Computational Economics (2023) 62:1125–1154
https://doi.org/10.1007/s10614-022-10294-y

Importance Sampling for Calculating the Value-at-Risk and Expected Shortfall of the Quadratic Portfolio with *t*-Distributed Risk Factors

Huei-Wen Teng[1]

Accepted: 29 June 2022 / Published online: 22 July 2022

Abstract

In the banking industry, the calculation of regulatory capital by the Basel accords is directly related to the values of the Value-at-Risk (VaR) and expected shortfall (ES). The Monte Carlo simulation approach for calculating the VaR and ES is preferred, because it is able to incorporate a wide range of realistic models. Motivated by the gigantic size of the derivatives market, we consider the quadratic portfolio with *t*-distributed risk factors. To overcome the slow convergence of the Monte Carlo simulation approach, we propose a novel importance sampling scheme, which is applicable to the calculation of the VaR and ES. Numerical experiments confirm the superiority of our method in terms of substantial reduction in variance and computing time, particularly in calculating the ES.

Keywords Importance sampling · Value-at-Risk · Expected Shortfall · *t* distribution · Quadratic portfolio

1 Introduction

The Value-at-Risk (VaR) is the quantile of a portfolio's loss distribution, and the Expected Shortfall (ES) is the conditional expectation (or expected value) of a portfolio's loss when the portfolio loss exceeds the VaR. The VaR and ES are critical for computing regulatory capital by the Basel accords (Bank for International Settlements, 2006, 2009). In addition, they are frequently used in portfolio management, financial reporting, and non-financial applications (Duffie & Pan, 1997). As a result, accurate and efficient calculation of the VaR and ES is of considerable importance both in theory and practice.

✉ Huei-Wen Teng
venteng@gmail.com

[1] Department of Information Management and Finance, National Yang Ming Chiao Tung University, 1001 University Road, Hsinchu, Taiwan

There are three types of method to estimate the VaR and ES, depending on how the portfolio loss is modeled: (1) The variance-covariance approach simply assumes the portfolio loss follows a Normal distribution and calculates VaR and ES through the inverse Normal distribution function; (2) The historical simulation uses the empirical distribution from historical samples and calculates the VaR and ES as the empirical quantile and empirical conditional sample mean, respectively; (3) The Monte Carlo simulation approach assumes a suitable parametric model for the portfolio loss, and calculates the VaR and ES using Monte Carlo simulation. Because the Monte Carlo simulation approach enables to employ a wide range of more realistic models, it is preferred (Jorion, 2000; Duffie & Singleton, 2003).

Although the Monte Carlo simulation approach is simple to implement, it is notoriously known for its slow convergence. As a remedy, various variance reduction techniques, such as antithetic variates, control variables, importance sampling, stratified sampling, and low-descrepancy sequences, have been proposed to improve the standard Monte Carlo simulation. Reviews on variance reduction techniques can be found in Glasserman (2003), Asmussen and Glynn (2007), and Ross (2013).

In calculating the VaR, the control variates method is applied in Hsu and Nelson (1990) and Hesterberg and Nelson (1998) and correlation-induction techniques are used in Avramidis and Wilson (1998). Because the VaR is about rare event simulation, importance sampling is particularly useful. For example, the asymptotic properties of the importance sampling estimator of VaR have been analyzed in Bahadur (1966), Sun and Hong (2010) and Chu and Nakayama (2012). In addition, Glynn (1996) analyzes quantile estimator, Glasserman and Li (2005) estimate the VaR of portfolio credit risk, and Glasserman and Juneja (2008) estimates the VaR of a sum of independent and identically distributed random variables.

However, existing studies only focus on calculating the VaR, but not the ES, possibly for two reasons: First, the formula of the minimum capital requirement in the earlier version of Basel accords is only involved with the VaR. However, the ES gains more attentions in recent years because it is able to capture the tail behavior of the portfolio loss (Jorion, 2000; Duffie & Singleton, 2003). Indeed, the formula of the minimum capital requirement under the internal models approach has been updated to include the ES (Bank for International Settlements, 2019).

Second, the calculation of ES is more difficult, because it is involved with an expectation. In contrast, the VaR is only involved with a probability. Note that a probability could be written as an expectation of an indicator function, and hence can be considered a simplified case of an expectation. The above two reasons motivate us to provide an efficient simulation scheme that is applicable to calculate both the VaR and ES.

On the other hand, motivated by the gigantic size of financial derivative markets, we follow Glasserman et al. (2000) and Glasserman et al. (2002) to approximate the portfolio loss using the delta-gamma method and focus on the so-called quadratic portfolio (Hull, 2014). For calculating the VaR of the quadratic portfolio, Glasserman et al. (2000) assume Normally distributed risk factors, and Glasserman et al. (2002) and Fuh et al. (2011) assume *t*-distributed risk factors. To capture the stylized feature of heavy-tails of asset returns, we focus on the *t*-distributed risk factors Cont (2001).

However, how to apply importance sampling to calculate a general expectation for the t distribution remains a challenging task. For a general expectation, importance sampling is only applicable to the Normal distribution and Lévy processes (Fu & Su, 2002; Kawai, 2009; Jiang et al., 2016; Teng et al., 2016). Recently, Botev and lÉcuyer (2015) propose an importance sampling, which is only applicable for calculating the probability of a truncated multivariate t distribution.

From a methodological aspect, we propose a generally applicable importance sampling scheme for the t distribution, and apply it to the calculation of the VaR and ES for the quadratic portfolio with t-distributed risk factors. The key idea is to represent the t distribution as a ratio of the Normal and Gamma distributions, and employ exponential titling with two tilting parameters for each distribution. Then, we characterize the optimal tilting parameters by minimizing the variance of the importance sampling estimator. We also provide a useful stochastic fixed-point-Newton algorithm to numerically search the optimal tilting parameter.

The rest of this paper is organized as following. Section 2 reviews the VaR, and ES, and the model for the portfolio loss. Section 3 proposes a novel importance sampling estimator for the t distribution. Section 4 illuminates the rationale of the proposed importance sampling approach. Section 5 summarizes simulation experiments, and Sect. 6 concludes. All proofs are deferred to Appendices.

2 Preliminaries

Section 2.1 introduces the VaR and ES, and Sect. 2.2 introduces the model assumptions on the portfolio loss.

2.1 The Value-at-Risk and Expected Shortfall

Let L be the random variable to denote the portfolio loss. Let $P(A)$ denote the probability of event A. Let $F(y) = P(L \leq y)$ be the cumulative distribution function of L. Given a confidence level $\alpha \in (0, 1)$, the α-VaR of the portfolio loss at the confidence α, denoted by v_α, is the smallest number such that the probability that the portfolio L exceeds it is at least α. In other words, the α-VaR satisfies,

$$v_\alpha = \inf\{y : F(y) \geq \alpha\}.$$

In practice, α is set to be 1% for calculating adequate capital requirement, and α is set to be 0.1% for conducting stress testing. See for example Tsay (2013).

A key step in calculating the VaR is to calculate the probability that L exceeds a given threshold, q,

$$P(L > q). \tag{1}$$

In practice, a set of thresholds are given to calculate the above threshold probabilities. When these probabilities are calculated accurately, v_α can be obtained with interpolation. The threshold probability in Eq. (1) can be written as an expectation,

$$P(L > q) = E\left[I_{\{L > q\}}(L)\right], \tag{2}$$

where $E[\cdot]$ is the expectation and $I_{\{A\}}(\cdot)$ denotes the indicator function with the support set A.

The ES is also known as the average value at risk, conditional Value-at-Risk, and expected tail loss. The α-ES of the portfolio loss at the confidence α, denoted as c_α, is the conditional expectation of the portfolio loss conditional on it exceeding the α-VaR,

$$c_\alpha = E[L|L > v_\alpha],$$

where $E[L|L > v_\alpha]$ denotes the conditional expectation of L with respect to the event $\{L > v_\alpha\}$. By the definition of conditional expectation, we have

$$c_\alpha = \frac{E[LI_{\{L > v_\alpha\}}(L)]}{P(L > v_\alpha)} = \frac{E[LI_{\{L > v_\alpha\}}(L)]}{\alpha}. \tag{3}$$

To obtain c_α, we only need to calculate the numerator,

$$E[LI_{\{L > v_\alpha\}}(L)], \tag{4}$$

because the above denominator is simply α. We note that Eqs. (2) and (4) are critical values for calculating VaR and ES, and their calculation involves expectations with rare event $\{L > q\}$. Rare event simulation makes the standard Monte Carlo estimator inefficient.

2.2 Quadratic Portfolio by the Delta–Gamma Method

Following Glasserman et al. (2002) closely, we assume that the portfolio value $V(t, S)$ at time t is exposed to d underlying risk factors $S = (S_1, \ldots, S_d)'$. Let ΔS denote the change in S from the current time t to the end of the horizon time $t + \Delta t$. L is approximated by the delta-gamma method,

$$L = V(t, S) - V(t + \Delta t, S) = a_0 + a'\Delta S + \Delta S' A \Delta S, \tag{5}$$

where $a_0 = -\partial V(t, S)/\partial t$ is a scalar, $a = (a_1, \ldots, a_d)'$ is a d-dimensional vector with $a_i = -\partial V(t, S)/\partial S_i$, and A is a $d \times d$ matrix. Let $A[i, j]$ denote the i-th row and j-th column of a matrix A. Then, we have $A[i,j] = -\partial^2 V(t, S)/(2\partial S_i \partial S_j)$. Here, all derivatives are evaluated at the initial point (t, S). In practice, parameters a_0, a, and A are given as known values. Equation (5) approximates L by a quadratic function in ΔS. As a result, the portfolio loss modelled this way is called a quadratic portfolio. The non-linearity in risk factors in the quadratic portfolio makes it popular when modelling a portfolio consisting of financial derivatives (Duffie & Pan, 1997).

Let $t_{d,v}$ denote the d-dimensional t distribution with degrees of freedom v, and $\sim$ denote "distributed as". Suppose $X \sim t_{d,v}$. To capture the feature of heavy tails for the change of underlying risk factors, we follow Glasserman et al. (2002) to assume

$$\Delta S = CX,$$

where C is the square root of the positive definite covariance matrix Σ, such that $\Sigma = C'C$ and $C'AC = \Lambda$ is diagonalized with elements $\lambda_1, \ldots, \lambda_d$. Standard algebra yields

$$\begin{aligned} L &= a_0 + a'\Delta S + \Delta S' A \Delta S \\ &= a_0 + a'CX + (CX)'A(CX) \\ &= a_0 + b'X + X'\Lambda X, \end{aligned} \tag{6}$$

where $b = a'C$.

With the model for the portfolio loss defined in Eq. (6), Eq. (2) becomes

$$P(L > q) = E[I_{\{L > q\}}(L)] = E[I_{\{(a_0 + b'X + X'\Lambda X) > q\}}(X)].$$

In addition, Eq. (4) becomes

$$E[LI_{\{L > v_\alpha\}}(L)] = E[(a_0 + b'X + X'\Lambda X)I_{\{(a_0 + b'X + X'\Lambda X) > v_\alpha\}}(X)].$$

Both these two values are expectations of a performance function for the d-variate t distribution.

3 Our Methodology

We propose a novel and generally applicable importance sampling estimator for the t distribution in Sect. 3.1. Then, we provide mathematical analysis to find the optimal tilting parameter in Sect. 3.2. We summarize a fixed-point Newton algorithm to numerically search the optimal tilting parameter 3.3. Finally, we summarise the methodological workflow in Sect. 3.4.

