Comput Econ (2018) 51:973–990
https://doi.org/10.1007/s10614-017-9654-z

Efficient Simulation of Value-at-Risk Under a Jump Diffusion Model: A New Method for Moderate Deviation Events

Cheng-Der Fuh[1] · **Huei-Wen Teng**[2] · **Ren-Her Wang**[3]

Accepted: 19 January 2017 / Published online: 17 February 2017

Abstract Importance sampling is a powerful variance reduction technique for rare event simulation, and can be applied to evaluate a portfolio's Value-at-Risk (VaR). By adding a jump term in the geometric Brownian motion, the jump diffusion model can be used to describe abnormal changes in asset prices when there is a serious event in the market. In this paper, we propose an importance sampling algorithm to compute the portfolio's VaR under a multi-variate jump diffusion model. To be more precise, an efficient computational procedure is developed for estimating the portfolio loss probability for those assets with jump risks. And the tilting measure can be separated for the diffusion and the jump part under the assumption of independence. The simulation results show that the efficiency of importance sampling improves over the naive Monte Carlo simulation from 9 to 277 times under various situations.

Keywords Importance sampling · Exponential tilting · Moderate deviation · Jump diffusion · VaR

Mathematics Subject Classification 90-08 · 65K10

✉ Ren-Her Wang
138230@mail.tku.edu.tw

Cheng-Der Fuh
cdfuh@cc.ncu.edu.tw

Huei-Wen Teng
venteng@gmail.com

1 Graduate Institute of Statistics, National Central University, Taoyuan, Taiwan, ROC

2 Department of Information Management and Finance, National Chiao Tung University, Hsinchu, Taiwan, ROC

3 Department of Banking and Finance, Tamkang University, No. 151, Yingzhuan Rd., Tamsui Dist., New Taipei City 25137, Taiwan, ROC

1 Introduction

We are interested in efficient calculation on probabilities of tail events where large jumps occurs in the dynamic of the portfolio, which relate to the Value-at-Risk (VaR) calculation in financial risk management. VaR is a measure of the potential losses due to movements in the underlying market.

To capture the empirical phenomenon on the dynamic of the portfolio, Merton (1976) proposed a jump diffusion model to describe discontinuous change of the asset price when abnormal information arrives in the market. Events and their corresponding jumps can occur at random or scheduled times. Therefore, the amplitude of the response in either case can be unpredictable or random. While the volatility of portfolios is often modeled by continuous Brownian motion processes, discontinuous jump processes are more appropriate for modeling the response to important external events that significantly affect the prices of financial assets. Discrete jump processes are modeled by compound Poisson processes for random events or scheduled events.

The jump diffusion model can reveal the empirical phenomena of both the leptokurtic feature and volatility smile, and therefore is suitable for VaR calculating in financial risk management. There are many literature in the setting of jump diffusion models, to which explicit option price formulas for stock market have been established by Merton (1976), Naik and Lee (1990) and Kou (2002).

In this paper, we propose a new method to compute a portfolio's VaR under a multi-variate jump diffusion model. To provide an efficient simulation scheme for VaR calculation under a jump diffusion model, this paper consider the problem of estimating small probabilities by Monte Carlo simulations. Specifically, tilting measures of importance sampling will be obtained for the continuous Brownian motion process and the compound Poisson processes, respectively.

That is, we estimate $z = P(A)$ when z is small, say of the order 10^{-2} or 10^{-3} or so; i.e., A is a moderate deviation rare event. Such problems appear in the construction of confidence region for asymptotically normal statistics; cf. (Beran 1987; Beran and Millar 1986; Fuh and Hu 2004; Hall 1987, 1992), and in the computation of VaR in risk management; cf. (Duffie and Singleton 2003; Fuh et al. 2011; Glasserman et al. 2000, 2002; Jorion 2001). For the calculation of VaR in financial risk management, the reduction in terms of the computational time would be essential, cf. Chapter 9 of Glasserman (2004). It is well known that importance sampling, where one uses samples from an alternative distribution Q other than the original distribution P, is a powerful tool in efficient simulation of events with small probabilities. Some general references are in Asmussen and Glynn (2007), Heidelberger (1995) and Liu (2001).

A useful tool in importance sampling for rare event simulation is exponential tilting, cf. (Bucklew 2004; Siegmund 1976) and references therein. The above mentioned algorithm is more efficient for a large deviation rare event (z is of the order 10^{-5} or less) than an arbitrary event (say, z is not rare). Examples of such events occur in telecommunications ($z =$ bit-loss rate, probability of buffer overflow) and reliability ($z =$ the probability of failure before time t). To be more precise, when a sequence of random vectors $\{X_n\}$ converge to a constant vector μ, for any event A not containing μ, the probability $P\{X_n \in A\}$ usually decays exponentially fast as $n \to \infty$. Efficient Monte Carlo simulation of such events has been obtained by Sadowsky and Bucklew

(1990) based on the large deviations theory given by Ney (1983). The dominating point of large deviations theory by Ney (1983), which is located at the boundary of the event, and it is differs from the mean of the optimal alternative distribution is usually in the interior of the event.

To provide a more accurate simulation algorithm for a moderately rare event ($p \sim 0.01$ or 0.001), efficient importance sampling has been studied by Davison (1988), Do and Hall (1991), Fuh and Hu (2004, 2007) and Johns (1988). However, those papers concern univariate and/or multivariate-normal distributions. Extension to heavy-tailed settings such as multivariate t distribution can be found in Fuh et al. (2011), 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.

It is worth mentioning that for events of large deviations $P\{X \in A = (a, \infty)\}$ for some $a > 0$, Sadowsky and Bucklew (1990) showed 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 Q-measure equals the dominating point, cf. page 83, Bucklew (2004), which is located at the boundary of A. However, for a moderate deviation rare event, we show that typically the tilting point of the optimal alternative distribution is different from that given by large deviations theory. Furthermore, by using the idea of conjugate measure of Q, $\bar{Q}(dx) = C\exp(-\theta x)P(dx)$, the general account of our approach characterizes the optimal tilting θ, by solving the equation that the expectation of X under Q-measure equals the conditional expectation of X under $\bar{Q}$-measure given the rare event.

There are three aspects in this study. To begin with, we provide a theoretical benchmark based on variance minimization criterion, and obtain a novel and explicit expression for the optimal alternative distribution under exponential embedding family, and propose numerical analysis to approximate the optimal tilting point. Second, the proposed algorithm is quite general to cover interesting examples, such as normal distribution and compound Poisson processes. Third, the derived tilting formula can be used to calculate portfolio VaR under jump diffusion models. It is shown that large deviation tilting importance sampling provides reasonable well estimation for light tail distribution. However, in the case of a moderate deviation event for VaR calculation under jump diffusion models, the improvement of the proposed method, compared with the importance sampling based on large deviations theory, seems to be substantial.

The rest of this paper is organized as follows. In Sect. 2, we present a general account of importance sampling that minimizes the variance of the importance sampling estimator within a parametric family, provide a recursive algorithm for finding the optimal alternative distribution, and approximate optimal tilting probability measure for moderate deviation events. Section 3 presents the tilting measure can be separated for the diffusion and the jump part and several interesting examples to which we can characterize the optimal tilting probability measures. Section 4 demonstrates the performance of the tilting formula by calculating portfolio VaR in financial risk management. Concluding remarks are given in Sect. 5. The proofs are deferred to the "Appendix 1".

