Efficient Importance Sampling Estimation for Joint Default Probability: the First Passage Time Problem

Chuan-Hsiang Han *

August 2, 2010

Abstract

Motivated from credit risk modeling, this paper extends the two-dimensional first passage time problem studied by Zhou (2001) to any finite dimension by means of Monte Carlo simulation. We provide an importance sampling method to estimate the joint default probability, and apply the large deviation principle to prove that the proposed importance sampling is asymptotically optimal. Our result is an alternative to the interacting particle systems proposed by Carmona, Fouque, and Vestal (2009).

1 Introduction

Estimation of the joint default probability under a structural-form models emerged pretty early in the presence of stochastic financial theory. In models of Black and Scholes [3] and Merton [15], default can only happen at expiration of debt when its issuer's asset value is less than the debt value. Black and Scholes modeled the asset value process by a geometric Brownian motion, then Merton incorporated an additional compound Poisson jump term. Black and Cox [4] generalized these models by allowing that default can occur at any time before the expiration of debt. They considered a first passage time

*Department of Quantitative Finance, National Tsing Hua University, Hsinchu, Taiwan, 30013, ROC, chhan@mx.nthu.edu.tw. Work supported by NSC 97-2115-M-007-002-MY2, Taiwan. We are also grateful to one anonymous referee and Professor Nicolas Privault . Other acknowledgments: NCTS, National Tsing-Hua University; TIMS, National Taiwan University; CMMSC, National Chiao-Tung University.

problem for the geometric Brownian motion in one dimension. Zhou [16] extended this one-dimensional geometric Brownian motion to the jump-diffusion model as Merton did. Later in Zhou [17], the joint default for two-dimensional geometric Brownian motions was treated. A comprehensive technical review can be found in [5].

In this paper, we focus on generalizing the joint default from a two-dimensional first passage time problem studied in [17] to higher dimensions through a Monte Carlo study. A high-dimensional setup of the first passage time problem under correlated geometric Brownian motions is the following. We assume that each firm value process, S_{it} $1 \leq i \leq n$, has the dynamics,

$$dS_{it} = \mu_i S_{it} dt + \sigma_i S_{it} dW_{it}, \tag{1}$$

where σ_i is a constant volatility and Brownian motions W_is are correlated by $d\left\langle W_i, W_j \right\rangle_t = \rho_{ij} dt$ with $< \cdot, \cdot >_t$ the quadratic variation at time t. Each firm also has a constant default barrier B_i for $1 \leq i \leq n$, and its default happens at the first time when the asset value S_{it} falls below the barrier level. Therefore, default time τ_i of the i-th firm is defined as

$$\tau_i = \inf\{t \geq 0 : S_{it} \leq B_i\}. \tag{2}$$

Let $\mathcal{F}$ be the filtration generated by all $S_{i\cdot}, i = 1, \cdots, n$, under a probability measure P. At time 0, the joint default probability with a terminal time T is defined by

$$DP = E\left\{\Pi_{i=1}^n \mathbf{I}\left(\tau_i \leq T\right) | \mathcal{F}_0 \right\}. \tag{3}$$

In general, there is no closed-form solution for the probability of joint default (3) in high dimension, so one has to rely on numerical methods. Using a deterministic approach such as numerical PDE (partial differential equation) or binomial tree to calculate the joint default probability typically suffers from the curse of dimensionality. Due to the property of dimension independence, Monte Carlo simulation becomes a feasible approach. Moreover, it is crucial to reduce the variance of Monte Carlo estimators for convergence improvement.

Recently, Carmona, Fouque, and Vestal [8] study a first passage time problem and estimate the loss density function for a credit portfolio under a stochastic volatility model. They use interacting particle systems for variance reduction. Alternatively, we propose an efficient Monte Carlo method which incorporates importance sampling for variance reduction in order to accurately estimate joint default probabilities under the classical Black-Cox model in high dimension. Han et al. [12] apply the same importance sampling for risk management applications with empirical studies and backtesting under stochastic volatility models.

There are many ways to analyze the variance of an importance sampling estimator. In a strong sense, one can minimize the variance, say over a parametrized space, by solving an optimization problem. See for example [1, 2], in which authors proposed an adaptive scheme, namely Robbins Monro algorithms, to solve a class of importance sampling estimators. They utilize constant change of drift in high dimension and solve nontrivial optimization problems. See also [10] for using an asymptotic result to approximate the optimal change of measure. In a weak sense, one seeks to minimize asymptotically the rate of variance, instead of the variance itself, of an importance sampling estimator. To demonstrate fundamental ideas, we provide an introduction in Section 2 that include constructions of importance sampling schemes for the standard normal random variable, and an asymptotic variance analysis of our proposed estimator by means of the large deviation principle. Details about the general definition of rate of variance and its analysis can be found in Chapter 5 of [7].

