Quantitative Finance

ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: http://www.tandfonline.com/loi/rquf20

On an automatic and optimal importance sampling approach with applications in finance

Huei-Wen Teng, Cheng-Der Fuh & Chun-Chieh Chen

To cite this article: Huei-Wen Teng, Cheng-Der Fuh & Chun-Chieh Chen (2016): On an automatic and optimal importance sampling approach with applications in finance, Quantitative Finance, DOI: 10.1080/14697688.2015.1136077

To link to this article: http://dx.doi.org/10.1080/14697688.2015.1136077

Published online: 07 Mar 2016.

Quantitative Finance, 2016
http://dx.doi.org/10.1080/14697688.2015.1136077

On an automatic and optimal importance sampling approach with applications in finance

HUEI-WEN TENG*† , CHENG-DER FUH† and CHUN-CHIEH CHEN‡

†Graduate Institute of Statistics, National Central University, Taoyuan, Taiwan
‡EnTie Commercial Bank Ltd, Taipei, Taiwan

(*Received 26 January 2015; accepted 17 December 2015; published online 7 March 2016*)

Calculating high-dimensional integrals efficiently is essential and challenging in many scientific disciplines, such as pricing financial derivatives. This paper proposes an exponentially tilted importance sampling based on the criterion of minimizing the variance of the importance sampling estimators, and its contribution is threefold: (1) A theoretical foundation to guarantee the existence, uniqueness, and characterization of the optimal tilting parameter is built. (2) The optimal tilting parameter can be searched via an automatic Newton's method. (3) Simplified yet competitive tilting formulas are further proposed to reduce heavy computational cost and numerical instability in high-dimensional cases. Numerical examples in pricing path-dependent derivatives and basket default swaps are provided.

Keywords: Importance sampling; Exponential tilting; Conjugate measures; Newton's method; Exotic options; Caps; Basket default swaps; Gaussian copula

JEL Classification: C150, C63

1. Introduction

Numerous financial derivatives have been designed to meet various needs in financial markets, and therefore pricing financial derivatives correctly and efficiently is an essential and indispensable task. It is known that asset pricing theory gives the price of a financial derivative equal to the discounted expected payoff function under the risk-neutral measure. Nevertheless, closed-form formulas of option prices rarely exist in general cases, and this leads to the requirement of additional numerical methods.

Monte Carlo simulation is a useful numerical method to price financial derivatives, because it is flexible and can avoid the curse of dimensionality (Boyle *et al*. 1997). However, the convergence of the Monte Carlo estimator is proportional to the reciprocal of the square root of the Monte Carlo sample size, and a large number of trials of Monte Carlo simulation are usually required to achieve the desired level of precision (Ross 2013). As a result, additional variance reduction techniques are proposed to improve the efficiency of the crude Monte Carlo estimators, including control variates, antithetic variates, moment matching, stratified sampling, importance sampling, conditional Monte Carlo method and quasi-Monte Carlo methods, among others. The reader is referred to Glasserman (2004) and references therein for details.

This paper focuses on the importance sampling method. Nonparametric importance sampling approaches considered in Zhang (1996), Givens and Raftery (1996) and Kim *et al*. (2000) apply histograms or kernel estimates to approximate the optimal alternative probability measure. However in high-dimensional cases, nonparametric importance sampling is extremely computationally intensive, and additional procedures to reduce the dimension are involved (Neddermeyer 2011). Moreover, nonparametric importance sampling does not exploit the probabilistic features of the underlying process of a financial derivative.

By making use of the structure of the underlying process of a financial derivative, in this paper we will study parametric importance sampling techniques. It is known that exponential tilting is a useful class among parametric importance sampling techniques, to which large deviations theory has been applied to select the exponentially tilting parameter (Siegmund 1976). A general account can be found in Bucklew (2004) and Asmussen and Glynn (2007). Further applications of this device for derivatives pricing are in Glasserman *et al*. (1999), and Chen and Glasserman (2008), for example. Recently, Guasoni and Robertson (2008) employed large deviation techniques in continuous time so that the asymptotically optimal change of drift is the solution to an one-dimensional variational problem.

*Corresponding author. Email: wenteng@ncu.edu.tw

Despite its popularity and efficiency, importance sampling estimators using large deviations theory may not always enjoy lower variances compared with the crude Monte Carlo estimators, cf. Glasserman and Wang (1997). To resolve this difficulty, a modification based on certain mixtures of exponentially tilting probability measures can be found in Chan and Lai (2007). Another complexity of applying large deviation tilting is that it is rather mathematically demanding and may involve approximating high-dimensional integrals using Laplace methods or other techniques.

An alternative for choosing a tilting parameter is based on the criterion of minimizing the variance of the importance sampling estimator directly. For example, Vazquez-Abad and Dufresne (1998) applies a gradient estimation and stochastic approximation to minimize the variance of the estimator under the sampling probability measure. Instead, Su and Fu (2000, 2002) minimize the variance under the original probability measure. Capriotti (2008) introduces a least squares importance sampling method. More recently, Fuh *et al*. (2011) applies the importance sampling method for Value-at-Risk (VaR) computation under a multivariate t-distribution.

Along the lines of minimizing the variance of the importance sampling estimator, this paper proposes a general framework of importance sampling in pricing financial derivatives that casts new insights on the problem. There are three aspects in this study. To begin with, we characterize the optimal tilting θ as the root of a novel and simple equation via a device called conjugate measure. Second, we present an automatic Newton's method to search for the optimal tilting parameter. Third, we further propose useful approximations to the optimal tilting parameter that substantially reduce computational burden in searching for the optimal tilting parameter. To illustrate the applicability of our method, we study some interesting financial derivatives, including Asian options under the Black–Scholes model, interest rate caps under a two-factor Hull–White model, and basket default swaps under the normal copula model.

The rest of this paper is organized as follows. A general account of importance sampling for pricing financial derivatives is given in section 2. Section 3 presents an automatic Newton's method to search for the optimal tilting parameters. For numerical illustrations, section 4 prices Asian options, section 5 prices interest caps, and section 6 prices nth-to-default swaps. Useful simplified tilting formulas are presented and discussed for various examples. Section 7 concludes. The proofs are deferred to the appendix 1.

2. Importance sampling

Let $(\Omega, \mathcal{F}, P)$ be a given probability space. Let $X = (X_1, \dots, X_d)'$ be a d-dimensional random vector having $f(x) = f(x_1, \dots, x_d)$ as a probability density function (pdf), with respect to the Lebesgue measure $\mathcal{L}$, under the probability measure P. Here and throughout, a prime denotes vector or matrix transpose. Let $\wp(\cdot)$ be a real-valued function from $\Re^d$ to $\Re$. The problem of interest is to calculate the expectation of $\wp(X)$,

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

where $E_P[\cdot]$ is the expectation operator under the probability measure P.

The asset pricing theory gives the price of a financial derivative equals the discounted expected payoff under the risk-neutral probability measure. In this case, $\wp(\cdot)$ is the payoff function, X is the randomness governing the underlying process, which may be stock prices, interest rates, or default times, among others, and the price of a financial derivative equals (1). In the paradigm of financial derivatives pricing, the probability measure P is the risk-neutral probability measure.

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

$$\begin{aligned}&\int_{\Re^d} \wp(x) f(x)\mathrm{d}x \\ &\quad = \int_{\Re^d} \wp(x)\frac{f(x)}{q(x)} q(x)\mathrm{d}x = \mathrm{E}_Q\left[\wp(X)\frac{f(X)}{q(X)}\right],\end{aligned} \tag{2}$$

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

Here, we focus on the exponentially tilted probability measure of P, denoted by Q_θ, where the subscript $\theta = (\theta_1, \dots, \theta_d)'$ is the tilting parameter. Assume that the moment generating function of X exists and is denoted by $\Psi(\theta)$. Let $f_\theta(x)$ be the pdf of X under the exponentially tilted probability measure Q_θ, defined by

$$f_\theta(x) = f(x)\frac{\mathrm{e}^{\theta' x}}{\Psi(\theta)} = f(x)\mathrm{e}^{\theta' x - \psi(\theta)}, \tag{3}$$

where $\psi(\theta) = \ln \Psi(\theta)$ is the cumulant function. Then, (2) becomes

$$\begin{aligned}&\int_{\Re^d} \wp(x) f(x)\mathrm{d}x \\ &\quad = \int_{\Re^d} \wp(x)\frac{f(x)}{f_\theta(x)} f_\theta(x)\mathrm{d}x = E_{Q_\theta}\left[\wp(x)\mathrm{e}^{-\theta' X + \psi(\theta)}\right].\end{aligned}$$

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

$$\begin{aligned}&\mathrm{Var}_{Q_\theta}\left[\wp(X)\mathrm{e}^{-\theta' X + \psi(\theta)}\right] \\ &\quad = E_{Q_\theta}\left[\left(\wp(X)\mathrm{e}^{-\theta' X + \psi(\theta)}\right)^2\right] - m^2.\end{aligned} \tag{4}$$

For simplicity, we assume that the variance of the importance sampling estimator exists.

Define the first term of the right-hand side (RHS) of (4) by $G(\theta)$. Then, minimizing $\mathrm{Var}_{Q_\theta}\left[\wp(X)\mathrm{e}^{-\theta' X + \psi(\theta)}\right]$ is equivalent to minimizing $G(\theta)$. Standard algebra gives a simpler form of $G(\theta)$,

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

which will be used to find the optimal tilting parameter.