2 Importance Sampling

2.1 A General Account in Importance Sampling

Let $(\Omega, \mathcal{F}, P)$ be a given probability space, X be a random variable on Ω and A be a measurable set in $(\mathbf{R}, \mathcal{B}(\mathcal{R}))$, where $\mathcal{B}(\mathcal{R})$ is the Borel σ-algebra in $\mathbf{R}$. To estimate the probability of an event $\{X \in A\}$, we shall employ the importance sampling method. That is, instead of sampling from the target distribution P of X directly, we sample from an alternative distribution $Q := Q_\theta$. Suppose X has moment generating function $\Psi(\theta) = E[e^{\theta X}]$ under P for $\theta \in \mathbf{R}$. Then we consider the exponential tilting measure Q of P, which has the form

$$\frac{dQ}{dP} = \frac{\mathrm{e}^{\theta X}}{E[\mathrm{e}^{\theta X}]} = \mathrm{e}^{\theta X - \psi(\theta)},$$

where $\psi(\theta)$ is $\log \Psi(\theta)$, the cumulant generating function of X.

The question now focuses on how to choose an alternative distribution Q so that the importance sampling estimator has the minimum variance. Indeed, the criterion of minimizing the variance of the importance sampling estimator can be traced back to Siegmund (1976), in which the exponential tilting probability for a stopping time event had been introduced. The importance sampling estimator for $p = P\{X \in A\}$ based on a sample of size n is

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \in A\}} \frac{dP}{dQ}, \tag{1}$$

where $\mathbf{1}_B$ is the indicator function of an event B, X_i, $i = 1, \ldots, n$, are independent observations from Q, and dP/dQ is the Radon-Nikodym derivative assuming P is absolutely continuous with respect to Q. Assume P and Q are absolutely continuous with respect to the Lebesgue measure $\mathcal{L}$, and denote their densities as $dP/d\mathcal{L}$ and $dQ/d\mathcal{L}$, respectively. Observing that, since the estimator $\hat{p}_n$ is unbiased, the variance of the importance sampling estimator is

$$\mathrm{var}_Q(\hat{p}_n) = n^{-1}(G(\theta) - p^2).$$

Define the second moment of importance sampling estimator to be

$$G(\theta) \equiv E_Q\left[\mathbf{1}_{\{X \in A\}} \frac{dP}{dQ}\right]^2 = E\left[\mathbf{1}_{\{X \in A\}} \frac{dP}{dQ}\right] = E[\mathbf{1}_{\{X \in A\}} \mathrm{e}^{-\theta X + \psi(\theta)}],$$

where the expectation without any qualification is under the target probability measure P unless otherwise stated. Therefore, the criterion of minimizing the variance of the importance sampling estimator becomes to minimizing the second moment, $G(\theta)$.

We remark that when $A = \{\omega : X(\omega) > a\}$ for some constant $a > 0$, large deviation theory considers the following inequality

$$G(\theta) \leq e^{-\theta a + \psi(\theta)},$$

and minimizes the above upper bound, cf. (Glasserman et al. 2000). The first-order condition (after taking logarithm) gives

$$\psi'(\theta) = a. \tag{2}$$

Here, we denote θ^+ as the optimal θ by the large deviation theory that satisfies (2). Note that the approximation (2) is more accurate when $P\{X > a\}$ is small, i.e., a is sufficient large. In contrast to the large deviation theory, our goal is to solve the optimization problem for θ

$$\theta^* = \operatorname*{argmin}_{\theta} G(\theta). \tag{3}$$

θ^* is hence the desired quantity for selecting the tilting measure that minimizes the variance of the importance sampling estimator within a suitable parametric family.

To minimize $G(\theta)$, the first-order condition gives

$$\frac{dG(\theta)}{d\theta} = E[\mathbf{1}_{\{X \in A\}} e^{-\theta X + \psi(\theta)}(-X + \psi'(\theta))] = 0.$$

Dividing $\mathrm{e}^{\psi(\theta)}$ for both sides of the above equation, we have an equivalent condition,

$$E[\mathbf{1}_{\{X \in A\}} e^{-\theta X}(-X + \psi'(\theta))] = 0, \tag{4}$$

and therefore θ^* is the solution of

$$\psi'(\theta) = \frac{E[\mathbf{1}_{\{X \in A\}} X \mathrm{e}^{-\theta X}]}{E[\mathbf{1}_{\{X \in A\}} \mathrm{e}^{-\theta X}]}. \tag{5}$$

Standard numerical procedures (or a recursive algorithm presented in Sect. 2.2 can be applied to find θ^* satisfying (5). To prove that the RHS of (5) can be represented in a simple formula, we consider the conjugate measure $\bar{Q} := \bar{Q}_\theta$ of the measure Q, which is defined as

$$\frac{d\bar{Q}}{dP} = \frac{\mathrm{e}^{-\theta X}}{E[\mathrm{e}^{-\theta X}]} = \mathrm{e}^{-\theta X - \bar{\psi}(\theta)}, \tag{6}$$

where $\bar{\psi}(\theta)$ is $\log \bar{\Psi}(\theta)$ with $\bar{\Psi}(\theta) = E[\mathrm{e}^{-\theta X}]$.

To present a connection between $\bar{Q}$ and Q, we consider their probability densities with respect to Lebesgue measure $\mathcal{L}$. It is straightforward to see that $\bar{\psi}(\theta) = \psi(-\theta)$, which implies

$$\frac{d\bar{Q}_\theta}{d\mathcal{L}} = \mathrm{e}^{-\theta x - \bar{\psi}(\theta)} \frac{dP}{d\mathcal{L}} = \mathrm{e}^{-\theta x - \psi(-\theta)} \frac{dP}{d\mathcal{L}} = \mathrm{e}^{(-\theta)x - \psi(-\theta)} \frac{dP}{d\mathcal{L}} = \frac{dQ_{-\theta}}{d\mathcal{L}}. \tag{7}$$

In plain language, the conjugate measure $\bar{Q}_\theta$ equals to the alternative measure Q_θ by replacing θ with $-\theta$, and as a result, additional mathematical derivation for $\bar{Q}_\theta$ is not required.

The following theorem states the existence and uniqueness for the optimization procedure (3), and provides a simplification on the RHS of (5) using $\bar{Q}$.

Before that, we need a condition for $\Psi(\theta)$ being steep to ensure the finiteness of the moment generating function $\Psi(\theta)$. To define steepness, let $\theta_{\max} := \sup\{\theta : \Psi(\theta) < \infty\}$ (for light-tailed distributions, we have $0 < \theta_{\max} \leq \infty$). Then steepness means $\Psi(\theta) \to \infty$ as $\theta \to \theta_{\max}$.

Proposition 1 *Suppose the moment generating function $\Psi(\theta)$ of X exists and is second order continuously differentiable for $\theta \in \mathbf{I} \subset \mathbf{R}$. Furthermore, assume that $\Psi(\theta)$ is steep. Then, $\psi'(\theta)$ is strictly increasing, and $E_{\bar{Q}}[X|X \in A]$ is strictly decreasing.*

The proof of Proposition 1 will be given in the "Appendix 1".

Theorem 1 *Suppose the moment generating function $\Psi(\theta)$ of X exists and is second order continuously differentiable for $\theta \in \mathbf{I} \subset \mathbf{R}$. Furthermore, assume that $\Psi(\theta)$ is steep and $E[X|X \in A] > E(X) := \mu$. Then there exists a unique solution for the optimization problem (3), which satisfies*

$$\psi'(\theta) = E_{\bar{Q}_\theta}[X|X \in A]. \tag{8}$$

The proof of Theorem 1 will be given in the "Appendix 1".

