows that $(\breve{\xi}_m - \xi_m)/m \to 0$ as $m \to \infty$, which (92) and (98) below sharpen. The next lemma derives asymptotic ($m \to \infty$) properties of the mean μ_m, the p_m-quantile ξ_m, and the EC η_m.

Lemma 1. *Suppose that Assumption 1 holds for the i.i.d. sum model* (30) *with* $m \geq 1$ *summands. Then*

$$\mu_m = \mu_0 m \text{ with } |\mu_0| < \infty, \text{ so } \mu_m = \Theta(m) \text{ as } m \to \infty \text{ when } \mu_0 \neq 0; \tag{86}$$

$$\sigma_0^2 \in (0,\infty). \tag{87}$$

Also, for Δ° *as the interior of the domain of the MGF* $M_0(\theta) = E_0[e^{\theta X_j}]$ *of each summand* $X_j \sim G_0$*,*

$$\theta_1,\theta_2 \in \Delta^\circ \text{ implies } \varrho\theta_1 + (1-\varrho)\theta_2 \in \Delta^\circ \text{ for all } \varrho \in (0,1). \tag{88}$$

Moreover, for $Q_0(\theta) = \ln M_0(\theta)$ *as the CGF of* G_0*,*

$$M_0(\theta) \text{ and } Q_0(\theta) \text{ have derivatives of all orders for } \theta \in \Delta^\circ, \text{ with} \tag{89}$$

$$Q_0''(\theta) > 0, \text{ for each } \theta \in \Delta^\circ. \tag{90}$$

Furthermore, if Assumptions 1 and 2 hold, then

$$F_m \text{ has a (Lebesgue) density } f_m \text{ for all } m \geq q_0, \text{ for } q_0 \text{ in (31)}, \tag{91}$$

and as $m \to \infty$,

$$\xi_m = Q_0'(\theta_\star)m + o(\sqrt{m}), \ \ \textit{so} \ \xi_m = \Theta(m) \ \ \textit{when} \ \ Q_0'(\theta_\star) \neq 0; \tag{92}$$

$$\eta_m = [Q_0'(\theta_\star) - \mu_0]\, m + o(\sqrt{m}) = \Theta(m), \ \ \textit{where} \ \ Q_0'(\theta_\star) - \mu_0 > 0; \tag{93}$$

there exists a convex and compact set $\Psi \subset \Delta^\circ$ with $\theta_\star$ in its interior Ψ°, and

$$\textit{for all } m \textit{ sufficiently large, there exists } \theta_m \in \Psi^\circ \textit{ such that } \xi_m = mQ_0'(\theta_m); \tag{94}$$

$$\sqrt{m}(\theta_m - \theta_\star) \to 0. \tag{95}$$

Proof. The condition that $0 \in \Delta^\circ$ in Assumption 1 ensures G_0 has finite moments of all orders (Billingsley 1995, p. 278), so $|\mu_0| < \infty$ and $\sigma_0^2 < \infty$. As a consequence, (86) holds because $c(\mathbf{X})$ is the sum of m i.i.d. random variables, each with mean μ_0. Also, Assumption 1 stipulates that $\sigma_0^2 > 0$, verifying (87).

Now consider any $\theta_1, \theta_2 \in \Delta^\circ$ and any $\varrho \in (0,1)$. Hölder's inequality then implies that $M_0(\varrho\theta_1 + (1-\varrho)\theta_2) = E_0[(e^{\theta_1 X_j})^{\varrho}(e^{\theta_2 X_j})^{1-\varrho}] \leq (E_0[e^{\theta_1 X_j}])^{\varrho}(E_0[e^{\theta_2 X_j}])^{1-\varrho} < \infty$, so $\varrho\theta_1 + (1-\varrho)\theta_2 \in \Delta$ and also in Δ°, confirming (88); also see (Dembo and Zeitouni 1998, Lemma 2.2.5). As $0 \in \Delta^\circ$ by Assumption 1, it follows (Billingsley 1995, p. 278) that the MGF $M_0(\theta)$ has derivatives of all orders for $\theta \in \Delta^\circ$, so the same holds for the CGF $Q_0(\theta) = \ln M_0(\theta)$, establishing (89). For proving (90), the exponential twist $\widetilde{G}_{0,\theta}$ in (33) of G_0 has mean and variance (Durrett 1996, pp. 72–73)

$$\int x \, \mathrm{d}\widetilde{G}_{0,\theta}(x) = \frac{1}{M_0(\theta)} \int x e^{\theta x} \, \mathrm{d}G_0(x) = \frac{M_0'(\theta)}{M_0(\theta)} = Q_0'(\theta) \ \ \text{and} \tag{96}$$

$$\int x^2 \, \mathrm{d}\widetilde{G}_{0,\theta}(x) - [Q_0'(\theta)]^2 = \frac{1}{M_0(\theta)} \int x^2 e^{\theta x} \, \mathrm{d}G_0(x) - [Q_0'(\theta)]^2 = \frac{M_0''(\theta)}{M_0(\theta)} - \left[\frac{M_0'(\theta)}{M_0(\theta)}\right]^2 = Q_0''(\theta). \tag{97}$$

As G_0 is nondegenerate by (87), $\widetilde{G}_{0,\theta}$ also is (they share the same support by (33)), securing (90).

To prove (91), note that the i.i.d. sum $c(\mathbf{X})$ has characteristic function $C_m(\theta) = [C_0(\theta)]^m$ (Billingsley 1995, eq. (26.12)). Now $|C_0(\theta)| \leq 1$ always holds for all θ (Durrett 1996, p. 92), implying that $|C_0(\theta)|^m \leq |C_0(\theta)|^{q_0}$ for each $m \geq q_0$ and all θ. Thus, $\int_{\Re} |C_m(\theta)| \, \mathrm{d}\theta = \int_{\Re} |C_0(\theta)|^m \, \mathrm{d}\theta \leq \int_{\Re} |C_0(\theta)|^{q_0} \, \mathrm{d}\theta < \infty$ by (31), so the inversion theorem (Durrett 1996, p. 97) guarantees the density f_m exists (and is bounded and continuous) for $m \geq q_0$.

We next establish (92), which by (85) is equivalent to

$$\frac{\xi_m - \breve{\xi}_m}{\sqrt{m}} \to 0 \ \ \text{as} \ \ m \to \infty. \tag{98}$$

First fix any $\epsilon > 0$, and suppose for a contradiction that $|\xi_m - \breve{\xi}_m| > \epsilon\sqrt{m}$ infinitely often. Without loss of generality, suppose that

$$\exists \text{ infinite subsequence } m_1 < m_2 < m_3 < \cdots \text{ such that } \xi_{m_i} - \breve{\xi}_{m_i} > \epsilon\sqrt{m_i} \text{ for all } i = 1,2,3,\ldots. \tag{99}$$

Now (99) and (85) imply that $\xi_{m_i} > m_i Q_0'(\theta_\star) + \epsilon\sqrt{m_i}$ for all $i = 1, 2, \ldots$. Let P_G (resp., $P_{\widetilde{G}_{\theta_\star}}$) denote the probability measure when $\mathbf{X} = (X_1, X_2, \ldots, X_{m_i}) \sim G$ (resp., $\mathbf{X} \sim \widetilde{G}_{\theta_\star}$), and the likelihood ratio is given by (34) for $\theta = \theta_\star$ and $\mathbf{x} = \mathbf{X}$. Note that ξ_{m_i} is the true $(1 - e^{\beta m_i})$-quantile of $c(\mathbf{X}) = \sum_{j=1}^{m_i} X_j$, with

$F_{m_i}(\xi_{m_i}) = 1 - e^{\beta m_i}$ (rather than $\geq$) because F_{m_i} is continuous at ξ_{m_i} since the derivative f_{m_i} exists there by Assumption 2. We then get

$$\begin{aligned}
e^{-\beta m_i} &= P_G\left(c(\mathbf{X}) > \xi_{m_i}\right) \leq E_G\left[I\left(c(\mathbf{X}) > m_i Q_0'(\theta_\star) + \epsilon\sqrt{m_i}\right)\right] \qquad (100)\\
&= E_{\widetilde{G}_{\theta_\star}}\left[I\left(c(\mathbf{X}) > m_i Q_0'(\theta_\star) + \epsilon\sqrt{m_i}\right) L_{\theta_\star}(\mathbf{X})\right]\\
&= E_{\widetilde{G}_{\theta_\star}}\Big[I\Big(c(\mathbf{X}) > m_i Q_0'(\theta_\star) + \epsilon\sqrt{m_i}\Big)\exp\Big(m_i Q_0(\theta_\star) - \theta_\star c(\mathbf{X})\Big)\Big]\\
&\leq \exp\Big[m_i\Big(Q_0(\theta_\star) - \theta_\star Q_0'(\theta_\star)\Big) - \theta_\star\epsilon\sqrt{m_i}\Big] E_{\widetilde{G}_{\theta_\star}}\Big[I\Big(c(\mathbf{X}) > m_i Q_0'(\theta_\star) + \epsilon\sqrt{m_i}\Big)\Big]\\
&= e^{-\beta m_i} e^{-\theta_\star\epsilon\sqrt{m_i}} P_{\widetilde{G}_{\theta_\star}}\Big(c(\mathbf{X}) > m_i Q_0'(\theta_\star) + \epsilon\sqrt{m_i}\Big), \qquad (101)
\end{aligned}$$

where the second line applies a change of measure, the third line uses (34), and the fourth line follows because $\theta_\star > 0$ by (32), which also justifies the last line. For the probability term in (101), note that $X_1, \ldots, X_m$ are i.i.d. under $\widetilde{G}_{\theta_\star}$, each with mean $Q_0'(\theta_\star)$ and variance $Q_0''(\theta_\star) \in (0, \infty)$ by (96)–(97). Thus, their sum $c(\mathbf{X}_m)$ obeys a CLT

$$\frac{c(\mathbf{X}_m) - mQ_0'(\theta_\star)}{\sqrt{mQ_0''(\theta_\star)}} \overset{\widetilde{G}_{\theta_\star}}{\Rightarrow} N(0,1) \quad \text{as} \quad m \to \infty,$$

where $\overset{\widetilde{G}_{\theta_\star}}{\Rightarrow}$ denotes convergence in $\widetilde{G}_{\theta_\star}$-distribution, and the same holds along the subsequence m_i. Hence, by the portmanteau theorem (Billingsley 1995, Theorem 25.8), the probability in (101) satisfies

$$P_{\widetilde{G}_{\theta_\star}}\Big(c(\mathbf{X}_{m_i}) > m_i Q_0'(\theta_\star) + \epsilon\sqrt{m_i}\Big) = P_{\widetilde{G}_{\theta_\star}}\left(\frac{c(\mathbf{X}_{m_i}) - m_i Q_0'(\theta_\star)}{\sqrt{m_i Q_0''(\theta_\star)}} > \frac{\epsilon}{\sqrt{Q_0''(\theta_\star)}}\right) \to 1 - \Phi\left(\frac{\epsilon}{\sqrt{Q_0''(\theta_\star)}}\right)$$

as $i \to \infty$, with $\Phi(\cdot)$ denoting the $N(0,1)$ CDF, so the limit is a constant in $(0, 1/2)$. Moreover, $e^{-\theta_\star\epsilon\sqrt{m_i}} \to 0$ as $i \to \infty$ in (101) because $\theta_\star > 0$ by (32). We thus see a contradiction between (100) and (101), so (99) cannot hold; i.e., $\xi_m - \breve{\xi}_m \leq \epsilon\sqrt{m}$ for all m sufficiently large. An analogous argument also shows that $\xi_m - \breve{\xi}_m \geq -\epsilon\sqrt{m}$ for all m sufficiently large. As $\epsilon > 0$ was arbitrary, we have then verified (98) and (92).

To establish (93), note that $\eta_m = \xi_m - \mu_m = [Q_0'(\theta_\star) - \mu_0]m + o(\sqrt{m})$ by (92), where $\mu_0 = Q_0'(0)$ by (96) for $\theta = 0$. The strict convexity of $Q_0(\cdot)$ on Δ° by (90) ensures that $Q_0'(\cdot)$ is strictly increasing there, where both 0 and $\theta_\star$ belong to Δ° by Assumption 1 and (32), so $Q_0'(\theta_\star) > Q_0'(0) = \mu_0$ since $\theta_\star > 0$ by (32), securing (93).

For showing (94), note that (85) and (98) ensure

$$\frac{\xi_m}{m} - Q_0'(\theta_\star) = \frac{\xi_m - \breve{\xi}_m}{m} \to 0 \quad \text{as} \quad m \to \infty. \qquad (102)$$

Since $\theta_\star \in \Delta^\circ$ by (32), all θ close enough to $\theta_\star$ also lie in Δ°. Moreover, $Q_0'(\theta)$ is continuous and strictly increasing on Δ° by (89)–(90), so (102) secures the existence of $\theta_m \in \Delta^\circ$ such that $Q_0'(\theta_m) = \xi_m/m$ for all sufficiently large m. Consequently, using $\breve{\xi}_m = mQ_0'(\theta_\star)$ by (85) and (98) yield

$$\sqrt{m}\,[Q_0'(\theta_m) - Q_0'(\theta_\star)] = \frac{\xi_m - \breve{\xi}_m}{\sqrt{m}} \to 0 \quad \text{as} \quad m \to \infty, \qquad (103)$$

implying $Q_0'(\theta_m) - Q_0'(\theta_\star) \to 0$ as $m \to \infty$. Now $Q_0'(\theta)$ is strictly increasing on Δ° by (90), leading to

$$\theta_m \to \theta_\star \quad \text{as} \quad m \to \infty, \qquad (104)$$

so by (88), there is a convex and compact set $\Psi \subset \Delta^\circ$ with both $\theta_\star$ and θ_m in its interior Ψ° for all large enough m, verifying (94).

To strengthen (104) to (95), note that $\theta_\star \in \Delta^\circ$ by (32) and $\theta_m \in \Delta^\circ$ for all sufficiently large m by (94), so (88) implies the line segment connecting them lies in Δ°, on which $Q_0'(\cdot)$ is differentiable by (89). Thus, the mean-value theorem ensures $\sqrt{m}[Q_0'(\theta_m) - Q_0'(\theta_\star)] = Q_0''(\theta_{m,\star})\sqrt{m}(\theta_m - \theta_\star)$ for some $\theta_{m,\star} \in \Delta^\circ$ between $\theta_\star$ and θ_m. Also, $Q_0''(\theta_{m,\star}) \to Q_0''(\theta_\star) > 0$ as $m \to \infty$ since $\theta_{m,\star} \to \theta_\star$ as $m \to \infty$ and $Q_0''(\theta)$ is continuous on Δ° by (89). Thus, (103) yields (95). $\square$

The asymptotic variances in (40) and (39) of the p_m-quantile estimator and the EC estimator, respectively, involve $f_m(\xi_m)$, where the density f_m exists by (91) for all $m \geq q_0$, for q_0 in (31). To get a handle on $f_m(\xi_m)$, we will approximate the true density $f(x) = f_m(x)$ of $c(\mathbf{X})$ using a *saddlepoint approximation* (Jensen 1995, Chapter 2), given by

$$\breve{f}(x) \equiv \breve{f}_m(x) = \frac{1}{\sqrt{2\pi m Q_0''(\theta_x)}} \exp\left[mQ_0(\theta_x) - x\theta_x\right], \text{ for } \theta_x \in \Delta^\circ \text{ satisfying } mQ_0'(\theta_x) = x. \tag{105}$$

LEMMA 2. *For the i.i.d. sum model* (30) *with* $m \geq 1$ *summands, suppose Assumptions 1 and 2 hold. Then for all* m *sufficiently large,*

$$f_m(\xi_m) = \breve{f}_m(\xi_m)\left[1 + O\left(\frac{1}{m}\right)\right] = \frac{\exp\left[mQ_0(\theta_m) - m\theta_m Q_0'(\theta_m)\right]}{\sqrt{2\pi m Q_0''(\theta_\star)}}\left[1 + o\left(\frac{1}{\sqrt{m}}\right)\right] \tag{106}$$

$$= \frac{e^{-\beta m}}{\sqrt{2\pi m Q_0''(\theta_\star)}} \Upsilon_m[1 + o(1/\sqrt{m})] = \Theta(m^{-1/2} e^{-\beta m} \Upsilon_m), \tag{107}$$

where θ_m *is from* (94) *and* $\Upsilon_m > 0$ *satisfies* $\Upsilon_m = e^{o(\sqrt{m})}$ *as* $m \to \infty$*, so for each* $t \in \Re$ *and each* $c_1 > 0$*,*

$$\Upsilon_m^t = e^{o(\sqrt{m})} = o(e^{c_1\sqrt{m}}) \quad \text{and} \quad \Upsilon_m^t = \omega(e^{-c_1\sqrt{m}}) \quad \text{as} \quad m \to \infty. \tag{108}$$

Proof. Assume throughout the proof that m is sufficiently large so that (94) holds. For each x such that there exists $\theta = \theta_x \in \Delta^\circ$ with $mQ_0'(\theta_x) = x$, eq. (2.2.4) of Jensen (1995) expresses the (relative) error of $\breve{f}_m(x)$ in (105) in terms of $\theta = \theta_x$, specifically, the cumulants of the twisted CDF $\widetilde{G}_{0,\theta_x}$ in (33). Proposition 2.3.1 and Lemmas 2.3.3–2.3.5 of Jensen (1995) establish that the error in the saddlepoint approximation is uniform for $\theta \in \Psi$ under our conditions $0 \in \Delta^\circ$ and (31) in Assumptions 1 and 2, so (94) secures the first equality in (106).