3.1 An Importance Sampling Estimator for the *t* Distribution

Let $N_d(\mu, \Sigma)$ denote the d-dimensional Normal distribution with mean vector μ and covariance matrix Σ. Let $\mathbf{0}$ and $\mathbb{I}$ denote the zero vector and identity matrix of size d, respectively. Then, $N_d(\mathbf{0}, \mathbb{I})$ denotes the d-dimensional standard Normal distribution. In addition, let χ^2_v denote the chi-squared distribution with degrees of freedom v, and $\Gamma(\kappa, \theta)$ denote the Gamma distribution with the shape parameter κ and scale parameter θ.

Consider the random vector $X \sim t_{d,v}$. Recall that X has the following stochastic representation,

$$X \stackrel{d}{=} \frac{Z}{\sqrt{Y/v}}, \tag{7}$$

where $Z \sim N_d(\mathbf{0}, \mathbb{I})$, $Y \sim \chi^2_v$, and Y and Z are independent. Here, $\stackrel{d}{=}$ denotes "equivalence in distribution". We are interesting in calculating the expectation of $\wp(X)$ for an arbitrary real-valued function $\wp(\cdot)$. As a result, we have

$$E[\wp(X)] = E\left[\wp\left(\frac{Z}{\sqrt{Y/v}}\right)\right]. \tag{8}$$

Abusing the notations slightly, we write

$$\wp(Y,Z) = \wp\left(\frac{Z}{\sqrt{Y/v}}\right)$$

to obtain

$$E[\wp(Y,Z)] = E\left[\wp\left(\frac{Z}{\sqrt{Y/v}}\right)\right] = E[\wp(X)].$$

The above identity allows us to estimate $E[\wp(X)]$ through Y and Z, but not from the original t-distributed X. Denote the standard estimator by

$$\wp_0(Y,Z) = \wp\left(\frac{Z}{\sqrt{Y/v}}\right). \tag{9}$$

Then, we estimate $E[\wp(X)]$ by the average of n realized values:

$$\hat{\wp}_0 = \frac{1}{n}\sum_{i=1}^{n}\wp(Y^{(i)}, Z^{(i)}), \tag{10}$$

where realizations $Y^{(i)}$ and $Z^{(i)}$ are generated independently according to $Y^{(i)} \sim \chi^2_v$ and $Z^{(i)} \sim N_d(\mathbf{0}, \mathbb{I})$ for $i = 1, \ldots, n$.

Importance sampling with exponential tilting is appealing for its mathematical tractability (Glynn, 1996), and it incorporates tilting parameters and sampling probability measure.With the stochastic representing of the t distribution in Eq. (7), we use $\eta \in \Re$ and $\vartheta = (\vartheta_1, \ldots, \vartheta_d) \in \Re^d$ as the tilting parameter for the Gamma and Normal distributions, respectively. Specifically, Lemma 1 proposes an importance sampling estimator for calculating $E[\wp(X)]$.

Lemma 1 *Suppose $X \sim t_{d,v}$. Let $\eta < \frac{1}{2}$ and $\vartheta = (\vartheta_1, \ldots, \vartheta_d)' \in \mathbb{R}^d$. The importance sampling estimator with exponential tilting for $E[\wp(X)]$ is*

$$\wp_{\eta,\vartheta}(Y,Z) = \wp(Y,Z)e^{-\eta Y - \vartheta' Z - v\log(1-2\eta)/2 + \vartheta'\vartheta/2}, \tag{11}$$

where $Y \sim \Gamma(\frac{v}{2}, \frac{2}{1-2\eta})$, $Z \sim N_d(\vartheta, \mathbb{I})$, and Y and Z are independent.

Proof See Appendix A. □

Here, the term $e^{-\eta Y - \vartheta' Z - v\log(1-2\eta)/2 + \vartheta'\vartheta/2}$ in the right-hand-side of Eq. (11) is also known as the importance sampling weight or the Radon-Nikodym derivative. Once the tilting parameters η and ϑ are given properly, $E[\wp(X)]$ can be evaluated by the average of n realized function values:

$$\hat{\wp}_{\eta,\vartheta} = \frac{1}{n}\sum_{i=1}^{n} \wp(Y^{(i)}, Z^{(i)}) e^{-\eta Y^{(i)} - \vartheta' Z^{(i)} - v\log(1-2\eta)/2 + \vartheta'\vartheta/2},$$

where realizations $Y^{(i)}$ and $Z^{(i)}$ are generated independently according to $Y^{(i)} \sim \Gamma\left(\frac{v}{2}, \frac{2}{1-2\eta}\right)$, and, $Z^{(i)} \sim N_d(\vartheta, \mathbb{I})$ for $i = 1, \ldots, n$.

3.2 Solutions for the Optimal Tilting Parameters

To search the optimal tilting parameter, similar to Fuh et al. (2011) and Teng et al. (2016), we minimize the variance of the importance sampling estimator:

$$\mathrm{V}(\wp_{\eta,\vartheta}(Y,Z)) = E_{\eta,\vartheta}[\wp^2_{\eta,\vartheta}(Y,Z)] - \left(E_{\eta,\vartheta}[\wp_{\eta,\vartheta}(Y,Z)]\right)^2.$$

Because the importance sampling estimator is unbiased, minimizing the variance of the importance sampling estimator is equivalent to minimize the second moment of the importance sampling estimator,

$$G(\eta,\vartheta) = E_{\eta,\vartheta}[\wp^2_{\eta,\vartheta}(Y,Z)].$$

Standard algebra simplifies $G(\eta,\vartheta)$ as an expectation under $\mathcal{P}$, which is defined as

$$\begin{aligned} G(\eta,\vartheta) &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^+} \left(\wp(y,z)e^{-\eta y - \vartheta z + \psi_Y(\eta) + \psi_Z(\vartheta)}\right)^2 f_Y(y) e^{\eta y - \psi_Y(\eta)} f_Z(z) e^{\vartheta' z - \psi_Z(\vartheta)} dydz \\ &= E[\wp^2(Y,Z)e^{-\eta Y - \vartheta Z - v\log(1-2\eta)/2 + \vartheta'\vartheta/2}], \end{aligned} \tag{12}$$

where $Y \sim \chi^2_v$, $Z \sim N_d(\mathbf{0}, \mathbb{I})$, and Y and Z are independent under $\mathcal{P}$. Theorem 1 guarantees the convexity and existence of a minimizer for $G(\eta, \theta)$.

Theorem 1 *$G(\eta,\vartheta)$ is strictly convex and has a unique minimizer.*

Proof See Appendix B. □

Let $f_Y(y)$ and $f_Z(z)$ denote the pdf of $Y \sim \chi^2_v$ and $Z \sim N_d(\mathbf{0}, \mathbb{I})$, respectively. To numerically calculate the optimal tilting parameter, we define the joint pdf for Y and Z for notational ease by

$$f_{\wp,\eta,\vartheta}(y,z) = f_Y(y) f_Z(z) \frac{\wp^2(y,z)e^{-\eta y - \vartheta' z}}{E[\wp^2(Y,Z)e^{-\eta Y - \vartheta' Z}]}, \tag{13}$$

where the subscripts of $f_{\wp,\eta,\vartheta}(y,z)$ emphasize its connections to the performance function $\wp$ and tilting parameters η and ϑ. Theory 2 shows that the tiling parameters minimizing $G(\eta,\vartheta)$ are indeed the solution of a system of $(d+1)$ non-linear equations.

Theorem 2 *The optimal η and ϑ that minimize the variance of the importance sampling estimator is the solution to the following system of $(d+1)$ non-linear equations*:

$$\frac{v}{1-2\eta} = E_{\wp,\eta,\vartheta}[Y], \tag{14}$$

$$\vartheta = E_{\wp,\eta,\vartheta}[Z], \tag{15}$$

where $E_{\wp,\eta,\vartheta}$ indicates that the expectation is taken with Y and Z distributed according to $f_{\wp,\eta,\vartheta}(y,z)$ defined in Eq. (13).

Proof See Appendix C. Note that Eq. (15) is in vector form and abbreviates the following d equations:

$$\begin{cases} \vartheta_1 & = & E_{\wp,\eta,\vartheta}[Z_1], \\ & \vdots & \\ \vartheta_d & = & E_{\wp,\eta,\vartheta}[Z_d]. \end{cases}$$

□

3.3 A Stochastic Fixed-Point-Newton Algorithm

Searching the optimal tilting parameters characterized in Theorem 2 is numerically difficult in two aspects: First, the solution needs to satisfy a system of d equations simultaneously. Second, the right-hand sides in Eqs. (14) and (15) are expectations without closed-form formulas and need to be evaluated by simulation.

To overcome the first difficulty, we provide a useful stochastic root solving algorithm with the following features: First, η for the Gamma distribution is searched via the fixed-point iteration; and second, ϑ for the Normal distribution is searched via the Newton method. Specifically, let $\eta^{(j)}$ and $\vartheta^{(j)}$ denote the values for η and ϑ at the j-th iteration, respectively.

To search η, we employ the fixed-point algorithm, so that η needs to satisfy

$$\frac{v}{1-2\eta} = E_{\wp,\eta,\vartheta}[Y].$$

Standard algebra gives us a simple formula for η:

$$\eta = \frac{1}{2}\left(1 - \frac{v}{E_{\wp,\eta,\vartheta}[Y]}\right).$$

As a result, the fixed-point algorithm updates $\eta^{(j+1)}$ by

$$\eta^{(j+1)} = \frac{1}{2}\left(1 - \frac{v}{E_{\wp,\eta^{(j)},\vartheta^{(j)}}[Y]}\right).$$

To search ϑ, we employ the Newton method. Now, define a function $g(\vartheta) = (g_1(\vartheta),\ldots,g_d(\vartheta))'$ from $\mathbb{R}^d$ to $\mathbb{R}^d$ by

$$g(\vartheta) = \vartheta - E_{\wp,\eta,\vartheta}[Z]. \tag{16}$$

Then, searching ϑ satisfying Eq. (15) is equivalent to search the root of Eq. (16). The Jacobian J_ϑ of $g(\vartheta)$ in Eq. (16) is a $d \times d$ matrix with

$$J_\vartheta[i,j] := \frac{\partial}{\vartheta_j} g_i(\vartheta) = \mathbb{I}[i,j] - E_{\wp,\eta,\vartheta}[Z_i]E_{\wp,\eta,\vartheta}[Z_j] + E_{\wp,\eta,\vartheta}[Z_iZ_j],$$

for $i,j = 1,\ldots,d$. In matrix form, we have

$$J_\vartheta = \mathbb{I} - E_{\wp,\eta,\vartheta}[Z]E_{\wp,\eta,\vartheta}[Z]' + E_{\wp,\eta,\vartheta}[ZZ']. \tag{17}$$

Newton method finds the root of Eq. (16) by updating

$$\vartheta = \vartheta - J_\vartheta^{-1} g(\vartheta),$$

Therefore, we update $\vartheta^{(j+1)}$ by

$$\vartheta^{(j+1)} = \vartheta^{(j)} - J_{\vartheta^{(j)}}^{-1} g(\vartheta^{(j)}),$$

where $J_{\vartheta^{(j)}}^{-1}$ is the inverse of $J_{\vartheta^{(j)}}$. In addition, $J_{\vartheta^{(j)}}$ is calculated with $\eta = \eta^{(j+1)}$ and $\vartheta = \vartheta^{(j)}$.

Moreover, Corollary 1 applies the importance sampling estimator in Eq. (11) with arbitrary values of the tilting parameters $\bar{\eta}$ and $\bar{\vartheta}$ to efficiently calculate $E_{\wp,\eta,\vartheta}[Y]$, $E_{\wp,\eta,\vartheta}[Z]$, and $E_{\wp,\eta,\vartheta}[ZZ']$.

