Statistics and Computing (2024) 34:65
https://doi.org/10.1007/s11222-023-10374-5

ORIGINAL PAPER

Efficient exponential tilting with applications

Cheng-Der Fuh[1] · Chuan-Ju Wang[2]

Received: 4 February 2023 / Accepted: 12 December 2023 / Published online: 13 January 2024

Abstract
To minimize the variance of Monte Carlo estimators, we develop a novel exponential embedding technique that extends the classical concept of sufficient statistics in importance sampling. Our method demonstrates bounded relative error and logarithmic efficiency when applied to normal and gamma distributions, especially in rare event scenarios. To illustrate this innovative technique, we address the problem of credit risk measurement in portfolios and present an efficient simulation algorithm to estimate the likelihood of significant portfolio losses, leveraging multi-factor models with a normal mixture copula. Finally, supported by comprehensive simulation studies, our approach offers a more effective and efficient way to simulate moderately rare events.

Keywords Simulation · Variance reduction · Importance sampling · Portfolio credit risk · Copula models

1 Introduction

Many scientific and statistical applications involve calculating a small probability and/or expectation of a complicated function of a given random variable, as discussed in Asmussen and Glynn (2007). Often, the statistic in question is too complex for an analytical solution, necessitating the use of Monte Carlo simulation methods. More specifically, consider a probability space $(\Omega, \mathscr{F}, P)$. Here, Ω is the sample space corresponding to the outcomes of a certain experiment (which may be hypothetical), $\mathscr{F}$ is the σ-algebra of subsets of Ω, and P is a probability measure defined on these subsets. Now let $X = (X_1, \ldots, X_d)^\intercal$ be a d-dimensional random vector on a probability space $(\Omega, \mathscr{F}, P)$ and let $\mathscr{P}(\cdot)$ be a real-valued function from $\mathbb{R}^d$ to $\mathbb{R}$, where $^\intercal$ denotes the transpose. The problem of interest is to calculate a functional of the form

$$m = E_P[\mathscr{P}(X)], \tag{1}$$

where $E_P[\cdot]$ is the expectation operator under the probability measure P. When $\mathscr{P}(X)$ is an indicator function $\mathbb{1}_{\{X \in A\}}$ for $A \in \mathscr{F}$, then $m = P\{X \in A\}$.

Due to the complexity of estimating such a complicated function, analytical calculation of $E_P[\mathscr{P}(X)]$ is often unavailable, and one must apply simulation methods. However, naive Monte Carlo methods are inefficient in computing small probabilities because the estimator's variance gradually dominates its mean, such that many repeated trials are needed to achieve acceptable accuracy. Consequently, importance sampling based on exponential tilting was proposed by Siegmund (1976) as a variance reduction technique in rare-event simulation for the case of independent and identically distributed (i.i.d.) random variables.[1] The basic idea is to sample under an alternative probability and adjust by the likelihood ratio (Radon–Nikodym derivative). Usually, the alternative distribution is designed based on a certain criterion within a family of parameterized distributions. Introducing this change of measure always yields an unbiased estimator. The likelihood ratio function must be positive and hence is usually constructed as the exponential function. This family of alternative distributions is also known as the

✉ Chuan-Ju Wang
cjwang@citi.sinica.edu.tw

Cheng-Der Fuh
cdffuh@gmail.com

1 Graduate Institute of Statistics, National Central University, No. 300 Zhongda Road, Zhongli District, Taoyuan City 320317, Taiwan

2 Research Center for Information Technology Innovation, Academia Sinica, No. 128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan

[1] The idea of importance sampling appeared even earlier, e.g., in Ulam etal. (1947), Kahn and Harris (1951).

exponential tilting family. For various extensions and applications, see, e.g., Glynn and Iglehart (1989), Sadowsky and Bucklew (1990), Ben Rached etal. (2018) ,Owen etal. (2017) and references therein. A comprehensive introduction can be found in Asmussen and Glynn (2007) and Owen (2013).

From a theoretical point of view for importance sampling, one must propose algorithms for the efficient simulation of the right tail of a given random variable. In the case of distributions with light right tails (i.e., those that decay at an exponential rate or faster), under some regularity assumptions, the standard exponential tilting importance sampling estimator, cf. Asmussen and Glynn (2007), satisfies the logarithmic efficiency property, which is a useful metric used to assess the efficiency of an estimator. In contrast, for heavy-tailed distributions, such as the example of log normals and Weibulls with shape parameters strictly less than 1, the exponential tilting method cannot be applied. Therefore, efficient algorithms have been developed to estimate tail probabilities involving heavy tails. In this regard, Asmussen and Binswanger (1997) provide the first logarithmic efficiency for such probabilities. Asmussen and Kroese (2006) propose an estimator with a bounded relative error under distributions with regularly varying tails.

In addition than the independent importance sampling algorithm, more complex state-dependent importance sampling has been proposed in the literature over the last few years to estimate probabilities for sums of heavy-tailed independent random variables, (e.g., Dupuis and Wang 2004; Dupuis etal. 2007; Blanchet and Liu 2008; Blanchet and Li 2011; Blanchet and Lam 2012; Amar etal. 2023). Some researchers address the left-tail region, that is, the probability that sums of nonnegative random variables fall below a sufficiently small threshold (e.g., Asmussen etal. 2016; Ben Issaid etal. 2017; Ben Rached etal. 2015, 2018, 2020a, b, 2021; Fuh etal. 2023a).

Applications of importance sampling have been given in other fields in the literature. To name a few, the improvement of the convergence rate in stochastic gradient-based optimization algorithms has been given in Zhao and Zhang (2015), Richtárik and Takáč (2016), Johnson and Guestrin (2018),Katharopoulos and Fleuret (2018),Csiba and Richtárik (2018), and Fuh etal. (2023b). Metelli etal. (2018) and Metelli etal. (2020) apply importance sampling in the target policy.

A useful tool in importance sampling for rare event simulation ($m = P\{X \in A\}$) is exponential tilting (e.g., Siegmund 1976; Sadowsky and Bucklew 1990), for which the above-mentioned algorithm is more efficient for large-deviation rare events (m of the order 10^{-5} or less) than arbitrary events (say, m is not rare). Examples of such events occur in telecommunications (m = bit-loss rate, probability of buffer overflow) and reliability (m = the probability of failure before time t). Efficient Monte Carlo simulation of such events is proposed by Sadowsky and Bucklew (1990) based on the large deviations theory given by Ney (1983).

Other than the above-mentioned extremely rare event simulation, in this paper, we are interested in simulating moderately rare events, that is, when $m = P\{X \in A\}$ is small, say, of the order 10^{-2} or 10^{-3} or so; i.e., $\{X \in A\}$ is a moderate-deviation rare event. Such problems appear in the construction of confidence regions for asymptotically normal statistics (e.g., Efron and Tibshirani 1994), and in the computation of value-at-risk (VaR) in risk management (e.g., Duffie and Singleton 2003; Glasserman etal. 2000, 2002; Fuh etal. 2011). For the calculation of VaR in financial risk management, the reduction in terms of the computational time is essential, cf. Chapter 9 of Glasserman (2004). For a more general $E_P[\mathscr{P}(X)]$ such as portfolio loss (e.g., Li 2000; Glasserman and Li 2005; Glasserman etal. 2007, 2008; Bassamboo etal. 2008; Chan and Kroese 2010), see Sect. 3 for details.

To provide a more accurate simulation algorithm for a moderately rare event, instead of using large-deviation exponential tilting, an alternative choice of the tilting parameter is based on the criterion of minimizing the variance of the importance sampling estimator directly. For example, Johns (1988), Do and Hall (1991), Fuh and Hu (2004) study efficient simulations for bootstrapping the sampling distribution of statistics of interest. Su and Fu (2000), Fu and Su (2002) minimize the variance under the original probability measure. Fuh etal. (2011) apply the importance sampling method for VaR computation under a multivariate t-distribution, in which the authors also show that their proposed method consistently outperforms, in the sense of variance reduction, existing methods derived from large deviations theory under various settings.

Our criterion in the search for desirable tilting parameters is to minimize the variance of the estimator. Numerically, this procedure is based on a pre-sampling stochastic approximation algorithm. This problem has been well-studied in Egloff etal. (2005), Kawai (2009), and Kawai (2018). Overall our algorithm has two stages. We first search the desired tilting parameters and then obtain samples under the alternative distribution. This is similar to adaptive importance sampling with stochastic approximation in Ahamed etal. (2006) in the sense that we both seek to use stochastic approximation to learn the optimal sampling distribution. However, our methods treat the search step as a separate stage other than an adaptive online procedure. Moreover, the proposed exponential tilting family is based on sufficient statistics.

Note that for events of large deviations $P\{X \in A = (a, \infty)\}$ for some $a > 0$, Sadowsky and Bucklew (1990) show that the asymptotically optimal alternative distribution is obtained through exponential tilting; that is, $Q(dx) = C \exp(\theta x) P(dx)$, where C is a normalizing constant and θ determines the amount of tilting. The optimal amount of tilting θ is such that the expectation of X under the Q-measure equals the dominating point, which is located at the boundary a of A. However, for a moderately deviating rare event, Do and Hall (1991) and Fuh and Hu (2004) show that typically the tilting point θ^* of the optimal alternative distribution is inside (a, ∞). This is different from that given by large deviations theory. However, it is shown in Fuh and Hu (2004) that $\theta^* - a$ approaches 0 as $a \to \infty$; see Example 1 in Sect. 2.

Along this line, the proposed exponential embedding is based on the idea of *sufficient statistics*. To illustrate the concept of *sufficient exponential embedding*, we consider a one-dimensional normal distributed random variable. Let $X \sim N(0, 1)$ be a random variable with the standard normal distribution, with probability density function (pdf) $e^{-x^2/2}/\sqrt{2\pi}$. Then the sufficient exponential embedding is

$$\begin{aligned}\frac{dQ}{dP} &= \frac{\exp\{\theta x + \eta x^2\}}{E[\exp\{\theta x + \eta x^2\}]} \\ &= \sqrt{1-2\eta}\exp\{\eta x^2 + \theta x - \theta^2/(2-4\eta)\}.\end{aligned}$$

Note that here for the normal distribution, the sufficient statistic is $T(x) = [x\ x^2]$. In contrast to the classical one-parameter exponential embedding, the sufficient exponential embedding is based on two parameters (θ, η) corresponding to the sufficient statistic $T(x) = [x\ x^2]$. A systematic study of this innovative technique will be given in Sect. 2, in which we also use the idea of the conjugate measure of Q, $\bar{Q}(dx) = C\exp(-(\theta x + \eta x^2))P(dx)$, to characterize the optimal tilting (θ, η) by solving the equation wherein the expectation of (X, X^2) under the Q-measure equals the conditional expectation of (X, X^2) under the $\bar{Q}$-measure given the rare event. To illustrate this innovative technique, we tackle the complex issue of assessing credit risk in portfolios that include financial instruments such as loans and bonds. We propose a streamlined simulation algorithm to efficiently estimate the probability of substantial portfolio losses.

There are two aspects in this study. To begin with, based on the criterion of minimizing the variance of the importance sampling estimator, we propose an innovative importance sampling algorithm based on an unconventional *sufficient (statistic) exponential embedding*. Here we term this *sufficient exponential tilting* because the form of the embedding is selected based on the sufficient statistic of the underlying distribution, for which more than one parameter (usually two parameters) can be tilted in our method; for instance, the tilting parameters can be the location and scale parameters in the multivariate normal distribution. Theoretical investigations, including bounded relative error and logarithmic efficiency analysis (Theorems 5–6), and numerical studies are given to support the proposed importance sampling method. Note that in Theorem 5 for normal distribution, to simulate $\{X > a\}$ for large $a > 0$, we show that optimal sufficient exponential tilting outperforms traditional optimal one-parameter exponential tilting in the sense of reducing asymptotic variance. In Theorem 6 for gamma distribution, to simulate $\{X > a\}$ and $\{X < 1/a\}$ for large $a > 0$, we show that optimal sufficient exponential tilting outperforms traditional optimal one-parameter exponential tilting. Furthermore, we show that for the normal, multivariate normal, and gamma distributions, our simulation study shows that sufficient exponential tilting performs 2 to 5 times better than classical one-parameter exponential tilting for some simple rare event simulations. Note that the innovative tilting formula is more suitable for the grouped normal mixture copula model, and is of independent interest. In particular, when applying sufficient exponential tilting in the normal mixture models, the tilting parameter can be either the shape or the rate parameter for the underlying gamma distribution, which results in a more efficient simulation.

Next, by utilizing a fast computational method for how the rare event occurs and the proposed importance sampling method, we provide an efficient simulation algorithm for a multi-factor model with the normal mixture copula model to estimate the probability that the portfolio incurs large losses. To be more precise, in this stage, we use the inverse Fourier transform to handle the distribution of total losses, i.e., the sum of n "weighted" independent but non-identically distributed Bernoulli random variables. An automatic variant of Newton's method is introduced to determine the optimal tilting parameters. Note that the proposed simulation device is based on importance sampling for a joint probability other than the conditional probability used in previous studies, which is also shown to achieve variance reduction in the asymptotic sense in terms of bounded relative error (Theorem 7). Finally, to illustrate the applicability of our method, we mention numerical results of the proposed algorithm under various copula settings, and highlight insights into the trade-off between the reduced variances and increased computational time.

The remainder of this paper is organized as follows. Section 2 presents a general account of importance sampling based on sufficient exponential embedding. Section 3 formulates the problem of estimating large portfolio losses, presents the normal mixture copula model, and studies the proposed optimal importance sampling for portfolio loss under the normal mixture copula model. Section 4 concludes. Some proofs and numerical results are deferred to "Appendices A–D".

2 Sufficient exponential tilting

To calculate the value of (1) using importance sampling, one selects a sampling probability measure Q under which X has a pdf $q(x) = q(x_1, \ldots, x_d)$ with respect to the Lebesgue measure $\mathscr{L}$. Q is assumed to be absolutely continuous with respect to the original probability measure P. Therefore, Eq. (1) can be written as

$$\int_{\mathbb{R}^d} \mathscr{P}(x) f(x) dx = \int_{\mathbb{R}^d} \mathscr{P}(x) \frac{f(x)}{q(x)} q(x) dx = \mathrm{E}_Q \left[\mathscr{P}(X) \frac{f(X)}{q(X)} \right], \tag{2}$$

where $\mathrm{E}_Q[\cdot]$ is the expectation operator under which X has a pdf $q(x)$ with respect to the Lebesgue measure $\mathscr{L}$. The ratio $f(x)/q(x)$ is called the importance sampling weight, the likelihood ratio, or the Radon–Nikodym derivative.

Here, we focus on the exponentially tilted probability measure of P. Instead of considering the commonly adopted one-parameter exponential tilting in the literature (Asmussen and Glynn 2007), we propose a sufficient exponential tilting algorithm. To the best of our knowledge, the use of sufficient exponential embedding is novel in the literature. As will be seen in the examples in Sect. 2.1, the tilting probabilities for existent two-parameter distributions, such as the gamma and normal distributions, can be obtained by solving simple formulas.

Let $Q_{\theta,\eta}$ be the tilting probability measure, where subscript $\theta = (\theta_1, \ldots, \theta_p)^\intercal \in \mathbb{R}^p$ and $\eta = (\eta_1, \ldots, \eta_\kappa)^\intercal \in \mathbb{R}^\kappa$ are the tilting parameters. Here p and κ denote the number of parameters, respectively. Let $h_1(x)$ be a function from $\mathbb{R}^d$ to $\mathbb{R}^p$, and $h_2(x)$ be a function from $\mathbb{R}^d$ to $\mathbb{R}^\kappa$. Denote Θ and H as the parameter spaces such that the moment-generating function $\Psi(\theta, \eta) := E[\mathrm{e}^{\theta^\intercal h_1(x) + \eta^\intercal h_2(x)}]$ of $(h_1(X), h_2(X))$ exists for $\theta \in \Theta \subset \mathbb{R}^p$ and $\eta \in H \subset \mathbb{R}^\kappa$. Let $f_{\theta,\eta}(x)$ be the pdf of X under the exponentially tilted probability measure $Q_{\theta,\eta}$, defined by

$$f_{\theta,\eta}(x) = \frac{\mathrm{e}^{\theta^\intercal h_1(x) + \eta^\intercal h_2(x)}}{\Psi(\theta, \eta)} f(x) = \mathrm{e}^{\theta^\intercal h_1(x) + \eta^\intercal h_2(x) - \psi(\theta,\eta)} f(x), \tag{3}$$

where $\psi(\theta, \eta) = \ln \Psi(\theta, \eta)$ is the cumulant function. Note that in (3), we present one type of parameterization which is suitable for our derivation. Explicit representations of $h_1(x)$ and $h_2(x)$ for specific distributions are given in the examples and remarks in Sect. 2.1, which include the normal and multivariate normal distributions as well as the gamma distribution.

Consider sufficient exponential embedding. Equation (2) becomes

$$\int_{\mathbb{R}^d} \mathscr{P}(x) f(x) dx = \int_{\mathbb{R}^d} \mathscr{P}(x) \frac{f(x)}{f_{\theta,\eta}(x)} f_{\theta,\eta}(x) dx = E_{Q_{\theta,\eta}} \left[\mathscr{P}(X) \mathrm{e}^{-(\theta^\intercal h_1(X) + \eta^\intercal h_2(X)) + \psi(\theta,\eta)} \right].$$

Because of the unbiasedness of the importance sampling estimator, its variance is

$$\begin{aligned} &\mathrm{Var}_{Q_{\theta,\eta}} \left[\mathscr{P}(X) \mathrm{e}^{-(\theta^\intercal h_1(X) + \eta^\intercal h_2(X)) + \psi(\theta,\eta)} \right] \\ &= E_{Q_{\theta,\eta}} \left[\left(\mathscr{P}(X) \mathrm{e}^{-(\theta^\intercal h_1(X) + \eta^\intercal h_2(X)) + \psi(\theta,\eta)} \right)^2 \right] - m^2, \end{aligned} \tag{4}$$

where $m = \int_{\mathbb{R}^d} \mathscr{P}(x) f(x) dx$. For simplicity, we assume that the variance of the importance sampling estimator exists. A simple example is that $\mathscr{P}(X)$ is bounded and the parameters (θ, η) lie in the domain discussed as above. For instance, in the one-dimensional Gaussian case with $\mathscr{P}(X) = \mathbb{1}_{\{X>a\}}$ or $\mathscr{P}(X) = \mathbb{1}_{\{X<a\}}$, it is possible to determine the exact domain on which the variance is finite. The same is true in the case where X is a gamma distribution. We will consider these two cases in Theorems 5 and 6.

Define the first term of the right-hand side (RHS) of (4) by $G(\theta, \eta)$. Then minimizing

$$\mathrm{Var}_{Q_{\theta,\eta}} \left[\mathscr{P}(X) \mathrm{e}^{-(\theta^\intercal h_1(X) + \eta^\intercal h_2(X)) + \psi(\theta,\eta)} \right]$$

is equivalent to minimizing $G(\theta, \eta)$. Standard algebra gives a simpler form of $G(\theta, \eta)$:

$$\begin{aligned} G(\theta, \eta) &:= E_{Q_{\theta,\eta}} \left[\left(\mathscr{P}(X) \mathrm{e}^{-(\theta^\intercal h_1(X) + \eta^\intercal h_2(X)) + \psi(\theta,\eta)} \right)^2 \right] \\ &= E_P \left[\mathscr{P}^2(X) \mathrm{e}^{-(\theta^\intercal h_1(X) + \eta^\intercal h_2(X)) + \psi(\theta,\eta)} \right], \end{aligned} \tag{5}$$

which is used to find the optimal tilting parameters. In the following theorem, we show that $G(\theta, \eta)$ in (5) is a convex function in θ and η. This property ensures that there exists no multi-mode problem in the search stage when determining the optimal tilting parameters.