To get the second equality of (106), use (105) with $x = \xi_m = mQ_0'(\theta_m)$ by (94) to arrive at

$$f_m(\xi_m) = \frac{\exp\left[mQ_0(\theta_m) - m\theta_m Q_0'(\theta_m)\right]}{\sqrt{2\pi m Q_0''(\theta_m)}}\left[1 + O\left(\frac{1}{m}\right)\right]. \tag{109}$$

In the denominator, $Q_0''(\cdot)$ has continuous derivative $Q_0'''(\cdot)$ on Ψ by (89), so for each m sufficiently large, the mean-value theorem ensures that there exists $\theta_{m,\star}$ between $\theta_\star$ and θ_m, all in Ψ° by (88), such that

$$Q_0''(\theta_m) = Q_0''(\theta_\star) + (\theta_m - \theta_\star)Q_0'''(\theta_{m,\star}) = Q_0''(\theta_\star)\left[1 + (\theta_m - \theta_\star)\frac{Q_0'''(\theta_{m,\star})}{Q_0''(\theta_\star)}\right]$$

as $Q_0''(\theta_\star) > 0$ by (90). Now (95) implies that $\theta_m - \theta_\star = o(1/\sqrt{m})$ and $Q_0'''(\theta_{m,\star}) = Q_0'''(\theta_\star)[1 + o(1)]$ as $m \to \infty$ since $|\theta_{m,\star} - \theta_\star| < |\theta_m - \theta_\star| \to 0$, so $Q_0''(\theta_m) = Q_0''(\theta_\star)[1 + o(1/\sqrt{m})]$ in the denominator of (109). Putting this into (109) yields the the second relation of (106) as $1/\sqrt{1 + o(1/\sqrt{m})} = 1 + o(1/\sqrt{m})$.

To show (107), write the numerator of (106) as $e^{mh(\theta_m)}$ for $h(\theta) = Q_0(\theta) - \theta Q_0'(\theta)$, which satisfies $h(\theta_\star) = -\beta$ by (32) and has derivative $h'(\theta) \equiv \frac{\mathrm{d}}{\mathrm{d}\theta} h(\theta) = -\theta Q_0''(\theta)$ for all $\theta \in \Delta^\circ$ by (89). Now $\theta_\star, \theta_m \in \Psi \subset \Delta^\circ$ by (94), so (88) and the convexity of Ψ ensure the line segment connecting them lies in Ψ, on which $h(\cdot)$ is differentiable. Thus, by the mean-value theorem, there is some $\theta'_{m,\star}$ between between $\theta_\star$ and θ_m such that $h(\theta_m) = h(\theta_\star) + (\theta_m - \theta_\star) h'(\theta'_{m,\star}) = -\beta + (\theta_\star - \theta_m)\theta'_{m,\star} Q_0''(\theta'_{m,\star})$. Multiplying by m, exponentiating, and writing $m = \sqrt{m}\sqrt{m}$ to apply (95) yields

$$e^{mh(\theta_m)} = \exp\Big(-\beta m\Big) \exp\Big(\sqrt{m}\Big[\sqrt{m}(\theta_\star - \theta_m)\Big]\theta'_{m,\star} Q_0''(\theta'_{m,\star})\Big) \equiv e^{-\beta m}\Upsilon_m. \tag{110}$$

Now $\sqrt{m}(\theta_\star - \theta_m) = o(1)$ as $m \to \infty$ by (95), implying $\theta'_{m,\star} = \theta_\star[1 + o(1/\sqrt{m})] > 0$ for all m sufficiently large because $|\theta_\star - \theta'_{m,\star}| \le |\theta_\star - \theta_m| = o(1/\sqrt{m})$ with $\theta_\star > 0$ by (32). Hence, we get $\Upsilon_m = e^{o(\sqrt{m})} > 0$ from the continuity of $Q_0''(\cdot) > 0$ on $\Psi \subset \Delta^\circ$ by (89) and (90), so (108) follows. Therefore, (107) holds by (106) and (110). $\square$

The next result establishes asymptotic upper bounds for

$$\vartheta_{j,k,m} \equiv E_G\Big[I(c(\mathbf{X}) > \xi_m) c^j(\mathbf{X}) L^k_{\theta_\star}(\mathbf{X})\Big] \tag{111}$$

for $j, k \in \{0, 1\}$. To see how $\vartheta_{j,k,m}$ arises, note that the asymptotic variance of the IS($\theta_\star$) estimator of the p_m-quantile ξ_m is $\kappa^2_{\mathrm{IS}(\theta_\star)} = \chi^2_{\mathrm{IS}(\theta_\star)}/f^2_m(\xi_m)$ in (40), with $\chi^2_{\mathrm{IS}(\theta_\star)} = \mathrm{Var}_{\widetilde{G}_{\theta_\star}}[I(c(\mathbf{X}) > \xi_m) L_{\theta_\star}(\mathbf{X})]$ by (19) and Theorem 7(ii). The numerator $\chi^2_{\mathrm{IS}(\theta_\star)} = E_{\widetilde{G}_{\theta_\star}}[I(c(\mathbf{X}) > \xi_m) L^2_{\theta_\star}(\mathbf{X})] - (1-p_m)^2 = E_G[I(c(\mathbf{X}) > \xi_m) L_{\theta_\star}(\mathbf{X})] - (1-p_m)^2$ by a change of measure, so $\vartheta_{0,1,m}$ is the second moment in the variance in the numerator. Further dividing by $f^2_m(\xi_m)$ as in $\kappa^2_{\mathrm{IS}(\theta_\star)}$ motivates studying (113) below. Also, for $j = 1$, $\vartheta_{1,k,m}$ with $k = 0$ (resp., $k = 1$) corresponds to the first term in the SRS (resp., IS($\theta_\star$)) covariance term γ_{SRS} in (10) (resp., $\gamma_{\mathrm{IS}(\theta_\star)}$ as in (20)), which are also further divided by $f_m(\xi_m)$, as in (113).

Lemma 3. *Under Assumptions 1 and 2 for the i.i.d. sum model* (30) *with* $m \ge 1$ *summands,* $\vartheta_{j,k,m}$ *in* (111) *satisfies the following for* $j, k \in \{0, 1\}$ *as* $m \to \infty$*:*

$$\vartheta_{j,k,m} = O\Big(m^j e^{-(k+1)\beta m}\Big) e^{o(\sqrt{m})}, \tag{112}$$

$$\frac{\vartheta_{j,k,m}}{f^{2-j}_m(\xi_m)} = O\Big(m^{(2+j)/2} e^{(1-j-k)\beta m}\Big)\Xi_m, \tag{113}$$

where

$$\Xi_m = \begin{cases} 1 & \text{if } \ j = 1 - k, \\ e^{o(\sqrt{m})} & \text{if } \ j = k. \end{cases} \tag{114}$$

Proof. We first establish the asymptotic upper bounds in (112). Note that for all m sufficiently large,

$$\begin{aligned}
\big|\vartheta_{j,k,m}\big| &\le E_G\big[I(c(\mathbf{X}) > \xi_m)\,\big|c^j(\mathbf{X})\big|\,L^k_{\theta_\star}(\mathbf{X})\big] = E_{\widetilde{G}_{\theta_\star}}\big[I(c(\mathbf{X}) > \xi_m)\,\big|c^j(\mathbf{X})\big|\,L^{k+1}_{\theta_\star}(\mathbf{X})\big] \\
&= E_{\widetilde{G}_{\theta_\star}}\Big[I(c(\mathbf{X}) > \xi_m)\,\big|c^j(\mathbf{X})\big| \exp\Big((k+1)\big[mQ_0(\theta_\star) - \theta_\star c(\mathbf{X})\big]\Big)\Big] \\
&\le \exp\Big[(k+1)m\Big(Q_0(\theta_\star) - \theta_\star Q_0'(\theta_m)\Big)\Big] E_{\widetilde{G}_{\theta_\star}}\big[I(c(\mathbf{X}) > \xi_m)\,\big|c^j(\mathbf{X})\big|\,\big] \equiv d_{1,k,m} d_{2,j,m},
\end{aligned} \tag{115}$$

where the second step applies a change of measure, the third step employs (34), and the last step uses $\xi_m = mQ_0'(\theta_m)$ by (94) and $\theta_\star > 0$ by (32). For $d_{2,j,m}$, the Cauchy-Schwarz inequality and $I^2(\cdot) \le 1$ imply

$$\begin{aligned} d_{2,j,m} &\le \left(E_{\widetilde{G}_{\theta_\star}} \left[I^2(c(\mathbf{X}) > \xi_m) \right] E_{\widetilde{G}_{\theta_\star}} \left[c^{2j}(\mathbf{X}) \right] \right)^{1/2} \\ &\le \left(E_{\widetilde{G}_{\theta_\star}} \left[c^{2j}(\mathbf{X}) \right] \right)^{1/2} \equiv d_{3,j,m} = \begin{cases} 1 & \text{for } j = 0, \\ \left(mQ_0''(\theta_\star) + \left[mQ_0'(\theta_\star) \right]^2 \right)^{1/2} & \text{for } j = 1, \end{cases} \end{aligned} \tag{116}$$

where the case $j = 1$ follows from (96)–(97) because $\mathrm{Var}_{\widetilde{G}_{\theta_\star}}[c(\mathbf{X})] = mQ_0''(\theta_\star)$ as $X_1, \ldots, X_m$ are i.i.d. with mean $Q_0'(\theta_\star)$ under $\widetilde{G}_{\theta_\star}$. Note that $d_{3,j,m} = O(m^j)$ for $j \in \{0, 1\}$. For $d_{1,k,m}$ in (115) and Ψ in (94), $Q_0'(\theta)$ is differentiable for all $\theta \in \Psi \subset \Delta^\circ$ by (89), where (94) ensures that Ψ contains $\theta_\star$, θ_m (for all m sufficiently large), and the line segment connecting them. Then by the mean-value theorem, there exists $\theta_{m,\star}$ between $\theta_\star$ and θ_m such that $Q_0'(\theta_m) = Q_0'(\theta_\star) + (\theta_m - \theta_\star) Q_0''(\theta_{m,\star})$. Thus, in (115), for all sufficiently large m,

$$d_{1,k,m} = \exp\left((k+1)m \left[Q_0(\theta_\star) - \theta_\star Q_0'(\theta_\star) \right] \right) \exp\left((k+1)m\theta_\star(\theta_\star - \theta_m) Q_0''(\theta_{m,\star}) \right) = e^{-(k+1)\beta m} d_{4,k,m} \tag{117}$$

by (32), where

$$d_{4,k,m} = \exp\left((k+1)\sqrt{m} \left[\sqrt{m}(\theta_\star - \theta_m) \right] \theta_\star Q_0''(\theta_{m,\star}) \right) = e^{o(\sqrt{m})} \tag{118}$$

as $m \to \infty$ by (95) and the facts that $\theta_{m,\star} \to \theta_\star > 0$ and $Q_0''(\cdot) > 0$ is continuous at $\theta_\star$ by (89) and (90). Using (116)–(118) in (115) gives $\left| \vartheta_{j,k,m} \right| \le e^{-(k+1)\beta m} e^{o(\sqrt{m})} d_{3,j,m}$ for all sufficiently large m, verifying (112) because $d_{3,j,m} = O(m^j)$ for $j \in \{0, 1\}$. Multiplying (112) by $1/f_m^{2-j}(\xi_m) = \Theta(m^{(2-j)/2} e^{(2-j)\beta m}) / \Upsilon_m^{2-j}$ from (107), with $\Upsilon_m = e^{o(\sqrt{m})}$ from (108), yields (113) for the case that $j = k$ in (114) as $e^{o(\sqrt{m})}/\Upsilon_m^{2-j} = e^{o(\sqrt{m}) - o(\sqrt{m})} = e^{o(\sqrt{m})}$.

Next we establish (113) when $j = 1 - k$ in (114), in which case $d_{1,k,m}$ in (115) has $k + 1 = 2 - j$. For $f_m(\xi_m)$ given by (106) and θ_m in (94) for all m sufficiently large, dividing $d_{1,k,m}$ by $f_m^{2-j}(\xi_m)$ leads to

$$\frac{d_{1,k,m}}{f_m^{2-j}(\xi_m)} = \left[2\pi m Q_0''(\theta_\star) \right]^{(2-j)/2} \exp\left((2-j)m \left[\left[Q_0(\theta_\star) - Q_0(\theta_m) \right] - Q_0'(\theta_m)(\theta_\star - \theta_m) \right] \right) \left[1 + o\Big(\frac{1}{\sqrt{m}} \Big) \right]^{j-2}. \tag{119}$$

For the exponential term in (119), Taylor's theorem with Lagrange remainder gives $Q_0(\theta_\star) - Q_0(\theta_m) = (\theta_\star - \theta_m) Q_0'(\theta_m) + \frac{1}{2}(\theta_\star - \theta_m)^2 Q_0''(\theta_{m,\star})$ for some $\theta_{m,\star}$ between $\theta_\star$ and θ_m, with $\theta_m, \theta_\star, \theta_{m,\star} \in \Psi \subset \Delta^\circ$ for all m sufficiently large by (94) and the convexity of Ψ. Combining with the rest of the exponent in (119) yields

$$\exp\left((2-j)m \left[\left[Q_0(\theta_\star) - Q_0(\theta_m) \right] - Q_0'(\theta_m)(\theta_\star - \theta_m) \right] \right) = \exp\left[\frac{2-j}{2} m (\theta_\star - \theta_m)^2 Q_0''(\theta_{m,\star}) \right] = 1 + o(1) \tag{120}$$

as $m \to \infty$ by (95) and because $Q_0''(\theta_{m,\star}) = Q_0''(\theta_\star)\,[1 + o(1)]$ by the continuity of Q_0'' on $\Psi \subset \Delta^\circ$ from (89) and $|\theta_{m,\star} - \theta_\star| \le |\theta_m - \theta_\star| \to 0$ as $m \to \infty$ by (95). Thus, putting (120) into (119) and using (115) and (116) yield

$$\frac{|\vartheta_{j,k,m}|}{f_m^{2-j}(\xi_m)} \le \frac{d_{1,k,m} d_{3,j,m}}{f_m^{2-j}(\xi_m)} = \left[2\pi m Q_0''(\theta_\star) \right]^{(2-j)/2} \left[1 + o(1) \right] d_{3,j,m} \tag{121}$$

as $m \to \infty$, verifying (113) for the case that $j = 1 - k$ in (114) because $d_{3,j,m} = O(m^j)$ for $j \in \{0, 1\}$. □

F.4. Proof of Theorem 6

Part (i): SRS. When we apply SRS, (66) holds because $c(\mathbf{X}) = \sum_{j=1}^m X_j$, where $X_1, X_2, \ldots, X_m$ are i.i.d. with variance $\sigma_0^2 \in (0, \infty)$ by (87). As $\mu = m\mu_0$ by (86), (67) follows, where the WNRE result uses in (37) that the expected CPU time (end of Section 6.1) to generate a single SRS output $c(\mathbf{X})$ is $m\tau_{\mathrm{SRS}}$ for a constant $\tau_{\mathrm{SRS}} \in (0, \infty)$, proving part (i).

Part (ii): IS(θ). Now consider IS(θ) with $\widetilde{G}_\theta$ as described in Section 6.2, with $\theta \neq 0$ and $\pm\theta \in \Delta^\circ$, as assumed. By (11) and a change of measure, we can write the variance in (68) as

$$\sigma^2_{\mathrm{IS}(\theta)} = E_{\widetilde{G}_\theta}[c^2(\mathbf{X}) L^2_\theta(\mathbf{X})] - \mu^2 = E_G[c^2(\mathbf{X}) L_\theta(\mathbf{X})] - (m\mu_0)^2,$$

giving the second term of (68). By (34), the second-moment term becomes

$$E_G[c^2(\mathbf{X}) L_\theta(\mathbf{X})] = [M_0(\theta)]^m E_G\left[c^2(\mathbf{X}) e^{-\theta c(\mathbf{X})}\right], \tag{122}$$

which we next show equals the first term in (68) via derivatives of the MGFs and CGFs of $X_j \sim G_0$ and $c(\mathbf{X}) \sim F_m$. Let $M_{F_m}(\theta) = E_G[e^{\theta c(\mathbf{X})}]$, $\theta \in \Re$, be the MGF of $Y = c(\mathbf{X}) = \sum_{j=1}^m X_j$. As the components of $\mathbf{X}$ are i.i.d., we have

$$M_{F_m}(\theta) = E_G\left[\prod_{j=1}^m e^{\theta X_j}\right] = \prod_{j=1}^m E_0\left[e^{\theta X_j}\right] = [M_0(\theta)]^m, \tag{123}$$

so $M_{F_m}(\theta) < \infty$ for $\theta \in \Delta^\circ$. We assumed that $\pm\theta \in \Delta^\circ$, in which case $M_0(\theta)$ and $M_0(-\theta)$ have derivatives of all orders by (89), and the same holds for $M_{F_m}(\theta)$ and $M_{F_m}(-\theta)$ by (123). The second derivative of M_{F_m} satisfies $M''_{F_m}(\theta) = E_G\left[\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} e^{\theta c(\mathbf{X})}\right] = E_G[c^2(\mathbf{X}) e^{\theta c(\mathbf{X})}]$, so (122) becomes

$$E_G[c^2(\mathbf{X}) L_\theta(\mathbf{X})] = [M_0(\theta)]^m M''_{F_m}(-\theta). \tag{124}$$

Next we use (123) to express the first derivative $M'_{F_m}(\theta) = m[M_0(\theta)]^{m-1} M'_0(\theta)$ and

$$\begin{aligned} M''_{F_m}(\theta) &= m(m-1)[M_0(\theta)]^{m-2}[M'_0(\theta)]^2 + m[M_0(\theta)]^{m-1} M''_0(\theta) \\ &= m[M_0(\theta)]^m \left[(m-1)\left(\frac{M'_0(\theta)}{M_0(\theta)}\right)^2 + \frac{M''_0(\theta)}{M_0(\theta)}\right] \\ &= m[M_0(\theta)]^m \left[m[Q'_0(\theta)]^2 + Q''_0(\theta)\right] \end{aligned} \tag{125}$$

by (96) and (97). As a consequence, substituting (125) in (124) yields

$$E_G[c^2(\mathbf{X}) L_\theta(\mathbf{X})] = m[M_0(\theta) M_0(-\theta)]^m \left[m[Q'_0(-\theta)]^2 + Q''_0(-\theta)\right],$$

giving the first term of (68). (When $\theta = 0$, which corresponds to SRS, we have $\widetilde{G}_\theta = G$ and $M_0(0) = \alpha(0) = 1$. Also, $M'_0(0) = \mu_0$ and $M''_0(0) = E_0[X_j^2]$, so (68) equals (66) by (96)–(97).)