Corollary 1 *Consider arbitrary and proper values of* $\bar{\eta} < \frac{1}{2}$ *and* $\bar{\vartheta} \in \mathbb{R}^d$. *Let* $Y \sim \Gamma(\frac{\nu}{2}, \frac{2}{1-2\bar{\eta}})$ *and* $Z \sim N_d(\bar{\vartheta}, \mathbb{I})$. *We have the following identities,*

$$\begin{aligned}
E_{\wp,\eta,\vartheta}[Y] &= \frac{E_{\bar{\eta},\bar{\vartheta}}\left[Y\wp^2(Y,Z)e^{-\eta Y - \vartheta' Z - \bar{\eta} Y - \bar{\vartheta}' Z}\right]}{E_{\bar{\eta},\bar{\vartheta}}\left[\wp^2(Y,Z)e^{-\eta Y - \vartheta' Z - \vartheta' Z - \bar{\eta} Y - \bar{\vartheta}' Z}\right]}, \\
E_{\wp,\eta,\vartheta}[Z] &= \frac{E_{\bar{\eta},\bar{\vartheta}}\left[Z\wp^2(Y,Z)e^{-\eta Y - \vartheta' Z}\right]}{E_{\bar{\eta},\bar{\vartheta}}[\wp^2(Y,Z)e^{-\eta Y - \vartheta' Z}]}, \\
E_{\wp,\eta,\vartheta}[ZZ'] &= \frac{E_{\bar{\eta},\bar{\vartheta}}\left[ZZ'\wp^2(Y,Z)e^{-\eta Y - \vartheta' Z}\right]}{E_{\bar{\eta},\bar{\vartheta}}[\wp^2(Y,Z)e^{-\eta Y - \vartheta' Z}]}.
\end{aligned}$$

Proof Rewriting the expectation $E_{\wp,\eta,\vartheta}[Y]$ with the joint pdf $f_{\wp,\eta,\vartheta}(y,z)$ in Eq. (13) yields the result. □

In practice, we set the most updated tilting parameter as the values for $\bar{\eta}$ and $\bar{\vartheta}$. Furthermore, we use common random numbers to reduce computational expenses of drawing new random samples from the Gamma and Normal distributions. With the scaling properties of the Gamma distribution, it is well-known that $\left(\frac{1}{1-2\bar{\eta}}\right) \times \chi_\nu^2$ follows $\Gamma\left(\frac{\nu}{2}, \frac{2}{1-\bar{\eta}}\right)$. Similarly, with the shifting property of the Normal distribution, it is well-known that $\bar{\vartheta} + N_d(\mathbf{0}, \mathbb{I})$ follows $N_d(\bar{\vartheta}, \mathbb{I})$. Algorithm 1 summarizes steps to calculate the values of $E_{\wp,\eta,\vartheta}[Y]$, $E_{\wp,\eta,\vartheta}[Z]$, and $E_{\wp,\eta,\vartheta}[ZZ']$ by combining Corollary 1 and the idea of common random numbers described above.

Algorithm 1 The following steps calculate the values of $E_{\wp,\eta,\vartheta}[Y]$, $E_{\wp,\eta,\vartheta}[Z]$, and $E_{\wp,\eta,\vartheta}[ZZ']$. Given the sample size n, η, ϑ, $\bar{\eta}$, and $\bar{\vartheta}$, we do the following steps:

1. Generate the initial random samples, $Y^{(i,0)} \sim \Gamma(\nu/2, 2)$, $Z^{(i,0)} \sim N_d(\mathbf{0}, \mathbb{I})$, for $i = 1, \ldots, n$.
2. Calculate $E_{\wp,\eta,\vartheta}[Y]$, $E_{\wp,\eta,\vartheta}[Z]$, and $E_{\wp,\eta,\vartheta}[ZZ']$ with the following steps:

   (a) Set $Y^{(i)} = \left(\frac{1}{1-2\bar{\eta}}\right) Y^{(i,0)}$ and $Z^{(i)} = \bar{\vartheta} + Z^{(i,0)}$ for $i = 1, \ldots, n$.

   (b) Set $c = \frac{1}{n}\sum_{i=1}^{n} \wp^2\left(Y^{(i)}, Z^{(i)}\right) e^{-\eta Y^{(i)} - \vartheta' Z^{(i)} - \bar{\eta} Y^{(i)} - \bar{\vartheta}' Z^{(i)}}$.

   (c) Calculate $E_{\wp,\eta,\vartheta}[Y]$ by the average of n realized function values:

   $$\hat{E}_{\wp,\eta,\vartheta}[Y] = \frac{1}{cn}\sum_{i=1}^{n} Y^{(i)} \wp^2\left(Y^{(i)}, Z^{(i)}\right) e^{-\eta Y^{(i)} - \vartheta' Z^{(i)} - \bar{\eta} Y^{(i)} - \bar{\vartheta}' Z^{(i)}}.$$

   (d) Calculate $E_{\wp,\eta,\vartheta}[Z]$ by the average of n realized function values:

   $$\hat{E}_{\wp,\eta,\vartheta}[Z] = \frac{1}{cn}\sum_{i=1}^{n} Z^{(i)} \wp^2\left(Y^{(i)}, Z^{(i)}\right) e^{-\eta Y^{(i)} - \vartheta' Z^{(i)} - \bar{\eta} Y^{(i)} - \bar{\vartheta}' Z^{(i)}}.$$

   (e) Calculate $E_{\wp,\eta,\vartheta}[ZZ']$ by the average of n realized function values:

   $$\hat{E}_{\wp,\eta,\vartheta}[ZZ'] = \frac{1}{cn}\sum_{i=1}^{n} Z^{(i)} Z^{(i)'} \wp^2\left(Y^{(i)}, Z^{(i)}\right) e^{-\eta Y^{(i)} - \vartheta' Z^{(i)} - \bar{\eta} Y^{(i)} - \bar{\vartheta}' Z^{(i)}}.$$

Finally, to terminate the iterations in searching the optimal tilting parameters, we define the sum of squared errors (SSE) of the system of non-linear equations at the j-th iteration:

$$\mathrm{SSE}^{(j)} = \left(\frac{\nu}{1 - 2\eta^{(j+1)}} - E_{\wp,\eta^{(j)},\vartheta^{(j)}}[Y]\right)^2 + \left(\vartheta^{(j+1)} - E_{\wp,\eta^{(j)},\vartheta^{(j)}}[Z]\right)'\left(\vartheta^{(j+1)} - E_{\wp,\eta^{(j)},\vartheta^{(j)}}[Z]\right).$$

We stop the iterations when the SSE achieves a given tolerance δ. Now, we provide the stochastic fixed-point Newton method to search the optimal exponential tilting parameters in Algorithm 2.

Algorithm 2 The stochastic fixed-point Newton algorithm to search the optimal tilting parameters characterized in Theorem2 is implemented as follows. Given the sample size n and tolerance $\delta > 0$, we do the following steps:

1. Generate the initial random samples $Y^{(i,0)} \sim \Gamma(v/2, 2)$ and $Z^{(i,0)} \sim N_d(\mathbf{0}, \mathbb{I})$ for $i = 1, \ldots, n$.
2. Set $j = 0$, $\eta^{(j)} = 0$, $\vartheta^{(j)} = \mathbf{0}$. Do the following iterations to find the optimal tilting parameters:

   (a) Set $\eta = \eta^{(j)}$, $\vartheta = \vartheta^{(j)}$, $\bar{\eta} = \eta^{(j)}$, and $\bar{\vartheta} = \vartheta^{(j)}$ to calculate $E_{\wp,\eta,\vartheta}[Y]$, $E_{\wp,\eta,\vartheta}[Z]$, and $E_{\wp,\eta,\vartheta}[ZZ']$ using Algorithm1.

   (b) Update $\eta^{(j+1)} = \frac{1}{2}\left(1 - \frac{v}{E_{\wp,\eta,\vartheta}[Y]}\right)$.

   (c) Set $\eta = \eta^{(j+1)}$ and $\bar{\eta} = \eta^{(j+1)}$ to calculate $E_{\wp,\eta,\vartheta}[Y]$, $E_{\wp,\eta,\vartheta}[Z]$, and $E_{\wp,\eta,\vartheta}[ZZ']$ using Algorithm1.

   (d) Calculate $g = \vartheta^{(j)} - E_{\wp,\eta,\vartheta}[Z]$ and $J = \mathbb{I} - E_{\wp,\eta,\vartheta}[Z]E_{\wp,\eta,\vartheta}[Z]' + E_{\wp,\eta,\vartheta}[Z'Z]$ to update $\vartheta^{(j+1)} = \vartheta^{(j)} - J^{-1}g$.

   (e) Set $\vartheta = \vartheta^{(j+1)}$ and $\bar{\vartheta} = \vartheta^{(j+1)}$ to calculate $E_{\wp,\eta,\vartheta}[Y]$, $E_{\wp,\eta,\vartheta}[Z]$, and $E_{\wp,\eta,\vartheta}[ZZ']$ using Algorithm1.

   (f) Calculate

$$SSE^{(j)} = \left(\frac{v}{1 - 2\eta^{(j+1)}} - E_{\wp,\eta,\vartheta}[Y]\right)^2 + \left(\vartheta^{(j+1)} - E_{\wp,\eta,\vartheta}[Z]\right)'\left(\vartheta^{(j+1)} - E_{\wp,\eta,\vartheta}[Z]\right).$$

   (g) If $SSE^{(j)} > \delta$, set $j = j + 1$ and return to step 2 (a). Otherwise, stop the algorithm and set the optimal tilting parameters: $\eta^* = \eta^{(j+1)}$ and $\vartheta^* = \vartheta^{(j+1)}$.

3.4 The Methodological Workflow

Recall the stochastic representation of $X \sim t_{d,v}$ in Eq. (7) enables us to calculate $E[\wp(X)]$ by $E[\wp(Y, Z)]$, where $Z \sim N_d(0, \mathbb{I})$, $Y \sim \chi_v^2$, and Z and Y are independent. Now, we summarize the methodological workflow to calculate $E[\wp(X)]$ by $E[\wp(Y, Z)]$ as follows.

1. Searching stage. Follow Algorithm 2 to search the optimal tilting parameters η^* and ϑ^*.
2. Calculating stage. Follow Lemma 1 to evaluate $E[\wp(X)]$ by

$$\hat{\wp}_{\eta^*,\vartheta^*} = \frac{1}{n}\sum_{i=1}^{n} \wp(Y^{(i)}, Z^{(i)})e^{-\eta^* Y^{(i)} - \vartheta^{*\prime} Z^{(i)} - v\log(1-2\eta^{ast})/2 + \vartheta^{*\prime}\vartheta^*/2}, \tag{18}$$

   where realizations $Y^{(i)}$ and $Z^{(i)}$ are generated independently according to $Y^{(i)} \sim \Gamma\left(\frac{v}{2}, \frac{2}{1-2\eta^*}\right)$, and, $Z^{(i)} \sim N_d(\vartheta^*, \mathbb{I})$ for $i = 1, \ldots, n$.

4 The Rationale of the Proposed Importance Sampling Estimator

To illuminate the rationale of importance sampling, we first outline the study plan in Sect. 4.1. We then report the optimal tilting parameters and explain why an importance sampling estimator has a less variance than the standard estimator in Sect. 4.2. With the optimized tilting parameters, we show that the importance sampling estimator indeed has a smaller variance than the standard estimator in Sect. 4.3.