In this entire paper, we adopt variance analysis in the weak sense. By means of the large deviation principle, we show that the proposed importance sampling estimator for the one-dimensional first passage time problem has a zero variance rate. That implies that our proposed importance sampling scheme is asymptotically optimal.

The organization of this paper is as follows: Section 2 provides a fundamental understanding of importance sampling in a simple Gaussian random variable model with a spatial scale. Approximation to variance rate of the efficient importance sampling estimator can be obtained by an application of Cramer's theorem. Section 3 explores the high-dimensional first passage time model, known as the Black-Cox model. We show that the proposed importance sampling method is asymptotically optimal (or called efficient) in one dimension by an application of Freidlin-Wentzell Theorem. Then we conclude in Section 4.

2 Efficient Importance Sampling for a One-Dimensional Toy Model

We start from a simple and static model to estimate the probability of default defined by $P_1^c := E\{\mathbf{I}\,(X > c)\}$, where the standard normal random variable $X \sim \mathcal{N}(0,1)$ and $c > 0$ stand for the loss of a portfolio and its loss threshold, respectively. Of course, P_1^c is simply a tail probability and admits a closed-form solution $\mathcal{N}(-c)$, where $\mathcal{N}(x)$

denotes the cumulative normal integral function. Using basic Monte Carlo method to estimate P_1^c is not accurate enough particularly when the loss threshold c is large. This is because the variance of $\mathbf{I}\,(X > c)$ is of the same order of the default probability P_1^c when a rare event occurs. For example, $E\{(\mathbf{I}\,(X > c))^2\} - (P_1^c)^2 \approx P_1^c(\ll 1)$ when c is large enough. This implies that the relative error of the basic Monte Carlo estimator is rather large. A problem of rare event simulation arises when the spatial scale c is sufficiently large.

Applying importance sampling as a variance reduction method to treat rare event simulation has been extensively studied [7]. A general procedure, known as exponential change of measure, for constructing an importance sampling scheme in order to estimate P_1^c is reviewed below.

Assume that the density function of a real-valued random variable X is $f(x) > 0$ for each $x \in \Re$. One can change the probability measure by incorporating a likelihood function $f(x)/f_\mu(x)$ so that the default probability P_1^c can be evaluated under the new probability measure P_μ:

$$E\left\{\mathbf{I}(X > c)\right\} = E_\mu \left\{ \mathbf{I}(X > c) \frac{f(X)}{f_\mu(X)} \right\},$$

where the new density function of X is $f_\mu(x) > 0$ for each $x \in \Re$. The twist or tilted probability measure refers to the choice of a new density

$$f_\mu(x) = \frac{exp(\mu\, x)\, f(x)}{M(\mu)}, \tag{4}$$

where $M(\mu) = E[\exp(\mu\, X)]$ denotes the moment generating function of X. We define the leading term of the estimator variance as $P_2^c(\mu) := E_\mu \left\{ \mathbf{I}(X > c) \frac{f^2(X)}{f_\mu^2(X)} \right\}$. By substituting $f_\mu(x)$ given above into $P_2^c(\mu)$, one can obtain the following results:

$$\begin{aligned} P_2^c(\mu) &= E \left\{ \mathbf{I}(X > c) \frac{f(X)}{f_\mu(X)} \right\} \\ &= M(\mu) E\left\{\mathbf{I}(X > c)\, \exp(-\mu\, X)\right\} \\ &\leq M(\mu)\exp(-\mu\, c), \end{aligned} \tag{5}$$

where μ and c are assumed positive numbers for this upper bound to hold. To minimize the logarithm of this upper bound, its first order condition satisfies

$$\begin{aligned} \frac{d \ln(M(\mu)\, \exp(-\mu\, c))}{d\mu} &= \frac{M'(\mu)}{M(\mu)} - c \\ &= 0. \end{aligned}$$