For the results in (69), the second one follows from (90) because we assumed that $-\theta \in \Delta^\circ$. We next show $\alpha(\theta) \equiv M_0(\theta) M_0(-\theta) \in (1, \infty)$ in (69) for all $\theta \neq 0$ with $\pm\theta \in \Delta^\circ$, which ensures that $\alpha(\theta) < \infty$. Also, (90) implies that $Q_0(\theta)$ is strictly convex on Δ°; also see, e.g., (Durrett 1996, p. 73). Hence, Jensen's inequality yields

$$\alpha(\theta) = \exp\left[2\left(\frac{1}{2} Q_0(\theta) + \frac{1}{2} Q_0(-\theta)\right)\right] > \exp\left[2 Q_0\left(\frac{1}{2}\theta - \frac{1}{2}\theta\right)\right] = e^{2Q_0(0)} = 1 \tag{126}$$

because $Q_0(0) = 0$, proving the first result of (69).

We next verify (70) for any $\theta \neq 0$ with $\pm\theta \in \Delta^\circ$, so $\alpha(\theta) > 1$ by (126). The asymptotic upper bound in (70) follows immediately from (68). It is possible in (68) to have $Q'_0(-\theta) = 0$, but $Q''_0(-\theta) > 0$ by (90), so $\sigma^2_{\mathrm{IS}(\theta)} \geq m Q''_0(-\theta)[\alpha(\theta)]^m - m^2 \mu_0^2$, securing the asymptotic lower bound in (70). Moreover, we have that $\mu = m\mu_0$. Thus, as $m \to \infty$, the RE of the IS(θ) estimator of μ with $\theta \neq 0$ asymptotically grows at rate (ignoring constants) at least $[\alpha(\theta)]^{m/2}/\sqrt{m}$ (and at faster rate $[\alpha(\theta)]^{m/2}$ when $M_0'(-\theta) \neq 0$ because then (96) implies $[Q_0'(-\theta)]^2 > 0$ in (68)), establishing the RE result in (71). Similarly, the WNRE result in (71) follows from (37) by multiplying the same lower bound for $\mathrm{Var}_{\widetilde{G}_\theta}[c(\mathbf{X})L_\theta(\mathbf{X})]$ by the expected computation time $m\tau_{\mathrm{IS}(\theta)}$ with constant $\tau_{\mathrm{IS}(\theta)} \in (0,\infty)$ to generate $(c(\mathbf{X}), L_\theta(\mathbf{X}))$ under $\mathbf{X} \sim \widetilde{G}_\theta$ (Section 6.2).

For $\theta = \theta_\star > 0$ in Assumption 2 as the root of (32) with $\pm\theta_\star \in \Delta^\circ$, it is clear that $\theta_\star$ remains fixed as m grows because the marginal CDF G_0 does not vary with m and β is a constant in (29). Moreover, (71) holds when $\theta = \theta_\star$ because $\theta_\star > 0$ so $\alpha(\theta_\star) > 1$ by (126), completing the proof of part (ii).

Part (iii): ISDM(θ). We apply arguments from Hesterberg (1995) to establish (72) for any $\theta \in \Delta^\circ$. Removing either term in the denominator of $L_{\mathrm{ISDM}(\theta)}(\mathbf{X}) = \frac{\mathrm{d}G(\mathbf{X})}{\delta \mathrm{d}\widetilde{G}_\theta(\mathbf{X}) + (1-\delta)\mathrm{d}G(\mathbf{X})}$ yields

$$L_{\mathrm{ISDM}(\theta)}(\mathbf{X}) \le \frac{L_\theta(\mathbf{X})}{\delta} \quad \text{and} \quad L_{\mathrm{ISDM}(\theta)}(\mathbf{X}) \le \frac{1}{1-\delta}. \tag{127}$$

Then use a change of measure and exploit the second relation in (127) to get

$$\begin{aligned}
\sigma^2_{\mathrm{ISDM}(\theta)} &= E_{\widetilde{G}_{\mathrm{ISDM}(\theta)}}\left[c^2(\mathbf{X})L^2_{\mathrm{ISDM}(\theta)}\right] - \mu^2 = E_G\left[c^2(\mathbf{X})L_{\mathrm{ISDM}(\theta)}\right] - \mu^2 &(128)\\
&\le \frac{1}{1-\delta}E_G\left[c^2(\mathbf{X})\right] - \mu^2 = \frac{1}{1-\delta}\left[\sigma^2_{\mathrm{SRS}} + \mu^2\right] - \mu^2 = \frac{\delta\mu_0^2}{1-\delta}m^2 + \frac{\sigma_0^2}{1-\delta}m &(129)
\end{aligned}$$

by (66) and because $\mu = m\mu_0$ by (86), proving (72) (previously shown in Hesterberg (1995)).

To establish (73) when $\mu_0 \neq 0$ and $\theta \neq 0$, we will show that $\sigma^2_{\mathrm{ISDM}(\theta)}$ is asymptotically bounded below by the m^2 term of the upper bound in (129). By (128), we have that $\sigma^2_{\mathrm{ISDM}(\theta)} = \nu^\#_{\mathrm{ISDM}(\theta),m} - \mu_0^2 m^2$ for $\nu^\#_{\mathrm{ISDM}(\theta),m} \equiv E_G\left[c^2(\mathbf{X})L_{\mathrm{ISDM}(\theta)}\right]$, on which we now focus. Recall that $\delta \in (0,1)$ is fixed, and consider any $\delta_0 \in (0,\delta)$. Also, consider any $\epsilon \in (0, |\mu_0|)$, and without loss of generality, assume $\mu_0 > 0$ and $\theta > 0$. Then

$$\begin{aligned}
\nu^\#_{\mathrm{ISDM}(\theta),m} &= E_G\left[c^2(\mathbf{X})\frac{\mathrm{d}G(\mathbf{X})}{\delta\mathrm{d}\widetilde{G}_\theta(\mathbf{X}) + (1-\delta)\mathrm{d}G(\mathbf{X})}\right] = E_G\left[c^2(\mathbf{X})\frac{1}{\delta\frac{\mathrm{d}\widetilde{G}_\theta(\mathbf{X})}{\mathrm{d}G(\mathbf{X})} + 1 - \delta}\right]\\
&= E_G\left[c^2(\mathbf{X})\frac{1}{\delta\exp[-mQ_0(\theta) + c(\mathbf{X})\theta] + 1 - \delta}\right]\\
&\ge E_G\left[c^2(\mathbf{X})\frac{I(c(\mathbf{X}) \in [m(\mu_0-\epsilon), m(\mu_0+\epsilon)])}{\delta\exp[-mQ_0(\theta) + c(\mathbf{X})\theta] + 1 - \delta}\right]\\
&\ge m^2(\mu_0-\epsilon)^2 E_G\left[\frac{I(c(\mathbf{X}) \in [m(\mu_0-\epsilon), m(\mu_0+\epsilon)])}{\delta\exp[-mQ_0(\theta) + c(\mathbf{X})\theta] + 1 - \delta}\right]\\
&\ge \frac{m^2(\mu_0-\epsilon)^2}{\delta\exp(-m[Q_0(\theta) - (\mu_0+\epsilon)\theta]) + 1 - \delta} P_G\left(c(\mathbf{X}) \in [m(\mu_0-\epsilon), m(\mu_0+\epsilon)]\right), \qquad (130)
\end{aligned}$$

where the second equality holds as $\widetilde{G}_\theta$ and G are absolutely continuous with respect to each other, the third step follows from (34), and the last step exploits the assumed $\theta > 0$.

For the exponential term in (130), the strict convexity of Q_0 by (90) along with $Q_0(0) = 0$, $Q_0'(0) = \mu_0$, and $\theta > 0$ ensures that $Q_0(\theta) > (\mu_0 + \epsilon)\theta$ for all sufficiently small $\epsilon > 0$. Thus, the exponential term in (130) vanishes as $m \to \infty$ for all small enough $\epsilon > 0$. Moreover, $\delta_0 \in (0,\delta)$ implies that $\frac{(\mu_0-\epsilon)^2}{1-\delta} > \frac{\mu_0^2}{1-\delta_0}$ for all sufficiently small $\epsilon > 0$. Using these results in (130) yields that the ratio (without the m^2) satisfies

$$\liminf_{m\to\infty} \frac{(\mu_0-\epsilon)^2}{\delta\exp\big(-m[Q_0(\theta) - (\mu_0+\epsilon)\theta]\big) + 1 - \delta} \ge \frac{(\mu_0-\epsilon)^2}{1-\delta} > \frac{\mu_0^2}{1-\delta_0} \tag{131}$$

for all sufficiently small $\epsilon > 0$.

For the probability term in (130), the weak law of large numbers (under G) ensures that for any $\epsilon > 0$, $P_G\left(c(\mathbf{X}) \in [m(\mu_0 - \epsilon), m(\mu_0 + \epsilon)]\right) \to 1$ as $m \to \infty$. Putting this and (131) into (130) implies that the second-moment term in (128) satisfies $\liminf_{m\to\infty} \nu^{\#}_{\mathrm{ISDM}(\theta),m}/m^2 > \mu_0^2/(1-\delta_0)$ as $m \to \infty$ since $\epsilon > 0$ was arbitrary (and small). For the full variance $\sigma^2_{\mathrm{ISDM}(\theta)} = \nu^{\#}_{\mathrm{ISDM}(\theta),m} - \mu_0^2 m^2$, it then follows from (129) that

$$\frac{\delta\mu_0^2}{1-\delta} = \lim_{m\to\infty} \frac{1}{m^2}\left[\frac{\delta\mu_0^2}{1-\delta}m^2 + \frac{\sigma_0^2}{1-\delta}m\right] \geq \limsup_{m\to\infty} \frac{\sigma^2_{\mathrm{ISDM}(\theta)}}{m^2} \geq \liminf_{m\to\infty} \frac{\sigma^2_{\mathrm{ISDM}(\theta)}}{m^2} > \frac{\mu_0^2}{1-\delta_0} - \mu_0^2 = \frac{\delta_0\mu_0^2}{1-\delta_0}. \quad (132)$$

As $\delta_0 \in (0, \delta)$ can be taken arbitrarily close to δ from below in (132), we see that $\lim_{m\to\infty} \frac{\sigma^2_{\mathrm{ISDM}(\theta)}}{m^2} = \frac{\delta\mu_0^2}{(1-\delta)}$, verifying (73) when $\mu_0 > 0$ and $\theta > 0$. The cases when $\mu_0 < 0$ or $\theta < 0$ (or both) can be handled similarly. Thus, for any $\mu_0 \neq 0$ and $\theta \neq 0$, (74) then follows from (86), where the WNRE result uses in (37) that the expected computation time to generate $(c(\mathbf{X}), L_{\mathrm{ISDM}(\theta)}(\mathbf{X}))$ for $\mathbf{X} \sim \widetilde{G}_{\mathrm{ISDM}(\theta)}$ as in (35) is $m\tau_{\mathrm{ISDM}(\theta)}$ with constant $\tau_{\mathrm{ISDM}(\theta)} \in (0, \infty)$ (Section 6.2).

When $\mu_0 = 0$, the bound in (72) shows that $\sigma^2_{\mathrm{ISDM}(\theta_\star)} = O(m)$ as $m \to \infty$, verifying the last statement of part (iii).

Part (iv): DE(θ)**.** Note (75) holds by (66) and (70), and (76) easily follows, also using (38). □

F.5. Proof of Theorem 7

First note that (90) implies $Q_0''(\theta_\star) > 0$ because $\theta_\star \in \Delta^\circ$ by (32). Also, for the WNRE in (37) of estimators of ξ_m, the end of Section 6.1 specifies that the expected CPU time to generate a single SRS output $c(\mathbf{X})$ is $m\tau_{\mathrm{SRS}}$ for a constant $\tau_{\mathrm{SRS}} \in (0, \infty)$. Similarly, Section 6.2 stipulates that for IS($\theta_\star$) (resp., ISDM($\theta_\star$)), the expected CPU time to generate $(c(\mathbf{X}), L_{\theta_\star}(\mathbf{X}))$ under $\mathbf{X} \sim \widetilde{G}_{\theta_\star}$ (resp., $(c(\mathbf{X}), L_{\mathrm{ISDM}(\theta_\star)}(\mathbf{X}))$ for $\mathbf{X} \sim \widetilde{G}_{\mathrm{ISDM}(\theta_\star)}$ as in (35)) is $m\tau_{\mathrm{IS}(\theta_\star)}$ with constant $\tau_{\mathrm{IS}(\theta_\star)} \in (0, \infty)$ (resp., $m\tau_{\mathrm{ISDM}(\theta_\star)}$ with constant $\tau_{\mathrm{ISDM}(\theta_\star)} \in (0, \infty)$). Also, (38) defines the form of the WNRE for the DE($\theta_\star$) estimator of ξ_m.

Part (i): SRS. The numerator of (40) is $\chi^2_{\mathrm{SRS}} = (1 - e^{-\beta m})e^{-\beta m}$ by (10) and (29). Using the square of (107) in the denominator with $\Upsilon_m^{-2} = e^{o(\sqrt{m})}$ as $m \to \infty$ by (108) gives the first two steps of (77). Taking logs and dividing by m shows $\lim_{m\to\infty} \frac{1}{m} \ln \kappa^2_{\mathrm{SRS}} = \beta$. Also, (108) ensures $\Upsilon_m^t = \omega(e^{-\sqrt{m}})$ as $m \to \infty$ for each $t \in \Re$, so $\kappa^2_{\mathrm{SRS}} = \omega(me^{\beta m - \sqrt{m}})$, securing the last part of (77). Finally, (78) holds by (92) because $Q_0'(\theta_\star) \neq 0$ was assumed.

Part (ii): IS($\theta_\star$)**.** We have that $\kappa^2_{\mathrm{IS}(\theta_\star)} = \chi^2_{\mathrm{IS}(\theta_\star)}/f_m^2(\xi_m)$, where

$$\begin{aligned}\chi^2_{\mathrm{IS}(\theta_\star)} &= \mathrm{Var}_{\widetilde{G}_{\theta_\star}}\Big[I(c(\mathbf{X}) > \xi_m)L_{\theta_\star}(\mathbf{X})\Big] = E_{\widetilde{G}_{\theta_\star}}\Big[I(c(\mathbf{X}) > \xi_m)L^2_{\theta_\star}(\mathbf{X})\Big] - (1-p)^2 \\ &\leq E_{\widetilde{G}_{\theta_\star}}\Big[I(c(\mathbf{X}) > \xi_m)L^2_{\theta_\star}(\mathbf{X})\Big] = E_G\Big[I(c(\mathbf{X}) > \xi_m)L_{\theta_\star}(\mathbf{X})\Big] = \vartheta_{0,1,m}\end{aligned}$$

by a change of measure and (111) for $j = 0$ and $k = 1$. Thus, (116) for $j = 0$ and (121) yield (79), and combining (79) with (92) establishes (80), completing the proof of (ii).

Part (iii): ISDM($\theta_\star$)**.** From the first relation in (127) and (34), we use the fact that $\theta_\star > 0$ from (32) to bound the numerator of (40) for all m sufficiently large as

$$\begin{aligned}\chi^2_{\mathrm{ISDM}(\theta_\star)} &\leq E_{\widetilde{G}_{\mathrm{ISDM}(\theta_\star)}}\left[L^2_{\mathrm{ISDM}(\theta_\star)}(\mathbf{X})\, I(c(\mathbf{X}) > \xi_m)\right] \\ &\leq \frac{1}{\delta^2} E_{\widetilde{G}_{\mathrm{ISDM}(\theta_\star)}}\Big[\exp\Big(2\big[mQ_0(\theta_\star) - \theta_\star c(\mathbf{X})\big]\Big)\, I(c(\mathbf{X}) > \xi_m)\Big] \\ &\leq \frac{1}{\delta^2}\exp\Big(2\big[mQ_0(\theta_\star) - \theta_\star \xi_m\big]\Big) = \frac{1}{\delta^2}\exp\Big(2m\big[Q_0(\theta_\star) - \theta_\star Q_0'(\theta_m)\big]\Big) \end{aligned} \quad (133)$$

by (94), where (133) equals $1/\delta^2$ times the bound for $\vartheta_{0,1,m}$ in (115) for $j=0$ and $k=1$ by (116). Thus, arguing as in the proof of (121) for $j=0$ and $k=1$ also shows (81) holds. Finally, (82) follows by (92).

Part (iv): DE($\theta_\star$)**.** Note that (83) is a simple consequence of (77) and (79), and (84) follows by (92). $\square$

F.6. Proof of Theorem 5

Before proving Theorem 5, we first explicitly define the WNREs for the MSIS($\theta_\star$) estimator and the DE($\theta_\star$) estimator of η. By similar reasoning used to define $\mathrm{WNRE}_{\mathrm{DE}(\theta),m}[\varphi]$ for $\varphi=\xi$ and μ in (38), we let

$$\begin{aligned}&\mathrm{WNRE}_{\mathrm{DE}(\theta_\star),m}[\eta]\\ &=\frac{1}{|\eta_m|}\left[\frac{m\tau_{\mathrm{IS}(\theta_\star)}}{\delta}\left[\upsilon_1^2\kappa^2_{\mathrm{IS}(\theta_\star)}+\upsilon_2^2\sigma^2_{\mathrm{IS}(\theta_\star)}-2\upsilon_1\upsilon_2\frac{\gamma_{\mathrm{IS}(\theta_\star)}}{f_m(\xi_m)}\right]+\frac{m\tau_{\mathrm{SRS}}}{1-\delta}\left[\upsilon_1'^2\kappa^2_{\mathrm{SRS}}+\upsilon_2'^2\sigma^2_{\mathrm{SRS}}-2\upsilon_1'\upsilon_2'\frac{\gamma_{\mathrm{SRS}}}{f_m(\xi_m)}\right]\right]^{1/2},\end{aligned} \tag{134}$$

where the allocation parameter $\delta\in(0,1)$ splits the CPU budget between IS($\theta_\star$) and SRS, and Sections 6.1 and 6.2 specify the constants $\tau_{\mathrm{SRS}},\tau_{\mathrm{IS}(\theta_\star)}\in(0,\infty)$ related to generating a single output. Analogously, define

$$\mathrm{WNRE}_{\mathrm{MSIS}(\theta_\star),m}[\eta]=\frac{1}{|\eta_m|}\left[m\left(\frac{\tau_{\mathrm{IS}(\theta_\star)}\kappa^2_{\mathrm{IS}(\theta_\star)}}{\delta}+\frac{\tau_{\mathrm{SRS}}\sigma^2_{\mathrm{SRS}}}{1-\delta}\right)\right]^{1/2}. \tag{135}$$

The WNREs for the SRS, IS($\theta_\star$), and ISDM($\theta_\star$) estimators of η are defined as in (37).