To minimize $G(\theta, \eta)$, the first-order condition requires the solution of θ, η, denoted by θ^*, η^*, to satisfy $\nabla_\theta G(\theta, \eta) \mid_{\theta=\theta^*} = 0$, and $\nabla_\eta G(\theta, \eta) \mid_{\eta=\eta^*} = 0$, where ∇_θ denotes the gradient with respect to θ and ∇_η denotes the gradient with respect to η. Under the condition that X is in an exponential family, and $\psi(\theta, \eta)$ are bounded continuously differentiable functions with respect to θ and η, by the dominated convergence theorem, simple calculation yields

$$\begin{aligned} \nabla_\theta G(\theta, \eta) &= E_P \Big[\mathscr{P}^2(X)(-h_1(X) \\ &\quad + \nabla_\theta \psi(\theta, \eta)) \mathrm{e}^{-(\theta^\intercal h_1(X) + \eta^\intercal h_2(X)) + \psi(\theta,\eta)} \Big], \\ \nabla_\eta G(\theta, \eta) &= E_P \Big[\mathscr{P}^2(X)(-h_2(X) \\ &\quad + \nabla_\eta \psi(\theta, \eta)) \mathrm{e}^{-(\theta^\intercal h_1(X) + \eta^\intercal h_2(X)) + \psi(\theta,\eta)} \Big]; \end{aligned}$$

therefore, (θ^*, η^*) is the root of the following system of nonlinear equations:

$$\nabla_\theta \psi(\theta,\eta) = \frac{E_P\left[\mathscr{P}^2(X)h_1(X)\mathrm{e}^{-(\theta^\intercal h_1(X)+\eta^\intercal h_2(X))}\right]}{E_P\left[\mathscr{P}^2(X)\mathrm{e}^{-(\theta^\intercal h_1(X)+\eta^\intercal h_2(X))}\right]}, \quad (6)$$

$$\nabla_\eta \psi(\theta,\eta) = \frac{E_P\left[\mathscr{P}^2(X)h_2(X)\mathrm{e}^{-(\theta^\intercal h_1(X)+\eta^\intercal h_2(X))}\right]}{E_P\left[\mathscr{P}^2(X)\mathrm{e}^{-(\theta^\intercal h_1(X)+\eta^\intercal h_2(X))}\right]}. \quad (7)$$

To simplify the RHS of (6) and (7), we define the conjugate measure $\bar{Q}^{\mathscr{P}}_{\theta,\eta}$ of the measure Q with respect to the payoff function $\mathscr{P}$ as

$$\frac{d\bar{Q}^{\mathscr{P}}_{\theta,\eta}}{dP} = \frac{\mathscr{P}^2(X)\mathrm{e}^{-(\theta^\intercal h_1(X)+\eta^\intercal h_2(X))}}{E_P[\mathscr{P}^2(X)\mathrm{e}^{-(\theta^\intercal h_1(X)+\eta^\intercal h_2(X))}]}$$
$$= \mathscr{P}^2(X)\mathrm{e}^{-(\theta^\intercal h_1(X)+\eta^\intercal h_2(X))-\bar{\psi}(\theta,\eta)},$$

where $\bar{\psi}(\theta,\eta)$ is $\log\bar{\Psi}(\theta,\eta)$ with $\bar{\Psi}(\theta,\eta) = E_P[\mathscr{P}^2(X)\mathrm{e}^{-(\theta^\intercal h_1(X)+\eta^\intercal h_2(X))}]$. Then the RHS of (6) equals $E_{\bar{Q}^{\mathscr{P}}_{\theta,\eta}}[h_1(X)]$, the expected value of $h_1(X)$ under $\bar{Q}^{\mathscr{P}}_{\theta,\eta}$, and the RHS of (7) equals $E_{\bar{Q}^{\mathscr{P}}_{\theta,\eta}}[h_2(X)]$, the expected value of $h_2(X)$ under $\bar{Q}^{\mathscr{P}}_{\theta,\eta}$.

The following theorem establishes the existence, uniqueness, and characterization for the minimizer of (5). Before that, to ensure the finiteness of the moment-generating function $\Psi(\theta,\eta)$, we add a condition that $\Psi(\theta,\eta)$ is steep, cf. Asmussen and Glynn (2007). To define steepness, let $\theta_i^- = (\theta_1,\ldots,\theta_i,\ldots,\theta_p) \in \Theta$ such that all θ_k are fixed for $k = 1,\ldots,i-1,i+1,\ldots,p$ except the i-th component. Denote $\eta_j^- \in H$ correspondingly for $j = 1,\ldots,\kappa$. Now, let $\theta_{i,\max} := \sup\{\theta_i : \Psi(\theta_i^-,\eta) < \infty\}$ for $i = 1,\ldots,p$, and $\eta_{j,\max} := \sup\{\eta_j : \Psi(\theta,\eta_j^-) < \infty\}$ for $j = 1,\ldots,\kappa$ (for light-tailed distributions, we have $0 < \theta_{i,\max} \le \infty$ for $i = 1,\ldots,p$, and $0 < \eta_{j,\max} \le \infty$ for $j = 1,\ldots,\kappa$). Then steepness means $\Psi(\theta,\eta) \to \infty$ as $\theta_i \to \theta_{i,\max}$ for $i = 1,\ldots,p$, or $\eta_j \to \eta_{j,\max}$ for $j = 1,\ldots,\kappa$.

The following conditions are used in Theorem 1:

i) $\bar{\Psi}(\theta,\eta)\Psi(\theta,\eta) \to \infty$ as $\theta_i \to \theta_{i,\max}$ for $i = 1,\ldots,p$, or $\eta_j \to \eta_{j,\max}$ for $j = 1,\ldots,\kappa$;
ii) $G(\theta,\eta)$ is a continuously differentiable function on $\Theta \times H$, and

$$\max_{i=1,\ldots,p,\ j=1,\ldots,\kappa}\left\{\lim_{\theta_i\to\theta_{i,\max}}\frac{\partial G(\theta,\eta)}{\partial\theta_i}, \lim_{\eta_j\to\eta_{j,\max}}\frac{\partial G(\theta,\eta)}{\partial\eta_j}\right\} > 0. \quad (8)$$

Note that condition i) or ii) is used to guarantee the existence of the minimum point. More details can be found in the proof of Theorem 1.

Theorem 1 *Suppose the moment-generating function $\Psi(\theta,\eta)$ of $(h_1(X), h_2(X))$ exists for $\theta \in \Theta \subset \mathbb{R}^p$ and $\eta \in H \subset \mathbb{R}^\kappa$. Assume $\Psi(\theta,\eta)$ is steep. Furthermore, assume either i) or ii) holds. Then $G(\theta,\eta)$ defined in (5) is a convex function in (θ,η), and there exists a unique minimizer of (5), which satisfies*

$$\nabla_\theta \psi(\theta,\eta) = E_{\bar{Q}^{\mathscr{P}}_{\theta,\eta}}[h_1(X)], \quad (9)$$

$$\nabla_\eta \psi(\theta,\eta) = E_{\bar{Q}^{\mathscr{P}}_{\theta,\eta}}[h_2(X)]. \quad (10)$$

Proof To prove Theorem 1, we require the following three propositions. Proposition 2 is taken from Theorem VI.3.4. of Ellis (1985), Proposition 3 is a standard result from convex analysis, and Proposition 4 is taken from Theorem 1 of Soriano (1994). Note that although the function domain is the whole space in Propositions 3 and 4, the results for a subspace still hold with similar proofs.

Proposition 2 *$f(x)$ is differentiable at $x \in int(\mathscr{X})$ if and only if the d partial derivatives $\frac{\partial f(x)}{\partial x_i}$ for $i = 1,\ldots,d$ exist at $x \in int(\mathscr{X}) \subset \mathbb{R}^d$ and are finite.*

Proposition 3 *Let $f : \mathbb{R}^d \to \mathbb{R}$ be continuous on all of $x \in \mathbb{R}^d$. If f is coercive (in the sense that $f(x) \to \infty$ if $\|x\| \to \infty$), then f has at least one global minimizer.*

Proposition 4 *Let $f : \mathbb{R}^d \to \mathbb{R}$, and let f, a continuously differentiable, convex function, satisfy (8). Then a minimum point for f exists.*

Below, we first show that $G(\theta,\eta)$ is a strictly convex function. For any given $\lambda \in (0,1)$, and $(\theta,\eta), (\theta',\eta') \in \Theta \times H \subset \mathbb{R}^p \times \mathbb{R}^\kappa$, by the convexity of $\psi(\cdot,\cdot)$, we have

$$\begin{aligned}\psi(\lambda\theta + (1-\lambda)\theta', \lambda\eta + (1-\lambda)\eta') &= \psi(\lambda(\theta,\eta)\\ &\quad+(1-\lambda)(\theta',\eta'))\\ &\le \lambda\psi(\theta,\eta) + (1-\lambda)\psi(\theta',\eta'). \quad (11)\end{aligned}$$

Note that $\lambda(\theta,\eta) + (1-\lambda)(\theta',\eta') = (\lambda\theta + (1-\lambda)\theta', \lambda\eta + (1-\lambda)\eta')$ is a $(p+\kappa)$-dimensional vector. Then

$$\begin{aligned}&G(\lambda(\theta,\eta) + (1-\lambda)(\theta',\eta'))\\ &= G(\lambda\theta + (1-\lambda)\theta', \lambda\eta + (1-\lambda)\eta')\\ &= E_p\Big[\mathscr{P}^2(X)\exp\big(-((\lambda\theta + (1-\lambda)\theta')^\intercal h_1(X)\\ &\quad+(\lambda\eta + (1-\lambda)\eta')^\intercal h_2(X))\\ &\quad+\psi(\lambda\theta + (1-\lambda)\theta', \lambda\eta + (1-\lambda)\eta')\big)\Big]\\ &\le E_p\Big[\mathscr{P}^2(X)\exp\big(-((\lambda\theta + (1-\lambda)\theta')^\intercal h_1(X)\\ &\quad+(\lambda\eta + (1-\lambda)\eta')^\intercal h_2(X))\\ &\quad+\lambda\psi(\theta,\eta) + (1-\lambda)\psi(\theta',\eta')\big)\Big] \quad \text{by (11)}\end{aligned}$$

$$
\begin{aligned}
&= E_p\Big[\mathscr{P}^2(X)\exp\big(-\lambda(\theta^\intercal h_1(X)+\eta^\intercal h_2(X)) \\
&\quad +\lambda\psi(\theta,\eta)-(1-\lambda)(\theta'^\intercal h_1(X)+\eta'^\intercal h_2(X)) \\
&\quad +(1-\lambda)\psi(\theta',\eta')\big)\Big] \\
&< E_P\Big[\lambda\mathscr{P}^2(X)e^{-(\theta^\intercal h_1(X)+\eta^\intercal h_2(X))+\psi(\theta,\eta)} \\
&\quad +(1-\lambda)\mathscr{P}^2(X)e^{-(\theta'^\intercal h_1(X)+\eta'^\intercal h_2(X))+\psi(\theta',\eta')}\Big] \\
&= \lambda G(\theta,\eta)+(1-\lambda)G(\theta',\eta').
\end{aligned}
$$

Next, we prove the existence of (θ, η) in the optimization problem (5). To obtain the global minimum of $G(\theta, \eta)$, we note that $G(\theta, \eta)$ is strictly convex from the above argument, and $\frac{\partial G(\theta,\eta)}{\partial \theta_i}$ and $\frac{\partial G(\theta,\eta)}{\partial \eta_i}$ exist for $i = 1, \ldots, p$, $j = 1, \ldots, \kappa$. Proposition 2 establishes that $G(\theta, \eta)$ is continuously differentiable for $(\theta, \eta) \in \Theta \times H$. By the definition of $G(\theta, \eta)$ in (5), it is easy to see that condition i) implies that $G(\theta)$ is coercive. Then by Proposition 3, $G(\theta, \eta)$ has a unique minimizer. Clearly ii) implies that conditions in Proposition 4 hold.

To prove (9) and (10), we simplify the right-hand side of (6) and (7) under $\bar{Q}^{\mathscr{P}}_{\theta,\eta}$. Standard algebra yields

$$
\frac{E_P\left[\mathscr{P}^2(X)h_1(X)\mathrm{e}^{-(\theta^\intercal h_1(X)+\eta^\intercal h_2(X))}\right]}{E_P\left[\mathscr{P}^2(X)\mathrm{e}^{-(\theta^\intercal h_1(X)+\eta^\intercal h_2(X))}\right]} = E_{\bar{Q}^{\mathscr{P}}_{\theta,\eta}}[h_1(X)],
$$
$$
\frac{E_P\left[\mathscr{P}^2(X)h_2(X)\mathrm{e}^{-(\theta^\intercal h_1(X)+\eta^\intercal h_2(X))}\right]}{E_P\left[\mathscr{P}^2(X)\mathrm{e}^{-(\theta^\intercal h_1(X)+\eta^\intercal h_2(X))}\right]} = E_{\bar{Q}^{\mathscr{P}}_{\theta,\eta}}[h_2(X)]
$$

for $i = 1, \ldots, p,\ j = 1, \ldots, \kappa$. This implies the desired result. □

Remark 1 (a) To keep the exponentially tilted probability measure $Q_{\theta,\eta}$ within the same exponential family as the original probability measure P, one possible selection of $h_1(x)$ and $h_2(x)$ is based on the sufficient statistic of the original probability distribution. For example, for the normal distribution, the sufficient statistic is $T(x) = [x\ x^2]$ and thus $h_1(x) = x$ and $h_2(x) = x^2$; for the gamma distribution, the sufficient statistic is $T(x) = [\log x\ x]$ and thus $h_1(x) = \log x$ and $h_2(x) = x$, and for the beta distribution, the sufficient statistic is $T(x) = [\log x\ \ \log(1-x)]$ and thus $h_1(x) = \log x$ and $h_2(x) = \log(1-x)$. Such a device can be applied to other distributions as well, such as the lognormal distribution, the inverse Gaussian distribution, and the inverse gamma distribution.

(b) Now, we provide a heuristic explanation for this device. The idea of using a sufficient statistic for exponential tilting is that we can treat this tilting as a *sufficient exponential tilting* within the same given parametric family. Furthermore, by using the Fisher–Neyman factorization theorem in the exponential family, we note that this device provides the maximum degree of freedom for exponential tilting within the same exponential family. We also expect to have an analogy parallel to the Rao–Blackwell theorem for sufficient exponential tilting: minimize the mean square loss among all possible importance sampling in the same exponential embedding.

2.1 Examples for multivariate normal and gamma distributions

To illustrate the proposed sufficient exponential tilting, we present examples of the multivariate normal distribution and the gamma distribution. We choose these two distributions to indicate the location and scale properties of the sufficient exponential tilting used in our general framework. In these examples, by using a suitable re-parameterization, we obtain neat tilting formulas for each distribution based on its sufficient statistic. Our simulation studies also show that the proposed sufficient exponential tilting performs 2 to 5 times better than classical one-parameter exponential tilting for some simple rare events.

We here check the validity of applying Theorem 1 for the example. First, we note that both the multivariate normal distribution and the gamma distribution are steep. Next, it is easy to see that the sufficient conditions $\bar{\Psi}(\theta)\Psi(\theta) \to \infty$ as $\theta_i \to \theta_{i,\max}$ for $i = 1, \ldots, p$, or $\eta_j \to \eta_{j,\max}$ for $j = 1, \ldots, \kappa$ in Theorem 1 hold in each example. For example, when $X \sim N_d(\mathbf{0}, \Sigma)$, then $\Psi(\theta) = O(e^{\|\theta\|^2})$ approaches ∞ sufficiently quickly. Another simple example illustrated here is when $\mathscr{P}(X) = \mathbb{1}_{\{X\in A\}}$ and $A := [a_1,\infty)\times\cdots\times[a_d,\infty)$, with $a_i > 0$ for all $i = 1, \ldots, d$, and X has a d-dimensional standard normal distribution; in this case it is easy to verify that the sufficient conditions in Theorem 1 hold. For gamma distribution $X \sim \mathtt{Gamma}(\alpha, \beta)$, we have $\Psi(\theta,\eta) = \frac{(1/\beta)^{-\alpha}\Gamma(\alpha+\theta)}{\Gamma(\alpha)(\beta-\eta)^{\alpha+\theta}}$ and $\bar{\Psi}(\theta,\eta) = \frac{(1/\beta)^{-\alpha}\Gamma(\alpha-\theta)}{\Gamma(\alpha)(\beta+\eta)^{\alpha-\theta}}$. Then it is easy to see that $\Psi(\theta,\eta)\bar{\Psi}(\theta,\eta) \to \infty$ as $\eta \uparrow \beta$ or $\eta \downarrow -\beta$.

Example 1 Multivariate normal distribution.

To illustrate the concept of sufficient exponential embedding, we first consider a one-dimensional normal distributed random variable as an example. Let X be a random variable with the standard normal distribution, denoted by $N(0, 1)$, with probability density function (pdf) $\frac{dP}{d\mathscr{L}} = e^{-x^2/2}/\sqrt{2\pi}$. By using the sufficient exponential embedding in (3) with $h_1(x) := x$ and $h_2(x) := x^2$, we have

$$
\begin{aligned}
\frac{dQ_{\theta,\eta}/d\mathscr{L}}{dP/d\mathscr{L}} &= \frac{\exp\{\theta h_1(x)+\eta h_2(x)\}}{E[\exp\{\theta h_1(X)+\eta h_2(X)\}]} \\
&= \sqrt{1-2\eta}\exp\{\eta x^2+\theta x-\theta^2/(2-4\eta)\}.
\end{aligned}
\tag{12}
$$

In this case, the tilting probability measure $Q_{\theta,\eta}$ is $N(\theta/(1-2\eta), 1/(1-2\eta))$, with $\eta < 1/2$, a location-scale family.

For the event of $\mathscr{P}(X) = \mathbb{1}_{\{X>a\}}$ for $a > 0$, define $\bar{Q}_{\theta,\eta}$ as $N(-\theta/(1+2\eta), 1/(1+2\eta))$ with $\eta > -1/2$. Applying Theorem 1, (θ^*, η^*) is the root of

$$\frac{\theta}{1-2\eta} = E_{\bar{Q}_{\theta,\eta}}[X \mid X > a] \quad \text{and}$$
$$\frac{1}{1-2\eta} + \frac{\theta^2}{(1-2\eta)^2} = E_{\bar{Q}_{\theta,\eta}}[X^2 \mid X > a], \tag{13}$$

which is equivalent to

$$\mu = E_{\bar{Q}_{\mu,\sigma^2}}[X \mid X > a] \quad \text{and} \quad \sigma^2 + \mu^2 = E_{\bar{Q}_{\mu,\sigma^2}}[X^2 \mid X > a].$$

under the standard parameterization.

Take the one-parameter exponential embedding case, with σ fixed. Standard calculation gives $\Psi(\theta) = e^{\theta^2/2}$, $\psi(\theta) = \theta^2/2$, and $\psi'(\theta) = \theta$. Using the fact that $X \mid \{X > a\}$ is a truncated normal distribution with minimum value a under $\bar{Q}$, θ^* must satisfy $\theta = \frac{\phi(a+\theta)}{1-\Phi(a+\theta)} - \theta$, cf. Fuh and Hu (2004).

Table 1 presents numerical results for the normal distribution. As demonstrated in the table, using sufficient exponential tilting for the simple event, $\mathbb{1}_{\{X>a\}}$, yields performance 2 to 3 times better than one-parameter exponential tilting in terms of variance reduction factors. Note that here we adopt the automatic Newton method (Teng etal. 2016) to search the optimal parameters for all experiments.

Due to the convex property $G(\theta, \eta)$ and the uniqueness of the optimal tilting parameters, the optimal tilting formula is robust and not sensitive to the initial value in most cases. To illustrate this phenomenon, we consider the $G(\theta, \eta)$ function for the simple event $X > 2$ under the standard normal distribution in Fig. 1. As shown in the figure, θ is not sensitive to the initial values; note that the range of the initial values becomes vital for finding optimal η due to the constraint $\eta < 1/2$. Moreover, there exists a flat area for simultaneously tilting both parameters θ, η.

We now proceed to a d-dimensional multivariate normal distribution. Let $X = (X_1, \ldots, X_d)^\intercal$ be a random vector with the standard multivariate normal distribution, denoted by $N(0, \mathbb{I})$, with pdf $\det(2\pi\mathbb{I})^{-1/2} \; e^{-(1/2)\, x^\intercal \mathbb{I}^{-1} x}$, where $\mathbb{I}$ is the identity matrix. By using the sufficient exponential embedding in (3), we have

$$\frac{dQ_{\theta,\eta}/d\mathscr{L}}{dP/d\mathscr{L}} = \frac{\exp\{\theta^\intercal x + x^\intercal M x\}}{E[\exp\{\theta^\intercal X + X^\intercal M X\}]} = \frac{e^{\theta^\intercal x + x^\intercal M x - \frac{1}{2}(\theta^\intercal (\mathbb{I}-2M)^{-1}\theta)}}{\sqrt{|(\mathbb{I}-2M)^{-1}|}},$$

where $|\cdot|$ denotes the determinant of a matrix, and $M = (a_{ij}) \in \mathbb{R}^{d\times d}$,

with $a_{ij} = \eta_i$ for $i = j$ and $a_{ij} = \eta_{d+1}$ for $i \neq j$. In this case, the tilting probability measure $Q_{\theta,\eta}$ is $N((\mathbb{I}-2M)^{-1}\theta, (\mathbb{I}-2M)^{-1})$.