4.1 The Study Plan

Abbreviate standard estimator by

$$\wp_0 := \wp(Y, Z),$$

where $Y \sim \chi^2_v$, $Z \sim N_d(\mathbf{0}, \mathbb{I})$, and Y and Z are independent. And abbreviate the importance sampling estimator with optimal tilting parameters by

$$\wp_{\eta^*,\vartheta^*} := \wp_{\eta^*,\vartheta^*}(Y, Z) = \wp(Y, Z)e^{-\eta^* Y - \vartheta^{*\prime} Z - v \log(1-2\eta^*/2) + \vartheta^{*\prime}\vartheta^*/2},$$

where $Y \sim \Gamma\left(\frac{v}{2}, \frac{2}{1-2\eta^*}\right)$, $Z \sim N_d(\vartheta^*, \mathbb{I})$, and Y and Z are independent.

Recall the portfolio loss L in Eq. (6):

$$L = a_0 + b'X + X'\Lambda X,$$

where $X \sim t_{d,v}$. For simplicity, we set parameters: $v = 3$, $a_0 = 0$, $b_i = 0.1 + 0.01i$, and, $\lambda_i = 0.05i$ for $i = 1, \ldots, d$. We consider the following two cases:

1. The VaR-related value, $P(L > q)$.
2. The ES-related value, $E(LI_{\{L>q\}})$.

In addition, we consider $\alpha = 5\%, 1\%, 0.5\%, 0.1\%$ to meet practical requirements.

In our numerical experiments, we use a moderate sample size of 10, 000 in estimating $E_{\wp,\eta,\vartheta}[Y]$, $E_{\wp,\eta,\vartheta}[Z]$, $E_{\wp,\eta,\vartheta}[ZZ']$, and we use a large sample size of $n = 100,000$ to obtain the estimated mean and variance of $\wp_0$ and $\wp_{\eta^*,\vartheta^*}$ in order to obtain more accurate estimates. The computing time is reported in seconds. We set tolerance $\delta = 0.1$ to terminate Algorithm 2. We remark all simulation results through out this paper are carried out using Matlab (R2019a) in a desktop (Intel Core i9 CPU with 64 GB Ram).

4.2 Rationale Behind the Importance Sampling Method

To illuminate the rationale of the proposed importance sampling scheme, we consider a simplified case of $d = 2$, and express the portfolio loss L explicitly as a quadratic function of t-distributed risk-factors:

$$L = 0.1X_1 + 0.11X_2 + 0.05X_1^2 + 0.1X_2^2.$$

To start, we first estimate the threshold q satisfying $P(L > q) = \alpha$ for $\alpha = 5\%, 1\%, 0.5\%, 0.1\%$ using simulation. The contour plot of L is depicted in Fig. 1. Because L is a quadratic function of x_1 and x_2, the contour of L forms an ellipse, and the support set of $\{(x_1, x_2) : L > q\}$ is outside of the corresponding ellipse.

Table 1 summarizes the α, q, and the optimal tilting parameters η^* and ϑ^* in calculating $P(L > q)$ and $E(LI_{\{L>q\}})$. Because η^* and ϑ^* are not zeros, the importance sampling estimator is different from the standard estimator. To contrast the importance sampling estimator with the standard estimator, we outline two major observations in Table 1: First, ϑ^* are not zeros. Second, a smaller α results in a more negative value of η^*.

Recall the stochastic representation of $X \stackrel{d}{=} Z/\sqrt{Y/v}$ in Eq. (7): A realization of X is obtained by combining a realization of Y and a realization of Z. Realizations of Z for the standard estimator are generated from $N_2(\mathbf{0}, \mathbb{I})$, whereas realizations of Z for the importance sampling estimator are generated from $N_(\vartheta^*, \mathbb{I})$. Compared with the later case, for the importance sampling estimaor, there are more realizations of Z close to the origin, and hence less realizations of X falling within the support set $\{L > q\}$. In other words, for the importance sampling estimator with non-zero ϑ^*, there will be more realizations of Z far way from the origin, and hence more realizations of X falling within the support set of $\{L > q\}$.

To see how η^* affect the importance sampling estimator, Fig. 2 depicts the probability density function of $Y \sim \Gamma\left(\frac{v}{2}, \frac{2}{1-2\eta^*}\right)$. Because Y is in the denominator of the stochastic representation of X in Eq. (7), a smaller realization of Y results in a realization of X far away from the origin. With a more negative value of η^*, there will

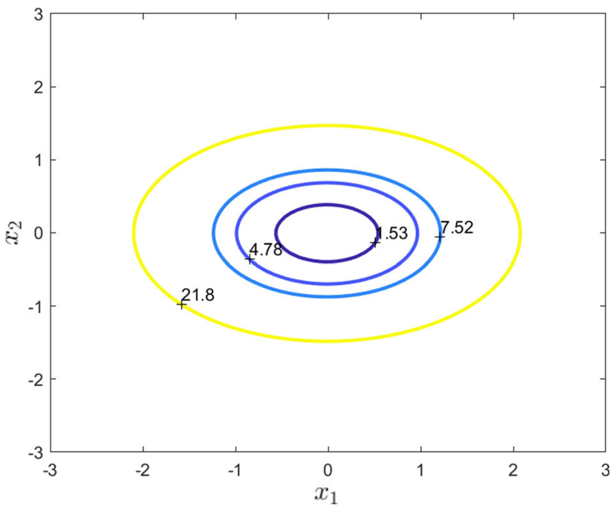

Fig. 1 The contour plot of $L(x_1, x_2) = 0.1x_1 + 0.11x_2 + 0.05x_1^2 + 0.1x_2^2$, where the contour lines are specified at $P(L > q) = \alpha$ for $\alpha = 5\%, 1\%, 0.5\%, 0.1\%$

Table 1 The optimal titling parameters η^* and θ^* of calculating $P(L > q)$ and $E(LI_{\{L>q\}})$ with q satisfying $P(L > q) = \alpha$ for $d = 2$

Case	α (%)	q	η^*	ϑ_1^*	ϑ_2^*
$P(L > q)$					
	5	1.53	−0.4	0.20	0.35
	1	4.78	−4.9	0.07	0.10
	0.5	7.52	−18.5	0.22	−0.23
	0.1	21.78	−28.3	0.02	0.02
$E(LI_{\{L>q\}})$					
	5	1.53	−7.8	0.21	−0.25
	1	4.78	−4.1	−0.19	0.86
	0.5	7.52	−15.5	−0.04	−0.03
	0.1	21.78	−99.4	0.05	−0.14

be more realizations of Y for the importance sampling estimator close to zero, and hence more realizations of X far away from the origin and falling within the support set of $\{L > q\}$.

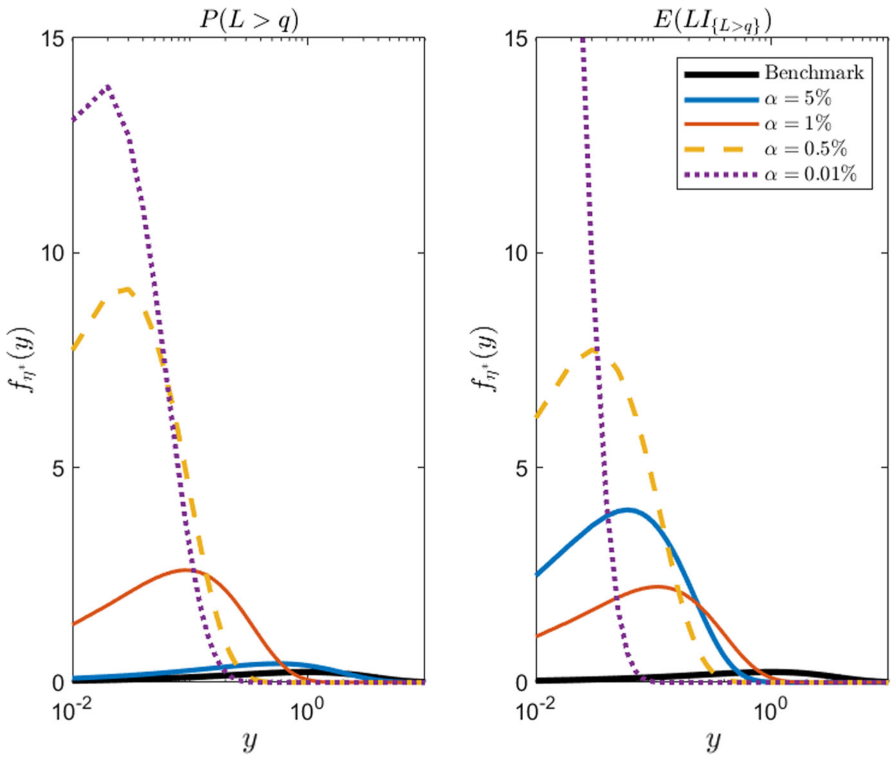

Fig. 2 The probability density plot of $Y \sim \Gamma\left(\frac{\nu}{2}, \frac{2}{1-2\eta^*}\right)$, where η^* is the optimal tilting parameter of calculating $P(L > q)$ and $E(LI_{\{L>q_\alpha\}})$ for $\alpha = 5\%, 1\%, 0.5\%, 0.1\%$

4.3 The Estimated Mean and Variance of $\wp_0$ and $\wp_{\eta^*,\vartheta^*}$

Table 2 compares the estimated mean and variance of $\wp_0$ and $\wp_{\eta^*,\vartheta^*}$ in calculating $P(L > q)$ and $E(LI_{\{L>q\}})$ for $\alpha = 5\%, 1\%, 0.5\%, 0.1\%$ and. To compare if the means of $\wp_0$ and $\wp_{\eta^*,\vartheta^*}$, we conduct an independent two-sample t test:

$$\begin{cases} H_0 : E(\wp_0) = E(\wp_{\eta^*,\vartheta^*}), \\ H_1 : E(\wp_0) \neq E(\wp_{\eta^*,\vartheta^*}). \end{cases}$$

The test statistic is

$$t_{stat} = \frac{\hat{\wp}_0 - \hat{\wp}_{\eta^*,\vartheta^*}}{\sqrt{\frac{\hat{\mathrm{V}}(\wp_0)+\hat{\mathrm{V}}(\wp_{\eta^*,\vartheta^*})}{n}}}.$$

With Eq. (10), the estimated variance of $\wp_0$ is

$$\hat{\mathrm{V}}(\wp_0) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\wp(Y^{(i)}, Z^{(i)}) - \hat{\wp}_0\right)^2.$$

Similarly, with Eq. (18), the estimated variance of $\wp_{\eta^*,\vartheta^*}$ is

$$\hat{\mathrm{V}}(\wp_{\eta^*,\vartheta^*}) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\wp(Y^{(i)}, Z^{(i)})e^{-\eta^* Y^{(i)} - \vartheta^{*\prime} Z^{(i)} - v\log(1-2\eta^{ast})/2 + \vartheta^{*\prime}\vartheta^*/2} - \hat{\wp}_{\eta^*,\vartheta^*}\right)^2.$$

The t-statistics in Table 2 indicates that the standard estimator appears to produce the same expectation as the importance sampling estimator.