F.6.1. Part (i): SRS.

To establish (42), we will separately analyze the three terms in (39) for $\mathfrak{M}=\mathrm{SRS}$, where (39) has $\Lambda_{\mathrm{SRS}}=\Lambda^\dagger_{\mathrm{SRS}}=1$ with χ^2_{SRS}, χ^2_{SRS}, and γ_{SRS} given in (9) and (10). We will show that as $m\to\infty$, the first term in (39) grows at a strictly faster rate than the other two terms, leading to

$$\zeta^2_{\mathrm{SRS}}=\frac{2\pi Q_0''(\theta_\star)\left[1+o\left(\frac{1}{\sqrt{m}}\right)\right]me^{\beta m}}{\Upsilon_m^2}=\frac{\Theta(me^{\beta m})}{\Upsilon_m^2}=\omega(me^{\beta m-\sqrt{m}}). \tag{136}$$

The first term in (39) is given by (77), with $\kappa^2_{\mathrm{SRS}}=\omega(me^{\beta m-\sqrt{m}})$ as $m\to\infty$. By (66), the second term in (39) equals $\sigma^2_{\mathrm{SRS}}=\sigma_0^2 m=o(\kappa^2_{\mathrm{SRS}})$ as $m\to\infty$, so $\sigma_{\mathrm{SRS}}=o(\kappa_{\mathrm{SRS}})$. For the third term in (39), use (9) and the Cauchy-Schwarz inequality to get $\left|-2\frac{\gamma_{\mathrm{SRS}}}{f_m(\xi_m)}\right|\le 2\kappa_{\mathrm{SRS}}\sigma_{\mathrm{SRS}}=o(\kappa^2_{\mathrm{SRS}})$ as $m\to\infty$. Thus, the second and third terms in (39) are asymptotically negligible compared to the first term, which verifies (136) and (42). Combining (42) with (93) establishes (43) because the expected computation time to generate $c(\mathbf{X})$ with $\mathbf{X}\sim G$ is $m\tau_{\mathrm{SRS}}$, which we use in (37), completing the proof of part (i).

F.6.2. Part (ii): IS($\theta_\star$).

To establish (44), we will show that

$$\zeta^2_{\mathrm{IS}(\theta_\star)}\ge Q_0''(-\theta_\star)\left[1+O(\alpha_\star^{-m})\right]m\alpha_\star^m=\Omega(m\alpha_\star^m). \tag{137}$$

We will accomplish this by separately analyzing the three terms in (39) for $\mathfrak{M}=\mathrm{IS}(\theta_\star)$ to show the second term grows at the strictly fastest rate, where we recall that $\Lambda_{\mathrm{IS}(\theta_\star)}=\Lambda^\dagger_{\mathrm{IS}(\theta_\star)}=1$, and $\chi^2_{\mathrm{IS}(\theta_\star)}$, $\chi^2_{\mathrm{IS}(\theta_\star)}$, and $\gamma_{\mathrm{IS}(\theta_\star)}$ are as in (19) and (20) with twisting parameter $\theta_\star\in\Delta^\circ$ from (32), where we also assumed $-\theta_\star\in\Delta^\circ$. The second term in (39) satisfies $\sigma^2_{\mathrm{IS}(\theta_\star)}=\Omega(m\alpha_\star^m)$ as $m\to\infty$ by (68) and (70), where $\theta_\star>0$ of (32) with $\pm\theta_\star\in\Delta^\circ$ implies that $\alpha_\star=\alpha(\theta_\star)>1$ by (69). For the first term in (39), we have by (79) that $\kappa^2_{\mathrm{IS}(\theta_\star)}=O(m)=o(\sigma^2_{\mathrm{IS}(\theta_\star)})$ as $m\to\infty$, so $\kappa_{\mathrm{IS}(\theta_\star)}=o(\sigma_{\mathrm{IS}(\theta_\star)})$. For the third term in (39), the Cauchy-Schwarz inequality implies $\left|-2\frac{\gamma_{\mathrm{IS}(\theta_\star)}}{f_m(\xi_m)}\right|\le 2\kappa_{\mathrm{IS}(\theta_\star)}\sigma_{\mathrm{IS}(\theta_\star)}=o(\sigma^2_{\mathrm{IS}(\theta_\star)})$ as $m\to\infty$. Thus, combining these results, including (68), yields (137) and (44). Using (44) with (93) verifies (45) as the expected time to generate $(c(\mathbf{X}),L_{\theta_\star}(\mathbf{X}))$ with $\mathbf{X}\sim\widetilde{G}_{\theta_\star}$ is $m\tau_{\mathrm{IS}(\theta_\star)}$ (Section 6.2), which we use in (37), completing the proof of part (ii).

F.6.3. Part (iii): $\mathrm{MSIS}(\theta_\star)$**.** By (23), $\mathfrak{M} = \mathrm{MSIS}(\theta_\star)$ leads to (39) having $\kappa^2_{\mathrm{MSIS}(\theta_\star)} = \kappa^2_{\mathrm{IS}(\theta_\star)}$, $\sigma^2_{\mathrm{MSIS}(\theta_\star)} = \sigma^2_{\mathrm{SRS}}$, $\gamma_{\mathrm{MSIS}(\theta_\star)} = 0$ (as ξ and μ are estimated independently), $\Lambda_{\mathrm{MSIS}(\theta_\star)} = 1/\delta$, and $\Lambda^\dagger_{\mathrm{MSIS}(\theta_\star)} = 1/(1-\delta)$. To prove (46), we separately analyze the two nonzero terms in (39) to show that each grows asymptotically at most linearly in m. The first term in (39) is $\kappa^2_{\mathrm{IS}(\theta_\star)}/\delta$, where (79) bounds $\kappa^2_{\mathrm{IS}(\theta_\star)} = O(m)$ and $\kappa^2_{\mathrm{IS}(\theta_\star)} > 0$ by (90). The second term in (39) is $\sigma^2_{\mathrm{SRS}}/(1-\delta) = [\sigma_0^2/(1-\delta)]m = \Theta(m)$ by (66), with $\sigma^2_{\mathrm{SRS}} > 0$ by (66). We then get

$$\zeta^2_{\mathrm{MSIS}(\theta_\star)} \le \left(\frac{2\pi Q_0''(\theta_\star)}{\delta} \big[1 + o(1)\big]^2 + \frac{\sigma_0^2}{1-\delta} \right) m, \tag{138}$$

verifying (46) since both terms in the large parentheses in (138) are positive. Combining (46) with (93) establishes the RE result in (47).

For the second result of (47), the boundedness of $\mathrm{WNRE}_{\mathrm{MSIS}(\theta_\star),m}[\eta]$ in (135) holds by putting (93), (79), and (66) into (135) and using the expected generation times $m\tau_{\mathrm{SRS}}$ and $m\tau_{\mathrm{IS}(\theta_\star)}$ for SRS and $\mathrm{IS}(\theta_\star)$, respectively, as specified in Sections 6.1 and 6.2.

F.6.4. Part (iv): $\mathrm{ISDM}(\theta_\star)$**.** The method $\mathfrak{M} = \mathrm{ISDM}(\theta_\star)$ is a special case of IS, sampling $\mathbf{X}$ as in (35) with $\theta = \theta_\star$, so (19) and (20) imply that (39) then has $\chi^2_{\mathrm{ISDM}(\theta_\star)} = \mathrm{Var}_{\widetilde{G}_{\mathrm{ISDM}(\theta_\star)}}[L_{\mathrm{ISDM}(\theta_\star)}(\mathbf{X}) I(c(\mathbf{X}) > \xi_m)]$, $\sigma^2_{\mathrm{ISDM}(\theta_\star)} = \mathrm{Var}_{\widetilde{G}_{\mathrm{ISDM}(\theta_\star)}}[c(\mathbf{X}) L_{\mathrm{ISDM}(\theta_\star)}(\mathbf{X})]$, $\gamma_{\mathrm{ISDM}(\theta_\star)} = \mathrm{Cov}_{\widetilde{G}_{\mathrm{ISDM}(\theta_\star)}}\big[I(c(\mathbf{X}) > \xi) L_{\mathrm{ISDM}(\theta_\star)}(\mathbf{X}), c(\mathbf{X}) L_{\mathrm{ISDM}(\theta_\star)}(\mathbf{X})\big]$, and $\Lambda_{\mathrm{ISDM}(\theta_\star)} = \Lambda^\dagger_{\mathrm{ISDM}(\theta_\star)} = 1$. To show (48), we will separately analyze the three terms in (39) to derive an upper bound for each. The first term $\kappa^2_{\mathrm{ISDM}(\theta_\star)}$ in (39) is bounded above as in (81), with the bound being $O(m)$. The second term obeys $\sigma^2_{\mathrm{ISDM}(\theta_\star)} \le \frac{\delta\mu_0^2}{1-\delta} m^2 + \frac{\sigma_0^2}{1-\delta} m$ by (72). For the third term in (39), the Cauchy-Schwarz inequality implies

$$\begin{aligned} \left| \frac{-2\gamma_{\mathrm{ISDM}(\theta_\star)}}{f_m(\xi_m)} \right| &\le 2\kappa_{\mathrm{ISDM}(\theta_\star)} \sigma_{\mathrm{ISDM}(\theta_\star)} \le 2 \left[\frac{2\pi Q_0''(\theta_\star)}{\delta^2} \big[1 + o(1)\big]^2 m \left(\frac{\delta\mu_0^2}{1-\delta} m^2 + \frac{\sigma_0^2}{1-\delta} m \right) \right]^{1/2} \\ &= 2 \left[\frac{2\pi Q_0''(\theta_\star)}{(1-\delta)} \left(\frac{\mu_0^2}{\delta} + \frac{\sigma_0^2}{m\delta^2} \right) \right]^{1/2} \big[1 + o(1)\big] m^{3/2} \end{aligned} \tag{139}$$

by (81). Applying the upper bounds from (81), (72), and (139) in (39) yields

$$\begin{aligned} \zeta^2_{\mathrm{ISDM}(\theta_\star)} \le{}& \left[\frac{\delta\mu_0^2}{1-\delta} \right] m^2 + \left[\frac{8\pi Q_0''(\theta_\star)}{(1-\delta)} \left(\frac{\mu_0^2}{\delta} + \frac{\sigma_0^2}{m\delta^2} \right) \right]^{1/2} \big[1 + o(1)\big] m^{3/2} \\ &+ \left[\frac{2\pi Q_0''(\theta_\star)}{\delta^2} \big[1 + o(1)\big]^2 + \frac{\sigma_0^2}{1-\delta} \right] m, \end{aligned} \tag{140}$$

which secures (48). Combining (48) with (93) establishes (49), which holds for any μ_0.

Now suppose that $\mu_0 \ne 0$. Then (73) gives $\sigma^2_{\mathrm{ISDM}(\theta)} = \Theta(m^2)$ as $m \to \infty$, and (81) ensures $\kappa^2_{\mathrm{ISDM}(\theta_\star)} = O(m)$. Also, the Cauchy-Schwarz inequality implies $\left| \frac{-2\gamma_{\mathrm{ISDM}(\theta_\star)}}{f_m(\xi_m)} \right| \le 2\kappa_{\mathrm{ISDM}(\theta_\star)} \sigma_{\mathrm{ISDM}(\theta_\star)} = O(m^{3/2})$. Thus, $\sigma^2_{\mathrm{ISDM}(\theta)}$ in (73) is the highest-order term of $\zeta^2_{\mathrm{ISDM}(\theta_\star)}$, so

$$\zeta^2_{\mathrm{ISDM}(\theta_\star)} = \left[\frac{\delta\mu_0^2}{1-\delta} \right] \big[1 + o(1)\big] m^2 = \Theta(m^2), \tag{141}$$

establishing (50) for $\mu_0 \ne 0$, and (51) follows from (93) and (141), where the WNRE result in (51) uses in (37) that the expected computation time to generate $(c(\mathbf{X}), L_{\mathrm{ISDM}(\theta_\star)}(\mathbf{X}))$ for $\mathbf{X} \sim \widetilde{G}_{\mathrm{ISDM}(\theta_\star)}$ as in (35) is $m\tau_{\mathrm{ISDM}(\theta_\star)}$ with constant $\tau_{\mathrm{ISDM}(\theta_\star)} \in (0, \infty)$ (Section 6.2).

When $\mu_0 = 0$, the bound yielding (48) shows that $\zeta^2_{\mathrm{ISDM}(\theta_\star)} = O(m)$. Thus, (93) ensures (52) for $\mu_0 = 0$, completing the proof of part (iv).

F.6.5. Part (v): $\mathrm{DE}(\theta_\star)$ **with Fixed Weights.** To show (53), we will provide lower bounds for each of the two terms in large parentheses for $\zeta^2_{\mathrm{DE}(\theta_\star)}$ in (41). Because $\delta, \upsilon_1, \upsilon_2 \in (0,1)$ are fixed, the asymptotic rate (in m) at which the first term in (41) grows is determined by $\upsilon_2^2 \sigma^2_{\mathrm{IS}(\theta_\star)}/\delta$, as in (44), so the first term in (41) is bounded below by $[\upsilon_2^2 Q_0''(-\theta_\star)/\delta]\big[1+o(1)\big] m\alpha_\star^m$, where $\alpha_\star = \alpha(\theta_\star) \in (1,\infty)$ by (69) since $\pm\theta_\star \in \Delta^\circ$. Also, as in (42) the asymptotic rate at which the second term in (41) increases is governed by $\frac{\upsilon_1'^2}{1-\delta}\kappa^2_{\mathrm{SRS}}$, so the second term in (41) is $\omega(me^{\beta m - \sqrt{m}})$ by (77). Combining these two results yields

$$\begin{aligned}
\zeta^2_{\mathrm{DE}(\theta_\star)} &\geq \frac{\upsilon_1'^2}{1-\delta}\,\frac{2\pi Q_0''(\theta_\star)\big[1+o\big(\frac{1}{\sqrt{m}}\big)\big] m e^{\beta m}}{\Upsilon_m^2} + \left[\frac{\upsilon_2^2}{\delta} Q_0''(-\theta_\star)\big[1+O(\alpha_\star^{-m})\big]\right] m\alpha_\star^m \\
&= \frac{\Theta(me^{\beta m})}{\Upsilon_m^2} + \Omega(m\alpha_\star^m) = \omega(me^{\beta m-\sqrt{m}}) + \Omega(m\alpha_\star^m) = \Omega(m s_0^m e^{-\sqrt{m}}), \qquad (142)
\end{aligned}$$

securing (53). Moreover, the first result of (54) follows from (142) and (93). Putting (93), (77), and (70) into (134) verifies the second part of (54).

F.6.6. Part (vi): $\mathrm{DE}_*(\theta_\star)$ **Optimal Weights Varying with** m **Satisfy Equation** (55). The optimal weights $(\upsilon^*_{1,m}, \upsilon^*_{2,m}) = \left(\frac{a_{1,m}}{a_{0,m}}, \frac{a_{2,m}}{a_{0,m}}\right)$ to minimize the asymptotic variance are defined in (28) and two paragraphs before Theorem 5, where

$$a_{0,m} = \mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m} + \mathrm{V}^{(\xi)}_{\mathrm{IS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m} + \mathrm{V}^{(\xi)}_{\mathrm{IS},m}\mathrm{V}^{(\mu)}_{\mathrm{SRS},m} + \mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{SRS},m} - \mathrm{C}^2_{\mathrm{IS},m} - \mathrm{C}^2_{\mathrm{SRS},m} - 2\mathrm{C}_{\mathrm{IS},m}\mathrm{C}_{\mathrm{SRS},m}, \quad (143)$$

$$a_{1,m} = \mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m} + \mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{SRS},m} - \mathrm{V}^{(\mu)}_{\mathrm{IS},m}\mathrm{C}_{\mathrm{SRS},m} + \mathrm{V}^{(\mu)}_{\mathrm{SRS},m}\mathrm{C}_{\mathrm{IS},m} - \mathrm{C}^2_{\mathrm{SRS},m} - \mathrm{C}_{\mathrm{IS},m}\mathrm{C}_{\mathrm{SRS},m}, \quad \text{and} \quad (144)$$

$$a_{2,m} = \mathrm{V}^{(\xi)}_{\mathrm{IS},m}\mathrm{V}^{(\mu)}_{\mathrm{SRS},m} + \mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{SRS},m} - \mathrm{V}^{(\xi)}_{\mathrm{IS},m}\mathrm{C}_{\mathrm{SRS},m} + \mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{C}_{\mathrm{IS},m} - \mathrm{C}^2_{\mathrm{SRS},m} - \mathrm{C}_{\mathrm{IS},m}\mathrm{C}_{\mathrm{SRS},m}. \quad (145)$$

We will prove (55) by analyzing growth rates of the terms in (143)–(145) as $m \to \infty$ for fixed $\delta \in (0,1)$. Theorem 6(ii), (32), and (29) yield $\alpha_\star \equiv \alpha(\theta_\star) > 1$ and $e^\beta > 1$. The first term in both (143) and (144) satisfies

$$\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m} = [\Theta(me^{\beta m})/\Upsilon_m^2]\Omega(m\alpha_\star^m) = \Omega(m^2[e^\beta\alpha_\star]^m)/\Upsilon_m^2 \qquad (146)$$

as $m \to \infty$ by (70) and (77), where (108) gives $\Upsilon_m^t = e^{o(\sqrt{m})} = o(e^{d_1\sqrt{m}})$ and $\Upsilon_m^t = \omega(e^{-d_1\sqrt{m}})$ as $m \to \infty$ for each $t \in \Re$ and each $d_1 > 0$. The rest of the proof will show that each other term in (143)–(145) is exponentially smaller than $\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}$ as $m \to \infty$, which will eventually secure (55).