For the event of $\mathscr{P}(X) = \mathbb{1}_{\{X\in A\}}$, define $\bar{Q}_{\theta,\eta}$ as $N((\mathbb{I}+2M)^{-1}(-\theta), (\mathbb{I}+2M)^{-1})$. Similar to the above one-dimensional normal distribution, we consider the standard parameterization by letting $\mu := (\mathbb{I}-2M)^{-1}\theta$, $\Sigma := (\mathbb{I}-2M)^{-1}$, and define $\bar{Q}_{\mu,\Sigma}$ as $N(-(2\mathbb{I}-\Sigma^{-1})^{-1}(\Sigma^{-1})\mu, (2\mathbb{I}-\Sigma^{-1})^{-1})$. Applying Theorem 1, (μ^*, Σ^*) is the root of

$$\mu = E_{\bar{Q}_{\mu,\Sigma}}[X \mid X \in A], \tag{14}$$
$$\mathscr{K}(\mu, \Sigma) = E_{\bar{Q}_{\mu,\Sigma}}\left[X^\intercal (\nabla_{\eta_i} M) X \mid X \in A\right] \quad \text{for } i = 1, 2, \ldots, d+1, \tag{15}$$

where

$$\nabla_{\eta_i} M = (b_{jk}) \in \mathbb{R}^{d\times d} \quad \text{for } i = 1, 2, \ldots, d+1, \tag{16}$$
$$\mathscr{K}(\mu, \Sigma) = \frac{1}{2}\operatorname{Tr}\left(-(\nabla_{\eta_i}\Sigma^{-1})(\Sigma)\right) - \frac{1}{2}\mu^\intercal(\nabla_{\eta_i}\Sigma^{-1})\mu \tag{17}$$

Here in (16), $\operatorname{Tr}(A)$ is the trace of matrix A; the value of b_{jk} is defined as follows: for each $i = 1, \ldots, d$,

$$b_{jk} = \begin{cases} 1, & \text{if } i = j = k, \\ 0, & \text{otherwise,} \end{cases}$$

and for $i = d+1$,

$$b_{jk} = \begin{cases} 1, & \text{if } i \neq k, \\ 0, & \text{otherwise.} \end{cases}$$

Remark 2 The left-hand sides (LHSs) of (14) and (15) are derivatives of the cumulant function $\psi_{MN}(\theta, \eta) := \log(\sqrt{|\,(\mathbb{I}-2M)^{-1}\,|}) + \frac{1}{2}(\theta^\intercal(\mathbb{I}-2M)^{-1}\theta)$ for a multivariate normal with respect to parameters θ and η, respectively. Note that in the RHS of (17), Jacobi's formula is adopted for the derivative of the determinant of the matrix $(\mathbb{I}-2M)^{-1}$ and $\nabla_{\eta_i}\Sigma^{-1}$ is

$$\nabla_{\eta_i}\Sigma^{-1} = \nabla_{\eta_i}(\mathbb{I}-2M) = (m_{jk}) \in \mathbb{R}^{d\times d} \quad \text{for } i = 1, 2, \ldots, d+1,$$

where for each $i = 1, \ldots, d$,

$$m_{jk} = \begin{cases} -2, & \text{if } i = j = k, \\ 0, & \text{otherwise,} \end{cases}$$

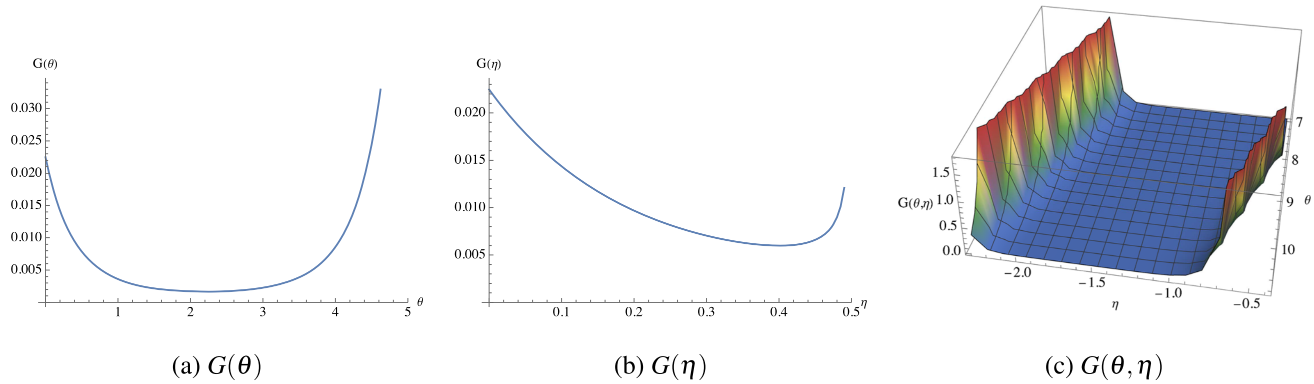


Fig. 1 $G(\theta, \eta)$ function for simple event $X > 2$ under standard normal distribution

Table 1 Sufficient exponential importance sampling for normal distribution

$[X \sim N(0, 1)]$		Variance reduction factors		
a	Crude for $P(X > a)$	θ^*	η^*	(θ^*, η^*)
1	1.566×10^{-1}	4	1	9
2	2.300×10^{-2}	19	4	58
3	1.370×10^{-3}	226	36	785
4	3.000×10^{-5}	6616	832	30,807

and for $i = d + 1$,

$$m_{jk} = \begin{cases} -2, & \text{if } i \neq k, \\ 0, & \text{otherwise.} \end{cases}$$

Table 2 presents the numerical results for the standard bivariate normal distribution. As shown in the table, for event types $\mathbb{1}_{\{X_1+X_2>a\}}$, $\mathbb{1}_{\{X_1>a, X_2>a\}}$, and $\mathbb{1}_{\{X_1X_2>a, X_1>0, X_2>0\}}$, tilting different parameters results in different performance in variance reduction. Although sometimes tilting the variance parameter or the correlation parameter alone provides poor performance, combining them as mean parameter tilting yields 2 to 3 times better performance than one-parameter exponential tilting.

Note that for easy implementation, we here consider the standard parameterization and directly tilt $\mu = (\mu_1, \mu_2)$, $\sigma = (\sigma_1, \sigma_2)$, and ρ in this example; that is, in this case, $(\mathbb{I} - 2M)^{-1}\theta = \mu$ and $(\mathbb{I} - 2M)^{-1} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$. In addition, we illustrate the $G(\mu, \sigma, \rho)$ function for simple event $X + Y > 3$ under the standard bivariate normal distribution in Fig. 2, where neither μ nor σ are sensitive to the initial values; although the range of the initial values becomes vital for finding optimal ρ, $G(\rho)$ is flat for most ρ.[2]

Example 2 Gamma distribution. Let X be a random variable with a gamma distribution, denoted by $\texttt{Gamma}(\alpha, \beta)$, with pdf $\frac{dP}{d\mathscr{L}} = (\beta^\alpha/\Gamma(\alpha))\, x^{\alpha-1}e^{-\beta x}$. By using the sufficient exponential embedding in (3) with $h_1(x) := \log(x)$ and $h_2(x) := x$, we have

$$\frac{dQ_{\theta,\eta}/d\mathscr{L}}{dP/d\mathscr{L}} = \frac{\exp\{\theta h_1(x) + \eta h_2(x)\}}{E[\exp\{\theta h_1(X) + \eta h_2(X)\}]} = e^{x\eta+\theta\log(x)} \frac{\Gamma(\alpha)}{(1/\beta)^{-\alpha}(\beta-\eta)^{-\alpha-\theta}\Gamma(\alpha+\theta)} \quad (18)$$

In this case, the tilting probability measure $Q_{\theta,\eta}$ is $\texttt{Gamma}(\alpha+\theta, \beta-\eta)$. For the event of $\mathscr{P}(X) = \mathbb{1}_{\{X>a\}}$ for $a > 0$, define $\bar{Q}_{\theta,\eta}$ as $\texttt{Gamma}(\alpha-\theta, \beta+\eta)$. Applying Theorem 1, (θ^*, η^*) is the root of

$$-\log(\beta-\eta) + \Upsilon(\alpha+\theta) = E_{\bar{Q}_{\theta,\eta}}\left[\log(X) \mid X > a\right] \quad (19)$$

$$\frac{\alpha+\theta}{\beta-\eta} = E_{\bar{Q}_{\theta,\eta}}[X \mid X > a], \quad (20)$$

where $\Upsilon(\alpha+\theta)$ is a digamma function equal to $\Gamma'(\alpha+\theta)/\Gamma(\alpha+\theta)$.

Table 3 presents the numerical results for the gamma distribution. Note that the commonly used one-parameter exponential tilting involves a change only for the parameter β (i.e., changing β to $\beta-\eta^*$) in the case of the gamma distribution. However, observe that tilting the other parameter α (i.e., $\alpha \to \alpha+\theta^*$) for some cases yields 2 to 3 times better performance than the one-parameter exponential tilting in terms of variance reduction factors. This is due to the $-\theta < \alpha$ and $\eta < \beta$ constraint. For instance, consider the case in which we tilt only one parameter, either θ or η, as follows. For the

[2] Note that to draw the figure, we assume $\mu = \mu_1 = \mu_2$ and $\sigma = \sigma_1 = \sigma_2$ for easy presentation.

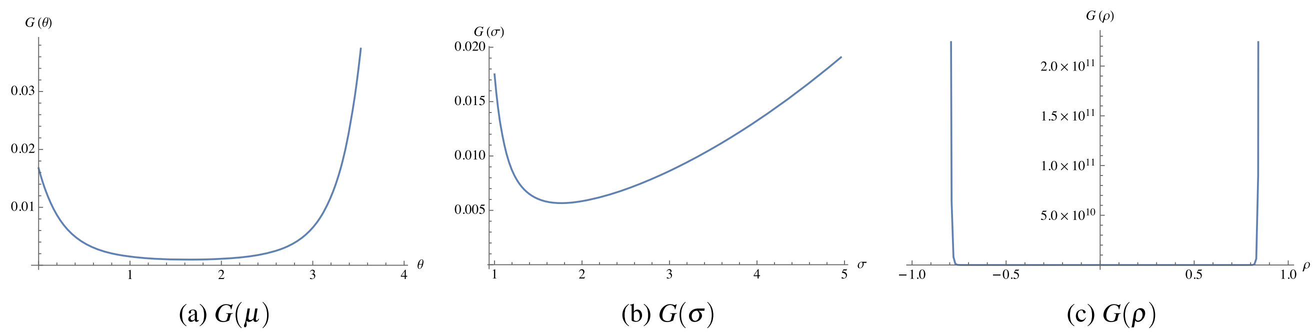


Fig. 2 $G(\mu, \sigma, \rho)$ function for simple event $X + Y > 3$ under standard bivariate normal distribution

Table 2 Sufficient exponential importance sampling for standard bivariate normal distribution

$[X \sim N_2(0, \mathbb{I})]$	k	Crude	Variance reduction factors μ^*	σ^*	ρ^*	$(\mu^*, \sigma^*, \rho^*)$
$P(X_1 + X_2 > a)$	3	1.663×10^{-2}	24	2	2	43
	4	2.400×10^{-3}	138	4	2	354
	5	1.800×10^{-4}	1064	8	4	4036
$P(X_1 > a, X_2 > a)$	1	2.532×10^{-2}	9	1	2	16
	1.5	4.600×10^{-3}	34	2	7	68
	2	5.800×10^{-4}	227	5	18	504
$P(X_1 X_2 > a, X_1 > 0, X_2 > 0)$	2	1.538×10^{-2}	21	2	3	46
	3	4.800×10^{-3}	57	2	3	145
	5	5.600×10^{-4}	425	5	3	1213

simple event $\mathbb{1}_{\{X>a\}}$, we can either choose parameter η such that $-\beta < \eta < \beta$ or parameter θ such that $-\alpha < \theta$ and $\alpha - \theta \in \mathbb{R}\backslash\{0, -1, -2, \ldots\}$ (see "Appendix B"), to obtain a larger mean for the tilted gamma distribution. In this case, it is clear that θ-tilting yields a larger parameter search space and thus achieves better performance than η-tilting. For simple event $\mathbb{1}_{\{1/X>a\}}$, the optimal tilting parameters θ^* and η^* can be obtained by solving the revised version of (19) and (20), where the condition is revised from $X > a$ to $1/X > a$. Note that event $\mathbb{1}_{\{1/X>a\}}$ shows the opposite effect.

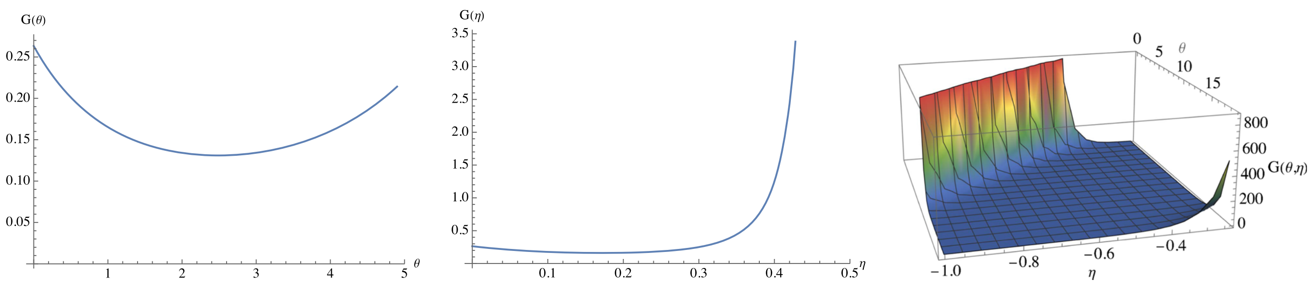
Additionally, Fig. 3 illustrates the $G(\theta, \eta)$ function for simple event $X > 10$ under the gamma distribution. The range of the initial values becomes vital for finding optimal θ, η due to the $\theta > -\alpha$ and $\eta < \beta$ constraint. Moreover, similar to the case in Fig. 1c, there exists a flat area for simultaneously tilting both parameters θ, η.

Remark 3 Observed from the above examples, the proposed sufficient exponential tilting yields improvements in terms of variance reduction in two aspects. First, in some cases, tilting multiple parameters simultaneously via our sufficient exponential tilting algorithm greatly improves the variance reduction performance; for example, in the case of simulating simple moderate-deviation rare events, for the normal distribution, although changing σ or ρ alone may not help much in variance reduction, changing them together with μ can yield 3 to 4 times better performance than traditional mean-shift one-parameter tilting. Second, in some cases, tilting other parameters results in better performance; for instance, for the gamma distribution, changing the shape parameter α leads to better performance for some rare events, though traditional one-parameter tilting always changes the rate parameter β. Moreover, for the simple cases, the computational time of such two-parameter tilting is almost the same as that of one-parameter tilting since our algorithm always converges in 3 or 4 iterations when locating optimal tilting parameters. Analyses for more complex mixture distributions regarding the computational cost are provided in Sect. 4.

For other events of interest, however, e.g., $E(X\mathbb{1}_{\{0<X<a\}})$, other than the presented rare event cases, the proposed sufficient exponential tilting method always yields better performance than one-parameter tilting because optimal two-parameter tilting includes the solution of one-parameter tilting. For example, in the case of simulating $E(X\mathbb{1}_{\{0<X<1\}})$ for the standard normal distribution, tilting the standard deviation together with the mean via our method yields 10 times better variance reduction performance than tilting the mean only.

Table 3 Sufficient exponential importance sampling for gamma distribution

$[X \sim \texttt{Gamma}(\alpha = 4, \beta = 1/2)]$									
		Variance reduction factors					Variance reduction factors		
a	Crude for $P(X > a)$	θ^*	η^*	(θ^*, η^*)	a	Crude for $P(1/X > a)$	θ^*	η^*	(θ^*, η^*)
10	2.613×10^{-1}	3	2	6	0.2	2.438×10^{-1}	2	4	9
20	1.050×10^{-2}	46	23	129	0.5	1.864×10^{-2}	15	41	79
30	1.800×10^{-4}	1311	562	4190	1.5	3.100×10^{-4}	301	1326	3,407
35	3.000×10^{-5}	11,718	4788	22,015	2.5	6.000×10^{-5}	2121	11,886	20,584



(a) $G(\theta)$ (b) $G(\eta)$ (c) $G(\theta, \eta)$

Fig. 3 $G(\theta, \eta)$ function for simple event $X > 10$ under gamma distribution $X \sim \texttt{Gamma}(\alpha = 4, \beta = 1/2)$

2.2 Bounded relative error and logarithmic efficiency analysis

Here we state the theoretical results of both bounded relative error and logarithmic efficiency for the normal and gamma distributions under the case of simple rare events. Note that the optimal tilting parameters in Theorems 5 and 6 are approximately optimal tilting parameters, in the standard large deviation sense, to be defined in "Appendices A and B".

Theorem 5 *Let X be a random variable from the standard normal distribution $N(0, 1)$ with pdf $\varphi(x)$. Denote $\varphi_{\theta,\eta}(x)$ as the pdf of $N(\theta/(1-2\eta), 1/(1-2\eta))$.*

For the simulation of the rare event probability $p = P(X > a)$, the domains of the tilting parameters θ and η are $\Theta = (-\infty, \infty)$ and $H = [-1/2, 1/2)$, respectively, and the approximated optimal tilting parameters are $\theta^* = a(1-2\eta^*)$ and $\eta^* \to -1/2$ as $a \to \infty$. Moreover, the approximated optimal sufficient exponential tilting entails logarithmic efficiency but not bounded relative error; i.e., for all $\varepsilon > 0$, we have

$$\lim_{a\to\infty} \frac{1}{p^{2-\varepsilon}} \int_a^\infty \frac{\varphi(x)}{\varphi_{\theta^*,\eta^*}(x)} \varphi(x) dx = 0.$$

Moreover, optimal sufficient exponential tilting outperforms traditional optimal one-parameter exponential tilting in the sense of reducing asymptotic variance.

Theorem 6 *Let X be a random variable from the gamma distribution $\texttt{Gamma}(\alpha, \beta)$ with pdf $f(x)$. Denote $f_{\theta,\eta}(x)$ as the pdf of $\texttt{Gamma}(\alpha + \theta, \beta - \eta)$.*

(i) For the simulation of rare event probability $p = P(X > a)$, the domains of the tilting parameters θ and η are $\Theta = \{\theta \mid -\alpha < \theta \text{ and } \alpha - \theta \in \mathbb{R}\backslash\{0, -1, -2, \ldots\}\}$ and $H = (-\beta, \beta)$, and the approximated optimal tilting parameters are $\eta^* = (a\beta - \alpha - \theta^*)/a$ and $\theta^* \to \infty$ as $a \to \infty$, such that $\theta^*/a \to c$ with $c \in (0, 2\beta)$.

Moreover, the approximated optimal sufficient exponential tilting entails bounded relative error; i.e.,

$$\lim_{a\to\infty} \frac{1}{p^2} \int_a^\infty \frac{f(x)}{f_{\theta^*,\eta^*}(x)} f(x) dx = 0.$$

Therefore, optimal sufficient exponential tilting outperforms traditional optimal one-parameter exponential tilting, which has only logarithmic efficiency for simulating $\{X > a\}$.

(ii) For the simulation of rare event probability $p = P(X < 1/a)$, the domains of the tilting parameters θ and η are $\Theta = \{\theta \mid -\alpha < \theta < \alpha \text{ and } \alpha - \theta \in \mathbb{R}\backslash\{0, -1, -2, \ldots\}\}$ and $H = (-\infty, \beta)$, and the approximated optimal tilting parameters are $\eta^* = \beta - a(\alpha + \theta^*)$ and $\theta^* \to \alpha$ as $a \to \infty$.

Moreover, approximated optimal sufficient exponential tilting entails bounded relative error; i.e.,

$$\lim_{a\to\infty} \frac{1}{p^2} \int_0^{1/a} \frac{f(x)}{f_{\theta^*,\eta^*}(x)} f(x) dx = 0.$$

Furthermore, the ratio of the expected squared estimators between optimal sufficient exponential tilting and optimal one-parameter tilting is less than one, indicating that the former outperforms the latter.

The proofs of Theorems 5 and 6 are given in "Appendices A and B", respectively.

Remark 4 Note that two matters affect the domain of tilting parameter spaces Θ and H. First, it is necessary to ensure realistic parameters for the tilted distribution; e.g., the $\eta < 1/2$ constraint for the normal distribution ($\eta < \beta$ for the gamma distribution) guarantees the variance (the ratio parameter) of the tilted distribution to be positive. Second, other constraints are needed to ensure the finiteness of the expected squared estimator for the given event; e.g., in the simulation of rare event $\{X > a\}$, $\eta \geq -1/2$ is needed for the normal distribution and $\eta > -\beta$ for the gamma distribution.

3 An application: portfolio loss under the normal mixture copula model

Consider a portfolio of loans consisting of n obligors, each of whom has a small probability of default. We further assume that the loss resulting from the default of the k-th obligor, denoted as c_k (monetary units), is known. In copula-based credit models, dependence among default indicator for each obligor is introduced through a vector of latent variables $X = (X_1, \ldots, X_n)$, where the k-th obligor defaults if X_k exceeds some chosen threshold χ_k. The total loss from defaults is then denoted by

$$L_n = c_1 \mathbb{1}_{\{X_1 > \chi_1\}} + \cdots + c_n \mathbb{1}_{\{X_n > \chi_n\}}, \tag{21}$$

where $\mathbb{1}$ is the indicator function. Particularly, the problem of interest is to estimate the probability of losses, $P(L_n > \tau)$, especially at large values of τ.

As mentioned earlier in the introduction, the widely-used normal copula model might assign an inadequate probability to the event of many simultaneous defaults in a portfolio. In view of this, Bassamboo etal. (2008) and Chan and Kroese (2010) set forth the t-copula model for modeling portfolio credit risk. In this paper, we further consider the normal mixture model (McNeil etal. 2015), including the normal copula and t-copula models as special cases, for the generalized d-factor model of the form

$$X_k = \rho_{k1} V_1 Z_1 + \cdots + \rho_{kd} V_d Z_d + \rho_k V_{d+1} \epsilon_k, \quad k = 1, \ldots, n, \tag{22}$$

in which

- $Z = (Z_1, \ldots, Z_d)^\intercal$ follows a d-dimensional multivariate normal distribution with zero mean and covariance matrix Σ, where $^\intercal$ denotes vector transpose;
- $V = (V_1, \ldots, V_{d+1})$ are non-negative scalar-valued random variables which are independent of Z, and each V_j is a shock variable independent from each other, for $j = 1, \ldots, d+1$;
- $\epsilon_k \sim N(0, \sigma_\epsilon^2)$ is an idiosyncratic risk associated with the k-th obligor, for $k = 1, \ldots, n$;
- $\rho_{k1}, \ldots, \rho_{kd}$ are the factor loadings for the k-th obligor, and $\rho_{k1}^2 + \cdots + \rho_{kd}^2 \leq 1$;
- $\rho_k = \sqrt{1 - (\rho_{k1}^2 + \cdots + \rho_{kd}^2)}$, for $k = 1, \ldots, n$.