We introduce the idea of conjugate measure for a few reasons. Needless to say, it helps to rewrite (5) by (8) for a characterization of θ^*, and can be used to state and prove Theorem 1. Furthermore, it provides an insightful interpretation of θ^*, which can be used to compare with the large deviation tilting.

In summary, the three-step procedures to find θ^* are

1. Calculate the cumulant generating function $\psi(\theta)$ of X and its derivative $\psi'(\theta)$.
2. Find the exponential tilting measure Q.
3. Find θ^* as the solution of $\psi'(\theta) = E_{\bar{Q}_\theta}[X|X \in A]$.

2.2 Calculating the Optimal Alternative Distribution

Before employing the importance sampling method, it is first necessary to identify the optimal tilting parameter. Since the optimal θ satisfying (8) cannot be computed directly, to find θ in the third step, we consider a recursive procedure for a general equation,

$$g(\theta) = h(\theta), \tag{9}$$

for some functions $g(\theta)$ and $h(\theta)$. In our setting, $g(\theta) = \psi'(\theta)$ and $h(\theta) = E_{\bar{Q}}[X|X \in A]$.

A simple recursive procedure is implemented as follows.

1. Start with a suitable $\theta^{(0)}$. Set $i = 1$.
2. Find $\theta^{(i)}$ as the solution of $g(\theta) = h(\theta^{(i-1)})$.

3. Set $i = i + 1$. Return to 2 until $|\theta^{(i+1)} - \theta^{(i)}|/\theta^{(i)} < \varepsilon$, where ε is the desired tolerance.

There are two reasons that Step 2 finds $\theta^{(i)}$ satisfying $g(\theta) = h(\theta^{(i-1)})$ other than satisfying $h(\theta) = g(\theta^{(i-1)})$. First, the inverse function of $g(\theta)$ is easier to derive. Second, when $A = \{\omega : X(w) > a\}$ for some constant $a > 0$, $h(\theta)$ has a lower bound, i.e., $h(\theta) = E_{\bar{Q}_\theta}[X|X \in A] \geq a$. If Step 2 is replaced with finding $\theta^{(i)}$ as the solution of $h(\theta) = g(\theta^{(i-1)})$, it is very likely that $g(\theta^{(i-1)}) < a$ and hence leads to the non-existence of $\theta^{(i)}$.

Let θ^* be the solution to (9), i.e., $g(\theta^*) = h(\theta^*)$. Conditions for the convergence of the proposed recursive algorithm are provided in the following proposition.

Proposition 2 *Set a suitable initial value $\theta^{(0)}$. Suppose $g(\theta)$ is strictly increasing and $h(\theta)$ is strictly decreasing, and there exists θ_0 such that $g(\theta_0) = h(\theta_0)$. Then* (i) *the recursive algorithm either converges to θ^* or alternates in the limit between a pair of values $\theta \neq \bar{\theta}$ satisfying*

$$g(\theta) = h(\bar{\theta}) \quad \text{and} \quad h(\theta) = g(\bar{\theta}). \tag{10}$$

(ii) *If there does not exist $\theta \neq \bar{\theta}$ such that* (10) *holds, then the recursive algorithm converges to the solution of* (9).

The proof of Proposition 2 is similar to that of Theorem 2 of Fuh et al. (2011), and is omitted here.

Proposition 1 shows that $g(\theta) = \psi'(\theta)$ is strictly increasing and $h(\theta) = E_{\bar{Q}_\theta}[X|X \in A]$ is strictly decreasing. Moreover, there is an intersection between $\psi'(\theta)$ and $E_{\bar{Q}_\theta}[X|X \in A]$ under the condition of $E[X|X \in A] > E(X) := \mu$ in Theorem 1. In general, it is not easy to check if condition ii) in Proposition 2 holds without given any specific distribution. The convergence of the algorithm seems to be very fast, partially because $E_{\bar{Q}_\theta}[X|X > a]$ is like a straight line.

Choosing a suitable initial value $\theta^{(0)}$ is crucial for practical implementation. Our numerical experiment in Sect. 4 yield a satisfactory result that less than ten recursive steps are required to achieve the desired tolerance by simply setting $\theta^{(0)} = 0$. Indeed, setting a better value of $\theta^{(0)}$ can be tailored for specific cases. For example, motivated by the fact that θ^* is not far away from θ^+, $\theta^{(0)}$ can be set as θ^+. Following this line, in the case of the standard normal distribution, because θ^+ satisfies $\theta^+ = a$ ($\psi'(\theta) = \theta$), the initial value $\theta^{(0)}$ can be chosen as a, which is the dominating point of the event $\{X > a\}$. Nevertheless, our numerical experiments also show that the improvement by setting $\theta^{(0)} = \theta^+$ in the recursive algorithm is not substantial.

3 Examples

The jump diffusion model can be separated for the diffusion and the jump part. The diffusion part can be simply to a normal distribution and the jump part follows compound Poisson process. To illustrate the general account of importance sampling, in this subsection, we study two examples: normal distribution, compound Poisson process, and investigate several other interesting distributions. The simulation event is

$\{X > a\}$ for a given random variable X and some $a > 0$. In these examples, we explicitly calculate a closed-form formula of $E_{\bar{Q}}[X|X > a]$ when possible.

Example 1: Normal distribution (Diffusion Part)

Let X be a random variable with standard normal distribution, denoted by $N(0, 1)$, with probability density function (pdf) $\frac{dP}{d\mathcal{L}} = e^{-x^2/2}/\sqrt{2\pi}$. Standard calculation gives $\Psi(\theta) = \mathrm{e}^{\theta^2/2}$, $\psi(\theta) = \theta^2/2$ and $\psi'(\theta) = \theta$. In this case, the tilting measure Q is $N(\theta, 1)$, a location shift, and $\bar{Q}$ is $N(-\theta, 1)$. Applying Theorem 1 and using the fact that $X|\{X > a\}$ is a truncated normal distribution with minimum value a under $\bar{Q}$, θ^* needs to satisfy

$$\theta = \frac{\phi(a+\theta)}{1 - \Phi(a+\theta)} - \theta. \tag{11}$$

Alternatively, $G(\theta) = \mathrm{e}^{\theta^2}(1 - \Phi(a+\theta))$, and θ^* must satisfy the first-order condition, $2\theta(1 - \Phi(a+\theta)) = \phi(a+\theta)$, or equivalently, $\theta = \phi(a+\theta)/2(1 - \Phi(a+\theta))$. By using $\frac{1-\Phi(x)}{\phi(x)} \sim \frac{1}{x}$ as $x \to \infty$, it is easy to see from Eq. (11) that $\theta^* \sim a$ when a is large. This is the same as the large deviation tilting probability.

We remark the normal distribution has been analysed in Fuh and Hu (2004). For illustration, we consider this example from our general account and provide a simple and explicit characterization for θ^*, in the sense that the right-hand-side of (11) is a straightforward application of Theorem 1.