To minimize $G(\theta)$, the first-order condition requires the solution of θ, denoted by θ^*, to satisfy $\nabla G(\theta) \mid_{\theta=\theta^*} = 0$, where ∇ denotes the gradient. Simple calculation leads

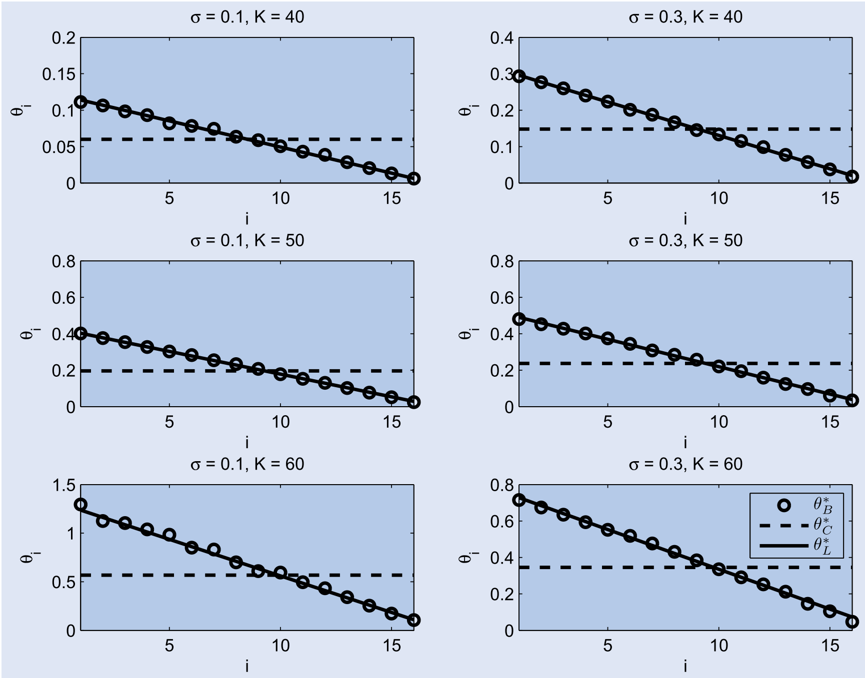

Figure 1. Estimated titling formulas θ_i against the ith time step in pricing Asian options under the Black–Scholes model in 16 dimensions at various trike price K and volatility σ. Parameter settings: $S_0 = 50$, $r = 0.05$, and $T = 1$.

$$\nabla G(\theta) = E_P\left[\wp^2(X)(-X + \nabla\psi(\theta))\mathrm{e}^{-\theta' X+\psi(\theta)}\right],$$

and, therefore, θ^* is the root of the following system of non-linear equations,

$$\nabla\psi(\theta) = \frac{E_P\left[\wp^2(X)X\mathrm{e}^{-\theta' X}\right]}{E_P\left[\wp^2(X)\mathrm{e}^{-\theta' X}\right]}. \tag{6}$$

To have a simple form on the RHS of (6), we define the conjugate measure $\bar{Q}_\theta := \bar{Q}_\theta^{\wp}$ of the measure Q with respect to the payoff function $\wp$ as

$$\frac{\mathrm{d}\bar{Q}_\theta}{\mathrm{d}P} = \frac{\wp^2(X)\mathrm{e}^{-\theta' X}}{E_P[\wp^2(X)\mathrm{e}^{-\theta' X}]} = \wp^2(X)\mathrm{e}^{-\theta' X-\bar{\psi}(\theta)}, \tag{7}$$

where $\bar{\psi}(\theta)$ is $\log\bar{\Psi}(\theta)$ with $\bar{\Psi}(\theta) = E_P[\wp^2(X)\mathrm{e}^{-\theta' X}]$. Then the RHS of (6) equals $E_{\bar{Q}_\theta}[X]$, which is the expected value of X under $\bar{Q}_\theta$.

The following theorem states the existence, uniqueness and characterization for the minimizer of (5).

THEOREM 2.1 *Suppose the moment generating function $\Psi(\theta)$ of X exists for $\theta \in \Re^d$. Furthermore, assume that $\bar{\Psi}(\theta)\Psi(\theta) \to \infty$ as $\|\theta\| \to \infty$, where $\|\cdot\|$ denotes the Euclidean norm in $\Re^d$. Then $G(\theta)$ defined in (5) is a convex function in θ, and there exists a unique minimizer of (5), which satisfies*

$$\nabla\psi(\theta) = E_{\bar{Q}_\theta}[X]. \tag{8}$$

The proof of theorem 2.1 will be given in the appendix.

Note that the condition $\bar{\Psi}(\theta)\Psi(\theta) \to \infty$ as $\|\theta\| \to \infty$ in theorem 2.1 is a sufficient condition to guarantee the existence of the minimizer of $G(\theta)$. We note that it is not the weakest condition, and the examples considered in this paper satisfy this condition. This is because of the fact that when $X \sim N_d(\mathbf{0}, \Sigma)$, equation (8) reduces to $\Sigma\theta = E_{\bar{Q}_\theta}[X]$ (see section 3.1 below), and $\Psi(\theta) = O(e^{\|\theta\|^2})$ approaches to ∞ sufficiently fast. A simple example illustrated here is when $\wp(X) = \mathbf{1}_{\{X\in A\}}$ and $A := [a_1, \infty)\times\cdots\times[a_d, \infty)$, with $a_i > 0$ for all $i = 1, \ldots, d$, and X has a d-dimensional standard normal distribution, then it is easy to check that the sufficient conditions in theorem 2.1 hold.

Although we focus on multivariate normal distributions and its applications in this paper, it is worth mentioning that theorem 2.1 can be applied to different random vectors. For instance, in a separate paper, we consider importance sampling for false alarm probabilities under the K-distributed sea clutter (Fuh and Teng 2015).

3. Implementation of the algorithm

For easy presentation, we divide this section into four parts. Section 3.1 provides sampling probability measure for the multivariate normal distribution, which will be used in pricing finance derivatives under various settings. Section 3.2 presents an automatic Newton's method to search the optimal tilting parameter. Section 3.3 summarizes an importance sampling algorithm for pricing financial derivatives. Section 3.4 presents the study plan.

3.1. *Exponentially tilted probability measure of* $N_d(0, \Sigma)$

Let X be a d-variate normal random variable with mean vector $\mu = (\mu_1, \ldots, \mu_d)'$ and covariance matrix Σ, written as $X \sim N_d(\mu, \Sigma)$. Then X has the pdf

$$\phi(x;\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}}\exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}.$$

Here, $|\cdot|$ is matrix determinant. When $d = 1$, we simply write $X \sim N(\mu, \sigma^2)$ with σ^2 being the variance.

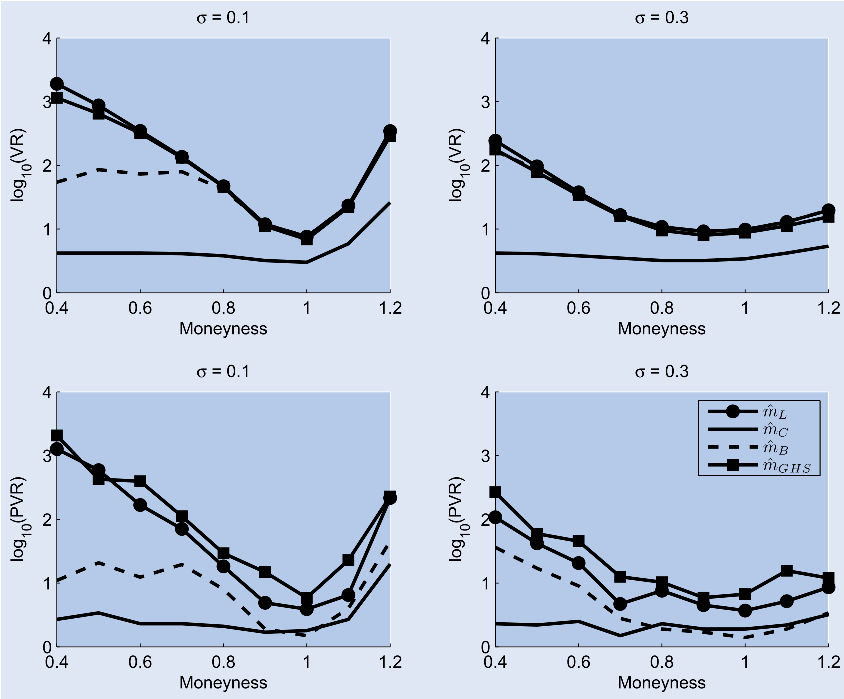

Figure 2. Estimated variance ratios and penalized variance ratios in based 10 logarithm against moneyness in pricing Asian options under the Black–Scholes model at various σ. Parameter settings: $S_0 = 50$, $r = 0.05$, $T = 1$, and $d = 64$.

The exponential tilting of the normal distribution is known for some spacial cases. For instance, Glasserman *et al*. (1999) consider the exponential tilting of $N_d(\mathbf{0}, \Sigma = I)$, and Fuh and Hu (2004) consider the exponential tilting of $N(\theta, 1)$. Motivated by various examples in pricing financial derivatives, we include the exponential tiling of $N_d(\mathbf{0}, \Sigma)$ for the sake of completeness. Here, $\mathbf{0}$ is the zero vector of size $d \times 1$. When $X \sim N_d(\mathbf{0}, \Sigma)$, its cumulant function is $\psi(\theta) = \theta'\Sigma\theta/2$. And, the gradient of the cumulant function is $\nabla\psi(\theta) = \Sigma\theta$.