Let $\mu^\star$ solve for $\frac{M'(\mu^\star)}{M(\mu^\star)} = c$. It follows that the expected value of X under the new probability measure $P_{\mu^\star}$ is exactly the loss threshold c. This is confirmed by evaluating $E_{\mu^\star}(X) = \int x\, f_{\mu^\star}(x)\, dx$ and substituting $f_{\mu^\star}(x)$ defined in (4) so that

$$
\begin{aligned}
E_{\mu^\star}(X) &= \frac{M'(\mu^\star)}{M(\mu^\star)} \\
&= c. \qquad (6)
\end{aligned}
$$

Remark that this whole procedure of exponential change of measure can be extended to high dimension.
From the simulation point of view, the final result shown in (6) is appealing because the rare event of default under the original probability measure is no longer rare under this new measure $P_{\mu^\star}$. In addition, even in the situation that moment generating functions are difficult or impossible to find, one can still possibly use other change of measure techniques to fulfill (6), i.e., "the expected value of a defaultable asset is equal to its debt value" in financial terms. We will see such example in Section 3 for high-dimensional Black-Cox model.
In the concrete case of X being a standard normal random variable, the minimizer $\mu^\star$ is exactly equal to c. This result is derived from a direct calculation of $c = M'(\mu^\star)/M(\mu^\star)$ given that the moment generating function of X is $M(\mu) = \exp(\mu^2/2)$. The twist or tilted density function $f_{\mu^\star}(x) = exp(\mu^\star x)\, f(x)/M(\mu^\star)$ becomes $\exp(-(x-c)^2/2)/\sqrt{2\,\pi}$. Hence, random samplings are generated from $X \sim \mathcal{N}(c,1)$ under this new density function, instead of $X \sim \mathcal{N}(0,1)$ under the original measure. The default probability P_1^c can be explicitly expressed by

$$
\begin{aligned}
P_1^c &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathbf{I}\,(x > c)\, \frac{e^{-x^2/2}}{e^{-(x-c)^2/2}} e^{-(x-c)^2/2} dx \\
&= E_c \left\{ \mathbf{I}\,(X > c)\; e^{c^2/2 - cX} \right\}, \qquad (7)
\end{aligned}
$$

and its second moment becomes

$$
P_2^c(c) \;:=\; E_c \left\{ \mathbf{I}\,(X > c)\; e^{c^2 - 2cX} \right\}. \qquad (8)
$$

Naturally, one can ask an optimization problem which minimizes variance of all possible importance sampling estimators associated with $f_\mu(x) = \exp(-(x-\mu)^2/2)/\sqrt{2\,\pi}$ for each $\mu \in \Re$. That is, given that

$$
\begin{aligned}
P_1^c &= E_\mu \left\{ \mathbf{I}\,(X > c)\, \frac{f(X)}{f_\mu(X)} \right\} \\
&= E_\mu \left\{ \mathbf{I}\,(X > c)\; e^{\mu^2/2 - \mu X} \right\}, \qquad (9)
\end{aligned}
$$

we seek to minimize its second moment

$$P_2^c(\mu) = E_\mu \left\{ \mathbf{I}\left(X > c\right) e^{\mu^2 - 2\mu X} \right\}. \tag{10}$$

The associated minimizer guarantees the minimal variance within the μ-parametrized measures but solving the minimizer via (10) requires numerical computation. In Section 2.2, we compare variance reduction performance between the optimal estimator (9) by minimizing (10), and the efficient estimator (7). We will see that in our numerical experiments these two estimators reach the same level of accuracy, but the computing times are very different. Efficient estimators perform more effectively then the optimal estimators.

2.1 Asymptotic Variance Analysis by Large Deviation Principle

It is known that the number of simulation to estimate a quantity in a certain level of accuracy should be proportional to $P_2^c(\mu)/\left(P_1^c\right)^2 - 1$. See for example Section 4.5 in [7]. If the decay rate of $P_2^c(\mu)$ and $\left(P_1^c\right)^2$ are the same asymptotically, we say that the asymptotic variance rate is zero and the corresponding importance sampling estimator is asymptotically optimal or efficient. In this section, we aim to prove that when the density parameter μ is particularly chosen as the default loss threshold c, the variance reduced by the importance sampling scheme defined in (7) is asymptotically optimal by an application of Cramer's theorem in large deviation theory. Recall that

Theorem 1. *(Cramer's theorem [7]) Let $\{X_i\}$ be real-valued i.i.d. random variables under P and $EX_1 < \infty$. For any $x \geq E\{X_1\}$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \ln P\left(\frac{S_n}{n} \geq x\right) = -\inf_{y \geq x} \Gamma^*(y), \tag{11}$$

where $S_n = \sum_{i=1}^n X_i$ denotes the sample sum of size n, $\Gamma(\theta) = \ln E\left\{e^{\theta X_1}\right\}$ denotes the cumulant function, and $\Gamma^*(x) = \sup_{\theta\in\Re}\left[\theta\, x - \Gamma(\theta)\right]$.