Model (22) is the so-called grouped normal mixture copula in McNeil etal. (2015), which is constructed by drawing randomly from this set of component multivariate normals based on a set of weights controlled by the distribution of V. This model enables us to blend in multiplicative shocks via the variables V, which could be interpreted as shocks that arise from new information. Note that in model (22), we consider a multi-factor model to capture the effect of different factors. In addition, rather than multiplying all components of a correlated Gaussian vector Z with a single V, we instead multiply different subgroups with different variates V_j; the V_j are themselves comonotonic [see Section 7.2.1 in McNeil etal. (2015)].

Let $W = (W_1, \ldots, W_{d+1})$ and $V_j = g(W_j)$ for $j = 1, \ldots, d+1$. Note that in this paper, we mainly focus on the class of V_j with $W_j \sim \texttt{Gamma}(\alpha_j, \beta_j)$, $j = 1, \ldots, d+1$. With $V_j = g(W_j) = \sqrt{\nu_j / W_j}$ and $W_j \sim \texttt{Gamma}(\nu_j/2, 1/2)$ (for $j = 1, \ldots, d+1$), we therefore create subgroups whose dependence properties are described by normal mixture copulas with different ν_j parameters. The groups may even consist of a single member for each ν_j parameter, as indicated in Section 7.3 of McNeil etal. (2015) and references therein. With the above setting, in the degenerated case that $W_1 = W_2 = \cdots = W_{d+1}$ is a common gamma random variable (i.e., $W_j \sim \texttt{Gamma}(\nu/2, 1/2)$ for $j = 1, \ldots, d+1$), X forms a multivariate t-distribution, which is the most popular form in financial modeling. In addition, we consider the case that $V_j = g(W_j) = \sqrt{W_j}$, $j = 1, \ldots, d+1$ and W_j is with a generalized inverse Gaussian (GIG) distribution. The GIG mixing distribution, a special case of the symmetric generalized hyperbolic (GH) distribution, is very flexible for modeling financial returns. Moreover, GH distributions also include the symmetric normal inverse Gaussian (NIG) distribution and a symmetric multivariate distribution with hyperbolic distribution for its one-dimensional marginal as interesting examples. Note that this class of distributions has become popular in the financial modeling literature. An important reason is their link to Lévy processes (such as Brownian motion or the compound

Poisson distribution) that are used to model price processes in continuous time. For example, the generalized hyperbolic distribution has been used to model financial returns in Eberlein and Keller (1995) and Eberlein etal. (1998). The reader is referred to McNeil etal. (2015) for more details.

Note that the number of factors used for model (22) usually depends on the number of obligors in the credit portfolio and their characteristics, e.g., the number of sectors that the obligors belong to. In addition, in practice we would expect the factor loadings to be fairly sparse, cf. Glasserman etal. (2008).

We now define the tail probability of total portfolio losses conditional on Z and V. Specifically, the tail probability of total portfolio losses conditional on the factors Z and V, denoted as $\varrho(Z, V)$ is defined as

$$\varrho(Z, V) = P(L_n > \tau \mid (Z, V)). \tag{23}$$

The desired probability of losses can be represented as

$$P(L_n > \tau) = E\left[\varrho(Z, V)\right]. \tag{24}$$

For an efficient Monte Carlo simulation of the probability of total portfolio losses (24), we apply importance sampling to the distributions of the factors $Z = (Z_1, \ldots, Z_d)^\intercal$ and $V = (V_1, \ldots, V_{d+1})^\intercal = (g(W_1), \ldots, g(W_{d+1}))^\intercal$ (see (22)). In other words, we attempt to choose importance sampling distributions for both Z and W that reduce the variance in estimating the integral $E[\varrho(Z, V)]$ against the original densities of Z and W.

As noted in Glasserman and Li (2005) for normal copula models, the simulation of (24) involves two rare events: the default event and the total portfolio loss event. For the normal mixture model (22), this makes the simulation of (24) even more challenging. For a general simulation algorithm for this type of problem, we simulate $P(L_n > \tau)$ as the expected value of $\varrho(Z, V)$ in (24). Our device is based on a joint probability simulation rather than the conditional probability simulation considered in the literature. Moreover, we note that the simulated distributions—the multivariate normal distribution Z and the commonly adopted multivariate gamma distribution for W—are both two-parameter distributions.[3] This motivates us to study a sufficient exponential tilting in the next section.

3.1 Sufficient exponential tilting for normal mixture distributions

Recall that in (22), the latent random vector X follows a multivariate normal mixture distribution. In this section, for simplicity, we consider a one-dimensional normal mixture distribution as an example to demonstrate the proposed sufficient exponential tilting. Let X be a one-dimensional normal mixture random variable with only one factor (i.e., $d = 1$) such that $X = \xi V Z = \xi g(W) Z$, where $\xi \in \mathbb{R}$, $Z \sim N(0, 1)$, $W \sim \texttt{Gamma}(\alpha, \beta)$, and V (as well as W) is a non-negative and scalar-valued random variable which is independent of Z. Since the random variable W is independent of Z, by using sufficient exponential embedding with $h_1(z)$ and $h_2(z)$ for Z and $\tilde{h}_1(w)$ and $\tilde{h}_2(w)$ for W, we have

$$\frac{dQ_{\theta_1,\eta_1,\theta_2,\eta_2}/d\mathscr{L}}{dP/d\mathscr{L}} = \frac{\exp\{\theta_1 h_1(z) + \eta_1 h_2(z)\}}{E[\exp\{\theta_1 h_1(Z) + \eta_1 h_2(Z)\}]} \frac{\exp\{\theta_2 \tilde{h}_1(w) + \eta_2 \tilde{h}_2(w)\}}{E[\exp\{\theta_2 \tilde{h}_1(W) + \eta_2 \tilde{h}_2(W)\}]}, \tag{25}$$

where θ_1, η_1 are the tilting parameters for Z and θ_2, η_2 are the tilting parameters for W.

As Z follows a standard normal distribution $N(0, 1)$ and W follows a gamma distribution $\texttt{Gamma}(\alpha, \beta)$, from (12) and (18), Eq. (25) becomes

$$\frac{dQ_{\theta_1,\eta_1,\theta_2,\eta_2}/d\mathscr{L}}{dP/d\mathscr{L}} = \frac{1}{\sigma} \exp\{\mu z - \mu^2/2 + \frac{\sigma^2 - 1}{2\sigma^2}(z - \mu)^2\} \times e^{w\eta_2 + \theta_2 \log(w)} \frac{\Gamma(\alpha)}{(1/\beta)^{-\alpha}(\beta - \eta_2)^{-\alpha-\theta_2}\Gamma(\alpha + \theta_2)},$$

where $\mu = \theta_1/(1-2\eta_1)$ and $\sigma^2 = 1/(1-2\eta_1)$.[4] The optimal tilting parameters θ_1^*, η_1^* for Z can be obtained by solving the modified version of (13) for the normal distribution; θ_2^* and η_2^* for W are the solutions of the modified version of (20) and (19) for the gamma distribution, where condition $X > a$ is modified to accommodate the event. For example, condition $X > a$ is modified to $\sqrt{W} Z > a$ in (13), (20), and (19) for the example in Table 4. Note that due to the fact that W is independent of Z, the optimal tilting can be done separately for the normal and gamma distributions. Moreover, for demonstration and to promote reproducibility, we release our implementation[5] for searching the optimal parameters for the normal, gamma, and normal mixture distributions (corresponding to the results in Tables 1, 3, and 4, respectively); the optimal parameters for each setting are also listed in the released code.

Table 4 shows the numerical results for this one-dimensi onal, one-factor normal mixture distribution with $V = g(W) = \sqrt{W}$. Since for a normal mixture random variable, the variance is associated with the random variable W, tilting the standard deviation-related parameter η_1 with θ_1 of the normal random variable Z is relatively insignificant in comparison to tilting the parameters of W (i.e., θ_2 or η_2) with

[3] Here we treat the mean vector and variance-covariance matrix of Z as two parameters.

[4] To simplify the notation, we here use μ and σ^2.

[5] https://github.com/jerewang/codes.twoparameters.is.git.

θ_1. This is shown in Table 4. In addition, similar to the case demonstrated in Example 2, in this case, tilting θ_2 with θ_1 also yields better performance than tilting η_2 with θ_1 in the sense of variance reduction, which is consistent with the theoretical results in "Appendices A and B" and the numerical results in Table 3 (see Remark 7 also).

Next, we summarize tilting for the event in which the k-th obligor defaults if X_k exceeds a given threshold χ_k as "ABC-event tilting," which involves the calculation of tail event

$$\{(\mathtt{A}+\mathtt{B})\,\mathtt{C} > \tau\},$$

where A denotes the normally distributed part of the systematic risk factors, B denotes the idiosyncratic risk associated with each obligor, and C denotes the non-negative and scalar-valued random variables which are independent of A and B,[6] For example, for the normal mixture copula model in (22), the d-dimensional multivariate normal random vectors $Z = (Z_1, \ldots, Z_d)$ are associated with A, ϵ_k is associated with B, and the non-negative and scalar-valued random variables W are associated with C. Table 5 summarizes the exponential tilting used in Glasserman and Li (2005), Bassamboo etal. (2008), Chan and Kroese (2010), Scott and Metzler (2015) and our paper. Note that Glasserman and Li (2005) consider the normal copula model so that there is no need for tilting C, and that Bassamboo etal. (2008), Chan and Kroese (2010), Scott and Metzler (2015) only consider the one-dimensional t-distribution, whereas we consider the multi-dimensional normal mixture distribution. Moreover, except for the proposed model, the other four methods adopt so-called one-parameter tilting. For example, even though Scott and Metzler (2015) consider the tilting of A and C, the same as our setting, only one parameter is tilted for each of the two distributions (i.e., mean for the normal distribution and shape for the gamma distribution); in our method, however, the tilting parameter can be either the shape or the rate parameter for the underlying gamma distribution, which results in a more efficient simulation.

Remark 5 When considering the t-distribution for X_k (i.e., $V_j = \sqrt{\nu_j/W_j}$ with $W_j \sim \mathtt{Gamma}(\nu_j/2, 1, 2)$ for $j = 1, \ldots, d+1$), traditional one-parameter tilting for the gamma distribution has bounded relative error, whereas it has only logarithmic efficiency for the normal distribution. This fact explains why Bassamboo etal. (2008) tilt only the gamma distribution (i.e., event C in Table 5). However, in this paper, we additionally tilt the normal distribution to improve the "second order" efficiency via the proposed sufficient exponential tilting (see Theorem 6 ii)). Moreover, when considering $V_j = \sqrt{W_j}$, $j = 1, \ldots, d+1$ with $W_j \sim \mathtt{Gamma}(\alpha_j, \beta_j)$, sufficient exponential tilting outperforms traditional one-parameter exponential tilting, which has only logarithmic efficiency (see Theorem 6 i)).

3.2 Exponential tilting for $\varrho(Z, V)$

In this subsection, we use notation similar to that in Sect. 2. Let $Z = (Z_1, \ldots, Z_d)^\intercal$ be a d-dimensional multivariate normal random variable with zero mean and identity covariance matrix $\mathbb{I}$,[7] and denote $V = (V_1, \ldots, V_{d+1})^\intercal = (g(W_1), \ldots, g(W_{d+1}))^\intercal$ as non-negative scalar-valued random variables, which are independent of Z. Under the probability measure P, let $f_z(z) = f_z(z_1, \ldots, z_d)$ and $f_w(w) = f_w(w_1, \ldots, w_{d+1})$ be the probability density functions of Z and W, respectively, with respect to the Lebesgue measure $\mathscr{L}$. As alluded to earlier, our aim is to calculate the expectation of $\varrho(Z, V)$,

$$m = E_P\left[\varrho(Z, V)\right] \tag{26}$$

under the probability measure P.[8]

To evaluate (26) via importance sampling, we choose a sampling probability measure Q, under which Z and W have the corresponding probability density functions $q_z(z) = q_z(z_1, \ldots, z_d)$ and $q_w(w) = q_w(w_1, \ldots, w_{d+1})$. Assume that Q is absolutely continuous with respect to P; Eq. (26) can then be written as

$$E_P\left[\varrho(Z, V)\right] = E_Q\left[\varrho(Z, V)\frac{f_z(Z)}{q_z(Z)}\frac{f_w(W)}{q_w(W)}\right]. \tag{27}$$

Let $Q_{\mu,\Sigma,\theta,\eta}$ be the sufficient exponential tilted probability measure of P. Here subscripts $\mu = (\mu_1, \ldots, \mu_d)^\intercal$, and Σ, constructed via ρ and $\sigma = (\sigma_1, \ldots, \sigma_d)^\intercal$, are the tilting parameters for random vector Z,[9] and $\theta = (\theta_1, \ldots, \theta_{d+1})^\intercal$ and $\eta = (\eta_1, \ldots, \eta_{d+1})^\intercal$ are the tilting parameters for W. Define the likelihood ratios

$$r_{z,\mu,\Sigma}(z) = \frac{f_z(z)}{q_{z,\mu,\Sigma}(z)} \quad \text{and} \quad r_{w,\theta,\eta}(w) = \frac{f_w(w)}{q_{w,\theta,\eta}(w)}, \tag{28}$$

where $q_{z,\mu,\Sigma}(z)$ and $q_{w,\theta,\eta}(w)$ denote the probability density functions corresponding to $q_z(z)$ with tilting parameters

[6] Here we omit the coefficients before A B, and C for simplicity.

[7] Although here we consider the identity covariance matrix for simplicity, it is straightforward to extend this to any valid covariance matrix Σ.

[8] Note that Theorem 1 works for a random variable from an exponential family. In this section, we simply apply the proposed importance sampling to simulate the portfolio loss. In the case of simulating a rare event probability, in which X is in a non-convex set, further decomposition based on mixture distribution tilting is required, cf. Fuh and Hu (2004), Glasserman etal. (2008). This line will be further studied in a separate paper.

[9] To simplify the notation, we use μ to denote one dimensional and high dimensional parameters.

Table 4 Sufficient exponential importance sampling for normal mixture distribution

$[Z \sim N(0,1),\, W \sim \texttt{Gamma}(\alpha=2, \beta=1/2)]$	a	Crude	Variance reduction factors θ_1^*	η_1^*	(θ_1^*, η_1^*)	θ_2^*	η_2^*	(θ_1^*, θ_2^*)	(θ_1^*, η_2^*)
$P(\sqrt{W}Z > a)$	2	1.344×10^{-1}	3	1	5	1	1	4	4
	4	2.808×10^{-2}	7	2	10	2	1	15	12
	8	9.500×10^{-4}	30	8	44	4	3	281	206
	12	2.000×10^{-5}	105	18	178	11	4	6691	3100

Table 5 ABC-event tilting

	One-parameter tilting (traditional) Glasserman and Li (2005) Multivariate normal dist	Bassamboo etal. (2008) t-dist.	Chan and Kroese (2010)	Scott and Metzler (2015)	Sufficient exponential tilting (proposed) Our paper Normal mixture dist
A	✓	✗	✓	✓	✓
B	✓	✗	✓	✗	✗
C	NA	✓	✗	✓	✓

μ and Σ and $q_w(w)$ with tilting parameters θ and η, respectively. Then, combined with (28), Eq. (27) becomes

$$E_Q\left[\varrho(Z,V)\frac{f_z(Z)}{q_z(Z)}\frac{f_w(W)}{q_w(W)}\right] = E_{Q_{\mu,\Sigma,\theta,\eta}}\left[\varrho(Z,V) r_{z,\mu,\Sigma}(Z) r_{w,\theta,\eta}(W)\right].$$

Denote $G(\mu,\Sigma,\theta,\eta) = E_P\left[\varrho^2(Z,V) r_{z,\mu,\Sigma}(Z) r_{w,\theta,\eta}(W)\right]$, which is assumed to be finite. By using the same argument as that in Sect. 2, we minimize $G(\mu,\Sigma,\theta,\eta)$ to get the tilting formula. That is, tilting parameters μ^*, Σ^*, θ^*, and η^* are chosen to satisfy[10]

$$\frac{\partial G(\mu,\Sigma,\theta,\eta)}{\partial \mu} = \vec{0}, \quad \frac{\partial G(\mu,\Sigma,\theta,\eta)}{\partial \Sigma} = \vec{0}, \tag{29}$$

$$\frac{\partial G(\mu,\Sigma,\theta,\eta)}{\partial \theta} = \vec{0}, \quad \frac{\partial G(\mu,\Sigma,\theta,\eta)}{\partial \eta} = \vec{0}. \tag{30}$$

(Fast) inverse Fourier transform for non-identical c_k

To obtain the optimal tilting parameters in (29) and (30), we must calculate the conditional default probability $\varrho(z,v) = P(L_n > \tau \mid Z = z, V = v)$ in (21). We here apply the fast inverse Fourier transform (FFT) method described as follows. Note that using the definition of the latent factor X_k for the k-th obligor in (22), the conditional default probability $P(X_k > \chi_k \mid Z = z, V = v)$ given $Z = (z_1, \ldots, z_d)^\intercal$ and $V = (v_1, \ldots, v_{d+1})^\intercal$ becomes

$$p_{z,v,k} = P\left(\epsilon_k > \frac{\chi_k - \sum_{i=1}^d \rho_{ki} V_i Z_i}{\rho_k V_{d+1}} \,\middle|\, Z = z, V = v\right). \tag{31}$$

With non-identical c_k, the distribution of the sum of n independent but non-identically distributed weighted Bernoulli random variables becomes difficult to evaluate. Here we adopt the inverse Fourier transform to calculate $\varrho(z,v)$ (Oberhettinger 2014). Recall that $L_n \mid (Z = z, V = v)$ equals

$$L_n^{z,v} = \sum_{\ell=1}^n c_\ell H_\ell^{z,v},$$

where $H_\ell^{z,v} \sim \texttt{Bernoulli}(p_{z,v,\ell})$ (see Eq. (31)), and the support of $L_n^{z,v}$ is a discrete set with a finite number of values. Its Fourier transform is

$$\phi_{L_n^{z,v}}(t) = E[e^{itL_n^{z,v}}] = E[e^{it(\sum_{i=1}^n c_\ell H_\ell^{z,v})}] = \prod_{\ell=1}^n E[e^{itc_\ell H_\ell^{z,v}}] = \prod_{\ell=1}^n \phi_{H_\ell^{z,v}}(tc_\ell),$$

where $\phi_{H_\ell^{z,v}}(s) = 1 - p_{z,v,\ell} + p_{z,v,\ell}e^{is}$. For random variable $L_n^{z,v}$, we can recover $q_k^{z,v} = P(L_n^{z,v} = k)$ by inverting the Fourier series:

$$q_k^{z,v} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ikt} \prod_{\ell=1}^n \phi_{H_\ell^{z,v}}(tc_\ell)dt, \tag{32}$$

where $k = 1, 2, \ldots, \infty$.

An FFT algorithm computes the discrete Fourier transform (DFT) of a sequence, or its inverse. To reduce the computational time, this paper uses the FFT to approximate the probability in (32). With Euler's relation $e^{i\theta} = \cos\theta + i\sin\theta$, we can confirm that $\phi_{L_n^{z,v}}(t)$ has a period of

[10] This partial derivative is componentwise.

2π; i.e., $\phi_{L_n^{z,v}}(t) = \phi_{L_n^{z,v}}(t + 2\pi)$ for all t, due to the fact that $e^{i(t+2\pi)k} = e^{itk}$.

With this periodic property, we now evaluate the characteristic function $\phi_{L_n^{z,v}}$ at N equally spaced values in the interval $[0, 2\pi]$ as

$$b_m^{z,v} = \phi_{L_n^{z,v}}\left(\frac{2\pi m}{N}\right), \quad m = 0, 1, \ldots, N-1,$$

which defines the DFT of the sequence of probabilities $q_k^{z,v}$. By using the corresponding sequence of characteristic function values above, we can recover the sequence of probabilities; that is, we aim for the sequence of $q_k^{z,v}$'s from the sequence of $b_m^{z,v}$'s, which can be achieved by employing the inverse DFT operation

$$\tilde{q}_k^{z,v} = \frac{1}{N}\sum_{m=0}^{N-1} b_m^{z,v} e^{-i2\pi km/N}, \quad k = 0, 1, \ldots, N-1.$$

Finally, the approximation of $\varrho(z, v)$ can be calculated as

$$\tilde{\varrho}(z, v) = 1 - P_{\mathrm{FFT}}(L_n^{z,v} \le \tau) = 1 - \sum_{\ell=0}^{\tau} \tilde{q}_\ell^{z,v}, \tag{33}$$

where $P_{\mathrm{FFT}}(\cdot)$ denotes the probability approximated using a fast inverse Fourier transform.