We first obtain asymptotic upper bounds for $\mathrm{C}_{\mathrm{SRS},m}$ and $\mathrm{C}_{\mathrm{IS},m}$. Theorem 4 defines $\mathrm{C}_{\mathrm{SRS},m} = \gamma_{\mathrm{SRS}}/[(1-\delta)f_m(\xi_m)]$, with $\gamma_{\mathrm{SRS}} = E_G[I(c(\mathbf{X}) > \xi_m)c(\mathbf{X})] - (1-p_m)\mu_m$ from (10). By (29), (86), (107), and (108),

$$\left|\frac{(1-p_m)\mu_m}{f_m(\xi_m)}\right| = \frac{e^{-\beta m} m|\mu_0|}{e^{-\beta m}\Upsilon_m[1+o(1/\sqrt{m})]/\sqrt{2\pi m Q_0''(\theta_\star)}} = \frac{\Theta(m^{3/2})}{\Upsilon_m} = o(m^{3/2}e^{d_1\sqrt{m}})$$

as $m \to \infty$ for each $d_1 > 0$ when $\mu_0 \neq 0$; if $\mu_0 = 0$, then $\frac{(1-p_m)\mu_m}{f_m(\xi_m)} = 0$, which is also $o(m^{3/2}e^{d_1\sqrt{m}})$. In addition, $E_G[I(c(\mathbf{X}) > \xi_m)c(\mathbf{X})] = \vartheta_{1,0,m}$ by (111) for $j=1$ and $k=0$, so when $\mu_0 \neq 0$, (113) yields

$$|\mathrm{C}_{\mathrm{SRS},m}| \leq \frac{|\vartheta_{1,0,m}|}{(1-\delta)f_m(\xi_m)} + \left|\frac{(1-p_m)\mu_m}{(1-\delta)f_m(\xi_m)}\right| = O(m^{3/2}) + \frac{\Theta(m^{3/2})}{\Upsilon_m} = o(m^{3/2}e^{d_1\sqrt{m}}) \qquad (147)$$

as $m \to \infty$ for each $d_1 > 0$; if $\mu_0 = 0$, then $\mathrm{C}_{\mathrm{SRS},m} = o(m^{3/2}e^{d_1\sqrt{m}})$ still holds. Similarly, Theorem 4 defines $\mathrm{C}_{\mathrm{IS},m} = \gamma_{\mathrm{IS}}/[\delta f_m(\xi_m)]$, with $\gamma_{\mathrm{IS}} = E_G[I(c(\mathbf{X}) > \xi_m)c(\mathbf{X})L_{\theta_\star}] - (1-p_m)\mu_m$ by (20) and $E_G[I(c(\mathbf{X}) > \xi_m)c(\mathbf{X})L_{\theta_\star}] = \vartheta_{1,1,m}$ by (111) for $j=k=1$. Thus, when $\mu_0 \neq 0$, (113) yields

$$|\mathrm{C}_{\mathrm{IS},m}| \leq \frac{|\vartheta_{1,1,m}|}{\delta f_m(\xi_m)} + \left|\frac{(1-p_m)\mu_m}{\delta f_m(\xi_m)}\right| = O(m^{3/2}e^{-\beta m})e^{o(\sqrt{m})} + \frac{\Theta(m^{3/2})}{\Upsilon_m} = o(m^{3/2}e^{d_1\sqrt{m}}) \qquad (148)$$

as $m \to \infty$ for each $d_1 > 0$; if $\mu_0 = 0$, then $\mathrm{C}_{\mathrm{IS},m} = o(m^{3/2} e^{d_1\sqrt{m}})$ still holds.

We next analyze each term in (143)–(145) divided by $\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}$ in (146). Now $\Upsilon^t_m = o(e^{c_1\sqrt{m}})$ as $m \to \infty$ for each $t \in \Re$ and each $c_1 > 0$ by (108), and recall $\alpha_\star = \alpha(\theta_\star) \in (1, \infty)$ by (69). Therefore, using (66), (70), (77), (79), (147), and (148) gives, as $m \to \infty$,

$$r_{1,m} \equiv \frac{\mathrm{V}^{(\xi)}_{\mathrm{IS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}}{\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}} = \frac{\mathrm{V}^{(\xi)}_{\mathrm{IS},m}}{\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}} = \frac{O(m)}{\Theta(me^{\beta m})/\Upsilon^2_m} = O\big([e^{\beta}]^{-m}\big)\Upsilon^2_m = o\big(e^{-\beta m+\sqrt{m}}\big); \tag{149}$$

$$r_{2,m} \equiv \frac{\mathrm{V}^{(\xi)}_{\mathrm{IS},m}\mathrm{V}^{(\mu)}_{\mathrm{SRS},m}}{\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}} = \frac{O(m)\Theta(m)}{\Theta(me^{\beta m}/\Upsilon^2_m)\Omega(m\alpha^m_\star)} = o\big([e^{\beta}\alpha_\star]^{-m}e^{\sqrt{m}}\big); \tag{150}$$

$$r_{3,m} \equiv \frac{\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{SRS},m}}{\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}} = \frac{\mathrm{V}^{(\mu)}_{\mathrm{SRS},m}}{\mathrm{V}^{(\mu)}_{\mathrm{IS},m}} = \frac{\Theta(m)}{\Omega(m\alpha^m_\star)} = O(\alpha_\star^{-m}); \tag{151}$$

$$r_{4,m} \equiv \frac{\mathrm{V}^{(\xi)}_{\mathrm{IS},m}\mathrm{C}_{\mathrm{SRS},m}}{\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}} = \frac{O(m)o(m^{3/2}e^{\sqrt{m}/2})}{\Theta(me^{\beta m}/\Upsilon^2_m)\Omega(m\alpha^m_\star)} = o(m^{1/2}[e^{\beta}\alpha_\star]^{-m}e^{\sqrt{m}/2}\Upsilon^2_m) = o(m^{1/2}[e^{\beta}\alpha_\star]^{-m}e^{\sqrt{m}}); \tag{152}$$

$$r_{5,m} \equiv \frac{\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{C}_{\mathrm{IS},m}}{\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}} = \frac{\mathrm{C}_{\mathrm{IS},m}}{\mathrm{V}^{(\mu)}_{\mathrm{IS},m}} = \frac{o(m^{3/2}e^{\sqrt{m}})}{\Omega(m\alpha^m_\star)} = o(m^{1/2}\alpha_\star^{-m}e^{\sqrt{m}}); \tag{153}$$

$$r_{6,m} \equiv \frac{\mathrm{V}^{(\mu)}_{\mathrm{IS},m}\mathrm{C}_{\mathrm{SRS},m}}{\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}} = \frac{\mathrm{C}_{\mathrm{SRS},m}}{\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}} = \frac{o(m^{3/2}e^{\sqrt{m}/2})}{\Theta(me^{\beta m}/\Upsilon^2_m)} = o(m^{1/2}[e^{\beta}]^{-m}e^{\sqrt{m}/2}\Upsilon^2_m) = o(m^{1/2}[e^{\beta}]^{-m}e^{\sqrt{m}}); \tag{154}$$

$$r_{7,m} \equiv \frac{\mathrm{V}^{(\mu)}_{\mathrm{SRS},m}\mathrm{C}_{\mathrm{IS},m}}{\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}} = \frac{\Theta(m)o(m^{3/2}e^{\sqrt{m}/2})}{\Theta(me^{\beta m}/\Upsilon^2_m)\Omega(m\alpha^m_\star)} = o(m^{1/2}[e^{\beta}\alpha_\star]^{-m}e^{\sqrt{m}/2}\Upsilon^2_m) = o(m^{1/2}[e^{\beta}\alpha_\star]^{-m}e^{\sqrt{m}}); \tag{155}$$

$$r_{8,m} \equiv \frac{\mathrm{C}^2_{\mathrm{IS},m}}{\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}} = \frac{o(m^3e^{\sqrt{m}/2})}{\Theta(me^{\beta m}/\Upsilon^2_m)\Omega(m\alpha^m_\star)} = o(m[e^{\beta}\alpha_\star]^{-m}e^{\sqrt{m}/2}\Upsilon^2_m) = o(m[e^{\beta}\alpha_\star]^{-m}e^{\sqrt{m}}); \tag{156}$$

$$r_{9,m} \equiv \frac{\mathrm{C}^2_{\mathrm{SRS},m}}{\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}} = \frac{o(m^3e^{\sqrt{m}/2})}{\Theta(me^{\beta m}/\Upsilon^2_m)\Omega(m\alpha^m_\star)} = o(m[e^{\beta}\alpha_\star]^{-m}e^{\sqrt{m}/2}\Upsilon^2_m) = o(m[e^{\beta}\alpha_\star]^{-m}e^{\sqrt{m}}); \tag{157}$$

$$r_{10,m} \equiv \frac{\mathrm{C}_{\mathrm{IS},m}\mathrm{C}_{\mathrm{SRS},m}}{\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}} = \frac{o(m^{3/2}e^{\sqrt{m}/4})o(m^{3/2}e^{\sqrt{m}/4})}{\Theta(me^{\beta m}/\Upsilon^2_m)\Omega(m\alpha^m_\star)} = o(m[e^{\beta}\alpha_\star]^{-m}e^{\sqrt{m}/2}\Upsilon^2_m) = o(m[e^{\beta}\alpha_\star]^{-m}e^{\sqrt{m}}). \tag{158}$$

Each upper bound includes a factor that is exponential in $-m$ with base greater than 1, so each $r_{j,m}$ shrinks exponentially fast in m.

We next use (149)–(158) to analyze $a_{0,m}$ in (143). Dividing $a_{0,m}$ by its first term $\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}$ in (143) leads to

$$\frac{a_{0,m}}{\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}} = 1 + r_{1,m} + r_{2,m} + r_{3,m} - r_{8,m} - r_{9,m} - 2r_{10,m} \equiv 1 + r_{11,m}, \tag{159}$$

where $r_{11,m} = r_{1,m} + r_{2,m} + r_{3,m} - r_{8,m} - r_{9,m} - 2r_{10,m}$ shrinks exponentially fast to 0 as $m \to \infty$. Thus,

$$a_{0,m} = \mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}(1 + r_{11,m}) \quad \text{and} \quad \frac{1}{1 + r_{11,m}} = 1 - r_{11,m} + \frac{r^2_{11,m}}{1 - r_{11,m}} = 1 - r_{11,m} + O(r^2_{11,m}) \tag{160}$$

as $m \to \infty$. Recall that $\alpha_\star > 1$ and $e^{\beta} > 1$, so the smallest base of the exponential terms (in $-m$) in the upper bounds in (149)–(158) for $r_{j,m}$ with $j \in J_{11} \equiv \{1, 2, 3, 8, 9, 10\}$ defining $r_{11,m}$ will lead to the exponential rate in an upper bound at which $r_{11,m}$ shrinks. For $\Upsilon_m = e^{o(\sqrt{m})}$ in (108), we then see that

$$r_{11,m} = O\big([e^{-\beta m}\Upsilon^2_m] \vee \alpha_\star^{-m}\big) \tag{161}$$

from the upper bounds for $r_{1,m}$ and $r_{3,m}$ in (149) and (151).

To analyze $\upsilon^*_{1,m}$, divide $a_{1,m}$ from (144) by $a_{0,m}$ in (160) and use (149)–(159) to get as $m\to\infty$,

$$\begin{aligned}
\upsilon^*_{1,m} &= \frac{a_{1,m}}{a_{0,m}} = \frac{a_{1,m}}{\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}}\left(\frac{1}{1+r_{11,m}}\right) = \big(1+r_{3,m}-r_{6,m}+r_{7,m}-r_{9,m}-r_{10,m}\big)\big[1-r_{11,m}+O(r^2_{11,m})\big]\\
&= 1-r_{11,m}+O(r^2_{11,m})+\big(r_{3,m}-r_{6,m}+r_{7,m}-r_{9,m}-r_{10,m}\big)\big[1-r_{11,m}+O(r^2_{11,m})\big]\\
&= 1-r_{1,m}-r_{2,m}-r_{6,m}+r_{7,m}+r_{8,m}+r_{10,m}-\big(r_{3,m}-r_{6,m}+r_{7,m}-r_{9,m}-r_{10,m}\big)r_{11,m}+O(r^2_{11,m})\\
&\equiv 1-\varepsilon_{1,m}, \quad \text{ with } \ \varepsilon_{1,m} = O\big((m^{1/2}[e^{\beta}]^{-m}e^{\sqrt{m}})\vee[\alpha^2_\star]^{-m}\big)\to 0 \text{ as } m\to\infty, \qquad (162)
\end{aligned}$$

from the bounds for $r_{6,m}=o(m^{1/2}[e^{\beta}]^{-m}e^{\sqrt{m}})$ in (154), $r_{3,m}r_{11,m}=O(\alpha_\star^{-m})O([e^{-\beta m}\Upsilon^2_m]\vee\alpha_\star^{-m}) = O([e^{\beta}\alpha_\star]^{-m}e^{\sqrt{m}}\vee\alpha_\star^{-2m})$ by (151) and (161), and $O(r^2_{11,m}) = O\big(([e^{2\beta}]^{-m}\Upsilon^4_m)\vee[\alpha^2_\star]^{-m}\big) = O\big(([e^{2\beta}]^{-m}e^{\sqrt{m}})\vee[\alpha^2_\star]^{-m}\big)$. Thus, $\upsilon^*_{1,m}\to 1$ exponentially fast as $m\to\infty$, establishing one part of (55).

To finally analyze $\upsilon^*_{2,m}$, dividing $a_{2,m}$ from (145) by $a_{0,m}$ in (160) and using (149)–(159) and (161) yield as $m\to\infty$,

$$\begin{aligned}
\upsilon^*_{2,m} &= \frac{a_{2,m}}{a_{0,m}} = \frac{a_{2,m}}{\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}}\left(\frac{1}{1+r_{11,m}}\right) = \big(r_{2,m}+r_{3,m}-r_{4,m}+r_{5,m}-r_{9,m}-r_{10,m}\big)\big[1-r_{11,m}+O(r^2_{11,m})\big]\\
&= \big(r_{2,m}+r_{3,m}-r_{4,m}+r_{5,m}-r_{9,m}-r_{10,m}\big)\big[1+o(1)\big]\\
&\equiv \varepsilon_{2,m}, \quad \text{ with } \ \varepsilon_{2,m} = o\big(m^{1/2}\alpha_\star^{-m}e^{\sqrt{m}}\big)\to 0 \text{ as } m\to\infty, \qquad (163)
\end{aligned}$$

by the bound for $r_{5,m}$ in (153). Thus, $\upsilon^*_{2,m}\to 0$ exponentially fast as $m\to\infty$, completing the proof of (55).

F.6.7. Part (vi): $\mathrm{DE}_*(\theta_\star)$ Property (57) for Optimal Weights Varying with m to Minimize $\mathrm{RE}_{\mathrm{DE}(\theta_\star),m}[\eta]$. To establish (57), we start by deriving an asymptotic expression for $\zeta^2_{\mathrm{DE}_*(\theta_\star)}$. By (55), write $\upsilon^*_{1,m} = 1-\varepsilon_{1,m}$ and $\upsilon^*_{2,m}=\varepsilon_{2,m}$, for $\varepsilon_{1,m}$ in (162) and $\varepsilon_{2,m}$ in (163), where both $\varepsilon_{1,m}$ and $\varepsilon_{2,m}$ shrink to 0 exponentially fast as $m\to\infty$, so $\varepsilon^2_{1,m}=o(\varepsilon_{1,m})$ and $\varepsilon^2_{2,m}=o(\varepsilon_{2,m})$. Then (27) implies

$$\begin{aligned}
\zeta^2_{\mathrm{DE}_*(\theta_\star)} &= \big(1-\varepsilon_{1,m}\big)^2\mathrm{V}^{(\xi)}_{\mathrm{IS},m}+\varepsilon^2_{2,m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}-2\big(1-\varepsilon_{1,m}\big)\varepsilon_{2,m}\mathrm{C}_{\mathrm{IS},m}\\
&\quad+\big(1-\varepsilon_{2,m}\big)^2\mathrm{V}^{(\mu)}_{\mathrm{SRS},m}+\varepsilon^2_{1,m}\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}-2\varepsilon_{1,m}\big(1-\varepsilon_{2,m}\big)\mathrm{C}_{\mathrm{SRS},m}\\
&= (1-2\varepsilon_{1,m}[1+o(1)])\,\mathrm{V}^{(\xi)}_{\mathrm{IS},m}+(1-2\varepsilon_{2,m}[1+o(1)])\,\mathrm{V}^{(\mu)}_{\mathrm{SRS},m}+\varepsilon_{3,m}, \qquad (164)
\end{aligned}$$

where

$$\varepsilon_{3,m}\equiv \mathrm{V}^{(\mu)}_{\mathrm{IS},m}\varepsilon^2_{2,m}+\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\varepsilon^2_{1,m}-2\mathrm{C}_{\mathrm{IS},m}(1-\varepsilon_{1,m})\varepsilon_{2,m}-2\mathrm{C}_{\mathrm{SRS},m}\varepsilon_{1,m}(1-\varepsilon_{2,m}). \qquad (165)$$