Let Q_θ be the exponentially tilted probability measure as defined in (3), then the pdf $f_\theta(x)$ of X under Q_θ equals

$$f_\theta(x) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left\{-\frac{1}{2}x'\Sigma^{-1}x\right\} \exp\left\{\theta'x - \frac{1}{2}\theta'\Sigma\theta\right\}$$
$$= \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left\{-\frac{x'\Sigma^{-1}x - 2x'\theta + \theta'\Sigma\theta}{2}\right\}.$$

Simple matrix calculation shows that the exponent can be factorized as

$$x'\Sigma^{-1}x - 2x'\theta + \theta'\Sigma\theta = (x - \Sigma\theta)'\Sigma^{-1}(x - \Sigma\theta).$$

As a result, we have the following proposition.

Proposition 3.1 *If $X \sim N_d(\mathbf{0}, \Sigma)$ under the original probability measure P, then*

$$X \sim N_d(\Sigma\theta, \Sigma),$$

under the exponentially tilted probability measure Q_θ.

3.2. An automatic Newton's method

The optimal tilting parameter θ^* needs to satisfy a system of nonlinear equations in (8). In the case of $X \sim N_d(\mathbf{0}, \Sigma)$, we have that $\psi(\theta) = \theta'\Sigma\theta/2$ and $\nabla\psi(\theta) = \Sigma\theta$, then (8) reduces to

$$\Sigma\theta = E_{\bar{Q}_\theta}[X]. \tag{9}$$

Define a function $g(\theta) = (g_1(\theta), \ldots, g_d(\theta))'$ from $\Re^d$ to $\Re^d$ by

$$g(\theta) = \Sigma\theta - E_{\bar{Q}_\theta}[X]. \tag{10}$$

Therefore, θ^* is the root of (10).

To use Newton's method to find the root of (10), the following proposition presents the Jacobian of $g(\theta)$. Let $A[i, j]$ be the element of ith raw and the j-column of a matrix A.

Proposition 3.2 *The Jacobian J_θ of $g(\theta)$ in (10) equals*

$$J_\theta[i, j] := \frac{\partial}{\partial\theta_j} g_i(\theta)$$
$$= \Sigma[i, j] - E_{\bar{Q}_\theta}[X_i]E_{\bar{Q}_\theta}[X_j] + E_{\bar{Q}_\theta}[X_i X_j],$$

for $i, j = 1, \ldots, d$, where $\bar{Q}_\theta$ is the conjugate measure defined in (7).

Based on proposition 3.2, the Jacobian can be estimated using Monte Carlo estimation easily for an arbitrary payoff function $\wp(\cdot)$. Applying Newton's method, the root of (10) is found iteratively by

$$\theta^{(n+1)} = \theta^{(n)} - J_{\theta^{(n)}}^{-1} g\left(\theta^{(n)}\right), \tag{11}$$

where J_θ^{-1} is the inverse of the matrix J_θ. To measure the precision of root to the solution in (10), we define the sum of the square error of $g(\theta)$ by

$$\|g(\theta)\| = g(\theta)'g(\theta),$$

and accept a $\theta^{(n)}$ when $\|g(\theta^{(n)})\| < \varepsilon$ where ε is a predetermined precision level.

3.3. Algorithms

To compare the complexity of the crude Monte Carlo and our proposed importance sampling, we first summarize steps in implementing the crude Monte Carlo method in the following algorithm.

Algorithm 1 The following steps implement the crude Monte Carlo method.

- Pricing stage:
  1. Generate independent samples $X^{(i)}$ from $N_d(\mu, \Sigma)$ for $i = 1, \ldots, N$.
  2. Estimate m by $\hat{m} = \frac{1}{N}\sum_{i=1}^{N} \wp\left(X^{(i)}\right)$.

Implementing the proposed importance sampling needs an additional searching stage. The searching stage employs the recursive formula in (11), in which the function $g(\theta)$ and the Jacobian J_θ do not have closed-form formulas and are approximated using Monte Carlo simulation.

Define $\hat{g}(\theta)$ as the estimated value of $g(\theta)$ using Monte Carlo simulation, then

$$\hat{g}(\theta) = \Sigma\theta - \frac{1}{N}\sum_{s=1}^{N} Y^{(s)},$$

where $Y^{(1)}, \ldots, Y^{(N)}$ are samples under $\bar{Q}_\theta$. However, generating samples under $\bar{Q}_\theta$ is difficult, because $\bar{Q}_\theta$ involves the payoff function of the financial derivative of interest. Recall that plugging (7) for $\bar{Q}_\theta$ in (10) yields

$$g(\theta) = \Sigma\theta - \frac{E_P[\wp^2(X)\, Xe^{-\theta' X}]}{E_P[\wp^2(X)\, e^{-\theta' X}]}.$$

Therefore, we estimate $\hat{g}(\theta)$ by

$$\hat{g}(\theta) = \Sigma\theta - \frac{\sum_{s=1}^{N} \wp^2\left(X^{(s)}\right) X^{(s)} e^{-\theta' X^{(s)}}}{\sum_{s=1}^{N} \wp^2\left(X^{(s)}\right) e^{-\theta' X^{(s)}}}, \tag{12}$$

where $X^{(1)}, \ldots, X^{(N)}$ are i.i.d. samples under P.

Note that $X^{(1)}, \ldots, X^{(N)}$ are i.i.d. random variables under P. Under the finiteness of the second moment assumption in (4), standard strong law of large numbers implies that the second term on the right-hand side of (12) converges P-almost surely to $E_P[\wp^2(X)\, Xe^{-\theta' X}]/E_P[\wp^2(X)\, e^{-\theta' X}]$, which is $E_{\bar{Q}_\theta}[X]$ by the definition of conjugate measure $\bar{Q}_\theta$ in (7).

Similarly, let $\hat{J}_\theta$ denote the estimated matrix for J_θ. Then $\hat{J}_\theta$ can be estimated by

$$\begin{aligned}\hat{J}_\theta[i,j] &= \Sigma[i,j] \\ &- \frac{\left(\sum_{s=1}^{N} \wp^2\left(X^{(s)}\right) X_i^{(s)} e^{-\theta' X^{(s)}}\right)\left(\sum_{s=1}^{N} \wp^2\left(X^{(s)}\right) X_j^{(s)} e^{-\theta' X^{(s)}}\right)}{\left(\sum_{s=1}^{N} \wp^2\left(X^{(s)}\right) e^{-\theta' X^{(s)}}\right)^2} \\ &+ \frac{\left(\sum_{s=1}^{N} \wp^2\left(X^{(s)}\right) X_i^{(s)} X_j^{(s)} e^{-\theta' X^{(s)}}\right)}{\left(\sum_{s=1}^{N} \wp^2\left(X^{(s)}\right) e^{-\theta' X^{(s)}}\right)},\end{aligned} \tag{13}$$

where $X^{(1)}, \ldots, X^{(N)}$ are samples under P. The following algorithm summarizes steps in implementing the proposed importance sampling method.

Algorithm 2 The following steps implement the proposed importance sampling via automatic Newton's method.

- Searching stage:
  1. Generate independent samples $X^{(i)}$ from $N_d(\mu, \Sigma)$ for $i = 1, \ldots, N$.
  2. Set $\theta^{(0)}$ properly and set $n = 1$.
  3. Calculate $\hat{g}\left(\theta^{(n-1)}\right)$ by (12).
  4. Calculate the Jacobian $\hat{J}_{\theta^{(n-1)}}$ by (13) and its inverse $\hat{J}^{-1}_{\theta^{(n-1)}}$.
  5. Calculate $\theta^{(n)} = \theta^{(n-1)} - \hat{J}^{-1}_{\theta^{(n-1)}}\hat{g}\left(\theta^{(n-1)}\right)$.
  6. Calculate $\hat{g}\left(\theta^{(n)}\right)$ by (12). If $\|g\left(\hat{\theta}^{(n)}\right)\| < \varepsilon$, set $\theta^* = \theta^{(n)}$ and stop. Otherwise, set $n = n + 1$ and return to step (4).
- Pricing stage:
  1. Generate $X^{(i)}$ from $N_d(\Sigma\theta^*, \Sigma)$ for $i = 1, \ldots, N$.
  2. Estimate m by $\hat{m}_B = \frac{1}{N}\sum_{i=1}^{N} \wp\left(X^{(i)}\right) e^{-\theta^{*\prime} X^{(i)} + \psi(\theta^*)}$.

Algorithm 2 possess several powerful advantages in real applications. First, because of the convexity on $G(\theta)$, $\theta^{(0)}$ can be simply set as a zero vector and the optimal tiling formula is not sensitive to the initial value. Second, as long as the quantity of interest is of the form $E_P[\wp(X)]$ for $X \sim N_d(\mu, \Sigma)$, the implementation of this algorithm is simple and automatic, and no additional mathematical treatments on the payoff function are required. In contrast, searching tilting parameters in Glasserman *et al.* (1999) requires the calculation on the Hessian of the payoff function, which is problem dependent and may be mathematically demanding.

3.4. Study plan

To measure the efficiency of the estimator, we report the variance ratio, which is the crude Monte Carlo estimator's variance divided by the associated importance sampling estimator's variance. When the variance ratio is larger than one, the estimator of interest is more efficient than the crude Monte Carlo estimator. And, the larger the variance ratio is, the more efficient the estimator is.