Note that even "exact" FFT algorithms have errors when using finite-precision floating-point arithmetic, but these errors are typically very small. Most FFT algorithms have outstanding numerical properties; for example, the bound on the relative error for the Cooley–Tukey algorithm is $O(\epsilon \log N)$. To attest the approximation performance, Table 6 provides several examples showing the approximation error and computational time of the inverse Fourier transform. In the table, we set the number of obligors $n = 250$ and assume $p_{z,v,\ell} = 0.1$ (i.e., $H_\ell^{z,v} \sim \texttt{Bernoulli}(0.1)$) for simplicity. To check the approximation performance, we first consider the case with equal $c_i = 1$, where the probability (denoted as $P_{\mathrm{Binomial}}(\cdot)$) is evaluated analytically via the cumulative density function of the binomial distribution with parameters $n = 250$ and $p = 0.1$. Observe that the differences between the approximated probabilities ($P_{\mathrm{FFT}}(\cdot)$) and the analytical ones ($P_{\mathrm{Binomial}}(\cdot)$) are negligible, i.e., the bias is extremely small. Moreover, we investigate the case with five different c_i, in which we compare the approximated probabilities with those generated via simulation with 500,000 samples; as shown in Table 6, the approximated probabilities all lie within the corresponding 95% confidence intervals. We note also that the computational time grows linearly with the number of different c_i.[11]

[11] All of the experiments were obtained by running programs via Mathematica 11 on a MacBook Pro with a 2.6 GHz Intel Core i7 CPU.

Bounded relative error analysis

To state that when one simulates the portfolio loss probability $P(L_n > \tau)$ in (24), the optimal sufficient exponential tilting has bounded relative error, we first define the following notation. To set the stage, recall L_n in (21). Let $u(x)$ be a function that increases at a subexponential rate such that $u(x) \to \infty$ as $x \to \infty$, and set the default thresholds for the k-th obligor to be $\chi_k = a_k u(n)$, where $a_k > 0$ is a positive constant; that is, we consider

$$L_n = c_1 \mathbb{1}_{\{X_1 > a_1 u(n)\}} + \cdots + c_n \mathbb{1}_{\{X_n > a_n u(n)\}}. \tag{34}$$

Let $W = (W_1, \ldots, W_d, W_{d+1}) := (\tilde{W}, W_{d+1})$. Due to the independence of $\{W_j, j = 1, \ldots, d+1\}$, we can define

$$\begin{aligned} L_* &= \left(\frac{f_z(Z)}{q_{z,\mu,\Sigma}(Z)}\right)\left(\frac{f_w(W)}{q_{w,\theta,\eta}(W)}\right) \\ &= \left(\frac{f_z(Z)}{q_{z,\mu,\Sigma}(Z)}\right)\left(\frac{f_{\tilde{w}}(\tilde{W})}{q_{\tilde{w},\theta,\eta}(\tilde{W})}\right)\left(\frac{f_{w_{d+1}}(W_{d+1})}{q_{w_{d+1},\theta,\eta}(W_{d+1})}\right). \end{aligned}$$

If W_{d+1} follows a gamma distribution $\texttt{Gamma}(\alpha, \beta)$, we have

$$\begin{aligned} L_* = &\left(\frac{f_z(Z)}{q_{z,\mu,\Sigma}(Z)}\right)\left(\frac{f_{\tilde{w}}(\tilde{W})}{q_{\tilde{w},\theta,\eta}(\tilde{W})}\right) \\ &\times e^{w\eta + \theta \log(w)} \frac{\Gamma(\alpha)}{(1/\beta)^{-\alpha}(\beta - \eta)^{-\alpha-\theta}\Gamma(\alpha+\theta)}. \end{aligned} \tag{35}$$

Here the domains of the tilting parameter spaces are the same as those in Theorems 5 and 6.

The asymptotic optimal tilting parameters μ^*, σ^* for Z can be obtained by using the same method of solving (14) and (15), in which the indicated set $\mathbb{1}_{\{X \in A\}}$ is replaced by $\varrho(Z, V)$ in (27); θ^* and η^* for W can be obtained by using the same method of solving (19) and (20), in which the indicated set $\mathbb{1}_{\{X > a\}}$ is replaced by $\varrho(Z, V)$ in (27).[12]

Theorem 7 *Under the setting in model (22) with $V_j = g(W_j)$ for $j = 1, \ldots, d+1$, and the assumption that $W_{d+1} \sim \texttt{Gamma}(\alpha, \beta)$, let the sequence $\{(c_k, a_k) : k \ge 1\}$, defined in (34), take values in a finite set $\mathscr{D}$. In addition, the proportion of each element $(c_k, a_k) \in \mathscr{D}$ in the portfolio converges to $q_j > 0$ as $n \to \infty$. Considering $\tau = nb$ for some $b > 0$, let $A_n = \{L_n > nb\}$. Then, we have*

$$\limsup_{n\to\infty} \frac{E^* L_*^2 \mathbb{1}_{\{A_n\}}}{(P(L_n > nb))^2} < \infty,$$

where E^* denotes the expectation under the probability measure Q in (35). In other words, optimal sufficient exponential tilting has bounded relative error.

[12] Note that here we do not tilt the correlation parameters.

The proof of Theorem 7 is given in "Appendix C".

3.3 Algorithms

This subsection summarizes the steps when we implement the proposed sufficient exponential importance sampling algorithm, which consists of two components: tilting parameter search and tail probability calculation. The aim of the first component is to determine the optimal tilting parameters. We implement the search phase using an automatic Newton method (Teng etal. 2016). We here define the conjugate measures $\bar{Q}_{\mu,\Sigma}$ for Z and $\bar{Q}_{\theta,\eta}$ for W of the measure Q.[13] With these two conjugate measures and the results in (14), (15), (20), and (19), we define functions $g_\mu(\mu)$, $g_\Sigma(\Sigma)$, $g_\theta(\theta)$, and $g_\eta(\eta)$ as

$$g_\mu(\mu) = \mu - E_{\bar{Q}_{\mu,\Sigma}}[Z \mid L_n > \tau], \tag{36}$$

$$g_\Sigma(\Sigma) = \mathscr{K}(\mu, \Sigma) - E_{\bar{Q}_{\mu,\Sigma}}\left[Z^\intercal(\nabla_{\eta_i} M)Z \mid L_n > \tau\right] \quad \text{for} \quad i = 1, 2, \ldots, d+1, \tag{37}$$

$$\begin{aligned} g_\theta(\theta) = {} & [-\log(\beta_1 - \eta_1) + \Upsilon(\alpha_1 + \theta_1), \ldots, \\ & -\log(\beta_{d+1} - \eta_{d+1}) + \Upsilon(\alpha_{d+1} + \theta_{d+1})]^\intercal \\ & -E_{\bar{Q}_{\theta,\eta}}[\ln(W) \mid L_n > \tau], \end{aligned} \tag{38}$$

$$g_\eta(\eta) = \left[\frac{\alpha_1+\theta_1}{\beta_1-\eta_1}, \ldots, \frac{\alpha_{d+1}+\theta_{d+1}}{\beta_{d+1}-\eta_{d+1}}\right]^\intercal - E_{\bar{Q}_{\theta,\eta}}[W \mid L_n > \tau], \tag{39}$$

where $\nabla_{\eta_i} M$ in (37) is defined in (16). To find the optimal tilting parameters, we must find the roots of the above four equations. With Newton's method, the roots of (36), (37), (38), and (39) are found iteratively by

$$\delta^{(k)} = \delta^{(k-1)} - J^{-1}_{\delta^{(k-1)}} g_\delta(\delta^{(k-1)}), \tag{40}$$

where the Jacobian of $g_\delta(\delta)$ is defined as

$$J_\delta[i, j] := \frac{\partial}{\partial \delta_j} g_{\delta,i}(\delta). \tag{41}$$

In (40) and (41), δ can be replaced with μ, Σ, θ, and η, and J_δ^{-1} is the inverse of the matrix J_δ.

To measure the precision of the roots to the solutions in (36), (37), (38), and (39), we define the sum of the square error of $g_\delta(\delta)$ as $\|g_\delta(\delta)\| = g'_\delta(\delta) g_\delta(\delta)$; a $\delta^{(n)}$ is accepted when $\|g_\delta(\delta^{(n)})\|$ is less than a predetermined precision level

[13] Note that these two conjugate measures are different from that defined in Sect. 2, which is additionally with respect to a general payoff function $\mathscr{P}$. As the payoff function here is the probability of losses in (24) and thus can be represented as the expectation of an indicator function (which is similar to the simple examples in Sect. 2.1), we here follow the conjugate measure in Fuh etal. (2018) for the following calculation—the event of the probability becomes the condition of the expectation (see Equations (36)–(39)).

Table 6 Approximation performance and computational time (seconds) of inverse Fourier transform

$c_i = 1$				$c_i = (\lceil 5i \rceil / n)^2$			
τ	$P_{\text{FFT}}(L_n \le \tau)$	$P_{\text{FFT}}(L_n \le \tau) - P_{\text{Binomial}}(L_n \le \tau)$	Time	τ	$P_{\text{FFT}}(L_n \le \tau)$	$P_{\text{MC}}(L_n \le \tau)$ (95% CI)	Time
20	1.72×10^{-1}	-7.19×10^{-15}	0.03	200	1.29×10^{-1}	1.29×10^{-1} $(1.28 \times 10^{-1}, 1.30 \times 10^{-1})$	0.13
10	3.53×10^{-4}	-7.69×10^{-17}	0.04	100	1.32×10^{-3}	1.31×10^{-3} $(1.21 \times 10^{-3}, 1.41 \times 10^{-3})$	0.14
5	5.84×10^{-7}	1.11×10^{-16}	0.04	50	1.20×10^{-5}	1.00×10^{-6} $(1.24 \times 10^{-6}, 1.88 \times 10^{-5})$	0.14

ϵ. To illustrate the algorithm, we here use the setting $V_j = g(W_j) = \sqrt{\nu_j/W_j}$ and $W_j \sim \texttt{Gamma}(\nu_j/2, 1/2)$ (for $j = 1, \ldots, d+1$) as an example. The detailed procedures of the first component are described as follows:

- Determine optimal tilting parameters:
  (1) Generate independent samples $z^{(i)}$ from $N(0, \mathbb{I})$ and $w_j^{(i)}$ from $\texttt{Gamma}(\nu_j/2, 1/2)$ for $i = 1, \ldots \mathscr{B}_1$; calculate $v_j^{(i)} = \sqrt{\nu_j/w_j^{(i)}}$ for $j = 1, \ldots, d+1$.
  (2) Set $\mu^{(0)}$, $\Sigma^{(0)}$, $\theta^{(0)}$, and $\eta^{(0)}$ properly; $k = 1$.
  (3) Calculate $g_{\delta^{(0)}}$ functions via (36), (37), (38), and (39).
  (4) Calculate the J_δ and their inverse matrices for the g_δ functions.
  (5) Calculate $\delta^{(k)} = \delta^{(k-1)} - J^{-1}_{\delta^{(k-1)}} g_\delta(\delta^{(k-1)})$ in (40) for $\delta = \mu, \Sigma, \theta, \eta$.
  (6) Calculate $g_{\delta^{(k)}}$ functions via (36), (37), (38), and (39). If $\forall\ \delta \in \{\mu, \Sigma, \theta, \eta\},\ \ \|g_\delta(\delta^{(k)})\| < \epsilon$, set $\delta^* = \delta^{(k)}$ and stop. Otherwise, return to step (4).

We proceed to describe the second component that calculates the probability of losses, in which optimal tilting parameters μ^*, Σ^*, θ^* and η^* are used (see step (6) for the first component above). The detailed procedures of the second component are summarized as follows:

- Calculate the probability of losses, $P(L_n > \tau)$:
  (1) Generate independent samples $z^{(i)}$ from $N(\mu^*, \Sigma^*)$ and $w_j^{(i)}$ from $\texttt{Gamma}(\nu_j/2 - \theta_j^*, 1/2 + \eta_j^*)$ for $i = 1, \ldots \mathscr{B}_2$; calculate $v_j^{(i)} = \sqrt{\nu_j/w_j^{(i)}}$ for $j = 1, \ldots, d+1$.
  (2) Estimate m by $\hat{m} = \frac{1}{\mathscr{B}_2}\sum_{i=1}^{\mathscr{B}_2} \tilde{\varrho}(z^{(i)}, v^{(i)})\, r_{z,\mu^*,\Sigma^*}(z)\, r_{w,\theta^*,\eta^*}(w)$ in (28), where $\tilde{\varrho}(z, v)$ is calculated by the analytical form from (33) and μ^*, Σ^*, θ^*, and η^* are obtained from step (6) of the first component of the algorithm.

As a side note, in the above sufficient exponential tilting algorithm, a component-wise Newton method is adopted to determine the optimal tilting parameters μ, Σ, θ, and η; this differs from Algorithm 2 in Teng etal. (2016), which involves only one-parameter tilting for μ.

Remark 6 To implement the proposed importance sampling, we require an additional searching stage. The searching stage employs the recursive formula in (40), in which the function $g_\delta(\cdot)$ and the Jacobian J_δ do not have closed-form formulas and must be approximated by Monte Carlo simulation. For easy presentation, as V (or W) dominates the performance [as stated in 3.3 of Bassamboo etal. (2008)], we here use the case for estimating η^* (see Eq. (39)) to illustrate the approximation.

Let $\hat{g}_\eta(\eta)$ be the Monte Carlo estimator of $g_\eta(\eta)$, defined as

$$\hat{g}_\eta(\eta) = \left[\tfrac{\alpha_1+\theta_1}{\beta_1-\eta_1}, \ldots, \tfrac{\alpha_{d+1}+\theta_{d+1}}{\beta_{d+1}-\eta_{d+1}}\right]^\top - \frac{1}{\mathscr{B}_1}\sum_{s=1}^{\mathscr{B}_1} Y^{(s)},$$

where $Y^{(1)}, \ldots, Y^{(\mathscr{B}_1)}$ are random samples under $\bar{Q}_{\theta,\eta}$. However, it is difficult to generate samples from $\bar{Q}_{\theta,\eta}$ because it involves the payoff function $\varrho(Z, V)$. Recall that substituting (8) for $\bar{Q}_{\theta,\eta}$ in (39) yields

$$g_\eta(\eta) = \left[\tfrac{\alpha_1+\theta_1}{\beta_1-\eta_1}, \ldots, \tfrac{\alpha_{d+1}+\theta_{d+1}}{\beta_{d+1}-\eta_{d+1}}\right]^\top - \frac{E_P[\varrho^2(Z,V)\, W \mathrm{e}^{-(\theta^\top \log(W) + \eta^\top W)}]}{E_P[\varrho^2(Z,V)\, \mathrm{e}^{-(\theta^\top \log(W) + \eta^\top W)}]}.$$

Therefore, we can estimate $g_\eta(\eta)$ by

$$\hat{g}_\eta(\eta) = \left[\tfrac{\alpha_1+\theta_1}{\beta_1-\eta_1}, \ldots, \tfrac{\alpha_{d+1}+\theta_{d+1}}{\beta_{d+1}-\eta_{d+1}}\right]^\top - \frac{\sum_{s=1}^{\mathscr{B}_1} \varrho^2\left(Z^{(s)}, V^{(s)}\right) W^{(s)} \mathrm{e}^{-(\theta^\top \log(W^{(s)}) + \eta^\top W^{(s)})}}{\sum_{s=1}^{\mathscr{B}_1} \varrho^2\left(Z^{(s)}, V^{(s)}\right) \mathrm{e}^{-(\theta^\top \log(W^{(s)}) + \eta^\top W^{(s)})}}, \tag{42}$$

where $Z^{(1)}, \ldots, Z^{(\mathscr{B}_1)}$ and $V^{(1)}, \ldots, V^{(\mathscr{B}_1)}$ (and $W^{(1)}, \ldots, W^{(\mathscr{B}_1)}$) are i.i.d. samples under P. Note that under the finiteness of the second moment assumption in (4), the standard strong law of large numbers implies that the second term on the right-hand side of (42) converges P-almost surely to

$$E_P[\varrho^2(Z,V)\, W \mathrm{e}^{-(\theta^\top \log(W) + \eta^\top W)}]/E_P[\varrho^2(Z,V)\, \mathrm{e}^{-(\theta^\top \log(W) + \eta^\top W)}],$$

which is $E_{\bar{Q}_{\theta,\eta}}[W \mid L_n > \tau]$ by the definition of conjugate measure $\bar{Q}_{\theta,\eta}$.

To approximate the second term on the right-hand side of (42), we apply a similar technique to that in Fuh and Hu (2004), which uses a small number of $\mathscr{B}_1$ to locate the optimal tilting parameters. By (39), (42), and a similar argument to that in Theorem 1 of Bassamboo etal. (2008), we have $\mathbb{1}_{\{L_n > \tau\}} \sim \mathbb{1}_{\{W_{d+1} > a\}}$ when $V_j = g(W_j) = \sqrt{W_j}$ for $j = 1, \ldots, d+1$. This implies that

$$E_{\bar{Q}_{\theta,\eta}}[W \mid L_n > \tau] \sim E_{\bar{Q}_{\theta,\eta}}\left[W \mid W_{d+1} > a\right] = \frac{\int_a^\infty x \cdot x^{\alpha-1}\exp(-\beta x)dx}{\int_a^\infty x^{\alpha-1}\exp(-\beta x)dx}.$$

Since $P(W_{d+1} > a)$ is small when a is large, the preceding equation indicates that we may lose numerical/simulation precision. Therefore, we multiply both the numerator and denominator of the preceding equation by $\exp(\eta x + \theta \log x)$, for suitable $\theta = \xi$, $\eta = \beta - (\alpha + \xi)/a$ for a large ξ as suggested in (B20) and (B21), and compute

$$E_{\bar{Q}_{\theta,\eta}}[W \mid L_n > \tau] \sim \frac{\int_a^\infty x \cdot x^{\alpha+\theta-1} \exp(-(\beta-\eta)x)dx}{\int_a^\infty x^{\alpha+\theta-1} \exp(-(\beta-\eta)x)dx}. \tag{43}$$

Then we run the simulation in (42) based on (43). Note that over an important part of the set $[a, \infty)$, $x^{\alpha+\theta-1} \exp(-(\beta - \eta)x)$ is a moderately sized number such that the conditional expectation in (43) can be simulated quite accurately with a reasonably small size of $\mathscr{B}_1$. On the other hand, similar techniques can be used for the case $V_j = g(W_j) = \sqrt{\nu_j/W_j}$; that is, we turn to estimate $E_{\bar{Q}_{\theta,\eta}}[W \mid L_n > \tau] \sim E_{\bar{Q}_{\theta,\eta}}\left[W \mid W_{d+1} < 1/a\right]$ (check "Appendix B").

3.4 Numerical results

We compare the performance between our method and those proposed in Bassamboo etal. (2008) and Chan and Kroese (2010). For comparison purposes, we adopt the same sets of parameter values as those in Table 1 of Bassamboo etal. (2008), where the latent variables X_k in 22 follow a t-distribution, i.e., $V_1 = V_2 = \sqrt{\nu_1/W_1}$ and $W_1 \sim \texttt{Gamma}(\nu_1/2, 1/2)$. The model parameters were chosen to be $n = 250$, $\rho_{11} = 0.25$, the default thresholds for each individual obligors $\chi_i = 0.5 \times \sqrt{n}$, each $c_i = 1$, $\tau = 250 \times b$, $b = 0.25$, and $\sigma_\epsilon = 3$. Table 7 reports the results of the exponential change of measure (ECM) proposed in Bassamboo etal. (2008) and conditional Monte Carlo simulation without and with cross entropy (CondMC and CondMC-CE, respectively) in Chan and Kroese (2010). As observed from the table, the proposed algorithm (the last four columns) offers substantial variance reduction compared with crude simulation; in general, it compares favorably to the ECM and CondMC estimators. Moreover, for a fair comparison with the results of CondMC-CE, we first follow the CondMC method by integrating out the shock variable analytically; then, instead of using the cross-entropy approach, we apply the proposed importance sampling method for variance reduction. Under this setting, the proposed method yields variance reductions comparable to those of CondMC-CE.

In addition to the above results, we additionally conducted experiments with the following setup, the results of which are reported and discussed in "Appendix D". First, we compare the performance of the proposed method with crude simulation, under three-factor normal mixture models, in which the t-distribution or the GIG distribution for X_k is considered. (Note that we compare the performance of our method only with crude Monte Carlo simulation henceforth, as most of the literature focuses on simulating one-dimensional cases.) The cases with different losses resulting from default of the obligors are also investigated. Finally, we compare the computational time of crude Monte Carlo simulation with that of the proposed importance sampling under several scenarios, and provide insight into the trade-off between reduced variance and increased computational time. Detailed setups for all the experiments for portfolio losses are also found in Appendix 3.4.

4 Conclusion

This paper introduces a comprehensive framework for a sufficient exponential tilting algorithm, supported by both theoretical foundations and empirical evidence from numerical experiments. We employ this approach in a multi-factor model featuring a normal mixture copula. Traditional methods for calculating portfolio credit risk—such as direct analysis or basic Monte Carlo simulation—are insufficient due to the large portfolio size, diverse obligor characteristics, and the interconnected yet infrequent nature of default events. To address these challenges, we offer an optimized simulation algorithm that more accurately estimates the likelihood of significant portfolio losses when a normal mixture copula is involved. In summary, our computational method brings two main innovations to the table. First, from a statistical standpoint, we introduce an unconventional sufficient (statistic) exponential embedding that serves as a versatile tool in importance sampling. Secondly, our work appears to be the first to specifically tackle high-dimensional importance sampling within multi-factor models that go beyond the standard normal copula assumptions. In the broader context, the relevance and application of our work extend beyond credit risk to extreme event modeling in various domains, such as pandemics, highlighting its importance and versatility. The computational challenges associated with extreme event modeling are well-known, and our paper contributes significantly to this arena.