Example 2: Compound Poisson process (Jump Part)

Let $R(t) = \sum_{n=1}^{N(t)} \log Y_n$ be the compound Poisson process, where $N(t)$ is the Poisson process with intensity λ, and the jump sizes Y_n are i.i.d. lognormal distribution with parameters of location η and scale δ^2. Let $A = \{R(t) > a\}$ for a given a, and compute

$$\begin{aligned} P(R(t) > a) &= \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} P\left(\sum_{i=1}^{n} Z_i > a|N(t) = n\right) \\ &= \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} P\left(Z > \frac{a - n\eta}{\sqrt{n\delta^2}}|N(t) = n\right), \end{aligned}$$

where $Z_i, i = 1, \ldots, n$, are i.i.d. normal with mean η and variance δ^2 and $Z = \frac{\sum_{i=1}^n Z_i - n\eta}{\sqrt{n\delta^2}}$. As a remark, the compound Poisson process is related to the calculation of VaR in risk management, and detailed numerical experiments will be investigated in Sect. 4.

Denote $f(R(t)) = R(t) - a$ and let the likelihood ratio

$$\frac{dQ}{dP} = e^{\theta f(R(t)) - \psi(\theta)}, \tag{12}$$

where $\psi(\theta) = \log E[\exp(\theta f(R(t)))]$. Here, the probability density of the original measure P with respect to Lebesgue measure $\mathcal{L}$ is

$$\frac{dP}{d\mathcal{L}} = \frac{e^{-\lambda t}(\lambda t)^n}{n!}\left(\frac{1}{\sqrt{2\pi\delta^2}}\right)^n e^{-\frac{\sum_{i=1}^n (z_i-\eta)^2}{2\delta^2}},$$

and the moment generating function is

$$\Psi(\theta) = E\left[e^{\theta f(R(t))}\right] = e^{-\theta a}E\left[e^{\theta\sum_{i=1}^{N(t)} Z_i}\right] = e^{-\theta a}\sum_{n=1}^{\infty}\frac{e^{-\lambda t}(\lambda t)^n}{n!}E\left[e^{\theta\sum_{i=1}^n Z_i}|N(t)=n\right]$$
$$= e^{-\theta a}\sum_{n=1}^{\infty}\frac{e^{-\lambda t}(\lambda t)^n}{n!}e^{\left(\theta\eta+\frac{1}{2}\theta^2\delta^2\right)n} = e^{\lambda t\left(e^{\theta\eta+\frac{1}{2}\theta^2\delta^2}-1\right)-\theta a}.$$

Therefore, $\psi(\theta) = \lambda t\left(e^{\theta\eta+\frac{1}{2}\theta^2\delta^2}-1\right)-\theta a$, and $\psi'(\theta) = \lambda t e^{\theta\eta+\frac{1}{2}\theta^2\delta^2}(\eta+\theta\delta^2)-a$. Hence, the likelihood ratio (12) becomes

$$\frac{dQ}{dP} = e^{\theta\sum_{i=1}^n z_i}e^{-\lambda t\left(e^{\theta\eta+\frac{1}{2}\theta^2\delta^2}-1\right)}.$$

For a given $N(t)=n$, the probability density of the alternative distribution Q with respect to Lebesgue measure $\mathcal{L}$ is

$$\frac{dQ}{d\mathcal{L}} = \frac{dQ}{dP}\frac{dP}{d\mathcal{L}} = e^{\theta\sum_{i=1}^n z_i}e^{-\lambda t\left(e^{\theta\eta+\frac{1}{2}\theta^2\delta^2}-1\right)}\frac{e^{-\lambda t}(\lambda t)^n}{n!}\left(\frac{1}{\sqrt{2\pi\delta^2}}\right)^n e^{-\frac{\sum_{i=1}^n (z_i-\eta)^2}{2\delta^2}}$$
$$= \frac{\left(\lambda t e^{\theta\eta+\frac{1}{2}\theta^2\delta^2}\right)^n e^{-\lambda t e^{\theta\eta+\frac{1}{2}\theta^2\delta^2}}}{n!}\left(\frac{1}{\sqrt{2\pi\delta^2}}\right)^n e^{-\frac{\sum_{i=1}^n (z_i-(\eta+\theta\delta^2))^2}{2\delta^2}}.$$

Under Q, it is recognized that, $R(t) = \sum_{i=1}^{N(t)} Z_i$, where $N(t)$ is the Poisson process with intensity $\lambda e^{\theta\eta+\theta^2\delta^2/2}t$, and Z_i are i.i.d. $N(\eta+\theta\delta^2, \delta^2)$. By Theorem 1, θ^* needs to satisfy

$$\lambda t e^{\theta\eta+\frac{1}{2}\theta^2\delta^2}(\eta+\theta\delta^2) - a = E_{\bar{Q}_\theta}\left[R(t)|R(t)>a\right],$$

where $R(t) = \sum_{i=1}^{N(t)} Z_i$ with $N(t)$ being the Poisson process with intensity $\lambda e^{-\theta\eta+\theta^2\delta^2/2}t$ and $R(t)|\{N(t)=n\} \sim N(\eta-\theta\delta^2, \delta^2)$ under $\bar{Q}_\theta$.

Table 1 reports importance sampling tilting probability for eight distributions: Binomial, Poisson, exponential, normal, χ^2, Gamma, noncentral χ^2, and uniform. The case of t distribution can be found in Fuh et al. (2011). This table includes the family of tilting probability Q and $\psi'(\theta)$. Note that in Table 1, when $A = \{X > a\}$ for some $a > 0$, then $E_{\bar{Q}}[X|X>a] = a + \frac{1}{1+2\theta}$ for exponential distribution $E(1)$ and $= \frac{\phi(a+\theta)}{1-\Phi(a+\theta)} - \theta$

Table 1 Summary of distributions and their tilting measures

Ex	P	Q	$\psi'(\theta)$
1	$B(n, p)$	$B\left(n, \frac{pe^{\theta}}{pe^{\theta}+(1-p)}\right)$	$\frac{npe^{\theta}}{1-p+pe^{\theta}}$
2	$Pois(\lambda)$	$Pois(\lambda e^{\theta})$	λe^{θ}
3	$N(0, \sigma^2)$	$N(\theta, \sigma^2)$	$\theta\sigma^2$
4	$E(1)$	$E\left(\frac{1}{1-\theta}\right)$	$\frac{1}{1-\theta}, \quad \theta < 1$
5	$\chi^2(\kappa)$	$\Gamma\left(\frac{\kappa}{2}, \frac{2}{1-2\theta}\right)$	$\frac{\kappa}{1-2\theta}, \quad ss\theta < 1/2$
6	$\Gamma(\alpha, \beta)$	$\Gamma\left(\alpha, \frac{\beta}{1-\beta\theta}\right)$	$\frac{\alpha\beta}{1-\beta\theta}, \quad \theta < 1/\beta$
7	$NC\chi^2(\kappa, \lambda)$	$X\vert\{N=i\} \sim \Gamma\left(\frac{\kappa+2i}{2}, \frac{2}{1-2\theta}\right), N \sim Pois\left(\frac{\lambda}{2(1-2\theta)}\right)$	$\frac{\lambda+\kappa(1-2\theta)}{(1-2\theta)^2}, \quad \theta < 1/2$
8	Uniform (0, 1)	$\propto e^{\theta X}\mathbf{1}_{\{0\le X\le 1\}}$	$\frac{e^{\theta}}{e^{\theta}-1} - \frac{1}{\theta}$

Table 2 Relative efficiency for the standard normal distribution, the exponential distribution with mean 1, the chi-squared distribution with degree of freedom 1, the gamma distribution with shape parameter 4 and scale parameter 10, and the non-central chi-squared distribution with shape parameter 2 and scale parameter 10, respectively

p	$N(0, 1)$	$E(1)$	$\chi^2(1)$	$\Gamma(4, 10)$	$NC\chi^2(2, 10)$
0.0001	2409.74	818.53	603.61	1282.87	1529.46
0.001	290.90	109.88	82.74	166.00	192.20
0.01	38.06	16.57	12.90	23.74	26.54
0.05	9.98	4.99	4.04	6.76	7.34
0.1	5.77	3.13	2.60	4.10	4.38

for standard normal distribution. By Theorem 1, we have that $\bar{Q} = Q_{-\theta}$ in Examples 1–8.