In addition, we report the penalized variance ratios to compromise computational times. The penalized variance by computational time is the product of the variance of an estimator and its computational time. Similarly, the penalized variance ratio is the penalized variance of the crude Monte Carlo estimator divided by the penalized variance of the estimator of interest. If the penalized variance ratio is larger than one, the estimator of interest is preferred after the computational time is taken into account. Likewise, the larger the penalized variance ratio is, the associated estimator is preferred.

For the rest of the paper, we set the Monte Carlo sample size N to be 100 000 to produce precise estimated variances, and set the precision level $\varepsilon = 0.0001$. Please see our discussions on the Monte Carlo sample size required in practice in section 4.4.

4. Pricing Asian options

To illustrate the implementation for the proposed importance sampling, we study a precise example. Section 4.1 considers

Asian options under the Black–Scholes model and discusses the pattern on the optimal tilting parameters and computational time. These observations motivate us to propose simplified tilting formulas in section 4.2. Section 4.3 reports the numerical results. Sensitivity analysis is given in section 4.4. Section 4.5 considers the issue of using control variates.

4.1. Asian options

To study Asian option under the Black–Scholes model, we need the following notations first. Denote the current stock price of the ith stock by S_0. Under the Black–Scholes assumptions, the stock price at time t under the risk-neutral probability measure can be generated by

$$S_t(X) = S_0 \exp\left\{\left(r - \frac{\sigma^2}{2}\right)t + \sigma\sqrt{t}X\right\},$$

where r is the risk-free interest rate, σ is the volatility of the stock and $X \sim N(0, 1)$.

Now, let T be the time to maturity and K be the strike price. For a path-dependent option with single underlying asset, we simulate a path of the stock price by discretization at time points $0 < t_1 < \cdots < t_d = T$, where $t_j = jT/d$ for $j = 0, \ldots, d$. Set S_0 as the current stock price and $\Delta = T/d$. The stock price at time t_i is iteratively simulated by $S_{t_i}(X) = S_{t_{i-1}} \mathrm{e}^{(r-\sigma^2/2)\Delta+\sigma\sqrt{\Delta}X_i}$, for $i = 1, \ldots, d$, and $X = (X_1, \ldots, X_d)' \sim N_d(\mathbf{0}, \mathbf{I})$. The discounted payoff function of the arithmetic Asian option is

$$\wp(X) = \mathrm{e}^{-rT}\left(\frac{1}{d}\sum_{i=1}^{d} S_{t_i}(X) - K\right)^+,$$

where $(\cdot)^+$ denotes the positive function.

Note that $X = (X_1, \ldots, X_d)' \sim N_d(\mathbf{0}, \mathbf{I})$, which implies that $\Psi(\theta) = O(e^{\|\theta\|^2})$. Next, consider $\wp(X)$ as the discounted payoff function of the arithmetic Asian option, we have

$$\begin{aligned}
&\bar{\Psi}(\theta)\\
&= E_P[\wp^2(X)e^{-\theta' X}]\\
&= \int_{\{\frac{1}{d}\sum_{i=1}^{d} S_{t_i}(X)>K\}} \mathrm{e}^{-2rT}\left(\frac{1}{d}\sum_{i=1}^{d} S_{t_i}(X) - K\right)^2 e^{-\theta' X}\phi_d(x)\mathrm{d}x\\
&> \int_{\{\prod_{i=1}^{d}(X_i>K_i)\}} \mathrm{e}^{-2rT}\left(\frac{1}{d}\sum_{i=1}^{d} S_{t_i}(X) - K\right)^2 e^{-\theta' X}\phi_d(x)\mathrm{d}x,
\end{aligned} \tag{14}$$

for some $K_i > 0$, $i = 1, \ldots, d$. Here $\phi_d(x)$ is a d-dimensional standard normal pdf. A simple calculation leads to the result that the last term of (14) is $O(e^{-\|\theta\|})$. Therefore $\Psi(\theta)\bar{\Psi}(\theta) \to \infty$ as $\|\theta\| \to \infty$.

Figure 1 plots optimal tilting parameters θ_i^* against the ith time step in pricing Asian options at various strike price and σ in 16 dimensions. These plots show a clearly linear decreasing tendency of θ^* against the time step. This is similar to the results which appeared in Fuh and Hu (2004) for simulating moderate deviation events with probability of around 10^{-3} (Kallenberg 1983). Moreover, θ^* appears to fluctuate around a linear line, possibly because of Monte Carlo discrepancy in simulating $\hat{g}(\theta)$ and $\hat{J}_\theta$. In addition, tilting parameters are of smaller absolute values for deep in-the-money options ($K = 40$) but of larger absolute values for deep out-of-the-money options ($K = 60$). Note that in the simulation of rare event, the exponential tilting for the tilting parameter in Fuh and Hu (2004) is of the form $\theta/\sqrt{n}$, with n as the simulated sample size. Furthermore, the tilting parameter in Fuh and Hu (2004) is different from the one which appeared in Guasoni and Robertson (2008), in which θ^* converges in some sense to the optimal continuous time change of measure.

4.2. Simplified tilting formulas

Motivated by the linear pattern of θ^* as demonstrated in figure 1, we propose two simplified tilting formulas to reduce computational burden in the searching stage.

To begin with, the constant tilting formula consists of just one single parameter α, and is defined as

$$\theta^C := \theta^C(\alpha) = (\alpha, \ldots, \alpha)' = \mathbf{1}\alpha,$$

where the superscript C abbreviates for a constant, and $\mathbf{1} = (1, \ldots, 1)'$ denotes the unit vector. The use of constant tilting formula means that all random variables $X_1, \ldots, X_n$ are tilted with the same parameter α. In a different context, Fu *et al.* (1999) consider the same tilting formula as θ^C.

Let α^* denote the optimal α that minimizes the variance of the associated importance sampling estimator. The optimal constant tilting formula is

$$\theta^{C*} := \theta^C(\alpha^*). \tag{15}$$

Next, we consider a linear approximation on the tilting formula with two parameters $\beta = (\beta_1, \beta_2)'$,

$$\begin{aligned}
\theta^L &:= \theta^L(\beta)\\
&= (\beta_1, \beta_1 + \beta_2, \beta_1 + 2\beta_2, \ldots, \beta_1 + (d-1)\beta_2)' = H\beta,
\end{aligned}$$

where $H = [H_{ij}]$ is a $d \times 2$ matrix with $H_{i1} = 1$ and $H_{i2} = i - 1$ for $i = 1, \ldots, d$. Similarly, let β^* denote the optimal β that minimizes the variance of the associated importance sampling estimator. The optimal linear tilting formula is

$$\theta^{L*} := \theta^L(\beta^*). \tag{16}$$

Searching the constant and linear tilting formulas is similar to searching the benchmark tilting formula exposited in section 3.3 and is omitted here for brevity. For the rest of the paper, we denote $\hat{m}_B$, $\hat{m}_C$ and $\hat{m}_L$, as the importance sampling estimators using the optimal benchmark tilting formula θ^*, the optimal constant tilting formula θ^{C*} and the optimal linear tilting formula θ^{L*}, respectively.

4.3. Numerical results

Figure 2 summarizes estimated variance ratios and penalized variance ratios against moneyness in pricing Asian options. Here, the moneyness is defined as the ratio of strike price over the current underlying stock price. Moreover, we compare the importance estimator given in Glasserman *et al.* (1999), denoted by $\hat{m}_{GHS}$.

When $\sigma = 0.1$, appropriate importance sampling estimators gain substantial improvements in efficiency particularly

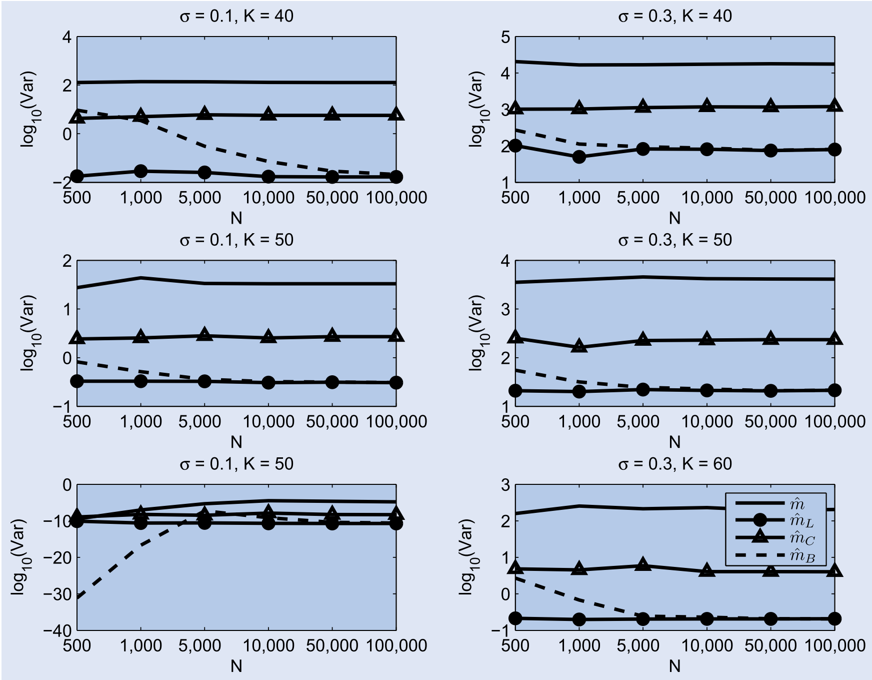

Figure 3. Estimated variances in base 10 logarithm against the number of paths N in pricing Asian options. Parameter settings: $S_0 = 50$, $r = 0.05$, $T = 1$, and $d = 64$.