There are several possible future directions based on this model. To name a few, first, we will explore more properties of the proposed sufficient exponential tilting, and apply it to more practical cases, to see how far it can go. Second, as suggested by a reviewer, it could be helpful to discuss computing the value-at-risk or expected shortfall for general financial risk management or for insurance company interests when managing the large number of obligors involved, as these remain demanding topics in practice. Third, although in this paper the default time is fixed and the default boundary is exogenous, the default time could be any time before a pre-fixed time T and the default boundary could depend on firm characteristics and may be state- and time-dependent. To capture these phenomena, we will consider more complicated dynamic models, for which importance sampling should be

Table 7 Performance of proposed algorithm with equal loss resulting from default of obligors for a one-factor model (t-distribution)

	Bassamboo etal. (2008)		Chan and Kroese (2010)		Importance sampling (IS)			
		ECM	CondMC	CondMC-CE	IS		CondMC-IS	
ν	$P(L_n > \tau)$	V.R. factor	V.R. factor		$P(L_n > \tau)$	V.R. factor	$P(L_n > \tau)$	V.R. factor
4	8.13×10^{-3}	65	271	2440	8.11×10^{-3}	338	8.09×10^{-3}	1600
8	2.42×10^{-4}	878	1690	20,656	2.36×10^{-4}	6212	2.47×10^{-4}	14,770
12	1.07×10^{-5}	7331	12,980	2.08×10^{5}	1.04×10^{-5}	16,100	1.10×10^{-5}	1.57×10^{5}
16	6.16×10^{-7}	52,185	81,170	1.30×10^{6}	6.34×10^{-7}	2.78×10^{5}	6.20×10^{-7}	1.89×10^{6}
20	4.38×10^{-8}	301,000	4.19×10^{5}	1.27×10^{7}	4.12×10^{-8}	5.44×10^{6}	4.14×10^{-8}	1.61×10^{7}

more sophisticated. Before that, it would be interesting to develop an importance sampling algorithm for first-passage time events (i.e., the probability that sums of non-negative random variables fall below a sufficiently small threshold), and study the performance of sufficient exponential tilting in the domain of state-dependent importance sampling. Finally, as credit derivatives are among the fastest growing contracts in the derivatives market (Chen and Sopranzetti 2003; Ericsson etal. 2009; Hirtle 2009), another interesting future research direction would be to apply the proposed technique to credit derivative valuation.

Author Contributions C-DF and C-JW wrote the main manuscript text and C-JW conducted the experiments and prepared the figures and tables. All authors reviewed the manuscript.

Declarations

Conflict of interest The authors declare no competing interests.

Appendix A Proof of Theorem 5

(I) Obtaining the expected squared estimator with optimal tilting parameters.

We first approximate the optimal tilting parameters θ^* and η^*.

With $\varphi(x)$ as the pdf of the standard normal distribution, the expected squared estimator for the event $\{X > a\}$ is

$$\int_a^\infty \frac{\varphi(x)}{\varphi_{\theta,\eta}(x)}\varphi(x)dx = \int_a^\infty \frac{1}{\sqrt{1-2\eta}} e^{-(\theta x+\eta x^2)+\theta^2/(2-4\eta)} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx = \frac{e^{\frac{\theta^2}{1-4\eta^2}}\operatorname{erfc}\left(\frac{a+2a\eta+\theta}{\sqrt{4\eta+2}}\right)}{\sqrt{2}\sqrt{1-2\eta}\sqrt{2+4\eta}}, \tag{A1}$$

where $\operatorname{erfc}(\cdot)$ denotes the complementary error function. Note that the above integral is finite if $\eta \geq -1/2$. Then, we approximate (A1) and define $g(\theta,\eta)$ as

$$\frac{e^{\frac{\theta^2}{1-4\eta^2}}\operatorname{erfc}\left(\frac{a+2a\eta+\theta}{\sqrt{2+4\eta}}\right)}{\sqrt{2}\sqrt{1-2\eta}\sqrt{2+4\eta}} \sim \frac{e^{\frac{\theta^2}{1-4\eta^2}}\, 2\frac{\phi\left(\sqrt{2}(a+2a\eta+\theta)/\sqrt{2+4\eta}\right)}{\sqrt{2}(a+2a\eta+\theta)/\sqrt{2+4\eta}}}{\sqrt{2}\sqrt{1-2\eta}\sqrt{2+4\eta}} = \frac{e^{\frac{\theta^2}{1-4\eta^2}-\frac{(a+2a\eta+\theta)^2}{2+4\eta}}}{\sqrt{2\pi}\sqrt{1-2\eta}(a+2a\eta+\theta)} := g(\theta,\eta). \tag{A2}$$

By taking the partial derivatives of (A2) with respect to θ and setting it to 0, we have

$$\frac{\partial g(\theta,\eta)}{\partial\theta} = 0 \Rightarrow \theta^* = \sqrt{-2\eta+a^2+1} - 2\eta a. \tag{A3}$$

Note that in (A3) we omit the other solution $\theta^* = -\sqrt{-2\eta+a^2+1} - 2\eta a$ as we here consider the event $\{X > a\}$ for a positive and large a. By substituting θ^* in (A3) into $\frac{\partial g(\theta,\eta)}{\partial\eta}$, we have $\frac{\partial \log g(\theta,\eta)}{\partial\eta} > 0$ for $-1/2 \leq \eta < 1/2$ (note that constraint $\eta < 1/2$ guarantees that the variance of the tilted distribution is positive; see the example for the normal distribution in Sect. 2.1.) Therefore, we have $\eta^* \to -1/2$ as $a \to \infty$.

From (A3), as $a \to \infty$, $\theta^* \to a(1-2\eta)$. By substituting $\theta^* = a(1-2\eta) + \Delta \sim a(1-2\eta)$, as $\Delta \to 0$, into (A2), we have

$$\mathscr{V}_\eta = \int_a^\infty \frac{\varphi(x)}{\varphi_{\theta^*,\eta}(x)}\varphi(x)dx \sim \frac{e^{\frac{(a(1-2\eta))^2}{1-4\eta^2}-\frac{(a+2a\eta+a(1-2\eta))^2}{2+4\eta}}}{\sqrt{2\pi}\sqrt{1-2\eta}(a+2a\eta+a(1-2\eta))} = \frac{e^{-c_1a^2}}{\sqrt{2\pi}2ac_2}, \tag{A4}$$

where

$$c_1 = 1, \quad c_2 = \sqrt{1-2\eta}. \tag{A5}$$

(II) Bounded relative error analysis.

Recall $p = P(X > a)$. Since $p = 1 - \Phi(a) \sim \frac{\phi(a)}{a}$ implies $\exp(-a^2) \sim 2\pi a^2 p^2$, from (A4) and (A5), we have

$$\frac{1}{p^2}\int_a^\infty \frac{\varphi(x)}{\varphi_{\theta^*,\eta}(x)}\varphi(x)dx \sim \frac{1}{p^2}\frac{e^{-c_1a^2}}{\sqrt{2\pi}2ac_2} = \frac{2\pi a^2}{e^{-a^2}}\frac{e^{-c_1a^2}}{\sqrt{2\pi}2ac_2} = \sqrt{\frac{\pi}{2}}\frac{a}{c_2}e^{-(c_1-1)a^2} = \sqrt{\frac{\pi}{2}}\frac{a}{c_2}. \tag{A6}$$

Equation (A6) clearly indicates that optimal sufficient exponential tilting does not exhibit bounded relative error.

(III) Logarithmic efficiency analysis.

Regarding logarithmic efficiency, we have

$$\begin{aligned}\frac{1}{p^{2-\varepsilon}}\int_a^\infty \frac{\varphi(x)}{\varphi_{\theta^*,\eta}(x)}\varphi(x)dx &\sim \frac{1}{p^{2-\varepsilon}}\frac{e^{-c_1a^2}}{\sqrt{2\pi}2ac_2}\\ &= \sqrt{\frac{\pi}{2}}\frac{a}{c_2}e^{-(c_1-1)a^2}\left(\frac{(\sqrt{2\pi}a)^{-\varepsilon}}{e^{\frac{a^2\varepsilon}{2}}}\right)\\ &= \pi^{\frac{1-\varepsilon}{2}}(1/2)^{\frac{1+\varepsilon}{2}}\frac{a^{1-\varepsilon}}{c_2}e^{-(\varepsilon/2)a^2}.\end{aligned} \tag{A7}$$

By (A7), we have logarithmic efficiency for optimal sufficient exponential tilting.

(IV) Ratio between $\mathscr{V}_\eta$ and $\mathscr{V}_0$.

By (A4), we have

$$\frac{\mathscr{V}_\eta}{\mathscr{V}_0} = \frac{\frac{e^{-c_1a^2}}{\sqrt{2\pi}2ac_2}}{\frac{e^{-c_1a^2}}{\sqrt{2\pi}2a}} = \frac{1}{c_2} = \frac{1}{\sqrt{1-2\eta}}. \tag{A8}$$

Note that the ratio in (A8) goes to $1/\sqrt{2}$ as $\eta^* \to -1/2$ when $a \to \infty$ (see the results two lines below (A3)). Therefore, the proposed optimal sufficient exponential tilting outperforms optimal one-parameter tilting in the sense of second (small) order asymptotic logarithmic efficiency.

Appendix B Proof of Theorem 6

B.1 For event $\{X > a\}$

(I) Obtaining the expected squared estimator with optimal tilting parameters.

For the gamma distribution with event $\{X > a\}$, we approximate θ^* and η^* as follows. From (18), we have

$$\frac{dP}{dQ_{\theta,\eta}} = e^{-x\eta-\theta\log(x)}\frac{(1/\beta)^{-\alpha}(\beta-\eta)^{-\alpha-\theta}\Gamma(\alpha+\theta)}{\Gamma(\alpha)}, \tag{B9}$$

where $\Gamma(\cdot)$ is the gamma function. As $a \to \infty$, plug (B9) into (5) and adopt Laplace's method to yield

$$\begin{aligned}&E_P\left[e^{-X\eta-\theta\log(X)}\frac{(1/\beta)^{-\alpha}(\beta-\eta)^{-\alpha-\theta}\Gamma(\alpha+\theta)}{\Gamma(\alpha)}\mathbb{1}_{\{X>a\}}\right]\\ &\sim e^{-a\eta-\theta\log(a)}\frac{(1/\beta)^{-\alpha}(\beta-\eta)^{-\alpha-\theta}\Gamma(\alpha+\theta)}{\Gamma(\alpha)}\\ &\times E_P[\mathbb{1}_{\{X>a\}}] := g(\theta,\eta).\end{aligned} \tag{B10}$$

Note that as $\mathbb{1}_{\{X>a\}}$ is a function of α, β, and a only, it is thus related to neither θ nor η. Then, we approximate the optimal tilting parameters θ^* and η^* by considering the log upper bound $\log g(\theta,\eta)$. By taking the partial derivatives of $\log g(\theta,\eta)$ with respect to θ and η and set to be 0, we have

$$\frac{\partial \log g(\theta,\eta)}{\partial\theta} = -\log(a) - \log(\beta-\eta) + \Upsilon(\alpha+\theta) \tag{B11}$$

$$\frac{\partial \log g(\theta,\eta)}{\partial\eta} = \frac{\alpha - a(\beta-\eta)+\theta}{\beta-\eta} = 0 \Rightarrow \eta^* = \frac{a\beta-\alpha-\theta}{a}, \tag{B12}$$

where $\Upsilon(\cdot)$ is the digamma function.

To obtain the optimal θ^* and η^*, we adopt the inequality $\log x - \frac{1}{x} \leq \Upsilon(x) \leq \log x - \frac{1}{2x}$, for all $x > 0$. Then we have

$$\Upsilon(\alpha+\theta) \sim \log(\alpha+\theta) - \frac{1}{\alpha+\theta}.$$

With the above approximation and the substitution of η^* in (B12), Eq. (B11) becomes

$$\begin{aligned}&-\log(a) - \log(\beta-\eta) + \log(\alpha+\theta) - \frac{1}{\alpha+\theta}\\ &= \log\frac{\alpha+\theta}{a(\beta-\eta)} - \frac{1}{\alpha+\theta}\\ &= \log\frac{\alpha+\theta}{a\beta-(a\beta-\alpha-\theta)} - \frac{1}{\alpha+\theta}\\ &= \log 1 - \frac{1}{\alpha+\theta} < 0.\end{aligned} \tag{B13}$$

Also, as $\alpha + \theta > 0$ (see the example for the gamma distribution in Sect. 2.1), we have

$$\eta^* = \frac{a\beta-\alpha-\theta}{a} < \beta. \tag{B14}$$

Note that the domain of θ is $-\alpha < \theta$ *and* $\alpha - \theta \in \mathbb{R}\backslash\{0,-1,-2,\ldots\}$, whereas $-\beta < \eta$ ensures that the expected squared estimator for the event $X > a$ is finite, and $\eta < \beta$ guarantees that the rate parameter of the tilted distribution is positive. By (B13) and (B14) and the above constraints, we have

1. $-\alpha < \theta^* \to \infty$ as $a \to \infty$;

2. $\eta^* = \beta - \frac{\alpha+\theta}{a} \to \beta - c$ as $\theta \to \infty$ and $a \to \infty$ such that $\theta/a \to c$, where c denotes a constant with the range $0 < c < 2\beta$.

With $f(x)$ as the pdf of the gamma distribution, Gamma(α, β), the expected square estimator with $\eta^* = (a\beta - \alpha - \theta)/a$ is

$$\begin{aligned}\mathscr{V}_\theta &= \int_a^\infty \frac{f(x)}{f_{\theta,\eta^*}(x)} f(x)dx \\ &= \int_a^\infty e^{-x\eta-\theta\log x} \frac{(1/\beta)^{-\alpha}(\beta-\eta)^{-\alpha-\theta}\Gamma(\alpha+\theta)}{\Gamma(\alpha)} \\ &\quad (\beta^\alpha/\Gamma(\alpha))x^{\alpha-1}e^{-\beta x}dx \\ &= \frac{(1/\beta)^{-\alpha}(\beta)^\alpha((\alpha+\theta)/a)^{-\alpha-\theta}\Gamma(\alpha+\theta)}{(\Gamma(\alpha))^2} \\ &\quad \times \frac{((2\beta a-\alpha-\theta)/a)^{-\alpha+\theta}\Gamma(\alpha-\theta, 2\beta a-\alpha-\theta)}{(\Gamma(\alpha))^2},\end{aligned} \tag{B15}$$

where $\Gamma(s,x) = \int_x^\infty t^{s-1}e^{-t}dt$ is the upper incomplete gamma function.

Let $\theta = 0$; we have

$$\mathscr{V}_0 = \frac{(1/\beta)^{-\alpha}(\beta)^\alpha((\alpha)/a)^{-\alpha}\Gamma(\alpha)((2\beta a-\alpha)/a)^{-\alpha}\Gamma(\alpha, 2\beta a-\alpha)}{(\Gamma(\alpha))^2}. \tag{B16}$$

(II) Bounded relative error analysis.

Using the asymptotic behavior

$$\frac{\Gamma(s,x)}{x^{s-1}e^{-x}} \to 1 \; as \; x \to \infty, \tag{B17}$$

from (B15), for optimal sufficient exponential tilting, we have

$$\lim_{a\to\infty,\theta\to\infty} \frac{1}{p^2} \int_a^\infty \frac{f(x)}{f_{\theta,\eta^*}(x)} f(x)dx = 0,$$

where $p = P(X > a)$.

Proof. Let $\theta/a \to c$ for a constant $c > 0$ as $a \to \infty$ and $\theta \to \infty$.

$$\begin{aligned}&\frac{1}{p^2}\int_a^\infty \frac{f(x)}{f_{\theta,\eta^*}(x)} f(x)dx \\ &= C \times \frac{((\alpha+\theta)/a)^{-\alpha-\theta}((2\beta a-\alpha-\theta)/a)^{-\alpha+\theta}\Gamma(\alpha-\theta, 2\beta a-\alpha-\theta)}{\Gamma^2(\alpha, \beta a)} \\ &\sim C \times \frac{((\alpha+\theta)/a)^{-\alpha-\theta}((2\beta a-\alpha-\theta)/a)^{-\alpha+\theta}(2\beta a-\alpha-\theta)^{\alpha-\theta-1}e^{-2\beta a+\alpha+\theta}}{((\beta a)^{\alpha-1}e^{-\beta a})^2} \\ &= C \times \frac{((\alpha+\theta)/a)^{-\alpha-\theta}a^{\alpha-\theta}(2\beta a-\alpha-\theta)^{-1}e^{-2\beta a+\alpha+\theta}}{((\beta a)^{\alpha-1}e^{-\beta a})^2} \\ &= C \times \frac{a^{2-\alpha}\beta^{2-2\alpha}e^\alpha}{2\beta a-\alpha-\theta} \times ((\alpha+\theta)/a)^{-\alpha-\theta}e^{-\theta(\log a-1)} \\ &= C \times \frac{a^{2-\alpha}\beta^{2-2\alpha}e^\alpha}{2\beta a-\alpha-\theta} \times ((\alpha+ca)/a)^{-\alpha-ca}e^{-ca(\log a-1)} \\ &= C \times \frac{a^{2-\alpha}\beta^{2-2\alpha}e^\alpha}{2\beta a-\alpha-\theta} \times (c(1+(\alpha/c)/a))^{-\alpha-ca}e^{-ca(\log a-1)} \\ &= C \times \frac{a^{2-\alpha}\beta^{2-2\alpha}e^\alpha c^{-\alpha}}{2\beta a-\alpha-\theta} \times (1+(\alpha/c)/a)^{-\alpha-ca}e^{-ca(\log a-1+\log c)} \\ &= C' \times \frac{a^{2-\alpha}}{2\beta a-\alpha-\theta}e^{-ca(\log a-1+\log c)},\end{aligned} \tag{B18}$$

where both C and C' are constants unrelated to a. Note that as $a \to \infty$ and $\theta \to \infty$, $(1+(\alpha/c)/a)^{-ca}$ in (B18) converges to $\exp(-\alpha)$ and thus the last term in the above equations approaches 0. Therefore, when simulating the event $\{X > a\}$, we have bounded relative error for optimal sufficient exponential tilting of the gamma distribution.

For traditional optimal one-parameter tilting on parameter η^*, from (B16) and (B17), it is clear that

$$\lim_{a\to\infty,\theta\to\infty} \frac{1}{p^2} \int_a^\infty \frac{f(x)}{f_{\theta,\eta^*}(x)} f(x)dx = \infty,$$

which means there is no bounded relative error for optimal one-parameter tilting for the gamma distribution when simulating the event $\{X > a\}$.

(III) Logarithmic efficiency analysis for optimal one-parameter tilting.

Concerning logarithmic efficiency for optimal one-parameter tilting, from (B16), we have

$$\begin{aligned}&\frac{1}{p^{2-\varepsilon}} \int_a^\infty \frac{f(x)}{f_{\theta,\eta^*}(x)} f(x)dx \\ &= \frac{(1/\beta)^{-\alpha}(\beta)^\alpha((\alpha)/a)^{-\alpha}\Gamma(\alpha)((-\alpha+2a\beta)/a)^{-\alpha}\Gamma(\alpha, -\alpha+2a\beta)}{(\Gamma(\alpha,\beta a)/\Gamma(\alpha))^{2-\varepsilon}} \\ &\sim \frac{(1/\beta)^{-\alpha}(\beta)^\alpha((\alpha)/a)^{-\alpha}\Gamma(\alpha)((-\alpha+2a\beta)/a)^{-\alpha}}{(\Gamma(\alpha))^{2-\varepsilon}} \\ &\times \frac{(-\alpha+2a\beta)^{\alpha-1}e^{\alpha-2a\beta}}{((\beta a)^{\alpha-1}e^{-\beta a})^{2-\varepsilon}}.\end{aligned} \tag{B19}$$

From (B19), it is clear that for all $\varepsilon > 0$,

$$\lim_{a\to\infty} \frac{1}{p^{2-\varepsilon}} \int_a^\infty \frac{f(x)}{f_{\theta,\eta^*}} f(x)dx = 0,$$

which means that traditional optimal one-parameter tilting on η is of logarithmic efficiency. This is consistent with Example 1.3 of Chapter VI in Asmussen and Glynn (2007).

B.2 For event $\{X < 1/a\}$

(I) Obtaining the expected squared estimator with optimal tilting parameters.

For simulation of rare event $\mathbb{1}_{\{X<1/a\}}$, by using Laplace's method, we consider

$$\begin{aligned}&E_P\left[e^{-X\eta-\theta\log(X)}\frac{(1/\beta)^{-\alpha}(\beta-\eta)^{-\alpha-\theta}\Gamma(\alpha+\theta)}{\Gamma(\alpha)}\mathbb{1}_{\{X<1/a\}}\right] \\ &\sim e^{-\eta/a-\theta\log(1/a)}\frac{(1/\beta)^{-\alpha}(\beta-\eta)^{-\alpha-\theta}\Gamma(\alpha+\theta)}{\Gamma(\alpha)} \\ &E[\mathbb{1}_{\{X<1/a\}}].\end{aligned}$$

Denote

$$g(\theta, \eta) := e^{-\eta/a+\theta \log(a)} \frac{(1/\beta)^{-\alpha}(\beta-\eta)^{-\alpha-\theta}\Gamma(\alpha+\theta)}{\Gamma(\alpha)} E[\mathbb{1}_{\{X<1/a\}}].$$

Note that as $\mathbb{1}_{\{X<1/a\}}$ is a function of α, β, and a only and is thus related to neither θ nor η, we have

$$\frac{\partial \log g(\theta,\eta)}{\partial\theta} = \log(a) - \log(\beta-\eta) + \Upsilon(\alpha+\theta) \tag{B20}$$

$$\frac{\partial \log g(\theta,\eta)}{\partial\eta} = \frac{a(\alpha+\theta)-(\beta-\eta)}{a(\beta-\eta)} = 0 \Rightarrow \eta^* = \beta - a(\alpha+\theta), \tag{B21}$$

where $\Upsilon(\cdot)$ is the digamma function.