Simulation studies in Table 2 show that the proposed importance sampling produces reasonable good variance reduction for different distributions. Here the relative efficiency is the variance of the naive estimator divided by that of the importance sampling estimator.

4 Evaluating Value-at-Risk

To illustrate the applicability of our proposed tilting formula, we apply the importance sampling for portfolio VaR computation, to which the tilting formula is given under the multivariate jump diffusion model of the underlying assets.

As a standard benchmark for market risk disclosure, VaR is the loss in market value over a specified time horizon that will not be exceeded with probability $1 - p$. Hence define VaR as the quantile l_p of the loss L in portfolio value during a holding period of a given time horizon Δt. Since the confidence levels of VaR are usually between

95 and 99%, they are events with moderately small probabilities of occurrence. To be more specific, we express the portfolio value $V(t, S(t))$ as a function of risk factors and time, where $S(t) = (S_1(t), \ldots, S_d(t))'$ comprises the d risk factors to which the portfolio is exposed at time t and the prime denotes transpose. The loss of the portfolio over the time interval $[t, t + \Delta t]$ is $L = V(t, S(t)) - V(t + \Delta t, S(t + \Delta t))$. Therefore VaR, l_p, associated with a given probability p and a time horizon Δt, is given by

$$P(L > l_p) = p. \tag{13}$$

Although $S(t)$ is assumed to follow the Geometric Brownian motion under the celebrated Black-Scholes model, this assumption suffers the lack of explaining empirical phenomena. To capture the empirical phenomena of leptokurtosis and volatility smile, jump diffusion models has been used in quantitative finance literature. The earliest version of the jump diffusion models with Poisson-type jump diffusions has been studied in Kou (2002), Merton (1976) and Naik and Lee (1990). By using the same type of modelling, we assume $S(t)$ follows a d-dimensional jump diffusion model such that the return processes are described by the stochastic differential equations

$$dr_i(t) := \frac{dS_i(t)}{S_i(t)} = \mu_i dt + \sigma_i dW_i(t) + d\sum_{j=1}^{N(t)}(Y_{ij} - 1), \; i = 1, 2, \ldots, d, \tag{14}$$

where $(W_1(t),\ldots,W_d(t))$ is a standard d-variate Brownian motion with $dW_i(t)dW_j(t) = \rho_{ij}dt$ for $i \neq j$, $N(t)$ is the Poisson process with intensity λ, and jump sizes of the ith return $\log Y_{i1}, \log Y_{i2}, \ldots$ are i.i.d. $N(\eta_i, \pi_{ii})$. We also assume $\log Y_{ij}$ and $N(t)$ are independent. Furthermore, we assume that the Brownian motion and the Poisson process are independent. Let $MVN(m, \Sigma)$ denote a multivariate normal distribution with mean vector m and covariance matrix Σ. Then, at the jth jump, we have the d-variate jump size $(\log Y_{1j}, \ldots, \log Y_{dj}) \sim MVN(\eta, \Pi)$ with $\eta = (\eta_1, \ldots, \eta_d)'$ and $\Pi = [\pi_{ij}]_{i,j=1,\ldots,d}$.

Denote $r = (r_1, \ldots, r_d)'$, where $r_i = r_i(t+\Delta t) - r_i(t)$ is the change in the ith return within a given time horizon Δt. A discrete version of (14) by Euler discretisation[1] is

$$r_i = \mu_i \Delta t + \sigma_i \sqrt{\Delta t} X_i + \sum_{j=1}^{N(\Delta t)} \log Y_{ij}, \quad i = 1, \ldots, d, \tag{15}$$

where $X = (X_1, \ldots, X_d)' \sim MVN(\mathbf{0}_d, \Sigma)$ with a d-variate zero vector $\mathbf{0}_d$, and $\Sigma = [\sigma_{ij}]_{i,j=1,\ldots,d}$ satisfying $\sigma_{ii} = 1$ and $\sigma_{ij} = \rho_{ij}$ for $i, j = 1, \ldots, d$. For ease of notation, denote $\mu = (\mu_1 \Delta t, \ldots, \mu_d \Delta t)'$, $\sigma = (\sigma_1 \sqrt{\Delta t}, \ldots, \sigma_d \sqrt{\Delta t})'$, and $J =$

[1] Because the VaR calculation in this example is one period, we simply need to simulate the return processes at the last time of consideration, but not the whole paths of the return processes. As a result, Euler discretisation is not necessary. As a remark, it is known that when the jump frequency λ gets bigger and the discretisation period Δt gets longer, then the discretisation error gets bigger with the same sample size, cf. http://marcoage.usuarios.rdc.pucrio.br/sim_stoc_proc.html. The discretisation issue may arise for multi-period VaR calculation, which is beyond the scope of this example.

$(\sum_{j=1}^{N(\Delta t)} \log Y_{1j}, \ldots, \sum_{j=1}^{N(\Delta t)} \log Y_{dj})'$. Then, (15) can be rewritten in a vector form as

$$r = \mu + \sigma' X + J. \tag{16}$$

Next we shall describe a quadratic approximation to the loss L. Approximate $\Delta S = [S(t + \Delta t) - S(t)] \approx r$ as the change in S over the corresponding time interval. The delta-gamma method refines the relationship between risk factors and portfolio value by including quadratic as well as linear terms, and approximates the change in the portfolio value by

$$V(t + \Delta t, S + \Delta S) - V(t, S) \approx \frac{\partial V}{\partial t}\Delta t + \delta' \Delta S + \frac{1}{2}\Delta S' \Gamma \Delta S,$$

where $\delta = (\delta_1, \ldots, \delta_d)'$ with $\delta_i = \partial V / \partial S_i$ and $\Gamma = [\Gamma_{ij}]_{i,j=1,\ldots,d}$ with $\Gamma_{ij} = \partial^2 V / \partial S_i \partial S_j$. Here, all derivatives are evaluated at the initial point (t, S), and $S_i = S_i(0)$ abbreviates the initial value of the ith asset for $i = 1, \ldots, d$. Hence we can approximate the loss by

$$L \approx a_0 + a_1' r + r' B r, \tag{17}$$

where $a_0 = -\frac{\partial V}{\partial t}\Delta t$ is a scalar, $a_1 = -\delta$ is a d-vector and $B = -\Gamma/2$ is a $d \times d$ symmetric matrix. Plugging (16) for r into (17), we have

$$\begin{aligned} L &= a_0 + a_1'\mu + a_1'\sigma X + a_1' J + \mu' B \mu + \sigma' Z' B Z \sigma \\ &= b_0 + a_1' \sigma X + a_1' J + \sigma' Z' B Z \sigma, \end{aligned} \tag{18}$$

where $b_0 = a_0 + a_1'\mu + \mu' B \mu$. To have a simple approximation, we neglect the quadratic approximation of the jump part in (17) as it is very small compared to the return of portfolio.