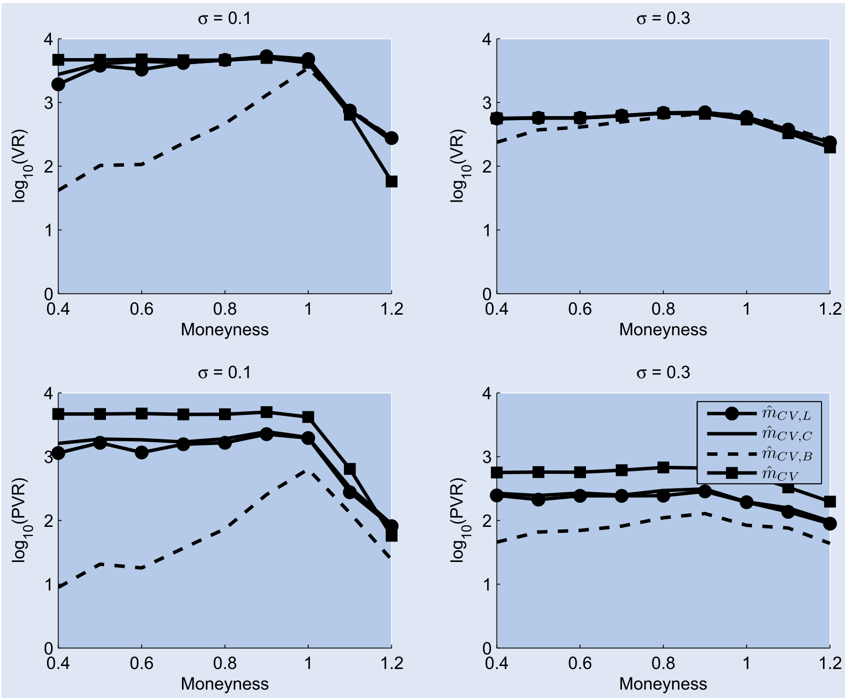

Figure 4. Estimated variance ratios and penalized variance ratios in based 10 logarithm against moneyness in pricing Asian options using geometric Asian options as control variates under the Black–Scholes model at various σ. Parameter settings: $S_0 = 50$, $r = 0.05$, $T = 1$, and $d = 64$.

for in-the-money or out-of-the-money options. In general, $\hat{m}_L$ produces the largerest variance ratios than $\hat{m}_B$ and $\hat{m}_C$. Indeed, θ^{L*} needs just two free parameters and appears to approximate θ^* nicely. Possibly because of the Monte Carlo discrepancy in the searching stage similar to the illustration in figure 1, $\hat{m}_B$ does not produce the largest variance ratio. $\hat{m}_C$ remains more efficient than the crude Monte Carlo estimator. In terms of penalized variance ratios, $\hat{m}_L$ remains the most competitive.

The same conclusion holds when $\sigma = 0.3$, but the improvements in efficiency using importance sampling estimators are less when σ is larger. Finally, $\hat{m}_L$ is competitive to $\hat{m}_{GHS}$, but $\hat{m}_{\mathrm{GHS}}$ is slightly less efficient than $\hat{m}_L$ for deep in-the-money options. It is possible because the Laplace method in Glasserman *et al.* (1999) fails to approximate (5) well.

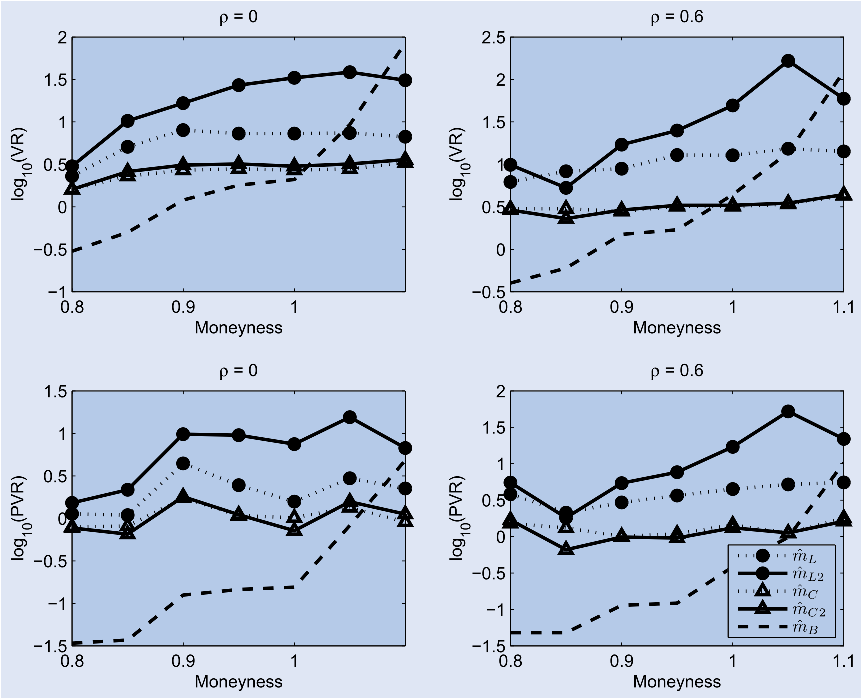

Figure 5. Estimated variance ratios and penalized variance ratios in base 10 logarithm against moneyness in pricing interest rate caps under a two-factor Hull–White model at various ρ. Parameter settings: $r_0 = 0.07$, $u_0 = 0.09$, $a = 1$, $b = 1$, $\sigma_1 = 0.01$, $\sigma_2 = 0.0165$, $\rho = 0.6$, $T = 1$, and $d = 64$.

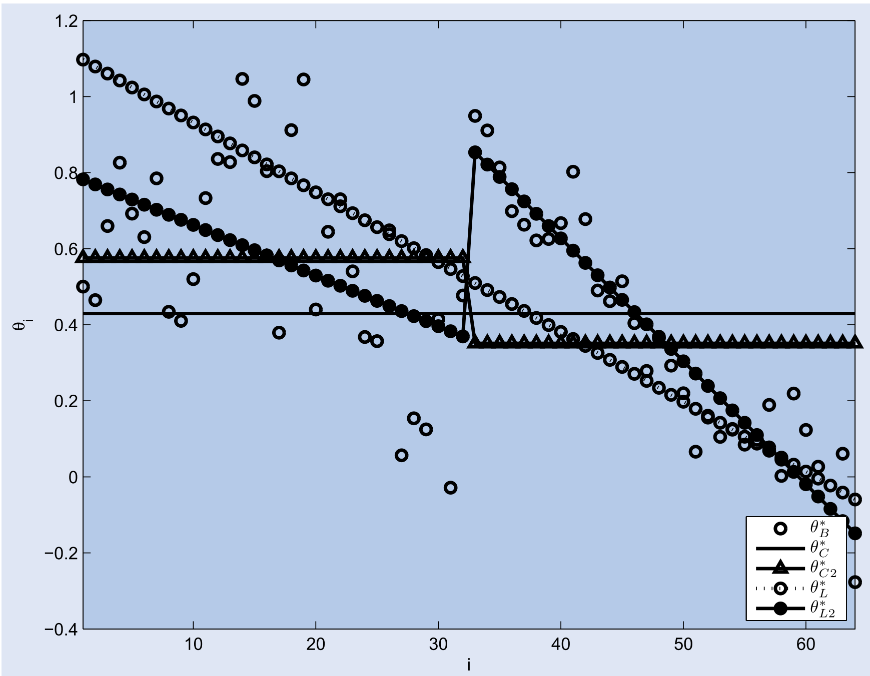

Figure 6. Estimated tilting formulas θ_i against the ith time step in pricing interest rate caps under a two-factor Hull–White model. Parameter settings: $r_0 = 0.07$, $u_0 = 0.09$, $a = 1$, $b = 1$, $\sigma_1 = 0.01$, $\sigma_2 = 0.0165$, $\rho = 0.6$, $K = 0.077$, $T = 1$, and $d = 64$.

4.4. Sensitivity analysis on the number of paths

To investigate how variance minimization depends on the number of paths (or the Monte Carlo sample size) N, figure 3 plots the estimated variance of various estimators against the number of paths for different correlations ρ and strike prices in 64 dimensions. Figure 3 provides a numerical evidence that the estimated variance of $\hat{m}_B$ appears to be stable when the number of paths is larger than 10 000. In contrast, for $\hat{m}_C$ and $\hat{m}_L$, the number of paths required to obtained stable estimated variances is simply 1000. In our simulation studies, we set N to be 100 000 in order to produce precise estimated variances, but fewer number of paths are sufficient for practical implementation.