For all $x > 0$, $\log x - \frac{1}{x} \le \Upsilon(x) \le \log x - \frac{1}{2x}$; thus we have

$$\Upsilon(\alpha+\theta) \sim \log(\alpha+\theta) - \frac{1}{\alpha+\theta}.$$

With the above approximation and the substitution of η^* in (B21), Eq. (B20) becomes

$$\begin{aligned}
&\log(a) - \log(\beta-\eta) + \log(\alpha+\theta) - \frac{1}{\alpha+\theta} \\
&= \log\frac{a(\alpha+\theta)}{\beta-\eta} - \frac{1}{\alpha+\theta} \\
&= \log\frac{a(\alpha+\theta)}{\beta-(\beta-a(\alpha+\theta))} - \frac{1}{\alpha+\theta} = \log 1 - \frac{1}{\alpha+\theta} < 0.
\end{aligned} \tag{B22}$$

From (B21), we have

1. With constraint $\alpha-\theta^*-1 > -1$, we have $-\alpha < \theta^* < \alpha$. Next, by (B22), $g(\theta,\eta)$ is a decreasing function of θ, and we have $\theta^* \uparrow \alpha$;
2. $\eta^* = \beta - a(\alpha+\theta) \to \beta - 2a\alpha \to -\infty$, as $a \to \infty$.

Note that constraint $\alpha-\theta-1 > -1$ ensures that the expected square estimator for simulating the event $\{X < 1/a\}$ is finite. A similar constraint can be found in Example 1.4 of Chapter VI in Asmussen and Glynn (2007).

For easier presentation and notational simplicity, we hereafter denote $z = 1/a$. The expected square estimator with $\eta^* = \frac{z\beta-(\alpha+\theta)}{z}$ thus becomes

$$\begin{aligned}
\mathscr{V}_\theta &= \int_0^z \frac{f(x)}{f_{\theta,\eta^*}(x)} f(x)dx \\
&= \int_0^z e^{-x\eta-\theta\log x} \frac{(1\beta)^{-\alpha}(\beta-\eta)^{-\alpha-\theta}\Gamma(\alpha+\theta)}{\Gamma(\alpha)} (\beta^\alpha/\Gamma(\alpha))x^{\alpha-1}e^{-\beta x}dx \\
&= \frac{(1/\beta)^{-\alpha}(\beta)^\alpha((\alpha+\theta)/z)^{-\alpha-\theta}\Gamma(\alpha+\theta)}{(\Gamma(\alpha))^2} \times \frac{((2\beta z-\alpha-\theta)/z)^{-\alpha+\theta}\gamma(\alpha-\theta, 2\beta z-\alpha-\theta)}{(\Gamma(\alpha))^2},
\end{aligned} \tag{B23}$$

where $\gamma(s,x) = \int_0^x t^{s-1}e^{-t}dt$ is the lower incomplete gamma function.

Letting $\theta = 0$, we have

$$\mathscr{V}_0 = \frac{(1/\beta)^{-\alpha}(\beta)^\alpha((\alpha)/z)^{-\alpha}\Gamma(\alpha)((2\beta z-\alpha)/z)^{-\alpha}\gamma(\alpha, 2\beta z-\alpha)}{(\Gamma(\alpha))^2}. \tag{B24}$$

(II) Bounded relative error analysis.

Using the asymptotic behavior

$$\frac{\gamma(s,x)}{x^s} \to \frac{1}{s}\ as\ x \to 0, \tag{B25}$$

and (B23), for optimal sufficient exponential tilting we have

$$\lim_{z\to 0,\theta\to\alpha} \frac{1}{p^2}\int_0^z \frac{f(x)}{f_{\theta,\eta^*}(x)} f(x)dx < \infty.$$

Proof.

$$\begin{aligned}
&\frac{1}{p^2}\int_0^z \frac{f(x)}{f_{\theta,\eta^*}(x)} f(x)dx \\
&= C \times \frac{((\alpha+\theta)/z)^{-\alpha-\theta}((2\beta z-\alpha-\theta)/z)^{-\alpha+\theta}\gamma(\alpha-\theta, 2\beta z-\alpha-\theta)}{\gamma^2(\alpha,\beta z)} \\
&\sim C \times \frac{((\alpha+\theta)/z)^{-\alpha-\theta}((2\beta z-\alpha-\theta)/z)^{-\alpha+\theta}(2\beta z-\alpha-\theta)^{\alpha-\theta}/(\alpha-\theta)}{(\beta z)^{2\alpha}/\alpha^2} \\
&\sim C' \times z^{2\alpha} \times z^{-2\alpha} = C',
\end{aligned} \tag{B26}$$

when replacing $\theta = \alpha - \Delta$ with $\Delta \to 0$ as $z \to 0$ in (B26); note that C' is a constant unrelated to z. Therefore, when simulating the event $\{X < 1/a\}$, we have bounded relative error for optimal sufficient exponential tilting of the gamma distribution.

For traditional optimal one-parameter tilting on parameter η, from (B24) and (B25), it is clear that

$$\lim_{z\to 0} \frac{1}{p^2}\int_0^z \frac{f(x)}{f_{\theta,\eta^*}} f(x)dx < \infty,$$

yielding bounded relative error.

(III) Ratio between $\mathscr{V}_\theta$ and $\mathscr{V}_0$.

As both optimal sufficient exponential tilting and optimal one-parameter tilting exhibit bounded relative error, we now compare their expected squared estimators as follows.

$$\begin{aligned}\frac{\mathscr{V}_\theta}{\mathscr{V}_0} &= \frac{(1/\beta)^{-\alpha}(\beta)^\alpha((\alpha+\theta)/z)^{-\alpha-\theta}\Gamma(\alpha+\theta)((2\beta z-\alpha-\theta)/z)^{-\alpha+\theta}\gamma(\alpha-\theta,2\beta z-\alpha-\theta)}{(1/\beta)^{-\alpha}(\beta)^\alpha((\alpha)/z)^{-\alpha}\Gamma(\alpha)((2\beta z-\alpha)/z)^{-\alpha}\gamma(\alpha,2\beta z-\alpha)}\\
&= \frac{(\alpha+\theta)^{-\alpha-\theta}\Gamma(\alpha+\theta)(2\beta z-\alpha-\theta)^{-\alpha+\theta}\gamma(\alpha-\theta,2\beta z-\alpha-\theta)}{(\alpha)^{-\alpha}\Gamma(\alpha)(2\beta z-\alpha)^{-\alpha}\gamma(\alpha,2\beta z-\alpha)}\\
&\sim \frac{(\alpha+\theta)^{-\alpha-\theta}\Gamma(\alpha+\theta)(2\beta z-\alpha-\theta)^{-\alpha+\theta}(2\beta z-\alpha-\theta)^{\alpha-\theta}/(\alpha-\theta)}{(\alpha)^{-\alpha}\Gamma(\alpha)(2\beta z-\alpha)^{-\alpha}(2\beta z-\alpha)^{\alpha}/\alpha}\\
&= \frac{(\alpha+\theta)^{-\alpha-\theta}\Gamma(\alpha+\theta)\alpha}{(\alpha)^{-\alpha}\Gamma(\alpha)(\alpha-\theta)} = \frac{(\alpha+\theta)^{-\alpha-\theta-1}\Gamma(\alpha+\theta)}{(\alpha)^{-\alpha-1}\Gamma(\alpha)}\\
&= \frac{(\alpha)^{\alpha+1}\Gamma(\alpha+\theta)}{(\alpha+\theta)^{\alpha+\theta+1}\Gamma(\alpha)}.\end{aligned}$$

Using Stirling's formula

$$\Gamma(z+1) \sim \sqrt{2\pi z}\left(\frac{z}{e}\right)^z,$$

we have

$$\begin{aligned}\frac{\mathscr{V}_\theta}{\mathscr{V}_0} &\sim \frac{(\alpha)^{\alpha+1}\sqrt{2\pi(\alpha+\theta-1)}\left(\frac{\alpha+\theta-1}{e}\right)^{\alpha+\theta-1}}{(\alpha+\theta)^{\alpha+\theta+1}\sqrt{2\pi(\alpha-1)}\left(\frac{\alpha-1}{e}\right)^{\alpha-1}}\\
&= \frac{(\alpha)^{\alpha+1}\sqrt{\alpha+\theta-1}(\alpha+\theta-1)^{\alpha+\theta-1}}{(\alpha+\theta)^{\alpha+\theta+1}\sqrt{\alpha-1}(\alpha-1)^{\alpha-1}e^\theta} < 1, \qquad \text{(B27)}\end{aligned}$$

for $-\alpha < \theta \uparrow \alpha$. From (B27), it thus can be said that there is an efficiency improvement in terms of relative bounded error for optimal sufficient exponential tilting over traditional optimal one-parameter tilting on η when simulating the event $\{X < 1/a\}$.

Appendix C Proof of Theorem 7

To prove Theorem 7, we must study the asymptotic behavior of the portfolio loss probability $P(L_n > nb)$ in the following proposition.

Proposition 8 *Under the assumptions of Theorem 7, consider $\tau = nb$ for some $b > 0$. Then*

$$\lim_{n\to\infty} u(n)^\alpha P(L_n > nb) = K, \qquad \text{(C28)}$$

for $0 < b < \bar{e}$, where K is a positive constant and $\bar{e}$ is the limiting average loss when all the obligors default.

Since the proof is similar in spirit to Theorem 1 of Bassamboo etal. (2008), it is omitted.

More specifically, as indicated in Bassamboo etal. (2008), rare event probability $P(L_n > nb)$ (or $p_{z,w,k}$ defined in (31)) occurs primarily when the shock variables $V_1, \ldots, V_d, V_{d+1}$ take large values, whereas Z_i, ρ_{ki} and ρ_k, for $i = 1, \ldots, d$ and $k = 1, \ldots, n$ exert little influence on the occurrence of the rare event. Put differently, only $V_1, \ldots, V_d, V_{d+1}$ are significantly affected by conditioning on the rare event, whereas all the other variables are not. Without loss of generality, we assume $W_1 = \cdots = W_d = 1$, $W_{d+1} = W = V$ with W_{d+1} following a gamma distribution in the proof.

Now we give a proof of Theorem 7 based on sufficient exponential tilting in "Appendix B i") and a Chernoff upper bound for $p_{z,w,k}$ defined in (31). Note that in the proof, we assume the tilting parameters are in the domain of the parameter spaces, like those in Theorems 5 and 6.

Proof of Theorem 7. We are given a constant $K_1 > 0$. To prove the theorem we re-express

$$E^* {L_*}^2 \mathbb{1}_{\{A_n\}} = E^*\left[{L_*}^2 \mathbb{1}_{\left\{A_n, W \le \frac{u(n)}{K_1}\right\}}\right] + E^*\left[{L_*}^2 \mathbb{1}_{\left\{A_n, W > \frac{u(n)}{K_1}\right\}}\right].$$

Recall from Theorem 5 that we can put $\theta_1^* = a(1-2\eta^*)$ and $\eta_1^* = 0$ for simplicity. This implies that $\mu^* = a$ and $\sigma^{*2} = 1$. By Theorem 6 i), we replace a by $u(n)$ to yield $\eta_2^* = \beta - (\alpha + \theta_2^*)/u(n)$ and $\theta_2^* \to \infty$ as $n \to \infty$. For notational simplicity, we replace (η_2^*, θ_2^*) by (η^*, θ^*).

The proof is divided into two steps.

Step 1. For a constant $K_1 > 0$, we establish that

$$\limsup_{n\to\infty} u(n)^{2\alpha} E^*\left[L_*^2 \mathbb{1}_{\left\{A_n, W > \frac{u(n)}{K_1}\right\}}\right] < \infty. \qquad \text{(C29)}$$

By (35) with suitable chosen θ^* such that $\theta^*/u(n) \to c < \beta$ as $n \to \infty$, and hence both $\theta^*, \eta^* > 0$, then we have on the set $\{A_n, W > \frac{u(n)}{K_1}\}$,

$$L_* = \left(\frac{f_z(Z)}{q_{z,\mu^*,\Sigma}(Z)}\right) \times e^{-\eta^* W - \theta^* \log W} \frac{(1/\beta)^{-\alpha}(\beta-\eta^*)^{-\alpha-\theta^*}\Gamma(\alpha+\theta^*)}{\Gamma(\alpha)} \leq \left(\frac{f_z(Z)}{q_{z,\mu^*,\Sigma}(Z)}\right) \times C(\eta^*, \theta^*) e^{-\eta^* \frac{u(n)}{K_1} - \theta^* \log \frac{u(n)}{K_1}}, \quad \text{(C30)}$$

where $C(\eta^*, \theta^*) = \frac{(1/\beta)^{-\alpha}(\beta-\eta^*)^{-\alpha-\theta^*}\Gamma(\alpha+\theta^*)}{\Gamma(\alpha)}$. Integrating L_*^2 over this set under P^*, we have

$$L_*^2 \leq K C^2(\eta^*, \theta^*) e^{-2\eta^* \frac{u(n)}{K_1} - 2\theta^* \log \frac{u(n)}{K_1}} = K C^2(\eta^*, \theta^*) e^{-2(\beta - \frac{1}{u(n)}(\alpha-\theta^*))\frac{u(n)}{K_1} - 2\theta^* \log \frac{u(n)}{K_1}}, \quad \text{(C31)}$$

for some $K > 0$. Since $\theta^* \uparrow \infty$ as $n \to \infty$, (C29) follows from (C31) by using the same procedure as that in the proof of Theorem 6 i).

Step 2. We show that for $K_1 > 0$ with K_1 sufficiently large,

$$\lim_{n\to\infty} u(n)^{2\alpha} E^* \left[L_*^2 \mathbb{1}_{\left\{A_n, W \leq \frac{u(n)}{K_1}\right\}} \right] < \infty. \quad \text{(C32)}$$

Recall from (31) with $\chi_k = a_k u(n)$, we have

$$p_{z,w,k} = P\left(\epsilon_k > \frac{a_k u(n)}{\rho_k W} - \sum_{i=1}^d \frac{\rho_{ki}}{\rho_k} Z_i \mid Z, W\right) \leq P\left(\epsilon_k > \frac{c_1 u(n)}{W} - \sum_{i=1}^d d_i Z_i \mid Z, W\right), \quad \text{(C33)}$$

where $c_1 = \min_k a_k/\rho_k$ and $d_i = \max_k \rho_{ki}/\rho_k$. As ϵ_k is light-tailed with an existing moment-generating function, there exist constants K_2 and c_2 such that

$$P\left(\epsilon_k > \frac{c_1 u(n)}{W} - \sum_{i=1}^d d_i Z_i \,\middle|\, Z, W\right) \leq K_2 \exp\left\{ - c_2 \left(\frac{c_1 u(n)}{W} - \sum_{i=1}^d d_i Z_i \right) \right\} \quad P\text{-a.s.} \quad \text{(C34)}$$

By (C34) and the existence of the moment-generating functions of $Z_i, i = 1, \ldots, d$, there exist constants K_3 and $c_3 > 0$ such that

$$L_* \mathbb{1}_{\{A_n\}} \leq \left(\frac{f_z(Z)}{q_{z,\mu^*,\Sigma}(Z)}\right) \times C(\eta^*, \theta^*) \exp\{-\eta^* W - \theta^* \log W\} K_3^n \times \exp\left\{ - n c_3 \left(\frac{c_1 u(n)}{W} \right) \right\} \mathbb{1}_{\{A_n\}} \quad P\text{-a.s.}$$

We restrict our discussion to the set $\left\{A_n, W \leq \frac{u(n)}{K_1}\right\}$. Then we have

$$\begin{aligned} L_* &\leq \left(\frac{f_z(Z)}{q_{z,\mu^*,\Sigma}(Z)}\right) C(\eta^*, \theta^*) \exp\{-\eta^* W - \theta^* \log W\} K_3^n \exp\left\{ - n c_3 \left(\frac{c_1 u(n)}{W} \right) \right\} \\ &\leq \left(\frac{f_z(Z)}{q_{z,\mu^*,\Sigma}(Z)}\right) C(\eta^*, \theta^*) \exp\left\{ - \left(\beta - \frac{(\alpha+\theta^*)}{u(n)} \right) \frac{u(n)}{K_1} - \theta^* \log \frac{u(n)}{K_1} \right\} \\ &\quad \times \exp\left\{ - \left(\beta - \frac{(\alpha+\theta^*)}{u(n)} \right) \left(W - \frac{u(n)}{K_1} \right) - \theta^* \left(\log W - \log \frac{u(n)}{K_1} \right) \right\} \quad P\text{-a.s.} \\ &\quad \times \exp\left\{ - n c_3 \left(c_1 u(n) (\frac{1}{W} - \frac{K_1}{u(n)}) \right) \right\} K_3^n \exp\left\{ - n c_3 \left(c_1 u(n) \frac{K_1}{u(n)} \right) \right\}. \end{aligned}$$

Note that the likelihood ratio $f_z(Z)/q_{z,\mu^*,\Sigma}(Z)$ is upper bounded by a constant for all Z. $\exp\{-(\beta - \frac{(\alpha+\theta^*)}{u(n)})\frac{u(n)}{K_1} - \theta^* \log \frac{u(n)}{K_1}\}$ has an upper bound for n large enough. Also note that for n large enough,

$$\begin{aligned} &(\beta - \frac{(\alpha+\theta^*)}{u(n)})(W - \frac{u(n)}{K_1}) + \theta^*(\log W - \log \frac{u(n)}{K_1}) \\ &\leq n c_3 c_1 u(n)(W - \frac{u(n)}{K_1}) \quad P\text{-a.s.} \end{aligned}$$

It follows that on the set $\{A_n, W \leq \frac{u(n)}{K_1}\}$, which is equivalent to $\{A_n, \frac{1}{W} \geq \frac{K_1}{u(n)}\}$, we have $\exp\left\{ - n c_3 \left(c_1 u(n)(\frac{1}{W} - \frac{K_1}{u(n)}) \right) \right\} \to 0$, as $n \to \infty$. Then for n large enough, there exists a constant $K_4 > 0$ such that $L_* \leq K_4 K_3^n \exp\{-K_1 c_3 c_1 n\}$. Now integrating L_*^2 over this set under P^*, we can select K_1 large enough so that (C32) holds. This completes the proof. ■

Appendix D Numerical results for modeling portfolio losses

We split the following discussion into three parts; the first two are presented in "Appendix D.1" and the last is presented in "Appendix D.2". First, we compare the performance of the proposed method with crude simulation, under three-factor normal mixture models, in which the t-distribution for X_k is considered. (Note that we compare the performance of our method only with crude Monte Carlo simulation

Table 8 Performance of proposed algorithm with two different losses resulting from default of obligors for a three-factor model with inverse FFT (t-distribution)

Two different losses ($c_i = (\lceil 2i \rceil/n)^2$)								
b	$P(L_n > \tau)$	V.R. factor	$\vec{\nu}$	$P(L_n > \tau)$	V.R. factor	n	$P(L_n > \tau)$	V.R. factor
0.7	4.79×10^{-3}	832	(4, 4, 4, 4)	4.78×10^{-3}	692	100	3.02×10^{-2}	325
1	2.91×10^{-4}	3078	(8, 6, 4, 4)	4.79×10^{-3}	832	250	4.79×10^{-3}	832
1.2	1.20×10^{-5}	14,471	(8, 8, 8, 8)	8.64×10^{-5}	7305	400	1.87×10^{-3}	739
ρ_{ki}	$P(L_n > \tau)$	V.R. factor	$\hat{\rho}$	$P(L_n > \tau)$	V.R. factor	$\sigma_1, \sigma_2, \sigma_3$	$P(L_n > \tau)$	V.R. factor
0.1	4.79×10^{-3}	832	−0.5	4.81×10^{-3}	1055	(0.6, 0.4, 0.1)	4.84×10^{-3}	700
0.3	2.96×10^{-3}	876	0	4.80×10^{-3}	745	(0.8, 0.6, 0.3)	4.79×10^{-3}	582
0.5	4.27×10^{-4}	739	0.5	4.79×10^{-3}	832	(1, 0.8, 0.5)	4.79×10^{-3}	832

Table 9 Performance of our algorithm with equal loss resulting from default of obligors for a three-factor model (t distribution)

	Crude		IS		
τ	$P(L_n > \tau)$	Variance	$P(L_n > \tau)$	Variance	V.R. factor
75	2.70×10^{-3}	2.69×10^{-3}	3.13×10^{-3}	8.6212×10^{-6}	313
100	2.60×10^{-4}	2.60×10^{-4}	2.44×10^{-4}	3.95×10^{-8}	6586

Table 10 Performance of our algorithm with five different losses resulting from default of obligors for three-factor model with inverse FFT (t-distribution)

Five different losses ($c_i = (\lceil 5i \rceil/n)^2$)								
b	$P(L_n > \tau)$	V.R. factor	$\vec{\nu}$	$P(L_n > \tau)$	V.R. factor	n	$P(L_n > \tau)$	V.R. factor
2	2.38×10^{-2}	141	(4, 4, 4, 4)	2.39×10^{-2}	139	100	1.09×10^{-1}	59
4	8.59×10^{-4}	3414	(8, 6, 4, 4)	2.38×10^{-2}	141	250	2.38×10^{-2}	141
6	4.15×10^{-7}	81,498	(8, 8, 8, 8)	1.84×10^{-3}	1131	400	9.77×10^{-3}	264
ρ_{ki}	$P(L_n > \tau)$	V.R. factor	$\hat{\rho}$	$P(L_n > \tau)$	V.R. factor	$\sigma_1, \sigma_2, \sigma_3$	$P(L_n > \tau)$	V.R. factor
0.1	2.38×10^{-2}	141	−0.5	2.39×10^{-2}	152	(0.6, 0.4, 0.1)	2.37×10^{-2}	149
0.3	1.51×10^{-2}	183	0	2.35×10^{-2}	125	(0.8, 0.6, 0.3)	2.40×10^{-2}	148
0.5	2.34×10^{-3}	143	0.5	2.38×10^{-2}	141	(1, 0.8, 0.5)	2.38×10^{-2}	141

henceforth, as most of the literature focuses on simulating one-dimensional cases.) Second, to evaluate the robustness of the proposed method, we also compare its performance with that of crude simulation, under three-factor normal mixture models, in which a generalized inverse Gaussian (GIG) distribution for X_k is considered. The cases with different losses resulting from default of the obligors are also investigated. Finally, in "Appendix D.2", we compare the computational time of crude Monte Carlo simulation with that of the proposed importance sampling under several scenarios, and provide insight into the trade-off between reduced variance and increased computational time.