From this theorem and the moment generating function $E\left\{\exp(\theta X)\right\} = \exp(\theta^2/2)$ for X being a standard normal univariate, we obtain the following asymptotic approximations.

Lemma 1. *(Asymptotically Optimal Importance Sampling) When c approaches infinity, the variance rate of the estimator, defined in (7), approaches zero. That is,*

$$\lim_{c\to\infty} \frac{1}{c^2} \ln P_2^c(c) = 2 \lim_{c\to\infty} \frac{1}{c^2} \ln P_1^c = -1.$$

Therefore, this importance sampling is asymptotically optimal or efficient.

Proof: Given that $X_i, i = 1, 2. \cdots$ are i.i.d. one-dimensional standard normal random variables, it is easy to obtain

$$\lim_{n\to\infty} \frac{1}{n} \ln P\left(\frac{\sum_{i=1}^n X_i}{n} \geq x\right) = -\frac{x^2}{2},$$

or equivalently $P\left(\frac{\sum_{i=1}^n X_i}{n} \geq x\right) \approx \exp(-n\frac{x^2}{2})$ by an application of Theorem 1. Introduce a rescaled default probability $P\left(X \geq \sqrt{n}x\right)$ for n large, then

$$\begin{aligned} P\left(X \geq \sqrt{n}x\right) &= P\left(\frac{\sum_{i=1}^n X_i}{\sqrt{n}} \geq \sqrt{n}x\right) \\ &= P\left(\frac{\sum_{i=1}^n X_i}{n} \geq x\right), \end{aligned}$$

in which each random variable X_i has the same distribution as X. Hence, let $1 \ll c := \sqrt{n}x$, the approximation to the first moment of $\mathbf{I}(X \geq c)$ or default probability P_1^c is obtained:

$$\lim_{n\to\infty} \frac{1}{\left(\sqrt{n}x\right)^2} \ln P\left(X \geq \sqrt{n}x\right) = -\frac{1}{2},$$

or equivalently

$$P\left(X \geq c\right) \approx \exp(-\frac{c^2}{2}). \tag{12}$$

Given the second moment defined in (10), it is easy to see that

$$\begin{aligned} P_2^c(\mu) &= E_{-\mu}\left\{\mathbf{I}\left(X > c\right)\right\} e^{\mu^2} \\ &= E_0\left\{\mathbf{I}\left(X > \mu + c\right)\right\} e^{\mu^2}. \end{aligned}$$

The first line is obtained by changing measure via $dP_\mu/dP_{-\mu}$, and the second line shifts the mean value of $X \sim \mathcal{N}(-\mu, 1)$ to 0. With the choice $\mu = c$, we get $P_2^c(c) = E_0\left\{\mathbf{I}\left(X > 2\,c\right)\right\} e^{c^2}$ and its approximation $P_2^c(c) \approx \exp\left(-c^2\right)$ can be easily obtained from the same derivation as in (12). As a result, we verify that the decay rate of the second moment $P_2^c(c)$ is two times the decay rate of the probability P_1^c, i.e.,

$$\lim_{c\to\infty} \frac{1}{c^2} \log P_2^c(c) = 2 \lim_{c\to\infty} \frac{1}{c^2} \log P_1^c = -1. \tag{13}$$

□

We have shown that the proposed importance sampling scheme defined in (7) is efficient. This zero variance rate should be understood as an optimal variance reduction in an asymptotic sense because the variance rate cannot be less than zero.

Remark: This lemma can be generalized to any finite dimension and possible extended to other generalized distribution such as Student's t. We refer to [14] for further details with applications in credit risk.

For our thematic topic of estimating joint default probabilities under high-dimensional Black-Cox model, this technique unfortunately does not work because the moment generating function of multivariate first passage times is unknown. We overcome this difficulty by considering a simplified problem, namely change measure for the joint distributions of underlying processes at maturity rather than their first passage times. Remarkably, we find that this measure change can still be proven asymptotically optimal for the original first passage time problem. Details can be found in Section 3.