Let F be the Cholesky decomposition of Σ so that $FF' = \Sigma$, and K be the orthonormal matrix (satisfying KK' equal to the identity matrix) so that $-\frac{1}{2}F'\Gamma F = KDK'$. Here, D is a diagonal matrix with diagonal elements ϱ_i for $i = 1, \ldots, d$. Set $C = FK$, then it is straightforward that $CC' = \Sigma$ and $-\frac{1}{2}C'\Gamma C = D$, cf. page 1351 of Glasserman et al. (2000). Denoting $L_b := L - b_0$, we have

$$\begin{aligned} L_b &= a_1'\sigma X + a_1' J + \sigma' X' B X \sigma = a_1' \sigma C Z + \sigma' Z' C' B C Z \sigma + a_1' J \\ &= b' Z + \sigma' Z' \Lambda Z \sigma + a_1' J = \sum_{j=1}^{d} b_j Z_j + \varrho_j \sigma_j^2 Z_j^2 + a_1' J, \end{aligned} \tag{19}$$

where $b' = a_1' \sigma C$.

Employing the tilting formula developed in Theorem 1, let the likelihood ratio of the alternative probability measures with respect to the target measure P be of the form

$$\frac{dQ}{dP} = e^{\theta(L_b - a) - \psi(\theta)}, \tag{20}$$

where L_b is defined in (19), a is the pth quantile, and $\psi(\theta) = \log E[\exp(\theta(L_b - a))]$ is the cumulant generating function of $L_b - a$ under the target distribution P. The domain of θ will be specified after (25). Standard algebra yields

$$\begin{aligned}\psi(\theta) &= \log E(e^{\theta L_b - \theta a}) \\ &= \frac{1}{2}\sum_{i=1}^{d}\left(\frac{(\theta b_i)^2}{1 - 2\theta\varrho_i\sigma_i^2} - \log(1 - 2\theta\varrho_i\sigma_i^2)\right) \\ &\quad + \lambda\Delta t\left(e^{\theta a_1'\eta + \frac{1}{2}\theta^2 a_1'\Pi a_1} - 1\right) - \theta a,\end{aligned} \tag{21}$$

and

$$\begin{aligned}\psi'(\theta) = &\sum_{i=1}^{d}\left(\frac{\theta(b_i)^2(1 - \theta\varrho_i\sigma_i^2)}{(1 - 2\theta\varrho_i\sigma_i^2)^2} + \frac{\varrho_i\sigma_i^2}{1 - 2\theta\varrho_i\sigma_i^2}\right) - a \\ &+ \lambda\Delta t e^{\theta a_1'\eta + \frac{1}{2}\theta^2 a_1'\Pi a_1}(a_1'\eta + \theta a_1'\Pi a_1).\end{aligned} \tag{22}$$

Given $N(\Delta t) = n$, the probability density of the alternative distribution Q with respect to the Lebesgue measure $\mathcal{L}$ is

$$\begin{aligned}\frac{dQ}{d\mathcal{L}} = \frac{dQ}{dP}\frac{dP}{d\mathcal{L}} = &\prod_{j=1}^{d}\frac{1}{\sqrt{2\pi}s_j(\theta)}\exp\left\{-\frac{(z_j - m_j(\theta))^2}{2s_j^2(\theta)}\right\} \times \frac{\left(l(\theta)\right)^n e^{-l(\theta)}}{n!} \\ &\times\left(\frac{1}{(2\pi)^{d/2}|\Pi|^{1/2}}\right)^n \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(v_i - e(\theta))'(\Pi)^{-1}(v_i - e(\theta))\right\},\end{aligned} \tag{23}$$

where

$$m_j(\theta) = \frac{\theta b_j}{1 - 2\theta\varrho_j\sigma_i^2}, \quad s_j^2(\theta) = \frac{1}{1 - 2\theta\varrho_j\sigma_i^2}, \tag{24}$$

and

$$l(\theta) = \lambda\Delta t e^{\theta a_1'\eta + \frac{1}{2}\theta^2 a_1'\Pi a_1}, \quad e(\theta) = \eta + \theta a_1'\Pi. \tag{25}$$

Based on (23), it is recognized that Z_j are independent $N(m_j(\theta), s_j^2(\theta))$, J_i are compound Poisson processes with intensity $l(\theta)$, and the d-variate jump size $V_j \sim MVN(e(\theta), \Pi)$ for $j = 1, \ldots, n$. To avoid negative variance in (24) for a normal distribution, θ must satisfy $1 - 2\theta\varrho_{(1)}\sigma_i^2 > 0$ and $1 - 2\theta\varrho_{(d)}\sigma_i^2 > 0$, where $\varrho_{(1)} = \max_{1\le i\le d}\varrho_i$ and $\varrho_{(d)} = \min_{1\le i\le d}\varrho_i$ for $i = 1, \ldots, d$.

To have an efficient importance sampling for approximating $P\{L_b > a\}$ for some $a > 0$, we need to characterize the optimal tilting θ via Theorem 1. Before stating the result, we define some quantities that facilitate the presentation of it. In view of (24), define

$$\bar{m}_j(\theta) = m_j(-\theta), \quad \bar{s}_j^2(\theta) = s_j^2(-\theta), \quad \bar{l}(\theta) = l(-\theta), \quad \bar{e}(\theta) = e(-\theta). \tag{26}$$

Let $\bar{V}_j \sim MVN(\bar{e}(\theta), \Pi)$ and $N(\Delta t)$ follows a $Pois(\bar{l}(\theta))$. Applying Theorem 1, it is straightforward to obtain the following

$$\psi'(\theta) = E_{\bar{Q}_\theta}[L_b | L_b > a]. \tag{27}$$

The θ^* satisfying (27) can be calculated by the recursive algorithm presented in Sect. 2.2 Next we present the algorithm for calculating the importance sampling estimator for VaR, denoted by $\hat{p}^{\theta^*}$, under the multi-variate jump diffusion model as follows:

1. Compute θ^* satisfying $\psi'(\theta^*) = E_{\bar{Q}_{\theta^*}}[L_b | L_b > a]$.
2. (i) Generate $Z_i \sim N(m_i(\theta^*), s_i^2(\theta^*))$ for $i = 1, \ldots, d$.
   (ii) Generate $N(\Delta t) \sim Pois(l(\theta^*))$.
   (iii) Given $N(\Delta t) = n$ generate

$$V_j = (\log Y_{1j}, \log Y_{2j}, \ldots, \log Y_{dj})' \sim MVN(e(\theta^*), \Pi), \ \ j = 1, \ldots, n.$$

3. Repeat step 2 k times to have $L_{b,i}$ defined in (19) for $i = 1, \ldots, k$. Compute

$$\hat{p}^{\theta^*} = \frac{1}{k} \sum_{i=1}^{k} \mathbf{1}_{\{L_{b,i} > a\}} e^{-\theta^*(L_{b,i} - a) + \psi(\theta^*)}.$$

To compare the efficiency of alternative estimators, we define the relative efficiency between two estimators as the variance ratio of these two estimators. That is, denote

$$\mathrm{RE}(\hat{p}^0, \hat{p}^{\theta^*}) = \frac{\mathrm{Var}(\hat{p}^0)}{\mathrm{Var}(\hat{p}^{\theta^*})} \text{ and } \mathrm{RE}(\hat{p}^+, \hat{p}^{\theta^*}) = \frac{\mathrm{Var}(\hat{p}^+)}{\mathrm{Var}(\hat{p}^{\theta^*})}.$$

Table 3 compares the relative efficiency of the proposed estimator $\hat{p}^{\theta^*}$ with respect to the crude Monte Carlo estimator $\hat{p}^0$ and the importance sampling estimator by the large deviation theory $\hat{p}^{\theta^+}$ in calculating the loss probability $p = P(L_b > a)$ for some $a > 0$.