Except for the experiments in Table 12, which compares the computational time under different numbers of samples,

Table 11 Performance of our algorithm with equal loss resulting from default of obligors for three-factor model (symmetric generalized hyperbolic distribution)

	Crude		IS (tilting η)			IS (tilting θ)		
b	$P(L_n > \tau)$	Variance	$P(L_n > \tau)$	Variance	V.R. factor	$P(L_n > \tau)$	Variance	V.R. factor
0.28	2.04×10^{-3}	2.04×10^{-3}	1.98×10^{-3}	6.99×10^{-6}	291	1.98×10^{-3}	2.79×10^{-6}	731
0.32	2.00×10^{-4}	2.00×10^{-4}	1.74×10^{-4}	8.09×10^{-8}	2473	1.75×10^{-4}	3.04×10^{-8}	6575
0.36	8.00×10^{-6}	8.00×10^{-6}	7.99×10^{-6}	3.77×10^{-10}	21,194	7.55×10^{-6}	1.11×10^{-10}	72,352

Table 12 Computational times (seconds)

	Crude			IS							
				Parameter search			Probability calculation ($\mathscr{B}_2 = 1000$)				
b	$\mathscr{B}_2$	Time (T_C)	$P(L_n > \tau)$	$\mathscr{B}_1$	Time (T_I^1)	#iter	Time (T_I^2)	$P(L_n > \tau)$	Variance	V.R. factor	$\mathrm{T}_C^*/(\mathrm{T}_I^1 + \mathrm{T}_I^2)$
0.3	100	2	–	100	191	9	200	3.20×10^{-3}	1.13×10^{-3}	3	0.46
	1000	19	1.00×10^{-3}	500	647	8	191	3.38×10^{-3}	1.42×10^{-4}	22	0.22
	10,000	181*	3.80×10^{-3}	1000	2049	8	201	3.01×10^{-3}	5.12×10^{-6}	602	0.08
0.4	1000	17	–	500	581	8	183	2.43×10^{-4}	1.37×10^{-6}	175	2.71
	10,000	184	1.00×10^{-4}	1000	1618	9	220	2.32×10^{-4}	8.08×10^{-7}	296	1.13
	100,000	2070*	2.00×10^{-4}	2000	2687	7	276	2.29×10^{-4}	2.83×10^{-7}	844	0.7
0.5	10,000	204	–	1000	1269	7	210	1.70×10^{-6}	2.88×10^{-10}	7411	12.77
	100,000	1812	–	2000	2442	8	209	1.98×10^{-6}	2.59×10^{-10}	8244	7.13
	1,000,000	18,896*	2.00×10^{-6}	5000	6001	8	228	1.90×10^{-6}	1.07×10^{-10}	19,845	3.03

The asterisk "*" in the column "Time (T_C)" denotes the corresponding value T_C^* used in the calculation of the last column

we generate $\mathscr{B}_1 = 5,000$ samples to locate the optimal tilting parameters and $\mathscr{B}_2 = 10,000$ samples to calculate the probability of losses in all of the remaining experiments. Following the settings in Bassamboo etal. (2008), variances under crude Monte Carlo simulation are estimated indirectly by exploiting the observation that for a Bernoulli random variable with success probability p, the variance equals $p(1-p)$. Note that as mentioned in Sect. 3.1 and stated in Theorem 7, for normal mixture random variables, tilting the variance of the normal random variable Z is relatively insignificant in comparison to tilting the parameters of W; therefore, in the experiments, we only conduct mean tilting for Z and θ- or η-tilting for W for multi-factor normal mixture models. Note that even in this case, our sufficient exponential importance sampling method differs from previous studies, as only θ in the normal distribution and η in the gamma distribution are considered as the tilting parameter in their methods.

Remark 7 When considering $V_j = \sqrt{\nu_j/W_j}$ for $j = 1, \ldots, d+1$ with $W_j \sim \mathtt{Gamma}(\nu_j/2, 1/2)$, we mainly deal with the event $\{W_{d+1} < 1/a\}$, whereas when considering $V_j = \sqrt{W_j}$, for $j = 1, \ldots, d+1$ with $W_j \sim \mathtt{Gamma}(\alpha_j, \beta_j)$, we turn to address $\{W_{d+1} > a\}$. As shown in both theoretical results in "Appendices A and B" and the simulation results in Table 3 in Sect. 2.1, when considering $\{W_{d+1} < 1/a\}$, tilting η only is fine because $\theta^* \to \alpha$ as $a \to \infty$. However, it is fine to tilt only θ when considering $\{W_{d+1} > a\}$, because as $a \to \infty$, $\eta^* \to \beta - c$ for some $c > 0$.

Computation and numerical experiments on different model settings

Second, Tables 9, 8, and 10 compare the performance of the proposed importance sampling method with that of crude simulation for a three-factor t-copula model, a special case of normal mixture copula models, where the latent variables X_k follow a multivariate t-distribution, i.e., $V_j = \sqrt{\nu_j/W_j}$ and $W_j \sim \mathtt{Gamma}(\nu_j/2, 1/2)$ for $j = 1, \ldots, 4$. The three tables list the results with equal losses (Table 9) and different losses (Table 8 and 10) resulting from the default of obligors with various parameter settings. Following the settings in Bassamboo etal. (2008) and Chan and Kroese (2010), the threshold for the i-th obligor χ_i is set to $0.5 \times \sqrt{n}$, the idiosyncratic risk ϵ_k is set to $N(0, 9)$, and the total loss τ is set to $n \times b$.

As shown in Tables 9, 8, and 10, our IS approach significantly outperforms crude simulation, especially when the loss threshold τ increases and the probability decreases. Moreover, in contrast to the results listed in Table 2 of Chan and Kroese (2010), which show that the performance of CondMC deteriorates when ρ_{ki} increases, the performance of our method is demonstrated to be stable. This is due to the fact that as ρ_{ki} increases, the factor Z gains importance in determining the occurrence of the rare event; CondMC simply ignores the contribution of Z, whereas the proposed method twists the distributions of both Z and W.

Third, in addition to the t-copula model, Table 11 shows the results for another type of three-factor normal mixture copula model, where ${V_j}^2$ follows a special case of the generalized inverse Gaussian (GIG) distribution, i.e., $V_j = \sqrt{W_j}$ with $W_j \sim \texttt{Gamma}(\nu_j/2, 1/2)$ for $j = 1, \ldots, 4$ and $\vec{\nu} = (8\ 6\ 4\ 4)$; except for b set to 0.28, 0.32, and 0.36, the other model parameters are the same as those for the base case in Table 9. From Table 11 we observe that our approach outperforms crude simulation, which attests the capability of the proposed algorithm for normal mixture copula models. Note also that in this setting, tilting the other parameter $\nu_j/2$ of the gamma distribution (i.e., $\nu_j/2 \to \nu_j/2 - \theta_j^*$) yields 2 to 4 times better performance than traditional one-parameter exponential tilting (tilting η) in terms of the variance reduction factors, which is consistent with theoretical results in "Appendices A and B" and the simulation demonstrated in Example 2 in Sect. 2.1 and Sect. 3.1 (see also Remark 7).

Computational time of proposed importance sampling algorithm

We now proceed to compare the computational time of crude Monte Carlo simulation with that of our importance sampling under several scenarios and provide insight into the trade-off between reduced variance and increased computational time.[14] Table 12 lists the computational time of crude simulation and our method for the three cases listed in the top-left corner of Table 9. We observe that using more samples ($\mathscr{B}_1$) to determine optimal tilting parameters greatly improves the variance reduction performance, which however linearly increases the computational time to determine the parameters[15]; note that the variance reduction factor grows nonlinearly with the number of samples $\mathscr{B}_1$, and that our search algorithm generally requires only 7 to 9 iterations to achieve convergence.[16] Despite the need for additional computational time to find suitable tilting parameters, we can use a mere $\mathscr{B}_2 = 1000$ samples to obtain rather good estimates with greatly reduced variances, especially when the tail probability is small. In the last column of Table 12, we also report the ratio of the computational time consumed by the crude simulation generating a fair estimate (T_C^*, the value with a star symbol on the left-hand side) to that consumed by our importance sampling algorithm, including the time for the parameter search (T_I^1) and probability calculation (T_I^2). From the table, observe that for the case $b = 0.5$, crude simulation with 10, 000 and even 100, 000 fails to generate the estimate, whereas our method yields a good estimate with a 7,411 variance reduction ratio, but the crude simulation requires 12.77 times more computational time than ours. The relation between the variance reduction factor and the time consumption ratio listed in the last two columns of Table 12 suggests that the proposed algorithm achieves good performance and thus makes a practical contribution to portfolio credit risk measurement in normal mixture copula models.

Remark 8 To find the optimal tilting parameters by solving the system of (6)–(7), we must evaluate the RHSs of (6)–(7). Table 12 shows that empirically, the proposed recursive algorithm is efficient in locating optimal tilting parameters, as it always converges within 10 iterations with a rather small number of samples, e.g., $\mathscr{B}_1 = 1000$ (see Remark 6 for the technique used for the searching stage).

[14] For our experiments, the software utilized was Mathematica 13.1; the hardware platform was a MacBook Pro (16-inch, 2021) (M1 Pro Chip with a 10-core CPU).

[15] Moreover, increasing the number of tilting parameters only affects the search time for the optimal tilting parameters, the time complexity of which grows linearly with the number of considered parameters.

[16] In all of the experiments reported here, the predetermined precision level ϵ was set to 10^{-4}.

References

Ahamed, T.I., Borkar, V.S., Juneja, S.: Adaptive importance sampling technique for Markov chains using stochastic approximation. Oper. Res. **54**(3), 489–504 (2006)

Amar, E.B., Rached, N.B., Haji-Ali, A.L., et al.: State-dependent importance sampling for estimating expectations of functionals of sums of independent random variables. Stat. Comput. **33**, Article No. 40 (2023)

Asmussen, S., Binswanger, K.: Simulation of ruin probabilities for subexponential claims. ASTIN Bull. J. IAA **27**(2), 297–318 (1997)

Asmussen, S., Glynn, P.W.: Stochastic Simulation: Algorithms and Analysis, vol. 57. Springer, Berlin (2007)

Asmussen, S., Kroese, D.P.: Improved algorithms for rare event simulation with heavy tails. Adv. Appl. Probab. **38**(2), 545–558 (2006)

Asmussen, S., Jensen, J., Rojas-Nandayapa, L.: Exponential family techniques for the lognormal left tail. Scand. J. Stat. **43**(3), 774–787 (2016)

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

Ben Issaid, C., Ben Rached, N., Kammoun, A., et al.: On the efficient simulation of the distribution of the sum of gamma-gamma variates with application to the outage probability evaluation over fading channels. IEEE Trans. Commun. **65**(4), 1839–1848 (2017)

Ben Rached, N., Kammoun, A., Alouini, M.S., et al.: Unified importance sampling schemes for efficient simulation of outage capacity over generalized fading channels. IEEE J. Sel. Top. Signal Process. **10**(2), 376–388 (2015)

Ben Rached, N., Benkhelifa, F., Kammoun, A.A., et al.: On the generalization of the hazard rate twisting-based simulation approach. Stat. Comput. **28**, 61–75 (2018)

Ben Rached, N., Kammoun, A., Alouini, M.S., et al.: An accurate sample rejection estimator of the outage probability with equal gain combining. IEEE Open J. Commun. Soc. **1**, 1022–1034 (2020a)

Ben Rached, N., MacKinlay, D., Botev, Z., et al.: A universal splitting estimator for the performance evaluation of wireless communications systems. IEEE Trans. Wirel. Commun. **19**(7), 4353–4362 (2020b)
Ben Rached, N., Haji-Ali, A.L., Rubino, G., et al.: Efficient importance sampling for large sums of independent and identically distributed random variables. Stat. Comput. **31**(6), 4353–4362 (2021)
Blanchet, J., Lam, H.: State-dependent importance sampling for rare-event simulation: an overview and recent advances. Surv. Oper. Res. Manag. Sci. **17**(1), 38–59 (2012)
Blanchet, J., Li, C.: Efficient rare event simulation for heavy-tailed compound sums. ACM Trans. Model. Comput. Simul. **21**(2), 1–23 (2011)
Blanchet, J., Liu, J.: State-dependent importance sampling for regularly varying random walks. Adv. Appl. Probab. **40**(4), 1104–1128 (2008)
Chan, J.C., Kroese, D.P.: Efficient estimation of large portfolio loss probabilities in t-copula models. Eur. J. Oper. Res. **205**(2), 361–367 (2010)
Chen, R.R., Sopranzetti, B.J.: The valuation of default-triggered credit derivatives. J. Financ. Quant. Anal. **38**(2), 359–382 (2003)
Csiba, D., Richtárik, P.: Importance sampling for minibatches. J. Mach. Learn. Res. **19**(1), 962–982 (2018)
Do, K.A., Hall, P.: On importance resampling for the bootstrap. Biometrika **78**(1), 161–167 (1991)
Duffie, D., Singleton, J.: Credit Risk. Princeton University Press, Princeton (2003)
Dupuis, P., Wang, H.: Importance sampling, large deviations, and differential games. Stochast. Int. J. Probab. Stochast. Process. **76**(6), 481–508 (2004)
Dupuis, P., Sezer, A.D., Wang, H.: Dynamic importance sampling for queueing networks. Ann. Appl. Probab. **17**(4), 1306–1346 (2007)
Eberlein, E., Keller, U.: Hyperbolic distributions in finance. Bernoulli **1**(3), 281–299 (1995)
Eberlein, E., Keller, U., Prause, K.: New insights into smile, mispricing, and value at risk: the hyperbolic model. J. Bus. **71**(3), 371–405 (1998)
Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. Chapman & Hall, Cambridge (1994)
Egloff, D., Leippold, M., Jöhri, S., et al.: Optimal importance sampling for credit portfolios with stochastic approximation. Available at SSRN 693441 (2005)
Ellis, R.S.: Entropy, Large Deviations, and Statistical Mechanics. Springer, New York (1985)
Ericsson, J., Jacobs, K., Oviedo, R.: The determinants of credit default swap premia. J. Financ. Quant. Anal. **44**(1), 109–132 (2009)
Fu, M., Su, Y.: Optimal importance sampling in securities pricing. J. Comput. Finance **5**, 27–50 (2002)
Fuh, C.D., Hu, I.: Efficient importance sampling for events of moderate deviations with applications. Biometrika **91**(2), 471–490 (2004)
Fuh, C.D., Hu, I., Hsu, Y.H., et al.: Efficient simulation of value at risk with heavy-tailed risk factors. Oper. Res. **59**(6), 1395–1406 (2011)
Fuh, C.D., Teng, H.W., Wang, R.H.: Efficient simulation of Value-at-Risk under a jump diffusion model: a new method for moderate deviation events. Comput. Econ. **51**, 973–990 (2018)
Fuh, C.D., Jia, Y., Kou, S.: A general framework for importance sampling with latent Markov processes (2023a) arXiv preprint arXiv:2311.12330
Fuh, C.D., Wang, C.J., Pai, C.H.: Markov chain importance sampling for minibatches. Mach. Learn. (2023b) (to appear)
Glasserman, P.: Monte Carlo Methods in Financial Engineering. Springer, New York (2004)
Glasserman, P., Li, J.: Importance sampling for portfolio credit risk. Manag. Sci. **51**(11), 1643–1656 (2005)
Glasserman, P., Heidelberger, P., Shahabuddin, P.: Variance reduction techniques for estimating value-at-risk. Manag. Sci. **46**(10), 1349–1364 (2000)
Glasserman, P., Heidelberger, P., Shahabuddin, P.: Portfolio value-at-risk with heavy-tailed risk factors. Math. Financ. **12**(3), 239–269 (2002)
Glasserman, P., Kang, W., Shahabuddin, P.: Large deviations in multifactor portfolio credit risk. Math. Financ. **17**(3), 345–379 (2007)
Glasserman, P., Kang, W., Shahabuddin, P.: Fast simulation of multifactor portfolio credit risk. Oper. Res. **56**(5), 1200–1217 (2008)
Glynn, P.W., Iglehart, D.L.: Importance sampling for stochastic simulations. Manag. Sci. **35**(11), 1367–1392 (1989)
Hirtle, B.: Credit derivatives and bank credit supply. J. Financ. Intermed. **18**(2), 125–150 (2009)
Johns, M.V.: Importance sampling for bootstrap conference intervals. J. Am. Stat. Assoc. **83**, 709–714 (1988)
Johnson, T.B., Guestrin, C.: Training deep models faster with robust, approximate importance sampling. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'18, pp. 7265–7275 (2018)
Kahn, H., Harris, T.E.: Estimation of particle transmission by random sampling. Natl. Bureau Stand. Appl. Math. Ser. **12**, 27–30 (1951)
Katharopoulos, A., Fleuret, F.: Not all samples are created equal: deep learning with importance sampling. In: Proceedings of the 35th International Conference on Machine Learning, ICML'18, pp. 2525–2534 (2018)
Kawai, R.: Optimal importance sampling parameter search for Lévy processes via stochastic approximation. SIAM J. Numer. Anal. **47**(1), 293–307 (2009)
Kawai, R.: Optimizing adaptive importance sampling by stochastic approximation. SIAM J. Sci. Comput. **40**(4), A2774–A2800 (2018)
Li, D.X.: On default correlation: a copula function approach. J. Fixed Income **9**(4), 43–54 (2000)
McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press, Princeton (2015)
Metelli, A.M., Papini, M., Faccio, F., et al.: Policy optimization via importance sampling. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'18, pp. 5447–5459 (2018)
Metelli, A.M., Papini, M., Montali, N., et al.: Importance sampling techniques for policy optimization. J. Mach. Learn. Res. **21**, 1–75 (2020)
Ney, P.: Dominating points and the asymptotics of large deviations for random walk on RD. Ann. Probab. **11**, 158–167 (1983)
Oberhettinger, F.: Fourier Transforms of Distributions and Their Inverses: A Collection of Tables. Academic Press, Inc, New York (2014)
Owen, A.B.: Monte Carlo Theory, Methods and Examples. https://artowen.su.domains/mc/. Online book, Accessed 2 Dec 2023 (2013)
Owen, A.B., Maximov, Y., Chertkov, M.: Moderate deviation principles for importance sampling estimators of risk measures. J. Appl. Probab. **54**(2), 490–506 (2017)
Richtárik, P., Takáč, M.: On optimal probabilities in stochastic coordinate descent methods. Optim. Lett. **10**(6), 1233–1243 (2016)
Sadowsky, J.S., Bucklew, J.A.: On large deviations theory and asymptotically efficient Monte Carlo estimation. IEEE Trans. Inf. Theory **36**(3), 579–588 (1990)
Scott, A., Metzler, A.: A general importance sampling algorithm for estimating portfolio loss probabilities in linear factor models. Insur. Math. Econ. **64**, 279–293 (2015)
Siegmund, D.: Importance sampling in the Monte Carlo study of sequential tests. Ann. Stat. **4**(4), 673–684 (1976)

Soriano, J.: Extremum points of a convex function. Appl. Math. Comput. **66**(2–3), 261–266 (1994)
Su, Y., Fu, M.C.: Importance sampling in derivative securities pricing. In: Proceedings of the 2000 Winter Simulation Conference, pp. 587–596 (2000)
Teng, H.W., Fuh, C.D., Chen, C.C.: On an automatic and optimal importance sampling approach with applications in finance. Quant. Finance **16**(8), 1259–1271 (2016)
Ulam, S., Richtmyer, R.D., von Neumann, J.: Statistical methods in neutron diffusion. Report LAMS-551, Los Alamos National Laboratory, pp. 1–22 (1947)
Zhao, P., Zhang, T.: Stochastic optimization with importance sampling for regularized loss minimization. In: Proceedings of the 32nd International Conference on Machine Learning, ICML'15, pp. 1–9 (2015)