Table 2 The estimated mean and variance of calculating $P(L > q)$ and $E(LI_{\{L>q\}})$ using $\wp_0$ and $\wp_{\eta^*,\vartheta^*}$ with q satisfying $P(L > q) = \alpha$, and the t and F statistics for comparing means and variances between $\wp_0$ and $\wp_{\eta^*,\vartheta^*}$ for $d = 2$

Case	α (%)	Mean		t_{stat}	Variance		F_{stat}
		$\wp_0$	$\wp_{\eta^*,\vartheta^*}$		$\wp_0$	$\wp_{\eta^*,\vartheta^*}$	
$P(L > q)$							
	5	0.0496	0.0487	1.1	4.7×10^{-2}	2.6×10^{-2}	1.8^{***}
	1	0.0100	0.0099	0.3	9.9×10^{-3}	7.0×10^{-4}	14.1^{***}
	0.5	0.0052	0.0052	0.0	5.1×10^{-3}	6.0×10^{-4}	8.5^{***}
	0.1	0.0012	0.0011	0.9	1.2×10^{-3}	1.0×10^{-5}	117.3^{***}
$E(LI_{\{L>q\}})$							
	5	0.2481	0.2421	0.3	37.7	5.6	6.8^{***}
	1	0.1698	0.1421	1.6	30.1	1.5	20.6^{***}
	0.5	0.1040	0.1160	−1.0	12.6	0.5	22.9^{***}
	0.1	0.0593	0.0674	−0.9	7.8	0.4	18.1^{***}

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Furthermore, we conduct a standard right-sided F test for comparing variances:

$$\begin{cases} H_0 : \mathrm{V}(\wp_0) \leq \mathrm{V}(\wp_{\eta^*,\vartheta^*}), \\ H_1 : \mathrm{V}(\wp_0) > \mathrm{V}(\wp_{\eta^*,\vartheta^*}). \end{cases}$$

The F statistic is

$$F_{stat} = \frac{\hat{\mathrm{V}}(\wp_0)}{\hat{\mathrm{V}}(\wp_{\eta^*,\vartheta^*})}.$$

From Table 2, we conclude that the importance sampling estimator has a smaller variance than the standard estimator. As an interesting note, the realized F-statistic equal the ratio between the estimated variances of $\wp_0$ and $\wp_{\eta^*,\vartheta^*}$. Details of the independent two-sample t test and right-sided F test can be found in a standard statistics textbook (Mendenhall et al., 2012).

5 Simulation Studies

Sect. 5.1 outlines the study plan, and Sect. 5.2 reports the estimated mean and variance of the standard and importance sampling estimators in calculating $P(L > q)$ and $E(L > q)$ for large dimensions. For practical comparisons, Sect. 5.3 calculates the computing time saved in percentage of the importance sampling estimator.

5.1 The Study Plan for Large Dimensions

To investigate if the importance sampling estimator works for large dimensions, we follow and extend the study plan in Sect. 4.1 for $d = 15, 30, 45$. When implementing the importance sampling estimator, we need to search the optimal tilting parameter of η and ϑ. Because ϑ is of d dimensions, in our experiences, its searching procedure is time consuming and unstable. To overcome these numerical issues, we extend Teng et al. (2016) to consider a linear specification on ϑ as follows.

Define a vector $\beta = (\beta_1, \beta_2)'$ and a $(d \times 2)$ matrix H:

$$H = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & (d-1) \end{bmatrix}.$$

Then, the linear specification for the tilting parameter ϑ is

$$\begin{aligned} \vartheta &= (\vartheta_1, \ldots, \vartheta_d)' \\ &= (\beta_1, \beta_1 + \beta_2, \ldots, \beta_1 + (d-1)\beta_2)' = H\beta. \end{aligned} \tag{19}$$

Here, β_1 refers to the first titling parameter of ϑ, and β_2 is a constant increment between consecutive tilting parameters in ϑ. As implicated from Lemma 1, the tilting parameter ϑ stands for the mean shift for the Normal distribution. Hence, there are no constrains on ϑ, and subsequently, no constrains on β. With this linear specification, we only need to search two tilting parameters, rather than the original d tilting parameters.

Now, the variance of the importance sampling estimator in Eq. (12) becomes

$$G(\eta, H\beta) = E[\wp^2(Y,Z)e^{-\eta Y - H\beta Z - v\log(1-2\eta)/2 + \beta' H' H\beta/2}].$$

Eq. (15) becomes

$$H'H\beta = H'E_{\wp,\eta,H\beta}[Z].$$

In addition, $g(\vartheta)$ in Eq. (16) is changed to

$$g(H\beta) = H'H\beta - H'E_{\wp,\eta,H\beta}[Z],$$

and the Jocobian in Eq. (17) is changed to

$$J_\beta = H'H - H'E_{\wp,\eta,H\beta}[Z]E_{\wp,\eta,H\beta}[Z]'H + H'E_{\wp,\eta,H\beta}[ZZ']H.$$

Thus, the optimal tilting parameter, β^*, can be searched in a similar manner:

$$\beta^{(j+1)} = \beta^{(j)} - J^{-1}_{\beta^{(j)}} g(H\beta^{(j)}).$$

We modify Algorithm 2 to search the optimal tilting parameters η^* and β^* in Algorithm 3.

Algorithm 3 The stochastic fixed-point Newton algorithm to search the optimal tilting parameters characterized in Theorem 2 with the linear specification in Eq. (19) is implemented as follows. Given the sample size n and tolerance $\delta > 0$, we do the following steps:

1. Generate the initial random samples $Y^{(i,0)} \sim \Gamma(v/2, 2)$ and $Z^{(i,0)} \sim N_d(\mathbf{0}, \mathbb{I})$ for $i = 1, \ldots, n$.
2. Set $j = 0$, $\eta^{(j)} = 0$, $\beta_1^{(j)} = 0$, $\beta_2^{(j)} = 0$.
3. Set $\vartheta^{(j)} = H\beta^{(j)}$.
4. Do the following iterations to find the optimal tilting parameters α^* and β^*:
   (a) Set $\eta = \eta^{(j)}$, $\vartheta = \vartheta^{(j)}$, $\bar{\eta} = \eta^{(j)}$, and $\bar{\vartheta} = \vartheta^{(j)}$ to calculate $E_{\wp,\eta,\vartheta}[Y]$, $E_{\wp,\eta,\vartheta}[Z]$, and $E_{\wp,\eta,\vartheta}[ZZ']$ using Algorithm 1.
   (b) Update $\eta^{(j+1)} = \frac{1}{2}\left(1 - \frac{v}{E_{\wp,\eta,\vartheta}[Y]}\right)$.
   (c) Set $\eta = \eta^{(j+1)}$ and $\bar{\eta} = \eta^{(j+1)}$ to calculate $E_{\wp,\eta,\vartheta}[Y]$, $E_{\wp,\eta,\vartheta}[Z]$, and $E_{\wp,\eta,\vartheta}[ZZ']$ using Algorithm 1.
   (d) Calculate $g = H'H\beta^{(j)} - H'E_{\wp,\eta,H\beta}[Z]$ and $J = H'H - H'E_{\wp,\eta,\vartheta}[Z]E_{\wp,\eta,\vartheta}[Z]'H + H'E_{\wp,\eta,\vartheta}[Z'Z]H$ to update $\beta^{(j+1)} = \beta^{(j)} - J^{-1}g$.
   (e) Set $\vartheta = H\beta^{(j+1)}$ and $\bar{\vartheta} = \vartheta^{(j+1)}$ to calculate $E_{\wp,\eta,\vartheta}[Y]$, $E_{\wp,\eta,\vartheta}[Z]$, and $E_{\wp,\eta,\vartheta}[ZZ']$ using Algorithm 1.
   (f) Calculate

$$SSE^{(j)} = \left(\frac{v}{1-2\eta^{(j+1)}} - E_{\wp,\eta,\vartheta}[Y]\right)^2 + \left(\vartheta^{(j+1)} - E_{\wp,\eta,\vartheta}[Z]\right)'\left(\vartheta^{(j+1)} - E_{\wp,\eta,\vartheta}[Z]\right).$$

(g) If $SSE^{(j)} > \delta$, set $j = j+1$ and return to step 3 (a). Otherwise, stop the algorithm and set the optimal tilting parameters: $\eta^* = \eta^{(j+1)}$ and $\vartheta^* = \vartheta^{(j+1)}$.

5.2 The Estimated Mean and Variance

We first estimate the threshold q satisfying $P(L > q) = \alpha$ for $\alpha = 5\%, 1\%, 0.5\%, 0.1\%$ and $d = 15, 30, 45$ using simulation. The optimal tilting parameters are listed in Table 3, which are not zeros in all cases. As a result, the importance sampling estimator is different from the standard estimator. Similar to the observation in Sect. 4.2, a smaller α results in a more negative value of η^*.

With the optimal tilting parameters, Table 4 compares the estimated mean and variance of the standard estimator and the importance sampling estimator in calculating $P(L > q)$ and $E(LI_{\{L>q\}})$ for $\alpha = 5\%, 1\%, 0.5\%, 0.1\%$ and $d = 15, 30, 45$. In calculating $P(L > q)$, the t test indicates that the standard estimator appears to produce the same expectation as the importance sampling estimator, and the F test indicates that the importance sampling estimator has a smaller variance than the standard estimator.

In calculating $E[LI_{\{L>q\}}]$, the t test however indicates a significant difference between the means of the standard and importance sampling estimators in a few cases. This is possibly because of the inaccurate estimated variances for the standard estimator. (As a note, Table 6 suggests to use a sample size of more than a million to produce accurate estimate for the mean of the standard estimator. Hence, to produce accurate estimated variance of the standard estimator, the current sample size 100,000 seems to be inadequate.) The F test indicates that the importance sampling estimator has a smaller variance than the standard estimator under the significance level, 5%.

5.3 The Computing Time

Table 5 compares the computing time of the standard and importance sampling estimators in calculating $P(L > q)$ and $E(LI_{\{L>q\}})$. For the importance sampling estimator, we split the computing time into a sum of the searching time for finding the optimal tilting parameters, and the calculating time to calculate the sample mean. In both cases, the computing time for the importance sampling estimator is longer than the standard estimator: The calculating time of the importance sampling estimator is slightly longer than the standard estimator, because it needs to deal with the additional importance sampling weight; However, the searching time for the importance sampling estimator does not exceed its calculating time.

Table 3 The optimal titling parameters η^* and θ^* of calculating $P(L > q)$ and $E(LI_{\{L>q\}})$ with q satisfying $P(L > q) = \alpha$ for $d = 15, 30, 45$

Case	d	α (%)	q	η^*	β_1^*	β_2^*
$P(L > q)$						
	15	5	53	−3.7	0.0193	−0.0003
		1	162	−12.3	−0.0035	0.0010
		0.5	259	−18.6	0.0034	0.0014
		0.1	763	−63.9	−0.0076	0.0018
	30	5	200	−4.4	0.0065	−0.0004
		1	600	−14.0	−0.0024	0.0005
		0.5	980	−22.6	−0.0048	0.0005
		0.1	2800	−66.3	0.0111	−0.0003
	45	5	438	−4.7	0.0005	0.0001
		1	1385	−15.1	0.0016	0.0001
		0.5	2185	−25.0	−0.0039	0.0001
		0.1	6285	−65.1	−0.0012	0.0003
$E(LI_{\{L>q\}})$						
	15	5	53	−14.2	−0.0497	−0.0033
		1	162	−33.3	−0.1429	0.0187
		0.5	259	−468.8	−0.1356	0.0177
		0.1	763	−257.4	−0.1118	0.0135
	30	5	200	−17.0	−0.1666	0.0122
		1	600	−17.3	−0.0202	0.0009
		0.5	980	−549.2	−0.0889	0.0114
		0.1	2800	−351.8	−0.0238	0.0005
	45	5	438	−49.1	−0.0657	0.0043
		1	1385	−132.7	−0.0286	0.0038
		0.5	2185	−183.3	−0.0221	0.0021
		0.1	6285	−806.0	0.0549	−0.0004

We define the time ratio as

$$\text{Time ratio} = \frac{\text{Computing time in implementing } \wp_{\eta,\vartheta}}{\text{Computing time in implementing } \wp_0}.$$

The importance sampling estimator takes about 1.2 to 1.6 (respectively, 1.5 to 1.8) times more than the standard estimator in calculating $P(L > q)$ (respectively, $E(LI_{\{L>q\}})$).