In this simulation study, we estimate the portfolio VaR by using equally weighted fifteen risk factors ($d = 15$). The purpose of this simulation study is to illustrate the applicability of our proposed importance sampling method. Therefore, we treat the jump diffusion model as a probability model with given parameters. In Table 3, the parameters are chosen based on financial practical issues. In practice, we can estimate the parameters based on a given real data set. Detailed analysis and real data applications of the proposed importance sampling algorithm will be studied in a separate paper.

To mimic practical scenarios, the parameters are chosen so that assets with high return are followed by high risk. To represent risk diversification among assets, a low

Table 3 Calculating the VaR-related probability $p = P(L_b > a)$ for some $a > 0$ with fifteen risk factors under the multivariate jump diffusion model. This table reports the mean, variance, relative efficiency, and amount of tilting for various estimators

a		0.824	1.166	1.549
p		0.0500	0.0100	0.00100
$\hat{p}^0$	Mean	0.05018	0.00994	0.00100
	Var	4.76E–05	9.88E–06	1.01E–06
$\hat{p}^{\theta^+}$	Mean	0.05008	0.01000	0.00100
	Var	6.51E–06	3.31E–07	4.03E–09
	θ^+	3.28	4.62	6.06
$\hat{p}^{\theta^*}$	Mean	0.05010	0.00999	0.00100
	Var	5.06E–06	2.69E–07	3.66E–09
	θ^*	4.81	5.72	6.58
	#	9	7	6
RE($\hat{p}^0, \hat{p}^{\theta^*}$)		9.41	36.70	277.18
RE($\hat{p}^{\theta^+}, \hat{p}^{\theta^*}$)		1.29	1.23	1.10

is the iteration number required for calculating θ^* with a tolerance of size 0.001. Sample size is $k = 1000$ and Monte Carlo sample size is 1,000,000

correlation is used among assets. The parameters are: $\lambda = 1, \eta_1 = \eta_2 = \ldots = \eta_{15} = 0, \delta_i = i/100, \mu_i = i/100, \sigma_i = 0.1 + i/100, \sigma_{ij} = 0.3$, and $\pi_{ij} = 0$ for $i \neq j$. These parameters lead to $b_1 = 0.044, b_2 = 0.0589891, b_3 = 0.0720326, b_4 = 0.0838734, b_5 = 0.0948957, b_6 = 0.105325, b_7 = 0.115304, b_8 = 0.124929, b_9 = 0.134271, b_{10} = 0.143378, b_{11} = 0.152289, b_{12} = 0.161034, b_{13} = 0.169635, b_{14} = 0.178112, b_{15} = 0.186479, \varrho_1 = 5.2$ and $\varrho_2 = \ldots = \varrho_{15} = 0.7$. In addition, we set sample size $k = 1000$ and Monte Carlo sample size 1,000,000.

From Table 3, we observe that $\hat{p}^{\theta^*}$ is significantly more efficient than $\hat{p}^0$ and moderately more efficient than $\hat{p}^{\theta^+}$ for all values of a. As expected, the efficiency gain of $\hat{p}^{\theta^*}$ is larger for smaller probabilities against $\hat{p}^0$ and the gain is smaller for smaller probabilities against $\hat{p}^{\theta^+}$. Table 3 also reports the values of θ^+ and θ^* for different importance sampling estimators. It indicates that the difference between θ^+ and θ^* becomes smaller as the probability p gets smaller. Note that the improvement of $\hat{p}^{\theta^*}$ with respect to $\hat{p}^{\theta^+}$ is around 25%, which is substantial for practical use of VaR calculation, cf. (Glasserman 2004).

Note that in this application, the improvement of our method over the large deviations tilting is substantial, to a ratio of 1.3. It is known that the VaR parameters are the choice of two quantitative factors: the length of the holding horizon or a greater confidence level. For instance, one illustration of the use of VaR as equity capital is the internal model approach. This period may be set in terms of hours, days or weeks. For traders, they will handle many portfolios to adjust positions immediately under intraday data. For bank managers, the regulator horizon is two weeks by Basel Accord. For investment managers, it may correspond to the regular reporting period, monthly or quarterly. Using higher-frequency data is generally more efficient because it uses more available information. To effectively meet the time constraint on calculating VaR is very important, and a moderate improvement on the efficiency of the importance sampling estimator is beneficial for practical purposes.

5 Conclusions

In this paper, we propose a general account in importance sampling with applications to portfolio VaR computation. It is shown that our method produces efficient approximation to the problem. For moderate deviation events, such as VaR, the performance of the proposed importance sampling compared with that based on large deviations is slightly better in simple case; while improve around 25% for portfolio VaR calculation under jump diffusion models.

The key features of our method are twofold. First, (8) characterizes the optimal alternative distribution for importance sampling under exponential tilting by using the idea of conjugate measures. And the recursive algorithm facilitates the computation of the optimal solution. The initial value of the recursive algorithm is the original point in most cases; while in the case of normal family, the initial point can be chosen as the dominating point of the large deviations tilting probability used previously by other authors; e.g., Sadowsky and Bucklew (1990). Due to the nature of the recursive algorithm, the additional programming effort and computing time are negligible. Second, the proposed tilting formula for normal distribution and its square, along with jump diffusion model, can be used to have more efficient simulation for portfolio VaR computation.

Our method highlights the two aforementioned key features in general settings. By using the idea of conjugate probability measure, we obtained the optimal tilting parameter via (8). Specific considerations can be found in several papers; see Fuh and Hu (2004) for multivariate normal distributions, Fuh and Hu (2007) for hidden Markov models, and Fuh et al. (2011) for multivariate t distribution. In the above works, the characterization of the optimal tilting is derived via change of variables and requires heavy algebraic manipulation. In contrast, the new idea of conjugate measure used in this paper allows to characterize the optimal tilting automatically without additional mathematical treatments, and hence makes the proposed importance sampling rather appealing for practical use.

Further applications to K-distributions and copula models will be published elsewhere. In this paper, we assume that the underlying random variables are independent over time. A more challenging project is to model the time dependence using, for example, Markov switching autoregression models or GARCH models. It is expected that our method can be applied to option pricing and correlated default probabilities calculation among others.

Acknowledgements The research of the first author is supported in part by a Grant NSC 100-2118-M-008-002-MY3 and NSC 101-3113-P-008-005, the second author is supported in part by a Grant NSC 102-2118-M-008-002, and the third author is supported in part by a Grant NSC 100-2410-H-032-024.