For $\varepsilon_{3,m}$ in (165), the last two terms have $|1-\varepsilon_{1,m}|\le 2$ and $|1-\varepsilon_{2,m}|\le 2$ for all m sufficiently large since $\varepsilon_{1,m}\to 0$ and $\varepsilon_{2,m}\to 0$ as $m\to\infty$. Moreover, we have as $m\to\infty$ that $\mathrm{V}^{(\mu)}_{\mathrm{IS},m}=O(m^2\alpha^m_\star)$ by (70), $\mathrm{C}_{\mathrm{IS},m} = o(m^{3/2}e^{\sqrt{m}})$ by (148), and $\mathrm{C}_{\mathrm{SRS},m}=o(m^{3/2}e^{\sqrt{m}})$ by (147). Theorem 4 and (9) give $\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}=\chi^2_{\mathrm{SRS}}/[(1-\delta)f^2_m(\xi_m)]$ with $\chi^2_{\mathrm{SRS}}=p_m(1-p_m)=\Theta(e^{-\beta m})$ by (29). Also, (107) shows that $1/f^2_m(\xi_m)=\Theta(me^{2\beta m}/\Upsilon^2_m)$ with $\Upsilon^{-2}_m = o(e^{\sqrt{m}})$ by (108), so $\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}=o(me^{\beta m+\sqrt{m}})$. Combining these with $\varepsilon_{1,m}=O\big((m^{1/2}[e^{\beta}]^{-m}e^{\sqrt{m}})\vee[\alpha^2_\star]^{-m}\big)$ and $\varepsilon_{2,m}=o\big(m^{1/2}\alpha_\star^{-m}e^{\sqrt{m}}\big)$ from (162)–(163) shows that for all large enough m,

$$\begin{aligned}
|\varepsilon_{3,m}| &\le \mathrm{V}^{(\mu)}_{\mathrm{IS},m}\varepsilon^2_{2,m}+\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\varepsilon^2_{1,m}+4|\mathrm{C}_{\mathrm{IS},m}||\varepsilon_{2,m}|+4|\mathrm{C}_{\mathrm{SRS},m}||\varepsilon_{1,m}|\\
&= O(m^2\alpha^m_\star)o(m\alpha_\star^{-2m}e^{2\sqrt{m}})+o(me^{\beta m+\sqrt{m}})O\big([me^{-2\beta m+2\sqrt{m}}]\vee\alpha_\star^{-4m}\big)\\
&\quad+o(m^{3/2}e^{\sqrt{m}})o(m^{1/2}\alpha_\star^{-m}e^{\sqrt{m}})+o(m^{3/2}e^{\sqrt{m}})O\big([m^{1/2}e^{-\beta m+\sqrt{m}}]\vee\alpha_\star^{-2m}\big)\\
&= o(m^3\alpha_\star^{-m}e^{2\sqrt{m}})+o\big((m^2e^{-\beta m+3\sqrt{m}})\vee(m[e^{-\beta}\alpha^4_\star]^{-m}e^{\sqrt{m}})\big)\\
&\quad+o(m^2\alpha_\star^{-m}e^{2\sqrt{m}})+o\big((m^2e^{-\beta m+2\sqrt{m}})\vee(m^{3/2}\alpha_\star^{-2m}e^{\sqrt{m}})\big)\\
&= o\big((m^3\alpha_\star^{-m}e^{2\sqrt{m}})\vee(m^2[e^{\beta}]^{-m}e^{3\sqrt{m}})\vee(m[e^{-\beta}\alpha^4_\star]^{-m}e^{\sqrt{m}})\big), \qquad (166)
\end{aligned}$$

where $e^{-\beta}\alpha_\star^4 > 1$ by (56), so $\varepsilon_{3,m} \to 0$ exponentially fast as $m \to \infty$.

We now prove for any fixed $\delta \in (0,1)$ that (57) holds. The MSIS variance is $\zeta^2_{\mathrm{MSIS}(\theta_\star)} = \mathrm{V}^{(\xi)}_{\mathrm{IS},m} + \mathrm{V}^{(\mu)}_{\mathrm{SRS},m}$ by (23), where as $m \to \infty$, $\mathrm{V}^{(\xi)}_{\mathrm{IS},m} = O(m)$ by (79), $\mathrm{V}^{(\mu)}_{\mathrm{SRS},m} = \Theta(m) \geq 0$ by (66), and $\zeta^2_{\mathrm{MSIS}(\theta_\star)} = \Theta(m)$ by (46). Hence, (164) reveals $\zeta^2_{\mathrm{DE}_*(\theta_\star)} = \zeta^2_{\mathrm{MSIS}(\theta_\star)} - 2\varepsilon_{1,m}[1+o(1)]\mathrm{V}^{(\xi)}_{\mathrm{IS},m} - 2\varepsilon_{2,m}[1+o(1)]\mathrm{V}^{(\mu)}_{\mathrm{SRS},m} + \varepsilon_{3,m}$, so

$$\begin{aligned}\frac{\zeta^2_{\mathrm{DE}_*(\theta_\star)}}{\zeta^2_{\mathrm{MSIS}(\theta_\star)}} &= 1 - \frac{2\varepsilon_{1,m}[1+o(1)]\mathrm{V}^{(\xi)}_{\mathrm{IS},m}}{\zeta^2_{\mathrm{MSIS}(\theta_\star)}} - \frac{2\varepsilon_{2,m}[1+o(1)]\mathrm{V}^{(\mu)}_{\mathrm{SRS},m}}{\zeta^2_{\mathrm{MSIS}(\theta_\star)}} + \frac{\varepsilon_{3,m}}{\zeta^2_{\mathrm{MSIS}(\theta_\star)}} \\ &= 1 - \frac{\varepsilon_{1,m}O(m)}{\Theta(m)} - \frac{\varepsilon_{2,m}\Theta(m)}{\Theta(m)} + \frac{\varepsilon_{3,m}}{\Theta(m)} = 1 - \varepsilon_{1,m}O(1) - \varepsilon_{2,m}\Theta(1) + \Theta(\varepsilon_{3,m}/m)\end{aligned}$$

as $m \to \infty$, validating (57) by (162), (163), and (166) under (56), completing the proof. □

F.6.8. Part (vi): $\mathrm{DE}_{**}(\theta_\star)$ Properties (55) and (57) for Optimal Weights Varying with m to Minimize $\mathrm{WNRE}_{\mathrm{DE}(\theta_\star),m}[\eta]$. Next we analyze the optimal value of $(\upsilon_1, \upsilon_2) = (\upsilon^{**}_{1,m}, \upsilon^{**}_{2,m})$ when minimizing $\mathrm{WNRE}_{\mathrm{DE}(\theta_\star),m}[\eta]$ in (134). In this case, we define $(\upsilon^{**}_{1,m}, \upsilon^{**}_{2,m}) = (\frac{a^{**}_{1,m}}{a^{**}_{0,m}}, \frac{a^{**}_{2,m}}{a^{**}_{0,m}})$, with $a^{**}_{0,m} = m^2\big[\tau_{\mathrm{SRS}}\tau_{\mathrm{IS}(\theta_\star)}\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m} - \tau^2_{\mathrm{IS}(\theta_\star)}\mathrm{C}^2_{\mathrm{IS},m} - 2\tau_{\mathrm{IS}(\theta_\star)}\tau_{\mathrm{SRS}}\mathrm{C}_{\mathrm{IS},m}\mathrm{C}_{\mathrm{SRS},m} - \tau^2_{\mathrm{SRS}}\mathrm{C}^2_{\mathrm{SRS},m} + \tau^2_{\mathrm{IS}(\theta_\star)}\mathrm{V}^{(\xi)}_{\mathrm{IS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m} + \tau_{\mathrm{IS}(\theta_\star)}\tau_{\mathrm{SRS}}\mathrm{V}^{(\xi)}_{\mathrm{IS},m}\mathrm{V}^{(\mu)}_{\mathrm{SRS},m} + \tau^2_{\mathrm{SRS}}\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{SRS},m}\big]$, $a^{**}_{1,m} = m^2\big[\tau_{\mathrm{SRS}}\tau_{\mathrm{IS}(\theta_\star)}\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{IS},m} + \tau^2_{\mathrm{SRS}}\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{SRS},m} - \tau_{\mathrm{IS}(\theta_\star)}\tau_{\mathrm{SRS}}\mathrm{V}^{(\mu)}_{\mathrm{IS},m}\mathrm{C}_{\mathrm{SRS},m} + \tau_{\mathrm{IS}(\theta_\star)}\tau_{\mathrm{SRS}}\mathrm{C}_{\mathrm{IS},m}\mathrm{V}^{(\mu)}_{\mathrm{SRS},m} - \tau_{\mathrm{IS}(\theta_\star)}\tau_{\mathrm{SRS}}\mathrm{C}_{\mathrm{IS},m}\mathrm{C}_{\mathrm{SRS},m} - \tau^2_{\mathrm{SRS}}\mathrm{C}^2_{\mathrm{SRS},m}\big]$, and $a^{**}_{2,m} = m^2\big[\tau_{\mathrm{IS}(\theta_\star)}\tau_{\mathrm{SRS}}\mathrm{V}^{(\xi)}_{\mathrm{IS},m}\mathrm{V}^{(\mu)}_{\mathrm{SRS},m} + \tau^2_{\mathrm{SRS}}\mathrm{V}^{(\xi)}_{\mathrm{SRS},m}\mathrm{V}^{(\mu)}_{\mathrm{SRS},m} - \tau_{\mathrm{IS}(\theta_\star)}\tau_{\mathrm{SRS}}\mathrm{V}^{(\xi)}_{\mathrm{IS},m}\mathrm{C}_{\mathrm{SRS},m} + \tau_{\mathrm{IS}(\theta_\star)}\tau_{\mathrm{SRS}}\mathrm{C}_{\mathrm{IS},m}\mathrm{V}^{(\xi)}_{\mathrm{SRS},m} - \tau_{\mathrm{IS}(\theta_\star)}\tau_{\mathrm{SRS}}\mathrm{C}_{\mathrm{IS},m}\mathrm{C}_{\mathrm{SRS},m} - \tau^2_{\mathrm{SRS}}\mathrm{C}^2_{\mathrm{SRS},m}\big]$. When minimizing $\mathrm{WNRE}_{\mathrm{DE}(\theta_\star),m}[\eta]$, the only difference between $(\upsilon^{**}_{1,m}, \upsilon^{**}_{2,m})$ and $(\upsilon^{*}_{1,m}, \upsilon^{*}_{2,m})$ in (28) that minimizes $\mathrm{RE}_{\mathrm{DE}(\theta_\star),m}[\eta]$ is that we further multiplied each product item in $a_{0,m}$, $a_{1,m}$ and $a_{2,m}$ from (143)–(145) by factors m^2 and a constant involving τ_{SRS} or $\tau_{\mathrm{IS}(\theta)}$ (or both). Thus, an argument like the one used to prove (55) shows that $(\upsilon^{**}_{1,m}, \upsilon^{**}_{2,m})$ converges exponentially fast to $(1,0)$ as $m \to \infty$. Similarly, applying arguments analogous to how we achieved (57) under (56) results in $\zeta^2_{\mathrm{DE}_{**}(\theta_\star)}/\zeta^2_{\mathrm{MSIS}(\theta_\star)} \to 1$ exponentially fast as $m \to \infty$.

G: Two-Step IS to Estimate Extreme Quantile and EC in PCRM with Random Loss Given Default

We now describe our IS approach to estimate an extreme quantile for the model in Section 7.2. We assume that the common shock $S \equiv 1$ in (58), as in Glasserman and Li (2005), but we extend their method to allow for the loss given default to be stochastic. Although our simulation experiments in Section 7.2 have LGD $C_k \sim \mathrm{Unif}(0, \beta_k)$, independent of $(\mathbf{Z}, \epsilon_1, \ldots, \epsilon_m)$, we develop the method for more general LGD satisfying certain conditions; see (168). Let V_k denote the marginal CDF (not necessarily uniform) of the LGD C_k.

Glasserman and Li (2005) developed a two-step IS method to estimate the tail probability $\lambda_x \equiv P(Y > x)$, where x is a given large threshold, and we adapt their ideas to estimate the p-quantile ξ, for $p \approx 1$. Their method critically depends on knowing the threshold x, which is explicitly used throughout their approach. We cannot simply apply their technique by letting $x = \xi$, as ξ is unknown. Instead, we first run pilot simulations for a few different values of the threshold x, and interpolate to obtain a crude approximation $\mathring{\xi}$ to ξ. Finally we use additional simulation runs with our modification of the two-step approach of Glasserman and Li (2005) with threshold $x = \mathring{\xi}$, and employ the resulting data to estimate ξ.

Before describing the two-step IS, we first consider a one-step IS conditional on $\mathbf{Z}$, which makes the obligors conditionally independent. In the following, Appendix G.1 first applies the one-step IS conditional on $\mathbf{Z}$ to estimate $\lambda_x(\mathbf{Z}) \equiv P(Y > x \mid \mathbf{Z})$. Next, Appendix G.2 extends the one-step IS to a two-step IS to estimate λ_x, by first using IS to sample $\mathbf{Z}$ from a different CDF than its original one and then applying the one-step IS given the observed $\mathbf{Z}$. Appendix G.3 finally adapts the two-step IS for λ_x to instead estimate ξ.

Recall that the total portfolio loss in Section 7.2 is $Y = \sum_{k=1}^m C_k D_k$, where $D_k = I(\epsilon_k > (w_k - \mathbf{a}_k \mathbf{Z})/b_k)$ is the indicator that obligor k defaults because we assumed that the common shock $S \equiv 1$. The mutual independence of $\mathbf{Z}, \epsilon_1, \epsilon_2, \ldots, \epsilon_m$ implies that

$$\text{given } \mathbf{Z}\text{, the default indicators } D_1, \ldots, D_m \text{ are conditionally independent.} \tag{167}$$

Moreover, we assume that for $\mathbf{D} = (D_1, \ldots, D_m)$,

$$C_1, \ldots, C_m, \mathbf{D} \text{ are conditionally independent, given } \mathbf{Z}. \tag{168}$$

The following methods will exploit properties (167) and (168).

G.1. One-Step IS Conditional on Z to Estimate $P(Y > x \mid \mathbf{Z})$

This section will modify the one-step IS of Glasserman and Li (2005) to estimate $\lambda_x(\mathbf{Z})$ when LGD is random and satisfies (168). The IS applies exponential twisting (Section 6.2). Let $F(\cdot \mid \mathbf{Z})$ be the *conditional CDF* (CCDF) of the loss Y given $\mathbf{Z}$. For each obligor k, define $T_k \equiv C_k D_k$, so $Y = \sum_{k=1}^m T_k$. Let $H_k(\cdot \mid \mathbf{Z})$ be the CCDF of T_k given $\mathbf{Z}$. We will see that given $\mathbf{Z}$, applying an exponential twist to the CCDF $F(\cdot \mid \mathbf{Z})$ of Y with twisting parameter θ is equivalent to twisting each $H_k(\cdot \mid \mathbf{Z})$ with the same θ. (We will later describe in (183) how to choose $\theta = \theta_x(\mathbf{Z})$ as a function of both the factor values $\mathbf{Z}$ and the threshold x.)

G.1.1. Exponential Twist to Each $H_k(\cdot \mid \mathbf{Z})$ This section will exponentially twist each $H_k(\cdot \mid \mathbf{Z})$ with the same $\theta \in \Re$, and gives details on how to generate T_k when applying IS conditional on $\mathbf{Z}$. The exponential twist $\widetilde{H}_{k,\theta}(\cdot \mid \mathbf{Z})$ of the CCDF $H_k(\cdot \mid \mathbf{Z})$ of T_k given $\mathbf{Z}$ using parameter $\theta \in \Re$ is defined by

$$\mathrm{d}\widetilde{H}_{k,\theta}(t \mid \mathbf{Z}) = \frac{e^{\theta t} \mathrm{d}H_k(t \mid \mathbf{Z})}{m_{H_k}(\theta, \mathbf{Z})}, \tag{169}$$

where $m_{H_k}(\theta, \mathbf{Z}) = \int e^{\theta t}\, \mathrm{d}H_k(t \mid \mathbf{Z})$ is the *conditional moment generating function* (CMGF) of $T_k \sim H_k(\cdot \mid \mathbf{Z})$. Let $\widetilde{E}_\theta[\,\cdot \mid \mathbf{Z}\,]$ denote conditional expectation given $\mathbf{Z}$, when each $T_k \sim \widetilde{H}_{k,\theta}(\cdot \mid \mathbf{Z})$. By (168) we can write

$$\begin{aligned}
\lambda_x(\mathbf{Z}) = E\left[I\left(\textstyle\sum_{k=1}^m T_k > x\right) \mid \mathbf{Z}\right] &= \int_{(t_1,\ldots,t_m)\in\Re^m} I\left(\textstyle\sum_{k=1}^m t_k > x\right) \prod_{k=1}^m \mathrm{d}H_k(t_k \mid \mathbf{Z}) \\
&= \int_{(t_1,\ldots,t_m)\in\Re^m} I\left(\textstyle\sum_{k=1}^m t_k > x\right) \prod_{k=1}^m \frac{\mathrm{d}H_k(t_k \mid \mathbf{Z})}{\mathrm{d}\widetilde{H}_{k,\theta}(t_k \mid \mathbf{Z})} \mathrm{d}\widetilde{H}_{k,\theta}(t_k \mid \mathbf{Z}) \\
&= \widetilde{E}_\theta\left[I\left(\textstyle\sum_{k=1}^m T_k > x\right) L_\theta(T_1, \ldots, T_m, \mathbf{Z}) \mid \mathbf{Z}\right],
\end{aligned} \tag{170}$$

where $L_\theta(t_1, \ldots, t_m, \mathbf{Z}) = \prod_{k=1}^m \mathrm{d}H_k(t_k \mid \mathbf{Z}) / \mathrm{d}\widetilde{H}_{k,\theta}(t_k \mid \mathbf{Z})$ is the conditional LR given $\mathbf{Z}$. Also, let $\psi_{H_k}(\theta, \mathbf{Z}) = \ln[m_{H_k}(\theta, \mathbf{Z})]$ be the *conditional cumulant generating function* (CCGF) of $T_k \sim H_k(\cdot \mid \mathbf{Z})$, so by (169),

$$L_\theta(t_1, \ldots, t_m, \mathbf{Z}) = \prod_{k=1}^m \frac{\mathrm{d}H_k(t_k \mid \mathbf{Z})}{e^{\theta t_k}\, \mathrm{d}H_k(t_k \mid \mathbf{Z}) / m_{H_k}(\theta, \mathbf{Z})} = \prod_{k=1}^m \frac{m_{H_k}(\theta, \mathbf{Z})}{e^{\theta t_k}} = \prod_{k=1}^m e^{\psi_{H_k}(\theta, \mathbf{Z}) - \theta t_k}. \tag{171}$$

We now give more details on the exponential twist $\widetilde{H}_{k,\theta}(\cdot\,|\,\mathbf{Z})$ defined by (169), which will require expressions for $H_k(\cdot\,|\,\mathbf{Z})$ and $m_{H_k}(\theta,\mathbf{Z})$. To compute $m_{H_k}(\theta,\mathbf{Z})$, let $\dot{p}_k(\mathbf{Z})$ be the conditional probability that obligor k defaults given $\mathbf{Z}$, which satisfies

$$\dot{p}_k(\mathbf{Z}) \equiv P(D_k=1\,|\,\mathbf{Z}) = P\left(\epsilon_k > \frac{w_k - \mathbf{a}_k\mathbf{Z}}{b_k}\,\Big|\,\mathbf{Z}\right) = \Phi\left(\frac{\mathbf{a}_k\mathbf{Z} + \Phi^{-1}(\dot{p}_k)}{b_k}\right) \tag{172}$$

because $\epsilon_k \sim N(0,1)$ is independent of $\mathbf{Z}$ and the $N(0,1)$ density is symmetric about the origin. Then we use (168) and (172) to compute the CMGF of $T_k = C_k D_k \sim H_k(\cdot\,|\,\mathbf{Z})$ as

$$\begin{aligned} m_{H_k}(\theta,\mathbf{Z}) &= E\left[e^{\theta C_k D_k}\,|\,\mathbf{Z}\right] = E\left[E\left[e^{\theta C_k D_k}\,|\,C_k,\mathbf{Z}\right]\,|\,\mathbf{Z}\right] = E\left[e^{\theta C_k\cdot 0}\,(1-\dot{p}_k(\mathbf{Z})) + e^{\theta C_k\cdot 1}\dot{p}_k(\mathbf{Z})\,|\,\mathbf{Z}\right] \\ &= 1 - \dot{p}_k(\mathbf{Z}) + \dot{p}_k(\mathbf{Z})E\left[e^{\theta C_k}\,|\,\mathbf{Z}\right] = 1 + \dot{p}_k(\mathbf{Z})[m_{V_k}(\theta,\mathbf{Z}) - 1], \end{aligned} \tag{173}$$

where $m_{V_k}(\theta,\mathbf{Z}) = E\left[e^{\theta C_k}\,|\,\mathbf{Z}\right]$ is the CMGF of $C_k \sim V_k(\cdot\,|\,\mathbf{Z})$, and $V_k(\cdot\,|\,\mathbf{Z})$ is the CCDF of C_k given $\mathbf{Z}$.