To address practical advantages in applying the importance sampling estimator, we report the minimum sample size and minimum computing time to achieve a predetermined precision level for each estimator. Let $\tilde{\wp}$ denote an estimator of interest. Recall that the margin of error of the $100(1-\alpha)\%$ confidence interval is $z_{\alpha/2}\sqrt{\mathrm{V}(\tilde{\wp})/n}$, where $z_{\alpha/2}$ is the upper quantile for the standard Normal distribution satisfying $P(Z > z_{\alpha/2}) = \alpha/2$.

Table 4 The estimated mean and variance of calculating $P(L > q)$ and $E(LI_{\{L>q\}})$ using $\wp_0$ and $\wp_{\eta^*,\vartheta^*}$ with q satisfying $P(L > q) = \alpha$, and the t and F statistics for comparing means and variances for $d = 15, 30, 45$

Case	d	α (%)	Mean		t_{stat}	Variance		F_{stat}
			$\wp_0$	$\wp_{\eta,\vartheta}$		$\wp_0$	$\wp_{\eta,\vartheta}$	
$P(L > q)$								
	15	5	0.0508	0.0499	1.2	4.8×10^{-2}	3.7×10^{-3}	13.1^{***}
		1	0.0101	0.0101	0.0	1.0×10^{-2}	1.7×10^{-4}	58.3^{***}
		0.5	0.0051	0.0050	0.0	5.0×10^{-3}	4.5×10^{-5}	111.6^{***}
		0.1	0.0010	0.0010	−0.4	9.7×10^{-4}	1.8×10^{-6}	541.3^{***}
	30	5	0.0507	0.0502	0.7	4.8×10^{-2}	2.7×10^{-3}	17.8^{***}
		1	0.0109	0.0105	1.3	1.1×10^{-2}	1.3×10^{-4}	81.4^{***}
		0.5	0.0050	0.0051	−0.5	4.9×10^{-3}	3.1×10^{-5}	157.1^{***}
		0.1	0.0011	0.0011	0.3	1.1×10^{-3}	1.4×10^{-6}	766.5^{***}
	45	5	0.0509	0.0514	−0.7	4.8×10^{-2}	2.5×10^{-3}	19.0^{***}
		1	0.0099	0.0099	−0.1	9.8×10^{-3}	1.0×10^{-4}	94.9^{***}
		0.5	0.0049	0.0051	−0.7	4.9×10^{-3}	2.7×10^{-5}	178.2^{***}
		0.1	0.0010	0.0010	−0.8	9.6×10^{-4}	1.2×10^{-6}	803.6^{***}
$E(LI_{\{L>q\}})$								
	15	5	8.2	8.1	0.2	2.5×10^{4}	6.2×10^{2}	41.2^{***}
		1	4.3	2.8	4.2^{***}	1.1×10^{4}	2.8×10^{3}	4.0^{***}
		0.5	3.8	3.9	−0.2	2.6×10^{4}	1.8×10^{2}	146.0^{***}
		0.1	2.8	2.4	0.6	4.4×10^{4}	4.7×10^{2}	94.0^{***}
	30	5	31.8	32.1	−0.2	1.6×10^{5}	5.7×10^{4}	2.7^{***}
		1	18.3	19.2	−0.6	2.3×10^{5}	3.7×10^{3}	62.1^{***}
		0.5	13.1	14.8	−1.8	8.5×10^{4}	5.9×10^{3}	14.5^{***}
		0.1	9.2	5.0	3.1^{***}	1.8×10^{5}	5.3×10^{3}	33.6^{***}
	45	5	63.7	70.1	-3.4^{***}	3.1×10^{5}	5.3×10^{4}	5.8^{***}
		1	37.3	17.2	5.1^{***}	1.4×10^{6}	9.6×10^{4}	14.9^{***}
		0.5	35.2	33.5	0.2	5.0×10^{6}	2.7×10^{4}	182.7^{***}
		0.1	18.7	19.5	−0.3	1.0×10^{6}	6.0×10^{3}	170.7^{***}

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

We require the margin of error for the $100(1-\alpha)\%$ confidence interval to be less than a multiple of ε and its estimated mean. Specifically, the sample size n needs to satisfy

$$z_{\alpha/2}\sqrt{\frac{\mathrm{V}(\tilde{\wp})}{n}} \leq \varepsilon E[\tilde{\wp}].$$

Equivalently, the above inequality can be simplified as

Table 5 The computing time and time ratio of calculating $P(L > q)$ and $E(LI_{\{L>q\}})$ using $\wp_0$ and $\wp_{\eta^*,\vartheta^*}$ with q satisfying $P(L > q) = \alpha$ for $d = 15, 30, 45$. The computing time of $\wp_{\eta^*,\vartheta^*}$ is split into a sum of searching time (1) and calculating time (2)

Case	d	α (%)	$\wp_0$	$\wp_{\eta,\vartheta}$			$\frac{\text{time}(\wp_{\eta,\vartheta})}{\text{time}(\wp_0)}$
			Time	(1)	(2)	Time	
$P(L > q)$							
	15	5	3	0.5	3.3	3.8	1.2
		1	3.1	1.4	3.1	4.5	1.5
		0.5	2.9	0.5	3.1	3.6	1.2
		0.1	2.9	1.4	3.2	4.6	1.5
	30	5	5	0.8	5.6	6.4	1.3
		1	4.9	0.9	5.4	6.3	1.3
		0.5	5.2	0.8	5.6	6.4	1.2
		0.1	5.1	2.7	5.4	8.1	1.6
	45	5	7.2	1.5	7.6	9.1	1.3
		1	7.1	1.7	7.5	9.2	1.3
		0.5	7.1	1.3	7.3	8.6	1.2
		0.1	7.1	1.2	7.4	8.6	1.2
$E(LI_{\{L>q\}})$							
	15	5	3.2	1.4	3.3	4.7	1.5
		1	3.0	1.4	3.2	4.6	1.6
		0.5	3.0	1.4	3.2	4.6	1.5
		0.1	2.9	1.4	3.2	4.6	1.6
	30	5	3.8	2.5	4.0	6.5	1.7
		1	3.6	2.3	4.0	6.3	1.8
		0.5	3.7	2.3	3.9	6.2	1.7
		0.1	3.6	2.3	3.9	6.2	1.8
	45	5	5.1	3.4	5.5	8.9	1.7
		1	5.0	3.3	5.4	8.7	1.7
		0.5	5.1	3.3	5.4	8.7	1.7
		0.1	5.0	3.3	5.3	8.6	1.7

$$n \geq \frac{z_{\alpha/2}^2 \mathrm{V}(\tilde{\wp})}{\varepsilon^2 E[\tilde{\wp}]^2}.$$

Therefore, the minimum sample size is the least positive integer satisfying the above equation. Theoretical values of $E[\tilde{\wp}]^2$ and $\mathrm{V}(\tilde{\wp})$ are unknown, and are approximated using simulation. We set $\alpha = 5\%$ and $\varepsilon = 0.1$ to meet practical requirements.

Table 6 reports the minimum sample size of the standard and importance sampling estimators in calculating $P(L > q)$ and $E(L > q)$. For both cases, the importance sampling estimator requires a smaller minimum sample size than the standard estimator. We define the minimum size ratio as

Table 6 The estimated minimum sample size, mean, variance, and standard error of calculating $P(L > q)$ and $E(LI_{\{L>q\}})$ using $\wp_0$ and $\wp_{\eta^*,\vartheta^*}$ with q satisfying $P(L > q) = \alpha$ for $d = 15, 30, 45$

Case	d		Size		Mean		Variance		Standard error	
		α (%)	$\wp_0$	$\wp_{\eta,\vartheta}$	$\wp_0$	$\wp_{\eta,\vartheta}$	$\wp_0$	$\wp_{\eta,\vartheta}$	$\wp_0$	$\wp_{\eta,\vartheta}$
$P(L > q)$										
	15	5	63,961	4,897	0.0508	0.0505	4.8×10^{-2}	3.9×10^{-3}	8.7×10^{-4}	8.9×10^{-4}
		1	66,205	1,135	0.0101	0.0098	1.0×10^{-2}	1.4×10^{-4}	3.9×10^{-4}	3.5×10^{-4}
		0.5	66,673	597	0.0047	0.0052	4.7×10^{-3}	3.6×10^{-5}	2.7×10^{-4}	2.5×10^{-4}
		0.1	64,295	119	0.0008	0.0012	7.8×10^{-4}	1.9×10^{-6}	1.1×10^{-4}	1.3×10^{-4}
	30	5	63,890	3,593	0.0495	0.0505	4.7×10^{-2}	3.1×10^{-3}	8.6×10^{-4}	9.3×10^{-4}
		1	71,661	881	0.0110	0.0098	1.1×10^{-2}	1.3×10^{-4}	3.9×10^{-4}	3.8×10^{-4}
		0.5	65,491	417	0.0049	0.0049	4.9×10^{-3}	3.4×10^{-5}	2.7×10^{-4}	2.8×10^{-4}
		0.1	72,903	95	0.0009	0.0009	8.8×10^{-4}	7.6×10^{-7}	1.1×10^{-4}	9.0×10^{-5}
	45	5	64,093	3,381	0.0497	0.0501	4.7×10^{-2}	2.3×10^{-3}	8.6×10^{-4}	8.3×10^{-4}
		1	64,709	682	0.0099	0.0095	9.8×10^{-3}	9.1×10^{-5}	3.9×10^{-4}	3.7×10^{-4}
		0.5	64,703	363	0.0047	0.0049	4.7×10^{-3}	2.7×10^{-5}	2.7×10^{-4}	2.7×10^{-4}
		0.1	63,633	79	0.0011	0.0010	1.1×10^{-3}	9.7×10^{-7}	1.3×10^{-4}	1.1×10^{-4}
$E(LI_{\{L>q\}})$										
	15	5	2,068,250	50,252	8.2	8.3	7.1×10^{4}	1.1×10^{3}	1.8×10^{-1}	1.5×10^{-1}
		1	1,696,810	428,145	4.8	4.9	2.8×10^{4}	3.5×10^{2}	1.3×10^{-1}	2.9×10^{-2}
		0.5	4,518,024	30,951	3.8	3.8	3.7×10^{4}	1.3×10^{2}	9.0×10^{-2}	6.5×10^{-2}
		0.1	10,310,439	109,722	2.2	2.2	3.3×10^{4}	2.3×10^{2}	5.7×10^{-2}	4.5×10^{-2}
	30	5	3,243,262	1,181,676	32.1	31.9	6.7×10^{5}	3.9×10^{5}	4.5×10^{-1}	5.7×10^{-1}
		1	7,934,635	127,795	18.9	18.6	8.7×10^{5}	1.5×10^{3}	3.3×10^{-1}	1.1×10^{-1}
		0.5	3,805,317	262,584	14.5	15.1	3.2×10^{5}	4.2×10^{3}	2.9×10^{-1}	1.3×10^{-1}