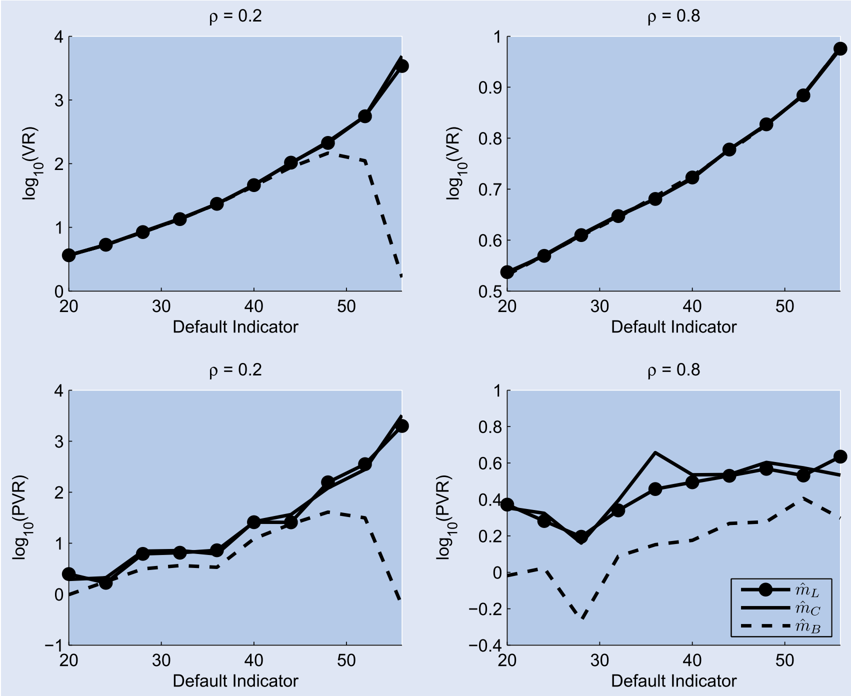

Figure 7. Estimated variance ratios and penalized variance ratios in base 10 logarithm against the default indicator n in pricing nth-to-default swaps under the normal copula model at various correlation ρ. Parameter settings: $h_i = 0.1$, $r = 0.1$, $T = 2$, and $d = 64$.

Table 1. Estimated second moment of various estimators.

K	$\hat{m}$	$\hat{m}_{CV}$	$\hat{m}_{CV,B}$	$\hat{m}_{CV,C}$	$\hat{m}_{CV,L}$
$\sigma = 0.1$					
20	894.390000	885.439000	885.241000	885.438000	885.436000
25	633.424000	625.105000	625.025000	625.104000	625.104000
30	418.454000	409.963000	409.887000	409.963000	409.963000
35	248.385000	240.041000	240.006000	240.041000	240.041000
40	123.652000	115.321000	115.305000	115.321000	115.321000
45	44.076300	35.943500	35.938100	35.943400	35.943200
50	7.973570	3.407570	3.406840	3.407550	3.407520
55	0.485918	0.031174	0.031034	0.031070	0.031070
60	0.007085	0.000141	0.000042	0.000047	0.000043
$\sigma = 0.3$					
20	964.385000	885.718000	885.523000	885.715000	885.714000
25	701.605000	625.445000	625.368000	625.445000	625.440000
30	487.272000	410.385000	410.327000	410.385000	410.385000
35	316.185000	241.255000	241.218000	241.255000	241.254000
40	191.332000	120.791000	120.768000	120.790000	120.789000
45	104.588000	49.382300	49.364300	49.380800	49.378300
50	53.020100	16.252900	16.238600	16.249700	16.246400
55	24.495500	4.391920	4.378640	4.387190	4.385170
60	11.008400	1.023860	1.011640	1.014960	1.014740

4.5. Control variates

In this subsection, we present numerical comparisons using geometric Asian options as control variates and associated importance sampling estimators. The discounted payoff function of the geometric Asian option is

$$\wp_g(X) = \mathrm{e}^{-rT}\left(\sqrt[d]{S_{t_1}(X)\cdots S_{t_d}(X)} - K\right)^+.$$

Standard algebra gives

$$\log\left(\sqrt[d]{S_{t_1}(X)\cdots S_{t_d}(X)}\right) \sim N(\mu, s^2),$$

where $\mu = \log S_0 + (r - \sigma^2/2)(d+1)\Delta/2$ and $s^2 = \sigma^2 (d+1)(2d+2)\Delta/(6d)$. Therefore, the closed-form solution of the price of the geometric Asian option is

$$g = \mathrm{e}^{-rT}\left(\mathrm{e}^{\mu+s^2/2}\bar{\Phi}(d_1) - K\bar{\Phi}(d_2)\right),$$

where $d_1 = (\log K - s^2 - \mu)/s$, $d_2 = (\log K - \mu)/s$ and $\bar{\Phi}(\cdot)$ is the survival function of the standard normal distribution.

The estimator of the arithmetic Asian option by using the geometric Asian option as the control variate is

$$\hat{m}_{CV} = \wp(X) - c(\wp_g(X) - g), \tag{17}$$

where $c = -\mathrm{Cov}(\wp(X), \wp_g(X))/\mathrm{Var}(\wp_g(X))$. For the rest of our analysis, c is estimated by Monte Carlo simulation in a pilot study and is considered as a fixed number. Because $\hat{m}_{CV}$ is a function of normal random vector X, it is straightforward to consider its associated importance estimators. The associated importance sampling estimators using optimal tilting formulas θ^{B*}, θ^{C*} and θ^{L*}, are denoted by $\hat{m}_{CV,B}$, $\hat{m}_{CV,C}$ and $\hat{m}_{CV,L}$, respectively.

Figure 4 depicts variance ratios and penalized variance ratios in base 10 logarithm against moneyness in pricing Asian options using control variates under the Black–Schole model for various σ. Figure 4 shows that applying importance sampling to the control variate gain greater improvements for out-of-money options than for in-the-money options.

Recall that section 4.3 concludes that $\hat{m}_B$ is the most competitive importance sampling estimator in pricing Asian options. A comparison between figures 2 and 4 finds that neither $\hat{m}_{CV}$ nor $\hat{m}_B$ dominates the others. Indeed, when $\sigma = 0.1$, $\hat{m}_{CV}$ outperforms $\hat{m}_B$ for in-the-money options, but $\hat{m}_B$ outperforms $\hat{m}_{CV}$ for out-of-the-money options.

Finally, $\hat{m}_{CV,B}$ seems less efficient than $\hat{m}_{CV,L}$ for in-the-money options. To explain why, we summarize estimated second moment of associated estimators in table 1. Let $M(\cdot)$ denote the second moment of an estimator.

5. Pricing interest rate caps

This section considers a two-factor Hull–White model for pricing interest caps. Section 5.1 presents the basic definitions. Section 5.2 proposes simplified tilting formulas. Numerical results are reported in section 5.3.

5.1. *A two-factor Hull–White model*

A two-factor Hull–White model (Hull and White 1994) is defined as

$$\mathrm{d}r_t = (u_t - ar_t)\mathrm{d}t + \sigma_1 \mathrm{d}w_t,$$
$$\mathrm{d}u_t = -bu_t \mathrm{d}t + \sigma_2 \mathrm{d}y_t,$$

where w_t and y_t are two standard Brownian motions with instantaneous correlation ρ. Let T be the time to maturity. To generate a path of interest rates at equally-spaced time points $t_i = i\Delta$ with $\Delta = 2T/d$ for $i = 1, \ldots, d/2$, we employ the first-order Euler discretization as follows,

$$r_{t_i}(X) = (u_{t_{i-1}}(X) - ar_{t_{i-1}}(X))\Delta + \sigma_1\sqrt{\Delta}W_i,$$
$$u_{t_i}(X) = -bu_{t_{i-1}}(X)\Delta + \sigma_2\sqrt{\Delta}Y_i,$$

where $X = (W_1, \ldots, W_{d/2}, Y_1, \ldots, Y_{d/2})' \sim N_d(\mathbf{0}, \Sigma)$. Here, $\Sigma = [\Sigma_{ij}]$ is a $d \times d$ covariance matrix, with $\Sigma_{ii} = 1$ for $i = 1, \ldots, d$, and $\Sigma_{d/2+i,i} = \Sigma_{i,d/2+i} = \rho$ for $i = 1, \ldots, d/2$.

An interest rate cap pays $(r_{t_{i-1}} - K)^+$ at time t_i for $i = 1, \ldots, d/2$ subject to strike price K, and has the discounted payoff function

$$\wp(X) = \sum_{i=1}^{d/2} \exp\left[-\Delta\sum_{j=0}^{i} r_{t_i}(X)\right]\left(r_{t_i}(X) - K\right)^+.$$

5.2. *Simplified tilting formulas*

In addition to the simplified linear and constant tilting formulas, because r_t and u_t are generated by W and Y, respectively, we further consider piecewise constant and linear tiling formulas. Specifically, let

$$\theta^{C2} := \theta^{C2}(\alpha) = (\alpha_1, \ldots, \alpha_1, \alpha_2, \ldots, \alpha_2)' = A\alpha,$$

where $\alpha = (\alpha_1, \alpha_2)'$, and $A = [A_{ij}]$ is a $d \times 2$ matrix with $A_{i1} = 1$ for $i = 1, \ldots, d/2$, $A_{i2} = 1$ for $i = d/2 + 1, \ldots, d$, and $A_{ij} = 0$ otherwise. The '2' in the superscript indicates that we set two piecewise constant approximations to the tilting formula. Intuitively, θ^{2C} set the first $2/d$ variates to be α_1 and the last $2/d$ variates to be α_2. Let α^* denote the optimal α that minimizes the variance of the associated importance sampling estimator. The optimal piecewise constant tilting formula is $\theta^{C2*} = \theta^{C2}(\alpha^*)$.