Appendix: Proofs

Proof (*Proof of Proposition* 1) We first note that $\psi(\theta)$ is the cumulant generating function of X, and therefore its second derivative $\psi''(\theta) > 0$ for $\theta \in \mathbf{I}$. This implies that $\psi'(\theta)$ is strictly increasing. Since $\Psi(\theta)$ is convex and steep by assumption, we have $\psi'(\theta) \to \infty$ as $\theta \to \theta_{\max}$. Furthermore, consider the conditional measure of $\bar{Q}$ on the set A, denoted by $\bar{Q}_A$, which is defined as

$$d\bar{Q}_A = \frac{\mathbf{1}_{\{X\in A\}} d\bar{Q}}{\int \mathbf{1}_{\{X\in A\}} d\bar{Q}}. \tag{28}$$

Then we have

$$\begin{aligned}\frac{dE_{\bar{Q}_\theta}[X|X\in A]}{d\theta} &= \frac{d}{d\theta}\left(\frac{E[\mathbf{1}_{\{X\in A\}} X \mathrm{e}^{-\theta X}]}{E[\mathbf{1}_{\{X\in A\}} \mathrm{e}^{-\theta X}]}\right) \\ &= -\frac{E[\mathbf{1}_{\{X\in A\}} X^2 \mathrm{e}^{-\theta X}]}{E[\mathbf{1}_{\{X\in A\}} \mathrm{e}^{-\theta X}]} + \frac{E^2[\mathbf{1}_{\{X\in A\}} X \mathrm{e}^{-\theta X}]}{E^2[\mathbf{1}_{\{X\in A\}} \mathrm{e}^{-\theta X}]} = -\mathrm{var}_{\bar{Q}_A}(X) < 0.\end{aligned} \tag{29}$$

Since under the assumption in Proposition 1, the second and third terms in (29) are bounded for all $\theta \in \mathbf{I}$, by Fubini's theorem, the second equality in (29) holds. This implies that $E_{\bar{Q}_\theta}[X|X \in A]$ is strictly decreasing. □

Proof (Proof of Theorem 1) The existence of the optimization problem (3) requires the following fact,

$$\psi'(\theta) \text{ is strictly increasing and } E_{\bar{Q}}[X|X \in A] \text{ is strictly decreasing,} \tag{30}$$

which has been proved in Proposition 1.

The existence of the optimization problem (3) follows from $E_{\bar{Q}_0}[X|X \in A] = E[X|X \in A] > \mu = \psi'(0)$.

To prove the uniqueness, we note that the second derivative of G equals

$$\begin{aligned}\frac{d^2 G(\theta)}{d\theta^2} &= \frac{d^2}{d\theta^2} E\left[\mathbf{1}_{\{X\in A\}} \frac{dP}{dQ}\right] = \frac{d^2}{d\theta^2} E\left[\mathbf{1}_{\{X\in A\}} e^{-\theta X+\psi(\theta)}\right] \\ &= \frac{d}{d\theta} E\Big[\mathbf{1}_{\{X\in A\}} (+\psi'(\theta)) e^{-\theta X+\psi(\theta)}\Big] \\ &= E\Big[\mathbf{1}_{\{X\in A\}} ((-X+\psi'(\theta))^2 + \psi''(\theta)) e^{-\theta X+\psi(\theta)}\Big].\end{aligned} \tag{31}$$

Since $\psi(\theta)$ is the cumulant generating function of X, its second derivative $\psi''(\theta) > 0$. It then follows from (31) that $\frac{d^2 G(\theta)}{d\theta^2} > 0$, which implies that there exists a unique minimum of $G(\theta)$.

To prove (8), we need to simplify the RHS of (5) under $\bar{Q}$. Standard algebra gives

$$\frac{E[\mathbf{1}_{\{X\in A\}} X \mathrm{e}^{-\theta X}]}{E[\mathbf{1}_{\{X\in A\}} \mathrm{e}^{-\theta X}]} = \frac{\int \mathbf{1}_{\{x\in A\}} x \mathrm{e}^{-\theta x} dP/\tilde{\Psi}(\theta)}{\int \mathbf{1}_{\{x\in A\}} \mathrm{e}^{-\theta x} dP/\tilde{\Psi}(\theta)} = \frac{\int \mathbf{1}_{\{x\in A\}} x d\bar{Q}}{\int \mathbf{1}_{\{x\in A\}} d\bar{Q}}. \tag{32}$$

As a result, (32) equals

$$\int \mathbf{1}_{\{x\in A\}} x d\bar{Q}_A = E_{\bar{Q}_A}[X] = E_{\bar{Q}}[X|X \in A],$$

which implies the desired result. □

References

Asmussen, S., & Glynn, P. (2007). *Stochastic simulation: Algorithms and analysis*. New York: Springer.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika*, *74*, 457–468.

Beran, R., & Millar, P. W. (1986). Confidence sets for a multivariate distribution. *Annals of Statistics*, *14*, 431–443.

Bucklew, J. A. (2004). *Introduction to rare event simulation*. New York: Springer.

Davison, A. C. (1988). Discussion of papers by D. V. Hinkley and by T. J. DiCiccio & J. P. Romano. *Journal of Royal Statistical Society*, *50*, 356–357.

Do, K. A., & Hall, P. (1991). On importance resampling for the bootstrap. *Biometrika*, *78*, 161–167.

Duffie, D., & Singleton, J. (2003). *Credit risk*. Princeton, New Jersey: Princeton University Press.

Fuh, C. D., & Hu, I. (2004). Efficient importance sampling for events of moderate deviations with applications. *Biometrika*, *91*, 471–490.

Fuh, C. D., & Hu, I. (2007). Estimation in hidden Markov models via efficient importance sampling. *Bernoulli*, *13*, 492–513.

Fuh, C. D., Hu, I., Hsu, Y. H., & Wang, R. H. (2011). Efficient simulation of value at risk with heavy-tailed risk factors. *Operations Research*, *59*, 1395–1406.

Glasserman, P. (2004). *Monte carlo methods in financial engineering*. New York: Springer.

Glasserman, P., Heidelberger, P., & Shahabuddin, P. (2000). Variance reduction techniques for estimating Value-at-Risk. *Management Science*, *46*, 1349–1364.

Glasserman, P., Heidelberger, P., & Shahabuddin, P. (2002). Portfolio Value-at-Risk with heavy-tailed risk factors. *Mathematical Finance*, *12*, 239–269.

Hall, P. (1987). On the bootstrap and likelihood-based confidence regions. *Biometrika*, *74*, 481–493.

Hall, P. (1992). *The bootstrap and edgeworth expansion*. New York: Springer.

Heidelberger, P. (1995). Fast simulation of rare events in queueing and reliability models. *ACM Transactions on Modeling and Computer Simulation*, *5*, 43–85.

Johns, M. V. (1988). Importance sampling for bootstrap conference intervals. *Journal of the American Statistical Association*, *83*, 709–714.

Jorion, P. (2001). *Value-at-Risk*. New York: McGraw-Hill.

Kou, S. G. (2002). A jump diffusion model for option pricing. *Management Science*, *48*, 1086–1101.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics*, *80*, 221–239.

Liu, J. (2001). *Monte carlo strategies in scientific computing*. New York: Springer.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, *3*, 125–144.

Naik, V., & Lee, M. (1990). General equilibrium pricing of options on the market portfolio with discontinuous return. *Review of Financial Studies*, *3*, 493–521.

Ney, P. (1983). Dominating points and the asymptotics of large deviations for random walk on $\mathcal{R}^d$. *Annals of Probability*, *11*, 158–167.

Sadowsky, J., & Bucklew, J. A. (1990). On large deviations theory and asymptotically efficient Monte Carlo simulation. *IEEE Transition of Information Theory*, *36*, 579–588.

Siegmund, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. *Annals of Statistics*, *4*, 673–684.