Table 6 continued

Case	d		Size		Mean		Variance		Standard error	
		α (%)	$\wp_0$	$\wp_{\eta,\vartheta}$	$\wp_0$	$\wp_{\eta,\vartheta}$	$\wp_0$	$\wp_{\eta,\vartheta}$	$\wp_0$	$\wp_{\eta,\vartheta}$
		0.1	23,368,497	695,541	8.7	9.0	5.9×10^{5}	2.9×10^{3}	1.6×10^{-1}	6.5×10^{-2}
	45	5	3,205,655	551,491	69.7	70.5	3.0×10^{6}	5.4×10^{4}	9.8×10^{-1}	3.1×10^{-1}
		1	25,504,462	1,706,842	41.0	17.7	3.1×10^{6}	4.5×10^{5}	3.5×10^{-1}	5.1×10^{-1}
		0.5	93,965,988	514,372	33.5	33.5	2.3×10^{7}	5.1×10^{4}	4.9×10^{-1}	3.1×10^{-1}
		0.1	36,371,568	213,015	19.2	19.6	6.4×10^{6}	4.9×10^{3}	4.2×10^{-1}	1.5×10^{-1}

Table 7 The computing time and computing time saved in percentage (Save (%)) of calculating $P(L > q)$ and $E(LI_{\{L>q\}})$ with the minimum sample size using $\wp_0$ and $\wp_{\eta^*,\vartheta^*}$ with q satisfying $P(L > q) = \alpha$. The computing time of $\wp_{\eta^*,\vartheta^*}$ is split into a sum of searching time (1) and calculating time (2)

Case	d	α (%)	$\wp_0$	$\wp_{\eta,\vartheta}$			Save (%)
			Time	(1)	(2)	Time	
$P(L > q)$							
	15	5	2.6	0.7	0.2	0.9	65.2
		1	2.6	0.6	0.0	0.7	73.3
		0.5	2.6	0.5	0.0	0.5	79.4
		0.1	2.5	1.6	0.0	1.6	37.0
	30	5	3.0	0.7	0.2	0.9	70.7
		1	3.4	2.7	0.0	2.7	20.5
		0.5	3.1	0.8	0.0	0.8	74.1
		0.1	3.4	0.8	0.0	0.8	77.3
	45	5	4.2	1.2	0.2	1.4	66.6
		1	4.2	1.2	0.0	1.3	70.1
		0.5	4.3	0.9	0.0	0.9	79.6
		0.1	4.2	1.1	0.0	1.1	74.0
$E(LI_{\{L>q\}})$							
	15	5	61.4	1.4	1.6	3.1	95.0
		1	50.3	1.4	13.5	14.8	70.5
		0.5	132.6	1.4	1.0	2.3	98.2
		0.1	303.8	1.3	3.5	4.8	98.4
	30	5	116.0	2.5	48.5	51.0	56.1
		1	284.9	2.4	4.9	7.3	97.4
		0.5	136.9	2.3	10.1	12.4	90.9
		0.1	838.2	2.5	28.4	31.0	96.3
	45	5	162.1	3.4	31.2	34.6	78.7
		1	1289.2	3.4	91.7	95.0	92.6
		0.5	4458.5	3.2	26.1	29.4	99.3
		0.1	1726.8	3.6	12.1	15.7	99.1

$$\text{minimum size ratio} = \frac{\text{minimum sample size of } \wp_0}{\text{minimum size of } \wp_{\eta,\vartheta}},$$

Standard algebra approximates the minimum size ratio,

$$\text{minimum size ratio} \approx \frac{\frac{z_{\alpha/2}^2 \mathrm{V}(\wp_0)}{\varepsilon^2 E[\wp_0]^2}}{\frac{z_{\alpha/2}^2 \mathrm{V}(\wp_{\eta,\vartheta})}{\varepsilon^2 E[\wp_{\eta,\vartheta}]^2}} \approx \frac{\mathrm{V}(\wp_0)}{\mathrm{V}(\wp_{\eta,\vartheta})},$$

which is equal to the ratio of the estimated variances of $\wp_0$ and $\wp_{\eta^*,\vartheta^*}$. In other words, the ratio of estimated variances between two estimators can be interpreted as the ratio of their minimum sample sizes to achieve the same precision. It is also interesting to note that the minimum sample size depends on α and ε, but the minimum size ratio is irrelevant of them.

With the minimum sample size, we implement each estimator in calculating $P(L > q)$ and $E[LI_{\{L>q\}}]$. Table 6 also records the estimated mean, variance, standard error, and computing time. We note the minimum computing time would be a realistic measure to evaluate the superiority among the two estimators. For both cases, the expected means of the standard and importance sampling estimators are about the same, and importance sampling estimator has a less variance.

To see if the minimum sample size is valid, we report the standard error:

$$\sqrt{\frac{\mathrm{V}(\tilde{\wp})}{\text{minimum sample size of } \tilde{\wp}}}.$$

The standard error of an arbitrary estimator is proportional to the width of its confidence interval. The standard errors of the standard and importance sampling estimators are of about the same, and hence the validity of the minimum sample size is numerically verified.

Table 7 compares the computing time for the standard and importance sampling estimators. The standard estimator in fact requires substantially longer minimum computing time in calculating $P(L > q)$ and $E(LI_{\{L>q\}})$ to achieve the precision level. Although the importance sampling estimator requires additional searching time, the need of less minimum sample size reduces its calculating time and results in a substantially shorter computing time. To furthermore evaluate the computing time saved by importance sampling, we calculate the computing time saved in percentage:

$$\left(\frac{\text{minimum computing time of } \wp_0 - \text{minimum computing time of } \wp_{\eta,\vartheta}}{\text{minimum computing time of } \wp_0}\right) \times 100\%.$$

In calculating $P(L > q)$, about 20% to 80% computing time is saved with the using importance sampling. In calculating $E(LI_{\{L>q\}})$, the standard estimator unfortunately needs formidably long computing time, and about 56% to 99% computing time is saved with the importance sampling estimator. The proposed importance sampling estimator enjoys a substantial saving in computing time, particularly in calculating $E(LI_{\{L>q\}})$!

6 Conclusion

This paper proposes an efficient importance sampling scheme for calculating the expectation of a general function for the t distribution. The importance sampling estimator is successfully applied to calculate the VaR and ES, which are critical measures for financial risk management. Our numerical experiments show that the proposed importance sampling estimator has a smaller variance compared with the standard estimator. With the proposed importance sampling estimator, substantial computing time is saved: about 20% to 80% in calculating $P(L > q)$ and about 56% to 99% computing time in calculating $E[LI_{\{L>q\}}]$.

The proposed method could be extended to other complex random vector, such as copulas. And, it is definitely worthy of further investigations if the importance sampling estimator could benefit other applications, particularly in the realm of machine learning and AI (Zhao & Zhang, 2015; Renand et al., 2019).

Appendix A: Proof of Lemma 1 To start, suppose $Y \sim \chi^2_v$. Then, Y has the probability density function (pdf)

$$f_Y(y) = \frac{1}{2^{v/2}\Gamma(v/2)} y^{\frac{v}{2}-1} e^{-\frac{y}{2}} \quad \text{for} \quad y \in \mathbb{R}^+.$$

It is known that Y has the moment generating function $\Psi_Y(\eta) = (1-2\eta)^{-v/2}$ for and cumulent function $\psi_Y(\eta) = \log \Psi_Y(\eta) = -v\log(1-2\eta)/2$ for $\eta < \frac{1}{2}$. Let $\eta < \frac{1}{2}$ denote the tilting parameter. Then, define the pdf $f_\eta(y)$ for Y by

$$f_\eta(y) = \frac{f_Y(y)e^{\eta y}}{E[e^{\eta Y}]} = \frac{1}{\left(\frac{2}{1-2\eta}\right)^{v/2}\Gamma(v/2)} y^{v/2-1} e^{-\frac{y}{2/(1-2\eta)}} \quad \text{for } y \in \mathbb{R}^+. \tag{20}$$

Hence, we recognize that $Y \sim \Gamma(\frac{v}{2}, \frac{2}{1-2\eta})$.

In a similar fashion, suppose $Z \sim N(0, \mathbb{I})$. Then, Z has the pdf

$$f_Z(z) = \frac{1}{\sqrt{(2\pi)^d}} e^{-z'z/2} \quad \text{for} \quad z \in \mathbb{R}^d.$$

It is known that Z has its moment generating function $\Psi_Z(\vartheta) = e^{\vartheta'\vartheta/2}$ and cumulent function $\psi_Z(\vartheta) = \log \Psi_Z(\vartheta) = \vartheta'\vartheta/2$ for $z \in \mathbb{R}^d$. Let $\vartheta \in \mathbb{R}^d$ denote the tilting parameter. Define the pdf for Z by

$$f_\vartheta(z) = \frac{f_Z(z)e^{\vartheta' z}}{E[e^{\vartheta' Z}]} = \frac{1}{\sqrt{(2\pi)^d}} e^{-(z-\vartheta)'(z-\vartheta)/2} \quad \text{for} \quad z \in \mathbb{R}^d. \tag{21}$$

We hence recognize that $Z \sim N_d(\vartheta, \mathbb{I})$.

With $f_\eta(y)$ and $f_\vartheta(y)$ in Eqs. (20) and (21), respectively, we obtain

$$\begin{aligned} E[\wp(Y,Z)] &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^+} \wp(w,z) f_Y(y) f_Z(x) dy dz \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^+} \wp(w,z) e^{-\eta y - \vartheta' z + \psi_Y(\eta) + \psi_Z(\vartheta)} f_\eta(y) f_\vartheta(z) dy dz \\ &= E_{\eta,\vartheta}\left[\wp(Y,Z) e^{-\eta Y - \vartheta' Z + \psi_Y(\eta) + \psi_Z(\vartheta)}\right], \end{aligned}$$

where $E_{\eta,\vartheta}[\cdot]$ indicates that the expectation is taken with Y and Z jointly distributed according to $f_\eta(y)f_\vartheta(z)$, respectively. Hence, the importance sampling estimator with exponential tilting for $E[\wp(X)]$ is

$$\wp_{\eta,\vartheta}(Y,Z) = \wp(Y,Z)\ \mathrm{e}\ (-\eta Y - \vartheta' Z - v\log(1-2\eta)/2 + \vartheta'\vartheta/2)$$

where $Y \sim \Gamma(\frac{v}{2}, \frac{2}{1-2\eta})$, $Z \sim N(\vartheta, I)$, and Y and Z are independent.