Similarly, we set a piecewise linear approximation to the tiling formula,

$$\theta^{L2} := \theta^{L2}(\beta) = H\beta,$$

where $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)'$ and H is a $d \times 4$ matrix, with $H_{i1} = 1$ for $i = 1, \ldots, d/2$, $H_{i2} = i - 1$ for $i = 1, \ldots, d/2$, $H_{i3} = 1$ for $i = d/2 + 1, \ldots, d$, $H_{i4} = (i - d/2 - 1)$ for $i = d/2 + 1, \ldots, d$, and $H_{ij} = 0$ otherwise. In plain words, θ^{L2} set two different linear approximations for $i = 1, \ldots, d/2$ and for $i = d/2 + 1, \ldots, d$, respectively. Let β^* denote the optimal β that minimizes the variance of associated importance sampling estimator. Then the optimal tilting formula for a piecewise linear approximation is $\theta^{L2*} = \theta^{L2}(\beta^*)$. For ease of notations, the importance sampling estimator based on θ^{C2*} and θ^{L2*} are denoted by $\hat{m}_{C2}$ and $\hat{m}_{L2}$, respectively.

We remark that an arbitrary functional form (higher order polynomials for instance) can be set to approximate the tilting formula. Or, a suitable piecewise linear approximation with more variables can be modified in pricing interest rate caps under a more sophisticated multi-factor model. A merit of employing a functional form is the substantial dimension reduction of the tilting parameter. And, this partially solves the numerical instability of $\hat{m}_B$ due to the Monte Carlo discrepancy in the searching stage.

5.3. *Numerical results*

Figure 5 depicts the variance ratio and penalized variance ratio of five importance sampling estimators: $\hat{m}_B$, $\hat{m}_C$, $\hat{m}_{C2}$, $\hat{m}_L$ and $\hat{m}_{L2}$ in base 10 logarithm against moneyness for various ρ. In terms of variance ratios, when $\rho = 0$ and the cap is in-the-money, $\hat{m}^{L2}$ outperforms other estimators, followed by $\hat{m}^L$, $\hat{m}^{C2}$, $\hat{m}^C$ and $\hat{m}^B$. But, when the cap is out-of-the-money, $\hat{m}^B$ outperforms other estimators. In terms of penalized variance ratios, θ^{L2} is the most competitive. The same conclusions can be drawn when $\rho = 0.6$.

The reason that $\hat{m}_{L2}$ and $\hat{m}_L$ performs well is explained by investigating the pattern of the optimal tilting formula given in figure 6. Although $\hat{m}_B$ is expected to produce the least variance, inevitable fluctuations of θ^{B*} caused by Monte Carlo discrepancy in the searching stage may deteriorate its efficiency.

Table 2. Estimated variance ratios of various swaps defined in Chen and Glasserman (2008) of different estimators.

T	0.5	1	2	3	4	5	10	15	20	30
Swap A1										
CP	1613.5	757.0	324.5			78.6	17.9	6.5	3.2	1.5
$\hat{\theta}_B$	1.4	1.4	1.3			1.2	1.2	1.2	1.3	1.4
$\hat{\theta}_C$	1.2	1.2	1.2			1.1	1.1	1.1	1.2	1.2
$\hat{\theta}_L$	1.3	1.2	1.2			1.2	1.1	1.2	1.2	1.3
Swap A2										
CP				10.5	7.9	6.5	3.9	3.1	2.2	1.2
$\hat{\theta}_B$				6.0	4.6	3.8	2.5	2.3	2.3	3.0
$\hat{\theta}_C$				5.6	4.3	3.5	2.4	2.2	2.2	2.8
$\hat{\theta}_L$				5.7	4.4	4.6	2.4	2.2	2.3	2.9
Swap A3										
CP		2.2	1.9			1.6	1.4	1.3	1.2	1.1
$\hat{\theta}_B$		13.7	7.6			4.0	2.9	2.6	2.7	3.3
$\hat{\theta}_C$		13.5	7.5			4.0	2.8	2.6	2.7	3.3
$\hat{\theta}_L$		13.5	7.5			4.0	2.8	2.6	2.7	3.3
Swap A4										
CP	32.3	25.0	20.7			16.8	10.1	5.1	2.8	1.4
$\hat{\theta}_B$	10.2	7.1	5.2			3.4	2.7	2.5	2.6	2.9
$\hat{\theta}_C$	6.2	5.8	4.2			2.9	2.4	2.3	2.3	2.6
$\hat{\theta}_L$	10.3	7.1	5.1			3.4	2.7	2.5	2.6	2.9
Swap A5										
CP	5.3	4.4	3.6			3.0	2.5	2.1	1.8	1.4
$\hat{\theta}_B$	29.4	16.5	9.8			5.3	3.8	3.4	3.4	3.9
$\hat{\theta}_C$	28.5	16.0	9.5			5.2	3.8	3.4	3.4	3.8
$\hat{\theta}_L$	29.4	16.5	9.8			5.3	3.8	3.4	3.4	3.9

6. Pricing *n*th-to-default swaps

Another example involves credit derivatives which have gained huge popularity in financial markets. Here, we consider the nth-to-default swap under the normal copula model. A widely used mechanism to build a joint distribution for the default times is through the normal copula model (Li 2000). Albeit its popularity and simplicity of the normal copula model for pricing credit derivatives, it is widely regarded as a model failed in the global financial crisis. However, to illustrate the applicability and competitiveness of our method, we include numerical studies in pricing basket default swaps based on the normal copula model in the following.

6.1. A simple payoff function

The nth-to-default swap gives the buyer one dollar if the nth default time is less than the maturity time T. Let $T_1, \ldots, T_d$ be the default time of the nth credit entity. As in Li (2000), each default time T_i is marginally exponentially distributed with rate λ_i, and a normal copula is used to link these default times. Specifically, the joint distribution function of $T_1, \ldots, T_d$ is

$$P\,(T_1 \le t_1, \ldots, T_d \le t_d) = C(F_1(t_1), \ldots, F_d(t_d); \Sigma), \tag{18}$$

where $F_i(t) = 1 - \exp^{-\lambda_i t}$ is the cumulative distribution function (cdf) of the exponential distribution with rate λ_i, and $C(u_1, \ldots, u_d; \Sigma)$ is the normal copula with correlation matrix Σ. Here, $\Phi(\cdot)$ is the cdf of the standard normal, and $\Phi^{-1}(\cdot)$ is the inverse function of $\Phi(\cdot)$.

Because $F_i^{-1}(u) = -\log(1-u)/\lambda_i$ is the inverse function of $F_i(\cdot)$, by change of variables, it is straightforward to obtain that default times defined in (18) have the following stochastic representation,

$$\begin{aligned}&(T_1(X_1), \ldots, T_d(X_d))\\ &\overset{d}{=} \left(-\frac{\log(1-\Phi(X_1))}{\lambda_1}, \ldots, -\frac{\log(1-\Phi(X_d))}{\lambda_d}\right),\end{aligned} \tag{19}$$

where $X \sim N_d(0, \Sigma)$. Let the order statistic $T_{(n)}$ denote the nth default time. The discounted payoff function of the nth-to-default swap considered here is

$$\wp(X) = e^{-rT_{(n)}} \mathbf{1}_{\{T_{(n)}<T\}}\,(T_1(X_1), \ldots, T_d(X_d)),$$

where $X \sim N_d(\mathbf{0}, \Sigma)$, and $T_i = -\log(1-\Phi(X_i))/\lambda_i$ for $i = 1, \ldots, d$.

To verify the sufficient condition for the existence of the optimal tilting parameter holds, note that $X = (X_1, \ldots, X_d)' \sim N_d(\mathbf{0}, \mathbf{I})$, which implies that $\Psi(\theta) = O(e^{\|\theta\|^2})$. $\wp(X) = e^{-rT_{(n)}} \mathbf{1}_{\{T_{(n)}<T\}}(X_1, \ldots, X_d)$ with $T_i = -\log(1-\Phi(X_i))/\lambda_i$ for $i = 1, \ldots, d$. This implies that $\bar{\Psi}(\theta) = O(e^{-\|\theta\|})$. Therefore $\Psi(\theta)\bar{\Psi}(\theta) \to \infty$ as $\|\theta\| \to \infty$.

For simplicity, we set the correlation matrix $\Sigma = [\Sigma_{ij}]$ to be $\Sigma_{ii} = 1$ and $\Sigma_{ij} = \rho$ for $i \ne j$ for $i, j = 1, \ldots, d$. Figure 7 provides estimated variance ratios and penalized variance ratios in base 10 logarithm against the default indicator in pricing the nth-to-default swap under the normal copula model for various ρ. Overall, the estimated variance ratio is larger as the default indicator n increases, possibly because a proper importance sampling shifts the mean of the sampling distribution and thus substantial improvement in efficiency is obtained. On the other hand, the relative efficiency is higher for smaller correlation. Our three importance sampling estimators produce similar variance ratios, possibly because θ^* has a saw-tooth horizontal trend in our unreported figures. When the computational time is taken into account, $\hat{m}_C$ is the most competitive.

6.2. *Comparisons with numerical examples in Chen and Glasserman (2008)*

Joshi and Kainth (2004) presents an innovative importance sampling method to price basket default swaps that forces all paths to produce at least n defaults. Chen and Glasserman (2008) improves and extends the Joshi-Kannth algorithm and proposes an alternative that is guaranteed to reduce variance. For comparison, we run five examples in Chen and Glasserman (2008) using our proposed importance sampling estimators, $\hat{m}_B$, $\hat{m}_C$ and $\hat{m}_L$. Detailed descriptions on the contract of the basket default swap and parameter settings can be seen in Capriotti (2008), and are omitted in this paper for brevity.

Table 2 reports estimated variance ratios of our importance sampling estimators and the conditional probability method (CP) in Chen and Glasserman (2008). To just investigate the effect of importance sampling, we do not compare the conditional probability with stratification (CPST) method in Chen and Glasserman (2008), because it extends the CP method with additional stratification. Estimated variance ratios of the CP method are duplicated from Chen and Glasserman (2008).