2.2 Numerical Results

Table 1 demonstrates performance of two importance sampling schemes, including optimal estimator and efficient estimator, to estimate the default probability $P(X > c)$ for $X \sim \mathcal{N}(0,1)$ with various loss threshold values c. In Column 2, exact solutions of $\mathcal{N}(-c)$ are reported. In each column of simulation, Mean and SE stand for the sample mean and the sample standard error, respectively. IS($\mu = c$) represents the scheme in (7) using the pre-determined choice of $\mu = c$ suggested from the asymptotic analysis in Lemma 1, while IS($\mu = \mu^\star$) represents the optimal scheme in (7) using $\mu = \mu^\star$, which minimizes $P_2^c(\mu)$ numerically. We observe that standard errors obtained from these two importance sampling schemes are comparable in terms of the same order of accuracy, while the computing time is not. From the last row, the optimal importance sampling scheme IS($\mu = \mu^\star$) takes about 50 times more than the efficient importance sampling scheme IS($\mu = c$). These numerical experiments are implemented in Matlab on a laptop PC with 2.40GHz Intel Duo CPU T8300.
There have been extensive studies and applications, see Chapter 9 of [11] for various applications in risk management, by using this concept of minimizing an upper bound of the second moment under a parametrized twist probability, then construct an importance sampling scheme. However, it remains to check whether this scheme is

Table 1: Estimation of default probability $P(X > c)$ with different loss threshold c when $X \sim \mathcal{N}(0,1)$. The total number of simulation is 10,000.

	DP	Basic MC		IS($\mu = c$)		IS($\mu = \mu^\star$)	
c	true	$Mean$	SE	$Mean$	SE	$Mean$	SE
1	0.1587	0.1566	0.0036	0.1592	0.0019	0.1594	0.0018
2	0.0228	0.0212	0.0014	0.0227	3.49E-04	0.0225	3.37E-04
3	0.0013	1.00E-03	3.16E-04	0.0014	2.53E-05	0.0014	2.51E-05
4	3.17e-05	-	-	3.13E-05	6.62E-07	3.11E-05	6.66E-07
time		0.004659		0.020904		1.060617	

efficient (with zero variance rate) or not.

3 Efficient Importance Sampling for High-Dimensional First Passage Time Problem

In this section, we review an importance sampling scheme developed by Han and Vestal [13] for the first passage time problem (3) in order to improve the convergence of Monte Carlo simulation. In addition, we provide a variance analysis to justify that the importance sampling scheme is asymptotic optimal (or efficient) in one dimension.
The basic Monte Carlo simulation approximates the joint default probability defined in (3) by the following estimator

$$DP \approx \frac{1}{N} \sum_{k=1}^{N} \Pi_{i=1}^{n} \mathbf{I}\left(\tau_i^{(k)} \leq T\right), \tag{14}$$

where $\tau_i^{(k)}$ denotes the k-th i.i.d. sample of the i-th default time defined in (2) and N denotes the total number of simulation.
By Girsanov Theorem, one can construct an equivalent probability measure $\tilde{P}$ defined by the following Radon-Nikodym derivative

$$\frac{dP}{d\tilde{P}} = Q_T(h.) = \exp\left(\int_0^T h(s,S_s)\cdot d\tilde{W}_s - \frac{1}{2}\int_0^T ||h(s,S_s)||^2 ds\right), \tag{15}$$

where we denote by $S_s = (S_{1s}, \cdots, S_{ns})$ the state variable (asset value process) vector and $\tilde{W}_s = \left(\tilde{W}_{1s}, \cdots, \tilde{W}_{ns}\right)$ the vector of standard

Brownian motions, respectively. The function $h(s, S_s)$ is assumed to satisfy Novikov's condition such that $\tilde{W}_t = W_t + \int_0^t h(s, S_s)ds$ is a vector of Brownian motions under $\tilde{P}$.
The importance sampling scheme proposed in [13] is to select a constant vector $h = (h_1, \cdots, h_n)$ which satisfies the following n conditions

$$\tilde{E}\{S_{iT}|\mathcal{F}_0\} = B_i, i = 1, \cdots, n. \tag{16}$$

These equations can be simplified by using the explicit log-normal density of S_{iT}, so we deduce the following sequence of linear equations for h_i's:

$$\Sigma_{j=1}^{i} \rho_{ij} h_j = \frac{\mu_i}{\sigma_i} - \frac{\ln B_i/S_{i0}}{\sigma_i T}, i = 1, \cdots, n. \tag{17}$$