Next we will work out the details of the conditional exponential twist given $\mathbf{Z}$ for each $T_k \sim H_k(\cdot\,|\,\mathbf{Z})$. To do this we need an expression for $H_k(\cdot\,|\,\mathbf{Z})$, which by (168) and (172) satisfies

$$\begin{aligned} H_k(t\,|\,\mathbf{Z}) &= P(C_kD_k \le t\,|\,\mathbf{Z}) = E[P(C_kD_k \le t\,|\,D_k,\mathbf{Z})\,|\,\mathbf{Z}] \\ &= P(D_k=0\,|\,\mathbf{Z})\,P(C_kD_k \le t\,|\,D_k=0,\mathbf{Z}) + P(D_k=1\,|\,\mathbf{Z})\,P(C_kD_k \le t\,|\,D_k=1,\mathbf{Z}) \\ &= (1-\dot{p}_k(\mathbf{Z}))I(t\ge 0) + \dot{p}_k(\mathbf{Z})P(C_k \le t\,|\,\mathbf{Z}) \end{aligned} \tag{174}$$

by (172), so $H_k(\cdot\,|\,\mathbf{Z})$ is a mixture of the CDFs $I(\cdot \ge 0)$ and $V_k(\cdot\,|\,\mathbf{Z})$. Thus, we have

$$\mathrm{d}H_k(t\,|\,\mathbf{Z}) = P(T_k \in \mathrm{d}t\,|\,\mathbf{Z}) = (1-\dot{p}_k(\mathbf{Z}))\delta_0(\{\mathrm{d}t\}) + \dot{p}_k(\mathbf{Z})v_k(t\,|\,\mathbf{Z})\,\mathrm{d}t, \tag{175}$$

where $v_k(\cdot\,|\,\mathbf{Z})$ is the density of $V_k(\cdot\,|\,\mathbf{Z})$, and δ_0 is a measure defined on measurable sets $A \subseteq \Re$, such that $\delta_0(A)=1$ if $0\in A$ and $\delta_0(A)=0$ if $0\notin A$. Then we can write $I(t\ge 0) = \int_{s=-\infty}^{t}\delta_0(\{\mathrm{d}s\}) = \delta_0((-\infty,t])$. Also, putting (175) into (169) yields

$$\begin{aligned} \mathrm{d}\widetilde{H}_{k,\theta}(t\,|\,\mathbf{Z}) &= \frac{e^{\theta t}(1-\dot{p}_k(\mathbf{Z}))\delta_0(\{\mathrm{d}t\})}{m_{H_k}(\theta,\mathbf{Z})} + \frac{e^{\theta t}\dot{p}_k(\mathbf{Z})v_k(t\,|\,\mathbf{Z})}{m_{H_k}(\theta,\mathbf{Z})}\,\mathrm{d}t \\ &= \frac{(1-\dot{p}_k(\mathbf{Z}))\delta_0(\{\mathrm{d}t\})}{m_{H_k}(\theta,\mathbf{Z})} + \frac{e^{\theta t}\dot{p}_k(\mathbf{Z})v_k(t\,|\,\mathbf{Z})}{m_{H_k}(\theta,\mathbf{Z})}\,\mathrm{d}t \\ &= \widetilde{q}_{k,\theta}(\mathbf{Z})\delta_0(\{\mathrm{d}t\}) + \widetilde{p}_{k,\theta}(\mathbf{Z})\widetilde{v}_{k,\theta}(t\,|\,\mathbf{Z})\,\mathrm{d}t, \end{aligned} \tag{176}$$

where (176) holds because $\delta_0(\{dt\})$ is nonzero only when $t=0$, in which case $e^{\theta t}=1$, and

$$\widetilde{v}_{k,\theta}(t\,|\,\mathbf{Z}) \equiv \frac{e^{\theta t}v_k(t\,|\,\mathbf{Z})}{m_{V_k}(\theta,\mathbf{Z})}, \quad \widetilde{q}_{k,\theta}(\mathbf{Z}) \equiv \frac{1-\dot{p}_k(\mathbf{Z})}{m_{H_k}(\theta,\mathbf{Z})}, \quad \widetilde{p}_{k,\theta}(\mathbf{Z}) \equiv 1 - \widetilde{q}_{k,\theta}(\mathbf{Z}) = \frac{\dot{p}_k(\mathbf{Z})m_{V_k}(\theta,\mathbf{Z})}{m_{H_k}(\theta,\mathbf{Z})}, \tag{177}$$

with (177) using (173). Note that $\widetilde{v}_{k,\theta}(\cdot\,|\,\mathbf{Z})$ is the exponential twist of $v_{k,\theta}(\cdot\,|\,\mathbf{Z})$. Given $\mathbf{Z}$, we have $\widetilde{q}_{k,\theta}(\mathbf{Z}) + \widetilde{p}_{k,\theta}(\mathbf{Z}) = 1$ with $\widetilde{q}_{k,\theta}(\mathbf{Z}) \ge 0$ and $\widetilde{p}_{k,\theta}(\mathbf{Z}) \ge 0$ by (173), so

$$\widetilde{H}_{k,\theta}(t\,|\,\mathbf{Z}) = \int_{s=0}^{t}\mathrm{d}\widetilde{H}_{k,\theta}(s\,|\,\mathbf{Z}) = \widetilde{q}_{k,\theta}(\mathbf{Z})I(t\ge 0) + \widetilde{p}_{k,\theta}(\mathbf{Z})\widetilde{V}_{k,\theta}(t\,|\,\mathbf{Z}) \tag{178}$$

is a mixture of the CDFs $I(t\ge 0)$ and $\widetilde{V}_{k,\theta}(t\,|\,\mathbf{Z}) \equiv \int_{s=-\infty}^{t}\widetilde{v}_{k,\theta}(s)\,\mathrm{d}s$. Compared to $H_k(\cdot\,|\,\mathbf{Z})$, the exponential twist $\widetilde{H}_{k,\theta}(\cdot\,|\,\mathbf{Z})$ shifts the original distribution's mass to the right when $\theta > 0$, making large losses more likely. Also, setting $\theta = 0$ leads to $\widetilde{H}_{k,0}(\cdot\,|\,\mathbf{Z}) = H_k(\cdot\,|\,\mathbf{Z})$.

For a given θ, we can generate an observation of $T_k \sim \widetilde{H}_{k,\theta}(\cdot\,|\,\mathbf{Z})$ as follows. With probability $\widetilde{q}_{k,\theta}(\mathbf{Z})$, we set $T_k = 0$; otherwise (with probability $\widetilde{p}_{k,\theta}(\mathbf{Z})$), we generate $T_k \sim \widetilde{V}_{k,\theta}(\cdot\,|\,\mathbf{Z})$.

G.1.2. Exponential Twist to $F(\cdot \mid \mathbf{Z})$ We will now show that given $\mathbf{Z}$, exponentially twisting the conditional distribution $F(\cdot \mid \mathbf{Z})$ of $Y = \sum_{k=1}^m T_k$ with twisting parameter θ is equivalent to twisting each $H_k(\cdot \mid \mathbf{Z})$ with the same θ. Note that (167) and (168) imply that the CMGF of Y given $\mathbf{Z}$ satisfies

$$m_F(\theta, \mathbf{Z}) = E\left[\prod_{k=1}^m e^{\theta T_k} \mid \mathbf{Z}\right] = \prod_{k=1}^m E[e^{\theta T_k} \mid \mathbf{Z}] = \prod_{k=1}^m m_{H_k}(\theta, \mathbf{Z}) = \prod_{k=1}^m (1 + \dot{p}_k(\mathbf{Z})[m_{V_k}(\theta, \mathbf{Z}) - 1])$$

by (173). Recall that $\psi_{H_k}(\theta, \mathbf{Z})$ is the CCGF of $T_k \sim H_k(\cdot \mid \mathbf{Z})$, so the CCGF of $Y \sim F(\cdot \mid \mathbf{Z})$ is

$$\psi_F(\theta, \mathbf{Z}) = \sum_{k=1}^m \psi_{H_k}(\theta, \mathbf{Z}) = \sum_{k=1}^m \ln\left(1 + \dot{p}_k(\mathbf{Z})[m_{V_k}(\theta, \mathbf{Z}) - 1]\right). \tag{179}$$

Then by (171), we can rewrite the conditional likelihood ratio in (170) as

$$L_\theta(T_1, \ldots, T_m, \mathbf{Z}) = e^{\psi_F(\theta, \mathbf{Z}) - \theta Y} \equiv L'_\theta(Y, \mathbf{Z}), \tag{180}$$

which depends on $T_1, \ldots, T_m$ through only their sum Y. Hence, given $\mathbf{Z}$, exponentially twisting $F(\cdot \mid \mathbf{Z})$ with θ is equivalent to applying an exponential twist to each $H_k(\cdot \mid \mathbf{Z})$ with the same θ.

Next we will show how we choose the twisting parameter θ given $\mathbf{Z}$. Let $\widetilde{F}_\theta(\cdot \mid \mathbf{Z})$ be the CCDF of Y given $\mathbf{Z}$ under an exponential twist with parameter θ. The conditional expectation $\widetilde{E}_\theta[Y \mid \mathbf{Z}]$ of $Y \sim \widetilde{F}_\theta(\cdot \mid \mathbf{Z})$ under IS given $\mathbf{Z}$ with twisting parameter θ satisfies (e.g., see p. 261 of Glasserman (2004))

$$\widetilde{E}_\theta[Y \mid \mathbf{Z}] = \psi'_F(\theta, \mathbf{Z}) \equiv \frac{\partial}{\partial \theta} \psi_F(\theta, \mathbf{Z}), \tag{181}$$

Also, by (179), we have that for $m'_{V_k}(\theta, \mathbf{Z}) = \frac{\partial}{\partial \theta} m_{V_k}(\theta, \mathbf{Z})$,

$$\psi'_F(\theta, \mathbf{Z}) = \sum_{k=1}^m \frac{\dot{p}_k(\mathbf{Z}) m'_{V_k}(\theta, \mathbf{Z})}{1 + \dot{p}_k(\mathbf{Z})(m_{V_k}(\theta, \mathbf{Z}) - 1)}. \tag{182}$$

Given $\mathbf{Z}$ and the threshold x in $P(Y > x)$ being estimated, we choose parameter $\theta = \theta_x(\mathbf{Z})$ as follows:

$$\begin{array}{c} \text{let } \theta_x(\mathbf{Z}) = 0 \text{ when } x \le \psi'_F(0, \mathbf{Z}); \\ \text{solve for } \theta_x(\mathbf{Z}) \text{ in } \psi'_F(\theta_x(\mathbf{Z}), \mathbf{Z}) = x \text{ when } x > \psi'_F(0, \mathbf{Z}). \end{array} \tag{183}$$

Here $\psi'_F(\theta_x(\mathbf{Z}), \mathbf{Z}) = \widetilde{E}_{\theta_x(\mathbf{Z})}[Y \mid \mathbf{Z}]$ in (181), and $\psi'_F(0, \mathbf{Z}) = \widetilde{E}_0[Y \mid \mathbf{Z}] = E[Y \mid \mathbf{Z}]$, the original conditional mean (without exponential twisting). The conditional event $\{Y > x \mid \mathbf{Z}\}$ is typically not rare when $x \le \psi'_F(0, \mathbf{Z})$, so we do not need IS in this case, and (183) lets $\theta_x(\mathbf{Z}) = 0$. But when the original conditional mean $\psi'_F(0, \mathbf{Z}) < x$, we choose the twisting parameter $\theta_x(\mathbf{Z})$ in (183) so that the conditional mean of Y given $\mathbf{Z}$ under IS equals the threshold x, making the event $\{Y > x \mid \mathbf{Z}\}$ not rare under the IS measure.

G.2. Two-Step IS to Estimate $P(Y > x)$

In this section, we will extend the one-step IS conditional on $\mathbf{Z}$ of Appendix G.1 to estimate the unconditional tail probability λ_x by adapting the two-step IS of Glasserman and Li (2005). To do this, Appendix G.2.1 will first specify a new joint CDF $\Gamma_x(\cdot)$ for sampling $\mathbf{Z}$ under IS. Then for a generated $\mathbf{Z} \sim \Gamma_x(\cdot)$, Appendix G.2.2 will apply the conditional IS from Appendix G.1 on the observed $\mathbf{Z}$, to estimate λ_x.

G.2.1. Specifying the Joint CDF of Z Under IS In this section we will discuss how to choose the new joint CDF for $\mathbf{Z}$ under IS. Let $\Phi_{\mathbf{0}}$ be original joint CDF of vector $\mathbf{Z}$, which has r i.i.d. $N(0,1)$ components, so $\mathrm{d}\Phi_{\mathbf{0}}(\mathbf{z}) = (2\pi)^{-r/2}\exp(-\frac{1}{2}\mathbf{z}^\top\mathbf{z})\,\mathrm{d}\mathbf{z}$. Define the new CDF (not necessarily joint normal) for $\mathbf{Z}$ as $\Gamma_x(\cdot)$, which may depend on the threshold x, satisfying $\mathrm{d}\Gamma_x(\mathbf{z}) > 0$ whenever $\lambda_x(\mathbf{z})\,\mathrm{d}\Phi_{\mathbf{0}}(\mathbf{z}) > 0$. Let $\widetilde{E}_{\Gamma_x}$ be the expectation operator when $\mathbf{Z} \sim \Gamma_x(\cdot)$. Applying a change of measure to $\lambda_x = E\left[\lambda_x(\mathbf{Z})\right]$ leads to

$$\lambda_x = \int_{\mathbf{z}\in\Re^r} \lambda_x(\mathbf{z})\,\mathrm{d}\Phi_{\mathbf{0}}(\mathbf{z}) = \int_{\mathbf{z}\in\Re^r} \lambda_x(\mathbf{z})\frac{\mathrm{d}\Phi_{\mathbf{0}}(\mathbf{z})}{\mathrm{d}\Gamma_x(\mathbf{z})}\,\mathrm{d}\Gamma_x(\mathbf{z}) = \widetilde{E}_{\Gamma_x}\left[\lambda_x(\mathbf{Z})\frac{\mathrm{d}\Phi_{\mathbf{0}}(\mathbf{Z})}{\mathrm{d}\Gamma_x(\mathbf{Z})}\right]. \tag{184}$$

Thus, sampling i.i.d. copies of $\mathbf{Z} \sim \Gamma_x(\cdot)$ and averaging the values of $\lambda_x(\mathbf{Z})\frac{\mathrm{d}\Phi_{\mathbf{0}}(\mathbf{Z})}{\mathrm{d}\Gamma_x(\mathbf{Z})}$ produces an unbiased estimator of λ_x. Ideally, we would like the optimal choice of $\Gamma_x(\cdot)$ to minimize the variance of the estimator.

Now consider $\Gamma^*_x(\cdot)$ defined by

$$\mathrm{d}\Gamma^*_x(\mathbf{z}) \equiv \frac{\lambda_x(\mathbf{z})\,\mathrm{d}\Phi_{\mathbf{0}}(\mathbf{z})}{\lambda_x}, \tag{185}$$

and $\Gamma^*_x(\cdot)$ is a CDF because $\lambda_x(\mathbf{z}) \geq 0$ and $\int_{\mathbf{z}\in\Re^r}\mathrm{d}\Gamma^*_x(\mathbf{z}) = 1$ by (184). If we let $\Gamma_x(\cdot) = \Gamma^*_x(\cdot)$ in (184) and sample $\mathbf{Z} \sim \Gamma^*_x(\cdot)$, then the quantity in the right-hand expectation in (184) always satisfies $\lambda_x(\mathbf{Z})\frac{\mathrm{d}\Phi_{\mathbf{0}}(\mathbf{Z})}{\mathrm{d}\Gamma^*_x(\mathbf{Z})} = \lambda_x$ by (185). Hence, the estimator has zero variance, making $\Gamma^*_x(\cdot)$ the optimal (minimum variance) choice of $\Gamma_x(\cdot)$ to estimate λ_x, as is well known (e.g., see p. 256 of Glasserman (2004)). But we cannot implement $\Gamma^*_x(\cdot)$ in practice because it requires knowing λ_x, which is what we want to estimate.