Table 2 indicates that neither our importance sampling estimators nor the CP method dominate the others under various parameter settings. In details, the CP method outperforms our importance sampling estimators for Swap A1, Swap A2, and Swap 4, whereas our methods outperform the CP method for Swap A3 and Swap A5. We therefore conclude that our method is a competitive alternative to the CP method.

7. Conclusion

This paper proposes an exponentially tilted importance sampling method based on the criterion to minimize the variance of the associated estimator. The contribution of this paper is threefold: (1) A theoretical foundation is built to guarantee the existence, uniqueness and characterization of the optimal tilting parameter. (2) The optimal tiling parameter can be searched for via an automatic Newton's method. (3) simplified yet competitive tilting formulas are further proposed to reduce heavy computational cost and numerical instability in high-dimensional cases. Extensive numerical examples in pricing financial derivatives confirm the above claims.

The current scope of this paper simply focuses on the multivariate normal distribution. It is of interest to extend the existing method to more complicated dynamic models, such as Lévy processes. It is also of interest to extend the proposed method to calculate the VaR under the incremental risk charge based on Basel III. Separate papers addressing these issues will be published elsewhere.

Acknowledgements

The authors are grateful for the two anonymous referees' valuable comments, which help to considerably improve the quality of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The research of Huei-Wen Teng is supported in part by [grant number MOST 103-2633-M-008-001], [grant number MOST 103-2410-H-008-039-MY2]; Cheng-Der Fuh is supported in part by [grant number MOST 103-2118-M-008-002-MY3], [grant number MOST 103-2410-H-008-039-MY2].

ORCID

Huei-Wen Teng http://orcid.org/0000-0002-3651-9934

References

Asmussen, S. and Glynn, P., *Stochastic Simulation: Algorithms and Analysis*, 2007 (Springer-Verlag: New York).
Boyle, P., Broadie, M. and Glasserman, P., Monte Carlo methods for security pricing. *J. Econ. Dyn. Control*, 1997, **21**, 1267–1321.
Bucklew, J.A., *Introduction to Rare Event Simulation*, 2004 (Springer-Verlag: New York).
Capriotti, L., Least-squares importance sampling for Monte Carlo security pricing. *Quant. Finance*, 2008, **8**, 485–497.
Chan, H.P. and Lai, T.L., Efficient importance sampling for Monte Carlo evaluation of exceedance probabilities. *Ann. Appl. Probab.*, 2007, **17**, 440–473.
Chen, Z. and Glasserman, P., Fast pricing of basket default swaps. *Oper. Res.*, 2008, **56**, 286–303.
Ellis, R.S., *Entropy, Large Deviations, and Statitical Mechanics*, 1985 (Springer: New York).
Fu, M.C., Madan, D.B. and Wang, T., Pricing continuous Asian options: A comparison of Monte Carlo and Laplace transform inversion methods. *J. Comput. Finance*, 1999, **2**, 49–74.
Fuh, C.-D. and Hu, I., Efficient importance sampling for events of moderate deviations with applications. *Biometrika*, 2004, **91**, 471–490.
Fuh, C.-D., Hu, I., Hsu, Y.-H. and Wang, R.-H., Efficient simulation of value at risk with heavy-tailed risk factors. *Oper. Res.*, 2011, **59**, 1395–1406.
Fuh C.D. and Teng H.W., Efficient importance sampling for false alarm probability under k-distributed sea clutter. Technical report, National Central University, 2015.
Givens, G.H. and Raftery, A.E., Local adaptive importance sampling for multivariate densities with strong nonlinear relationships. *J. Am. Stat. Assoc.*, 1996, **91**, 132–141.
Glasserman, P., *Monte Carlo Methods in Financial Engineering*, 2004 (Springer: New York).
Glasserman, P., Heidelberger, P. and Shahabuddin, P., Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Math. Finance*, 1999, **9**, 117–152.

Glasserman, P. and Wang, Y., Counterexamples in importance sampling for large deviations probabilities. *Ann. Appl. Probab.*, 1997, **7**, 731–746.
Guasoni, P. and Robertson, S., Optimal importance sampling with explicit formulas in continuous time. *Financ. Stoch.*, 2008, **12**, 1–19.
Hull, J. and White, A., Numerical procedures for implementing term structure models ii: Two-factor models. *J. Derivatives*, 1994, **2**, 37–48.
Joshi, M. and Kainth, D., Rapid and accurate development of prices and Greeks for *n*th-to-default credit swaps in the Li model. *Quant. Finance*, 2004, **4**, 266–275.
Kallenberg, W., On moderate deviation theory in estimation. *Ann. Stat.*, 1983, **11**, 498–504.
Kim Y.B., Roh D.S. and Lee M.Y., Nonparametric adaptive importance sampling for rare event simulation, Vol. 1, *Proceedings of the 2000 Winter Simulation Conference*, Orlando, FL, pp. 767–772, 2000.
Li, D., On default correlation: A copula function approach. *J. Fixed Income*, 2000, **9**, 43–54.
Neddermeyer, J.C., Non-parametric partial importance sampling for financial derivative pricing. *Quant. Finance*, 2011, **11**, 1193–1206.
Ross, S.M., *Simulation*, 2013 (Academic Press: New York).
Siegmund, D., Importance sampling in the Monte Carlo study of sequential tests. *Ann. Stat.*, 1976, **4**, 673–684.
Su Y. and Fu M.C., Importance sampling in derivative securities pricing, Vol. 1, *Proceedings of the 2000 Winter Simulation Conference*, Orlando, FL, pp. 587–596, 2000.
Su, Y. and Fu, M.C., Optimal importance sampling in securities pricing. *J. Comput. Finance*, 2002, **5**, 27–50.
Vazquez-Abad F.J. and Dufresne D., Accelerated simulation for pricing Asian options, Vol. 2, *Proceedings of the 1998 Winter Simulation Conference*, Washington, DC, pp. 1493–1500, 1998.
Zhang, P., Nonparametric importance sampling. *J. Am. Stat. Assoc.*, 1996, **91**, 1245–1253.

Appendix 1

Proof of Theorem 2.1 In the following, we first show that $G(\theta)$ is a strictly convex function. For any given $\lambda \in (0,1)$, and $\theta_1, \theta_2 \in \Theta$,

$$\begin{aligned}
&G(\lambda\theta_1 + (1-\lambda)\theta_2)\\
&= E_P\left[\wp^2(X)e^{-(\lambda\theta_1+(1-\lambda)\theta_2)'X+\psi(\lambda\theta_1+(1-\lambda)\theta_2)}\right]\\
&< E_P\Big[\wp^2(X)\big(\lambda e^{-\theta_1'X} + (1-\lambda)e^{-\theta_2'X}\big)\\
&\qquad + \wp^2(X)\big(\lambda e^{\psi(\theta_1)} + (1-\lambda)e^{\psi(\theta_2)}\big)\Big]\\
&= E_P\left[\lambda\wp^2(X)\big(e^{-\theta_1'X}\cdot e^{\psi(\theta_1)}\big) + (1-\lambda)\wp^2(X)\big(e^{-\theta_2'X}\cdot e^{\psi(\theta_2)}\big)\right]\\
&= \lambda E_P\left[\wp^2(X)e^{-\theta_1'X+\psi(\theta_1)}\right] + (1-\lambda)E_P\left[\wp^2(X)e^{-\theta_2'X+\psi(\theta_2)}\right]\\
&= \lambda G(\theta_1) + (1-\lambda)G(\theta_2).
\end{aligned}$$

Next, we will prove the existence of θ in the optimization problem (5). To get the global minimum of $G(\theta)$, we apply the standard result: *Let $f : \Re^d \to \Re$ be continuous on all of $\Re^d$. If f is coercive (in the sense that $f(\mathbf{x}) \to \infty$ if $\|\mathbf{x}\| \to \infty$), then f has at least one global minimizer*.

Note that $G(\theta)$ is strictly convex from above argument, and $\frac{\partial G(\theta)}{\partial \theta_i}$ exists for $i = 1, \ldots, d$. Then apply Theorem VI.3.4. of Ellis (1985), which states that $G(\theta)$ *is differentiable at* $\theta \in int(\Theta)$ *if and only if the d partial derivatives* $\frac{\partial G(\theta)}{\partial \theta_i}$ *for $i = 1, \ldots, d$ exist at $\theta \in int(\Theta)$ and are finite*. We have $G(\theta)$ is continuous for $\theta \in \Re^d$. By the definition of $G(\theta)$ in (5), it is easy to see that $\bar{\Psi}(\theta)\Psi(\theta) \to \infty$ as $\|\theta\| \to \infty$ implies that $G(\theta)$ is coercive. The uniqueness comes form the fact that $G(\theta)$ is strictly convex.

To prove (8), we need to simplify the right hand side of (6) under $\bar{Q}_\theta$. Standard algebra gives

$$\begin{aligned}
&\frac{E_P[\wp^2(x)X_i e^{-\theta'X}]}{E_P[\wp^2(x)e^{-\theta'X}]}\\
&\quad = \frac{\int \wp^2(x)x_i e^{-\theta'x}\,dP}{\int \wp^2(x)e^{-\theta'x}\,dP} = \int x_i\,d\bar{Q}_\theta = E_{\bar{Q}_\theta}[X_i],
\end{aligned}$$

for $i = 1, \ldots, d$. This implies the desired result. □