If the covariance matrix $\Sigma = (\rho_{ij})_{1\leq i,j,\leq n}$ is non-singular, the vector h exists uniquely so that the equivalent probability measure $\tilde{P}$ is uniquely determined. The joint default probability defined in (3) becomes

$$DP = \tilde{E}\left\{\Pi_{i=1}^{n} \mathbf{I}\left(\tau_i \leq T\right) Q_T(h)|\mathcal{F}_0\right\}. \tag{18}$$

Equation (16) requires that, under the new probability measure $\tilde{P}$, the expectation of asset's value at time T is equal to its debt level. When the debt level B of a company is much smaller than its initial asset value S_0 (see examples in Table 2), or returns of any two names are highly negative correlated (see examples in Table 4), joint default events are rare. By the proposed importance sampling scheme, random samples drawn under the new measure $\tilde{P}$ cause more defaults than those samples drawn under P.
Table 2 and Table 4 illustrate numerical results for estimating the (joint) default probabilities of a single-name case and a three-name case. The exact solution of the single name default probability

$$1 - \mathcal{N}(d_2^+) + \mathcal{N}(d_2^-)\left(\frac{S_0}{B}\right)^{1-2\mu/\sigma^2} \tag{19}$$

with $d_2^{\pm} = \frac{\pm\ln(S_0/B)+(\mu-\sigma^2/2)T}{\sigma\sqrt{T}}$ can be found in [8]. This result is obtained from the distribution of the running minimum of Brownian motion. However, there is no closed-form solution for the joint default probability of three names in Table 4 except for the case of zero correlation.

Table 2: Comparison of single-name default probability by basic Monte Carlo (BMC), exact solution, and importance sampling (IS). The number of simulation is 10^4 and an Euler discretization for (1) is used by taking time step size $T/400$, where T is one year. Other parameters are $S_0 = 100, \mu = 0.05$ and $\sigma = 0.4$. Standard errors are shown in parenthesis.

B	BMC	Exact Sol	IS
50	0.0886 (0.0028)	0.0945	0.0890 (0.0016)
20	- (-)	$7.7310 * 10^{-5}$	$7.1598 * 10^{-5}(2.3183 * 10^{-6})$
1	- (-)	$1.3341 * 10^{-30}$	$1.8120 * 10^{-30}(3.4414 * 10^{-31})$

Table 3: Comparison of three-name joint default probability by basic Monte Carlo (BMC), and importance sampling (IS). The number of simulation is 10^4 and an Euler discretization for (1) is used by taking time step size $T/100$, where T is one year. Other parameters are $S_{10} = S_{20} = S_{30} = 100, \mu_1 = \mu_2 = \mu_3 = 0.05$, $\sigma_1 = \sigma_2 = 0.4, \sigma_3 = 0.3$ and $B_1 = B_2 = 50, B_3 = 60$. Standard errors are shown in parenthesis.

ρ	BMC	IS
0.3	$0.0049(6.9832 * 10^{-4})$	$0.0057(1.9534 * 10^{-4})$
0	$3.0000 * 10^{-4}(1.7319 * 10^{-4})$	$6.4052 * 10^{-4}(6.9935 * 10^{-5})$
-0.3	$-(-)$	$2.2485 * 10^{-5}(1.1259 * 10^{-5})$

3.1 Asymptotic Variance Analysis by Large Deviation Principle

We provide a theoretical verification to show that the importance sampling developed above is asymptotic optimal for the one-dimensional first passage time problem under the geometric Brownian motion. This problem has also been considered in Carmona et al. [8].
Our proof is based on the Freidlin-Wentzell theorem [6, 9] in large deviation theory in order to approximate the default probability and the second moment of the importance sampling estimator defined in (18). We consider the scale $\varepsilon = -(\ln(B/S_0))^{-1}$ being small or equivalently $0 < B \ll S_0$, the current asset value S_0 is much larger than its debt value B. Our asymptotic results show that the second moment approximation is the square of the first moment (or default probability) approximation. Therefore, we attain the optimality of variance reduction in an asymptotic sense so that the proposed importance sampling scheme is efficient.

Theorem 2. *(Efficient Importance Sampling) Let S_t denote the asset value following the log-normal process $dS_t = \mu S_t dt + \sigma\, S_t dW_t$ with the*

Table 4: Comparison of multiname joint default probabilities by basic Monte Carlo (BMC), and importance sampling (IS) under high-dimensional Black-Cox model. Let n denote the dimension, the total number of firms. The number of simulation is $3 * 10^4$ and an Euler discretization for (1) is used by taking time step size $T/100$, where T is one year. Other parameters are $S_0 = 100, \mu = 0.05$, $\sigma = 0.3$, $\rho = 0.3$, and $B = 50$.