Appendix B: Proof for Theorem 1

Let $\Theta_Y = (-\infty, 1/2)$ and $\Theta_Z = \mathbb{R}^d$ denote the parameter space for η and ϑ, respectively. For notational ease, define a $(d+1)$-variate random vector $W = (Y, Z')'$ and a $(d+1)$-variate vector $\theta = (\eta, \vartheta')'$. Furthermore, the parameter space for θ is denoted by $\Theta_W = \Theta_Y \times \Theta_Z = (-\infty, 1/2) \times \mathbb{R}^d$. Then, we have

$$G(\theta) = G(\eta, \vartheta) = E\left[\wp^2(Y, Z)e^{-\eta Y - \vartheta' Z + \psi_Y(\eta) + \psi_Z(\vartheta)}.\right] \tag{22}$$

It is straightforward that η_1, η_2 and ϑ_1, ϑ_2 belong in Θ_Y and Θ_Z, respectively, if and only if $\theta_1 = (\eta_1, \vartheta_1')'$ and $\theta_2 = (\eta_2, \vartheta_2')'$ belong in Θ_W. To prove that $G(\theta) = G(\eta, \vartheta)$ is strictly convex, we need to show for any $\alpha \in (0, 1)$ and $\theta_1 = (\eta_1, \vartheta_1')'$ and $\theta_2 = (\eta_2, \vartheta_2')'$ in Θ_W, the inequality holds:

$$G(\alpha\theta_1 + (1-\alpha)\theta_2) < \alpha G(\theta_1) + (1-\alpha)G(\theta_2).$$

This is equivalent to

$$G(\alpha(\eta_1, \vartheta_1) + (1-\alpha)(\eta_2, \vartheta_2)) < \alpha G(\eta_1, \vartheta_1) + (1-\alpha)G(\eta_2, \vartheta_2).$$

With the definition of $G(\theta)$ in Eq. (22), we have

$$G(\alpha(\eta_1, \vartheta_1) + (1-\alpha)(\eta_2, \vartheta_2)) = G(\alpha\eta_1 + (1-\alpha)\eta_2, \alpha\vartheta_1 + (1-\alpha)\vartheta_2),$$

and the above inequality becomes

$$G(\alpha\eta_1 + (1-\alpha)\eta_2, \alpha\vartheta_1 + (1-\alpha)\vartheta_2) < \alpha G(\eta_1, \vartheta_1) + (1-\alpha)G(\eta_2, \vartheta_2). \tag{23}$$

Furthermore, define

$$g(\theta) = e^{-\theta' W + \psi_W(\theta)}.$$

By the strict convexity of the exponential and moment generating function, it is clear that

$$\begin{aligned}
&g(\alpha\theta_1 + (1-\alpha)\theta_2) \\
&\quad = e^{-(\alpha\theta_1 + (1-\alpha)\theta_2)' W + \psi_W(\alpha\theta_1 + (1-\alpha)\theta_2)} \\
&\quad < (\alpha e^{-\theta_1' W} + (1-\alpha)e^{-\theta_2' W})(\alpha \Psi_W(\theta_1) + (1-\alpha)\Psi_W(\theta_2)) \\
&\quad = \alpha^2 e^{-\theta_1' W}\Psi_W(\theta_1) + (1-\alpha)^2 e^{-\theta_2' W}\Psi_W(\theta_2) + \alpha(1-\alpha)\left(e^{-\theta_1' W}\Psi_W(\theta_2) + e^{-\theta_2' W}\Psi_W(\theta_1)\right) \\
&\quad \le \alpha e^{-\theta_1' W}\Psi_W(\theta_1) + (1-\alpha)e^{-\theta_2' W}\Psi_W(\theta_2) \\
&\quad = \alpha g(\theta_1) + (1-\alpha)g(\theta_2),
\end{aligned}$$

where the last inequality comes from the positivity of the exponential and moment generating function. Then, Eq. (23) is proved by the linearity of expectation.

To show that $G(\theta)$ has a solution, we first define the set

$$D = \{\theta : G(\theta) \leq \alpha\},$$

where α is selected so that D is non-empty. We would like to show that D is compact by showing that D is closed and bounded. Because $G(\theta)$ is continuous, D is closed. To show that D is bounded, we use contradiction. Suppose there exist a sequence $\{\theta_1, \theta_2, \ldots,\} \in D$, such that $f(\theta_n) \to \infty$ as $\|\theta_n\| \to \infty$, where $\|\cdot\|$ is the Euclidean norm. But this contradicts that $f(\theta_n) \leq \alpha$ for all $n = 1, 2, \ldots$. Then, by Weierstrass extreme value theorem, $G(\theta)$ attains its minimum in D. Therefore, $G(\theta)$ attains its minimum on Θ_W, or equivalently, there exists a minimizer for $G(\theta)$. By the convexity of $G(\theta)$, there exists a unique minimizer for $G(\theta)$.

Appendix C: Proof of Theorem 2

To minimize $G(\eta, \vartheta)$, the first-order-condition requires that the optimal tilting parameters to be the solution of the following equations:

$$E\left[\left(-Y + \frac{\partial \psi_Y(\eta)}{\partial \eta}\right) \wp^2(Y, Z) e^{-\eta Y - \vartheta' Z + \psi_Y(\eta) + \psi_Z(\vartheta)}\right] = 0, \tag{24}$$

$$E\left[(-Z + \nabla \psi_Z(\vartheta)) \wp^2(Y, Z) e^{-\eta Y - \vartheta' Z + \psi_Y(\eta) + \psi_Z(\vartheta)}\right] = 0. \tag{25}$$

With standard algebra, Eq. (24) becomes

$$\left[Y \wp^2(Y, Z) e^{-\eta Y - \vartheta' Z + \psi_Y(\eta) + \psi_Z(\vartheta)}\right] = E\left[\frac{\partial \psi_Y(\eta)}{\partial \eta} \wp^2(Y, Z) e^{-\eta Y - \vartheta' Z + \psi_Y(\eta) + \psi_Z(\vartheta)}\right],$$

which can be simplified as

$$E\left[Y \wp^2(Y, Z) e^{-\eta Y - \vartheta' Z}\right] = E\left[\frac{\partial \psi_Y(\eta)}{\partial \eta} \wp^2(Y, Z) e^{-\eta Y - \vartheta' Z}\right],$$

by dividing it with the common factor $e^{\psi_Y(\eta) + \psi_Z(\vartheta)}$. Similarly, Eq. (25) can be simplified as

$$E\left[Z \wp^2(Y, Z) e^{-\eta Y - \vartheta' Z}\right] = E\left[\nabla \psi_Z(\vartheta) \wp^2(Y, Z) e^{-\eta Y - \vartheta' Z}\right].$$

With cumulent functions of Y and Z, we have

$$\frac{\partial \psi_Y(\eta)}{\partial \eta} = \frac{v}{1 - 2\eta},$$
$$\nabla \psi_Z(\vartheta) = \vartheta.$$

As a result, we have

$$E\left[Y\wp^2(Y,Z)e^{-\eta Y-\vartheta' Z}\right] = \frac{v}{1-2\eta}E\left[\wp^2(Y,Z)e^{-\eta Y-\vartheta' Z}\right],$$
$$E\left[Z\wp^2(Y,Z)e^{-\eta Y-\vartheta' Z}\right] = \vartheta E\left[\wp^2(Y,Z)e^{-\eta Y-\vartheta' Z}\right].$$

Standard algebra gives

$$\frac{v}{1-2\eta} = \frac{E\left[Y\wp^2(Y,Z)e^{-\eta Y-\vartheta' Z}\right]}{E\left[\wp^2(Y,Z)e^{-\eta Y-\vartheta' Z}\right]},$$
$$\vartheta = \frac{E\left[Z\wp^2(Y,Z)e^{-\eta Y-\vartheta' Z}\right]}{E\left[\wp^2(Y,Z)e^{-\eta Y-\vartheta' Z}\right]}.$$

With the joint pdf $f_{\wp,\eta,\vartheta}(y,z)$ in Eq. (13), we obtain the following simplified formula:

$$\begin{aligned}\frac{E\left[Y\wp^2(Y,Z)e^{-\eta Y-\vartheta' Z}\right]}{E\left[\wp^2(Y,Z)e^{-\eta Y-\vartheta' Z}\right]} &= \frac{\int_{\mathbb{R}^d}\int_{\mathbb{R}^+} y\wp^2(y,z)e^{-\eta y-\vartheta' z}f_Y(y)f_Z(z)dydz}{E\left[\wp^2(Y,Z)e^{-\eta Y-\vartheta' Z}\right]} \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^+} y\left(\frac{\wp^2(y,z)e^{-\eta y-\vartheta' z}f_Y(y)f_Z(z)}{E\left[\wp^2(Y,Z)e^{-\eta Y-\vartheta' Z}\right]}\right)dydz \\ &= E_{\wp,\eta,\vartheta}[Y].\end{aligned}$$

In a similar manner, we obtain Eq. (15).

Author Contributions Teng is the single contributor of this paper: Teng developed the methodology, wrote codes, performed numerical comparisons, and completed writing and reviewing this manuscript.

Funding Financial support of the Higher Education Sprout Project of National Yang Ming Chiao Tung Unversity by the the Ministry of Education of Taiwan (ROC) and the Ministry of Science and Technology of Taiwan (ROC) under Grants 110-2118-M-A49-003, 111-2118-M-A49-007, and 108-2634-F-163-001 through Pervasive Artificial Intelligence Research Labs are greatly acknowledged.

Declarations

Conflict of interest The author declares that there is no conflict of interest.

Ethical Approval The author confirms that ethical approval has been obtained.

Informed Consent This research does not involve human participants or animals.

References

Asmussen, S., & Glynn, P. (2007). *Stochastic simulation: algorithms and analysis*. Springer-Verlag.

Avramidis, A. N., & Wilson, J. R. (1998). Correlation-induction techniques for estimating quantiles in simulation experiments. *Operations Research, 46*(4), 574–591.

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

Bank for International Settlements (2006). International convergence of capital measurement and capital standards. Basel Committee Publications, Bank For International Settlements.

Bank for International Settlements (2009). Revisions to the Basel II Market Risk Framework. Basel Committee Publications, Bank For International Settlements

Bank for International Settlements (2019). Minimum capital requirements for market risk. Basel Committee Publications, Bank For International Settlements

Botev, Z. I., & lÉcuyer, P. (2015). Efficient probability estimation and simulation of the truncated multivariate student-t distribution. In 2015 Winter Simulation Conference (pp. 380–391). IEEE.

Chu, F., & Nakayama, M. K. (2012). Confidence intervals for quantiles when applying variance-reduction techniques. *ACM Transactions on Modeling and Computer Simulation, 22*(2), 1–25.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance, 1*(2), 223–236.

Duffie, D., & Pan, J. (1997). An overview of Value at Risk. *Journal of Derivatives, 4*(3), 7–49.

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

Fu, M., & Su, Y. (2002). Optimal importance sampling in securities pricing. *Journal of Computational Finance, 5*(4), 27–50.

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*(6), 1359–1406.

Glasserman, P. (2003). *Monte Carlo methods in financial engineering*. Springer.

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

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

Glasserman, P., Heidelberger, P., & Shahabuddin, P. (2000). Variance reduction techniques for estimating value-at-risk. *Management Science, 46*(10), 1349–1364.

Glasserman, P., Heidelberger, P., & Shahabuddin, P. (2002). Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance, 12*(3), 239–269.

Glynn, P. W. (1996). Importance sampling for Monte Carlo estimation of quantiles. In Proceedings of the Second International Workshop on Mathematical Methods in Stochastic Simulation and Experimental Design (pp. 180–181).

Hesterberg, T. C., & Nelson, B. L. (1998). Control variates for probability and quantile estimation. *Management Science, 44*(9), 1295–1312.

Hsu, J. C., & Nelson, B. L. (1990). Control variates for quantile estimation. *Management Science, 36*(7), 835–851.

Hull, J. C. (2014). *Options, futures, and other derivatives* (9th ed.). Pearson.

Jiang, G., Xu, C., & Fu, M. C. (2016). On sample average approximation algorithms for determining the optimal importance sampling parameters in pricing financial derivatives on Lévy processes. *Operations Research Letters, 44*(1), 44–49.

Jorion, P. (2000). *Value-at-risk*. McGrall-Hill.

Kawai, R. (2009). Optimal importance sampling parameter search for Lévy processes via stochastic approximation. *SIAM Journal on Numerical Analysis, 47*(1), 293–307.

Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2012). *Introduction to probability and statistics* (15th ed.). Cengage Learning.

Renand, H., Zhao, S., & Ermon, S. (2019). Adaptive antithetic sampling for variance reduction. In: International Conference on Machine Learning, PMLR (pp. 5420–5428).

Ross, S. M. (2013). *Simulation*. Academic Press.

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

Teng, H. W., Fuh, C. D., & Chen, C. C. (2016). On an automatic and optimal importance sampling approach with applications in finance. *Quantitative Finance, 16*(8), 1259–1271.

Tsay, R. S. (2013). *An introduction to analysis of financial data with R*. Wiley.

Zhao, P., & Zhang, T. (2015). Stochastic optimization with importance sampling for regularized loss minimization. In International Conference on Machine Learning, PMLR (pp. 1–9).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.