However, $\Gamma^*_x(\cdot)$ defined by (185) suggests properties of a "good" choice for $\Gamma_x(\cdot)$. For example, we would like to select $\Gamma_x(\cdot)$ such that $\mathrm{d}\Gamma_x(\cdot)$ is large (resp., small) when $\lambda_x(\cdot)\,\mathrm{d}\Phi_{\mathbf{0}}(\cdot)$ is large (resp., small). A simple heuristic approach that roughly tries to achieve this chooses the CDF $\Gamma_x(\cdot)$ in (184) from within a particular parametric family so that its density has the same mode as $\mathrm{d}\Gamma^*_x(\cdot)$. Specifically, we let $\Gamma_x(\cdot) = \Phi_\nu(\cdot)$, which is the joint CDF of r independent normal components, with mean vector $\nu = (\nu_1, \ldots, \nu_r)^\top$ and unit marginal variances. We want to specify ν so that the mode of $\mathrm{d}\Phi_\nu(\mathbf{z})$, which is at $\mathbf{z} = \nu$, occurs at the same location as the mode of $\mathrm{d}\Gamma^*_x(\mathbf{z})$.

But another issue arises: $\lambda_x(\cdot)$ in (185) is also unknown. To handle this, Glasserman and Li (2005) consider replacing $\lambda_x(\cdot)$ with one of several different approximations. We use one of their approaches, which substitutes $\lambda_x(\mathbf{z})$ with the tail probability at threshold x of a (univariate) normal distribution $N(\eta(\mathbf{z}), \sigma^2_Y(\mathbf{z}))$, where $\eta(\mathbf{z}) \equiv E\left[Y \mid \mathbf{Z} = \mathbf{z}\right]$ and $\sigma^2_Y(\mathbf{z}) \equiv \mathrm{Var}[Y \mid \mathbf{Z} = \mathbf{z}]$. By (167), (168), and (172), we have

$$\eta(\mathbf{z}) = \sum_{k=1}^m E\left[C_k \,|\, \mathbf{Z} = \mathbf{z}\right] E\left[D_k \,|\, \mathbf{Z} = \mathbf{z}\right] = \sum_{k=1}^m E\left[C_k \,|\, \mathbf{Z} = \mathbf{z}\right] \dot{p}_k(\mathbf{z}), \quad \text{and}$$

$$\sigma^2_Y(\mathbf{z}) = \sum_{k=1}^m \mathrm{Var}[C_k D_k \,|\, \mathbf{Z} = \mathbf{z}] = \sum_{k=1}^m \left(E[C_k^2 \,|\, \mathbf{Z} = \mathbf{z}]\dot{p}_k(\mathbf{z}) - E^2[C_k \mid \mathbf{Z} = \mathbf{z}]\dot{p}_k^2(\mathbf{z})\right).$$

Thus, we approximate $\lambda_x(\mathbf{z})$ in (185) by $\lambda^\dagger_x(\mathbf{z}) \equiv 1 - \Phi\left(\frac{x - \eta(\mathbf{z})}{\sigma_Y(\mathbf{z})}\right)$. The mode-matching heuristic identifies

$$\mathbf{z}^\dagger_x \equiv \operatorname*{arg\,max}_{\mathbf{z}\in\Re^r}\left[\lambda^\dagger_x(\mathbf{z})\,\mathrm{d}\Phi_{\mathbf{0}}(\mathbf{z})\right] = \operatorname*{arg\,max}_{\mathbf{z}\in\Re^r}\left[\lambda^\dagger_x(\mathbf{z})e^{-\mathbf{z}^T\mathbf{z}/2}\right], \tag{186}$$

which we can try to compute using numerical optimization methods. (Our simulation experiments employed scipy.optimize with method COBYLA on a few (1 or 2) randomly generated starting points.) Finally, the new joint CDF for $\mathbf{Z}$ under IS is $\Gamma_x(\cdot) = \Phi_\nu(\cdot)$ with mean $\nu = \mathbf{z}^\dagger_x$.

G.2.2. Applying Two-Step IS to Estimate $P(Y > x)$ Now that the joint CDF of $\mathbf{Z}$ under IS has been specified as Φ_ν, this section will extend the one-step IS of Appendix G.1 to a two-step IS to estimate the unconditional tail probability λ_x. We sample $\mathbf{Z}$ under IS from Φ_ν, and $\mathrm{d}\Phi_\nu(\mathbf{z}) = (2\pi)^{-d/2}\exp(-\frac{1}{2}(\mathbf{z}-\nu)^\top(\mathbf{z}-\nu)\,\mathrm{d}\mathbf{z}$. Letting $\Gamma_x(\cdot) = \Phi_\nu(\cdot)$ in (184) results in

$$\lambda_x = \widetilde{E}_\nu[\,\lambda_x(\mathbf{Z})\,L^*_\nu(\mathbf{Z})\,], \tag{187}$$

where $\widetilde{E}_\nu$ is the expectation operator when $\mathbf{Z} \sim \Phi_\nu$, and the likelihood ratio

$$L^*_\nu(\mathbf{Z}) = \frac{\mathrm{d}\Phi_{\mathbf{0}}(\mathbf{Z})}{\mathrm{d}\Phi_\nu(\mathbf{Z})} = \exp\left(\frac{1}{2}\nu^\top\nu - \nu^\top\mathbf{Z}\right) \tag{188}$$

corresponds to IS for only $\mathbf{Z}$. Putting (170) with $\theta = \theta_x(\mathbf{Z})$ into (187) then gives

$$\begin{aligned}
\lambda_x &= \widetilde{E}_\nu\left[\widetilde{E}_{\theta_x(\mathbf{Z})}\left[I(\textstyle\sum_{k=1}^m T_k > x)\,L_{\theta_x(\mathbf{Z})}(T_1,\ldots,T_m,\mathbf{Z})\mid\mathbf{Z}\right]L^*_\nu(\mathbf{Z})\right]\\
&= \widetilde{E}_\nu\left[\widetilde{E}_{\theta_x(\mathbf{Z})}\left[I(Y > x)\,L'_{\theta_x(\mathbf{Z})}(Y,\mathbf{Z})\,L^*_\nu(\mathbf{Z})\mid\mathbf{Z}\right]\right]\\
&= \widetilde{E}^*_\nu\left[I(Y > x)\,L'_{\theta_x(\mathbf{Z})}(Y,\mathbf{Z})\,L^*_\nu(\mathbf{Z})\right]
\end{aligned} \tag{189}$$

by (180) and using iterated expectations, with $\widetilde{E}^*_\nu$ as the expectation corresponding to two-step IS, where we first generate $\mathbf{Z} \sim \Phi_\nu$, and then given $\mathbf{Z}$, we generate Y from the conditional distribution $\widetilde{F}_{\theta_x(\mathbf{Z})}(\cdot\mid\mathbf{Z})$ with twisting parameter $\theta_x(\mathbf{Z})$ in (183). Thus, the right side of (189) shows that using this two-step IS approach leads to $I(Y > x)\cdot L'_{\theta_x(\mathbf{Z})}(Y,\mathbf{Z})\cdot L^*_\nu(\mathbf{Z})$ as an unbiased estimator of $P(Y > x)$ based on a single run.

We now detail the two-step IS to estimate λ_x with multiple runs. We first

0. Compute the mean ν for the CDF Φ_ν of $\mathbf{Z}$ under IS as $\nu = \mathbf{z}^\dagger_x$ from (186).

We execute step 0 only once. For a sample size n, do the following in each run $i = 1, 2, \ldots, n$:

1. Generate $\mathbf{Z}_i \sim \Phi_\nu$.
2. Compute the twisting parameter $\theta_i = \theta_x(\mathbf{Z}_i)$ using (183), for $\psi'_F(0,\mathbf{Z}_i)$ in (182).
3. Given $\mathbf{Z}_i$, for each obligor $k = 1, 2, \ldots, m$, if $\theta_i = 0$, generate $T_{k,i}$ from $H_k(\cdot\mid\mathbf{Z}_i)$ in (174); else (when $\theta_i > 0$), generate $T_{k,i}$ from $\widetilde{H}_{k,\theta_i}(\cdot\mid\mathbf{Z}_i)$ in (178).
4. Compute $Y_i = \sum_{k=1}^m T_{k,i}$, which has CCDF $\widetilde{F}_{\theta_i}(\cdot\mid\mathbf{Z}_i)$.
5. Compute $L'_{\theta_i}(Y_i,\mathbf{Z}_i)$ and $L^*_\nu(\mathbf{Z}_i)$ using (180) and (188).

After completing all n runs, we obtain an unbiased estimator of λ_x as

$$\widehat{\lambda}^*_n \equiv \frac{1}{n}\sum_{i=1}^n I(Y_i > x)L'_{\theta_i}(Y_i,\mathbf{Z}_i)L^*_\nu(\mathbf{Z}_i).$$

G.3. Two-step IS to Estimate Quantile

Now we adapt the two-step IS method for estimating the unconditional tail probability $\lambda_x = P(Y > x)$ to instead estimate the p-quantile ξ, which equals the threshold x satisfying $P(Y > x) = 1 - p$. The two-step IS approach of Appendix G.2.2 to estimate λ_x for some fixed x critically depends on knowing the threshold x. As the p-quantile ξ is unknown, we cannot directly apply this IS method with $x = \xi$ to estimate ξ. Instead, we run pilot simulations (with small sample size n_0) with the two-step IS method at a small number j_0 of thresholds x_j, $j = 1, 2, \ldots, j_0$, estimating the tail probability λ_{x_j} for each j, and then interpolate to obtain a crude approximation $\mathring{\xi}$ to the quantile. Then we run additional simulations using the two-step IS approach of Appendix G.2.2 for estimating λ_x with $x = \mathring{\xi}$, and use the resulting data to estimate ξ.

In our experiment, we implemented the approximation method for the p-quantile via three steps:

1. Let $x_j = (1-\alpha_p^j)y^*$ for $j = 1,\ldots,j_0$, where $0 < \alpha_p < 1$ is a constant that may depend on p and other model parameters, and y^* is the maximum possible loss, which is assumed known. Our simulation experiments use $\alpha_p = 0.95$ and $j_0 = 5$. Also, $y^* = \sum_{k=1}^{m} \beta_k$ as our experiments have the LGD $C_k \sim \text{Unif}(0,\beta_k)$.

2. For each $j = 1,\ldots,j_0$, use x_j as the threshold in the two-step IS algorithm of Appendix G.2.2 with sample size n_0 to obtain an estimate $\widehat{\lambda}_{x_j}$ of the tail probability $\lambda_{x_j} = P(Y > x_j)$. We let $n_0 = 100$ in our simulation experiments.

3. Find the $j^* \in \{1,2,\ldots,j_0-1\}$ such that $\widehat{\lambda}_{x_{j^*}} \le 1-p < \widehat{\lambda}_{x_{j^*}+1}$, and use log-interpolation on $(x_{j_*}, \widehat{\lambda}_{x_{j^*}})$ and $(x_{j_*+1}, \widehat{\lambda}_{x_{j^*}+1})$ to obtain $\mathring{\xi}$ as our p-quantile approximation. If $1-p$ is not between any pair of the $\widehat{\lambda}_{x_j}$, we may need to alter α_p to end up with $\widehat{\lambda}_{x_{j^*}} \le 1-p < \widehat{\lambda}_{x_{j^*}+1}$ for some j^*.

After obtaining the quantile approximation $\mathring{\xi}$, to implement the two-step IS to estimate ξ, apply steps 0–5 of the algorithm in Appendix G.2.2 with threshold $x = \mathring{\xi}$ and sample size n, resulting in outputs $(Y_i, \mathbf{Z}_i, \theta_i)$, $i = 1,2,\ldots,n$. Then we can compute the IS p-quantile estimator via the algorithm described after (14) as follows. Let $Y_{1:n} \le Y_{2:n} \le \cdots \le Y_{n:n}$ be the sorted values of $Y_1, Y_2, \ldots, Y_n$. Also, let $L_{i::n} = L'_{\theta_j}(Y_j, \mathbf{Z}_j) L^*_\nu(\mathbf{Z}_j)$ for $(Y_j, \mathbf{Z}_j, \theta_j)$ corresponding to $Y_{i:n}$. Finally, our IS p-quantile estimator is $\widehat{\xi}_{\text{IS},n} = Y_{i_p:n}$, where i_p is the greatest integer for which $\sum_{\ell=i_p}^{n} L_{i::n} \ge n(1-p)$.

H: Main Notation and Definitions

SRS	: simple random sampling (§ 3).
$\text{IS}, \text{IS}(\theta)$	: importance sampling (§ 4), IS with twisting parameter θ (§ 6.2).
MSIS	: measure-specific importance sampling (§ 5.1).
ISDM	: IS with defensive mixture (§ 5.2).
DE	: double estimator (§ 5.3).
$\mathbf{X} \sim G$	: input random vector with joint CDF G (§ 2).
$Y = c(\mathbf{X}) \sim F$	: loss, with CDF F and density f, computed as function c of $\mathbf{X}$ (§ 2).
$\mu = E[Y]$	: mean loss (§ 2).
$\xi = F^{-1}(p)$	: p-quantile (value-at-risk) of F (§ 2).
$\eta = \xi - \mu$	: economic capital (EC) (§ 2).
$P_{G^\dagger}, E_{G^\dagger}, \text{Var}_{G^\dagger}, \text{Cov}_{G^\dagger},$	: probability, expectation, variance, and covariance operators when $\mathbf{X} \sim G^\dagger$.
$\mathfrak{M}$	: simulation method $\mathfrak{M}$ = SRS, IS, MSIS, ISDM, DE.
$L(\mathbf{x}), L_\theta(\mathbf{x})$	: likelihood ratio, LR with twisting parameter θ.
$\delta \in (0,1)$	: allocation or mixing parameter between IS and SRS for MSIS, ISDM, DE.
υ_1, υ_2	: DE weights $\in (0,1)$ in eq. (25) for IS estimators of ξ and μ.
υ'_1, υ'_2	: DE weights $\upsilon'_1 = 1-\upsilon_1, \upsilon'_2 = 1-\upsilon_2$ for SRS estimators of ξ and μ.
$\zeta^2_{\mathfrak{M}}$	: asymptotic variance of η estimator using method $\mathfrak{M}$; see (39).
$\kappa^2_{\mathfrak{M}} = \dfrac{\chi^2_{\mathfrak{M}}}{f^2(\xi)}$	: asymptotic variance of ξ estimator using method $\mathfrak{M}$; see (40).

$\sigma^2_{\mathfrak{M}}$: asymptotic variance of μ estimator using method $\mathfrak{M}$.

$\dfrac{\gamma_{\mathfrak{M}}}{f(\xi)}$: covariance in asymptotic variance of η estimator using method $\mathfrak{M}$.

m : number of summands in i.i.d. sum model (§ 6).

$p_m = 1 - e^{-\beta m}$: quantile level in (29) for i.i.d. sum model (§ 6).

$m\tau_{\mathfrak{M}}$: expected time to generate single output for method $\mathfrak{M}$ (§ 6.1 and 6.2).

$\mathrm{RE}_{\mathfrak{M},m}, \mathrm{WNRE}_{\mathfrak{M},m}$: relative error, work-normalized RE for method $\mathfrak{M}$ (§ 6.3).

$O(\cdot), o(\cdot), \Omega(\cdot), \omega(\cdot), \Theta(\cdot)$: asymptotic (as $m \to \infty$) upper bound, strictly dominant upper bound, lower bound, strictly subdominant lower bound, same order (§ 6.4).

G_0, μ_0, σ_0^2 : marginal CDF, mean, and variance of each i.i.d. summand X_j (§ 6.1).

$M_0(\theta), M_0'(\theta), M_0''(\theta)$: moment generating function (MGF) of G_0 and first, second derivatives (§ 6.1).

$Q_0(\theta), Q_0'(\theta), Q_0''(\theta)$: cumulant generating function (CGF) of G_0 and first, second derivatives (§ 6.1).

Δ, Δ° : domain of M_0 and its interior (Assumption 1).

$\widetilde{G}, \widetilde{G}_\theta, \widetilde{G}_{0,\theta}$: CDFs under IS and exponential twisting for $\mathbf{X}$ and X_j.

$\theta_\star, \theta_m$: twisting parameters in (32) and (94).

$\breve{\xi} = mQ_0'(\theta_\star), \breve{f}$: quantile approximation in (85), saddlepoint approximation to f in (105).

ARHW : average relative half width (§ 7.2).

RMSRE : root-mean-squared relative error (§ 7.2).

PCRM : portfolio credit risk model (§ 7.2).

$\mathbf{Z}$: systematic risk factors in PCRM (§ 7.2).

ϵ_k : idiosyncratic risk of obligor k (§ 7.2).

$\mathbf{a}_k$: loading factors of obligor k (§ 7.2).

$\dot{p}_k$: marginal default probability of obligor k (§ 7.2).

LGD C_k : loss given default of obligor k (§ 7.2).

D_k : default indicator of obligor k (§ 7.2).

$\lambda_x = P(Y > x)$: tail probability of total loss Y (§ 7.2, Appendix G).

$T_k = C_k D_k$: unconditional loss of obligor k (Appendix G.1).

$H_k(\cdot \mid \mathbf{Z}), \widetilde{H}_{k,\theta}(\cdot \mid \mathbf{Z})$: conditional CDF of T_k of obligor k given $\mathbf{Z}$, original and with twist (App. G.1).

$\dot{p}_k(\mathbf{Z}) = P(D_k = 1 \mid \mathbf{Z})$: conditional default probability of obligor k given $\mathbf{Z}$ (Appendix G.1).

$m_{H_k}(\theta, \mathbf{Z})$: conditional MGF of T_k given $\mathbf{Z}$ of obligor k given $\mathbf{Z}$ (Appendix G.1).