	Basic MC		Importance Sampling	
n	*Mean*	*SE*	*Mean*	*SE*
2	1.1E-03	3.31E-04	1.04E-03	2.83E-05
5	-	-	6.36E-06	3.72E-07
10	-	-	2.90E-07	2.66E-08
15	-	-	9.45E-09	1.16E-09
20	-	-	1.15E-09	1.98E-10
25	-	-	2.06E-10	3.84E-11
30	-	-	6.76E-11	2.36E-11
35	-	-	1.35E-11	2.89E-12
40	-	-	6.59E-12	1.58E-12
45	-	-	3.25E-12	1.08E-12
50	-	-	6.76E-13	2.26E-13

initial value S_0, and B denote the default boundary. We define the default probability and its importance sampling scheme by

$$\begin{aligned} P_1^\varepsilon &= E\left\{ \mathbf{I}\left(\min_{0\le t\le T} S_t \le B \right) \right\} \\ &= \tilde{E}\left\{ \mathbf{I}\left(\min_{0\le t\le T} S_t \le B \right) Q_T(h) \right\}, \end{aligned}$$

where the Radon-Nykodym derivative $Q(h)$ is defined in (15). The second moment of this estimator is denoted by

$$P_2^\varepsilon(h) = \tilde{E}\left\{ \mathbf{I}\left(\min_{0\le t\le T} S_t \le B \right) Q_T^2(h) \right\}.$$

By the choice of $h = (\mu T + 1/\varepsilon)/(\sigma T)$ with the scale ε defined from $-1/\varepsilon = \ln(B/S_0)$, the expected value S_T under $\tilde{P}$ is B. That is, $\tilde{E}\{S_T\} = B$. When ε is small enough or equivalently $B \ll S_0$, we obtain a zero variance rate, i.e., $\lim_{\varepsilon\to 0} \varepsilon^2 \ln\left(P_2^\varepsilon(h)/(P_1^\varepsilon)^2 \right) = 0$, so that the importance sampling scheme is efficient.

Proof: Recall that the one-dimensional default probability is defined by

$$\begin{aligned} & P\left[\inf_{0\le t\le T} S_t = S_0\, e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t} \le B \right] \\ = \; & E\left[\mathbf{I}\left(\inf_{0\le t\le T} \varepsilon\left(\mu - \frac{\sigma^2}{2}\right) t + \varepsilon\sigma W_t \le -1 \right) \right], \end{aligned} \tag{20}$$

where we have use the strictly monotonicity of the logarithmic transformation and we introduce a scaling $\ln(B/S_0) = \frac{-1}{\varepsilon}$. For small parameter ε, the default probability will be small in financial intuition because the *debt to asset value*, B/S_0, is small. By an application of Freidlin-Wentzell Theorem [6, 9], it is easy to prove that the rate function of (20) is $\frac{1}{2\sigma^2 T}$. That is the rescaled default probability

$$P\left[\inf_{0\le t\le T} S_t = S_0\, e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t} \le B \right] \approx \exp\left(\frac{-1}{\varepsilon^2\, 2\, \sigma^2\, T} \right), \tag{21}$$

when ε is small. Recall that under the measure change defined in (18), the price dynamics becomes $S_t = S_0\, e^{\left(\mu - \frac{\sigma^2}{2} - \sigma h\right)t + \sigma \tilde{W}_t}$ with $h = \frac{\mu}{\sigma} - \frac{\ln B/S_0}{\sigma T}$. The second moment $P_2^\varepsilon(h)$ becomes

$$\begin{aligned} & \tilde{E}\left[\mathbf{I}\left(\inf_{0\le t\le T} S_t \le B \right) e^{2h\tilde{W}_T - h^2 T} \right] \\ = \; & \hat{E}\left[\mathbf{I}\left(\inf_{0\le t\le T} S_0\, e^{\left(\mu - \frac{\sigma^2}{2} + \sigma h\right)t + \sigma \hat{W}_t} \le B \right) \right] e^{h^2 T} \\ = \; & \hat{E}\left[\mathbf{I}\left(\inf_{0\le t\le T} \left(\varepsilon\left(2\mu - \frac{\sigma^2}{2}\right) + \frac{1}{T} \right) t + \varepsilon\sigma\hat{W}_t \le -1 \right) \right] \times e^{\left(\frac{r}{\sigma} + \frac{1}{\varepsilon\sigma T}\right)^2 T}, \end{aligned}$$

where the measure change $d\hat{P}/d\tilde{P}$ is defined by $Q_T(2h)$ for the second line and we incorporate the same scaling $\ln(B/S_0) = \frac{-1}{\varepsilon}$ to rescale our problem for the last line above. By Freidlin-Wentzell Theorem, the rate function of the expectation is $\frac{2}{\sigma^2 T}$. Consequently, the approximation

$$\tilde{E}\left[\mathbf{I}\left(\inf_{0\leq t\leq T} S_t \leq B\right) e^{2\,h\,\tilde{W}_T - h^2\,T}\right] \approx \exp\left(\frac{-1}{\varepsilon^2\,\sigma^2\,T}\right) \qquad (22)$$

is derived. By $\lim_{\varepsilon\to 0}\varepsilon^2 \ln P_2^\varepsilon(h) = 2\lim_{\varepsilon\to 0}\varepsilon^2 \ln P_1^\varepsilon$, we confirm that the variance rate of this importance sampling is asymptotically zero so that this scheme is efficient. □

Remark: The same result can be obtained from a PDE argument studied in [13].

4 Conclusion

Estimation of joint default probabilities under a first passage time problem in the structural form model are tackled by importance sampling. Imposing "the expected asset value at debt maturity equals to its debt value" as a condition, importance sampling schemes can be uniquely determined within a family of parametrized probability measures. This approach overcomes the hurdle of exponential change of measure that requires existence of the moment generating function of first passage times. According to the large deviation principle, our proposed importance sampling scheme is asymptotically optimal or efficient in a rare event simulation.

References

[1] B. Arouna,"Robbins Monro algorithms and variance reduction in finance," *Journal of Computational Finance,* 7 (2), (Winter 2003/04).

[2] B. Arouna, "Adaptive Monte Carlo method, a variance reduction technique," *Monte Carlo Methods Appl,* 10 (1) (2004), 1-24.

[3] F. Black and M. Scholes, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy,* 81 (1973), 637-654.

[4] F. Black and J. Cox, "Valuing Corporate Securities: Some Effects of Bond Indenture Provisions," *Journal of Finance*, 31(2) (1976): 351-367.

[5] T.R. Bielecki and M. Rutkowski, *Credit Risk: Modeling, Valuation and Hedging*, Springer 2002.

[6] J. A. Bucklew, *Large Deviation Techniques in Decision, Simulation, and Estimation,* Wiley-Interscience, Applied Probability and Statistics Series, New York, 1990.

[7] J. A. Bucklew, *Introduction to rare event simulation*, Springer, 2003.

[8] R. Carmona, J.P. Fouque, and D. Vestal, " Interacting Particle Systems for the Computation of Rare Credit Portfolio Losses," *Finance and Stochastics*, 13(4), 2009 (p. 613-633).

[9] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2/e, Springer, 1998.

[10] J.-P. Fouque and C.-H. Han, "Variance Reduction for Monte Carlo Methods to Evaluate Option Prices under Multi-factor Stochastic Volatility Models," *Quantitative Finance*, Volume 4, number 5 (2004) 1-10.

[11] P. Glasserman, *Monte Carlo Methods for Financial Engineering*, Springer-Verlag, New York, 2003.

[12] C.-H. Han, W.-H. Liu, Z.-Y. Chen, "An Improved Procedure for VaR/CVaR Estimation under Stochastic Volatility Models," Submitted, 2010.

[13] C.-H. Han and D. Vestal, "Estimating the probability of joint default in structural form by efficient importance sampling," working paper, Natoinal Tsing-Hua University, 2010.

[14] C.-H. Han and C.-T. Wu, "Efficient importance sampling in estimation of default probability under factor copula model," working paper, National Tsing-Hua University, 2010.

[15] R. Merton, "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," *The Journal of Finance,* 29 (1974), 449-470.

[16] C. Zhou, "The Term Structure of Credit Spreads with Jump Risk", *Journal of Banking & Finance,* Vol. 25, No. 11, (November 2001), pp. 2015-2040.

[17] C. Zhou, "An Analysis of Default Correlations and Multiple Defaults," *The Review of Financial Studies*, 14(2) (2001), 555-576.