Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

IMPORTANCE SAMPLING FOR A MIXED POISSON MODEL OF PORTFOLIO CREDIT RISK

Paul Glasserman
Jingyi Li

Columbia Business School
Columbia University
New York, NY 10027, U.S.A.

ABSTRACT

Simulation is widely used to estimate losses due to default and other credit events in financial portfolios. The challenge in doing this efficiently results from (i) rare-event aspects of large losses and (ii) complex dependence between defaults of multiple obligors. We discuss importance sampling techniques to address this problem in two portfolio credit risk models developed in the financial industry, with particular emphasis on a mixed Poisson model. We give conditions for asymptotic optimality of the estimators as the portfolio size grows.

1 INTRODUCTION

Developments in risk management have led financial institutions to make greater use of probabilistic models to quantify their risks. Two main components of financial risk are market risk and credit risk. Whereas market risk results from changes in prices, credit risk refers to losses resulting from the failure of an obligor (a party under a legal obligation) to make a contractual payment. Credit risk includes, for example, the possibility that a company will fail to repay a loan or a bond issuer will miss a coupon payment. Increased interest in the modeling and management of credit risk has led to the development of various commercial models, now in widespread use. These include CreditMetrics, originally developed by JP Morgan, and CreditRisk$^+$, developed by Credit Suisse Financial Products. For an overview, see Crouhy, Galai, and Mark (2001). These models are designed for the credit risk banks face from other companies and differ from those used for consumer credit.

Given a credit risk model, the rapid and accurate construction of the portfolio loss distribution is at the heart of credit risk management. Monte Carlo simulation is frequently used to estimate this distribution. Each replication of such a simulation usually consists of determining which obligors default and the losses given default. For high-quality portfolios, most replications produce few if any defaults, so the computational cost required to obtain accurate credit risk estimates can be very large. This is particularly true for accurate estimation of small but important probabilities of large losses, which are usually the focus of risk measurement. Importance sampling (IS) is a natural technique to consider for rare event simulation; however, complex dependence between defaults of multiple obligors complicates the application IS. Capturing dependence between defaults is at the heart of a portfolio view of credit risk, so this issue is fundamental.

In most models of credit risk, dependence is introduced through a set of "risk factors" and defaults become independent conditional on the risk factors. This suggests a general approach to IS based on applying a change of distribution to the factors and a change of distribution to the default indicators conditional on the factors. This is the approach we follow. We have used this approach in Glasserman and Li (2003) for the "normal copula" model of Gupton et al. (1997). Here we show that a similar strategy can be used very conveniently in the mixed Poisson model of CreditRisk$^+$ (CSFP 1997). Indeed, the mixed Poisson model is sufficiently tractable that it is usually solved through numerical transform inversion, without simulation. Nevertheless, it provides an interesting illustration of a more general approach to IS for credit risk.

Section 2 reviews the normal copula and mixed Poisson models. In Section 3, we discuss IS for the normal copula model, based on Glasserman and Li (2003). In Section 4, we propose an IS method for the mixed Poisson model. We establish the asymptotic optimality of this method under alternative limiting regimes. Numerical examples illustrate the effectiveness of the method. Section 5 concludes the paper.

2 CREDIT RISK MODELS

We consider a portfolio with m obligors. Let Y_i denote the default indicator of the ith obligor for some fixed time horizon (e.g., one year). Thus, $Y_i = 1$ if this obligor defaults within the horizon and $Y_i = 0$ otherwise. Let c_i denote the loss resulting from default of the ith obligor. These are sometimes modeled as random variables, but for simplicity we take them to be (positive) constants. The portfolio loss over the horizon is

$$L = \sum_{i=1}^{m} c_i Y_i. \qquad (1)$$

Our goal is to measure the tail of the loss distribution $P(L > x)$, particularly at large values of x.

The marginal default probabilities $p_i = P(Y_i = 1)$ are usually assumed known (e.g., from published credit ratings). Different credit risk models differ in the mechanisms they use to capture dependence among the Y_i. Here we give a brief description of two models.

2.1 Normal Copula Model

In the CreditMetrics model of Gupton et al. (1997) (see also Li (2000)) the default indicators are modeled as

$$Y_i = \mathbf{1}\{X_i > x_i\}, \quad i = 1, \ldots, m, \qquad (2)$$

where $(X_1, \ldots, X_m)$ are correlated $N(0,1)$ random variables. Each threshold x_i is chosen to match the marginal default probability p_i for the ith obligor; thus, $x_i = \Phi^{-1}(1 - p_i)$, with Φ the cumulative normal distribution. This construction transfers correlations among the X_i to dependence among the Y_i. This is an instance of a normal copula construction of dependent random variables, or what Cario and Nelson (1997) call "normal to anything." In the credit risk context, the X_i are often given a financial interpretation.

Correlations among the X_i are usually specified through a factor model of the form

$$X_i = a_{i0}\epsilon_i + a_{i1}Z_1 + \cdots + a_{id}Z_d, \quad i = 1, \ldots, m, \qquad (3)$$

with ϵ_i and $Z_1, \ldots, Z_d$ independent $N(0,1)$ random variables and $a_{i1}^2 + \cdots + a_{id}^2 + a_{i0}^2 = 1$. Each ϵ_i represents risk affecting only the ith obligor, whereas the Z_j represent common risk factors affecting multiple obligors. For example, each Z_j may be associated with an industry, a geographic region, or a market-wide risk factor.

Normal copula models rely on simulation for the calculation of the portfolio loss distribution. In each replication, every X_i is generated from independent $N(0,1)$ random variables ϵ_i and $Z_1, \ldots, Z_d$ according to the model specification (3), and the portfolio loss is evaluated from (2) and (1).

2.2 Mixed Poisson Model

An alternative way of introducing dependence uses a mixed Poisson model, as in CSFP's (1997) CreditRisk$^+$. In this setting, each Y_i is (conditionally) Poisson distributed. This may be viewed as a Poisson approximation to a Bernoulli random variable (based on the fact that a Poisson random variable with a very small mean has a very small probability of taking a value other than 0 or 1); alternatively, it can be viewed as a reinterpretation of (1) in which i indexes groups of obligors with roughly equal exposure c_i, rather than individual obligors. In this reinterpretation, values of Y_i greater than 1 are meaningful.

The common risk factors in this model are independent gamma random variables $\Gamma_1, \ldots, \Gamma_d$. Conditional on these random variables, each Y_i has a Poisson distribution with mean R_i,

$$R_i = a_{i0} + a_{i1}\Gamma_1 + \cdots + a_{id}\Gamma_d, \qquad (4)$$

for some positive coefficients $a_{i0}, \ldots, a_{id}$. Thus, each Y_i may be viewed as a Poisson random variable with a random mean — a mixed Poisson random variable. We normalize $\Gamma_1, \ldots, \Gamma_d$ to have mean 1 and variances $\sigma_1^2, \ldots, \sigma_d^2$.

Mixed Poisson models have long been used in many applications; see Section 3.2 of Johnson et al. (1993). Using gamma random variables for the mixing variables leads to some tractability and allows calculation of the distribution of L through numerical inversion of its probability generating function (see CSFP 1997). The model nevertheless provides an interesting setting for rare event simulation. Also, for the IS method we develop the gamma random variables could be replaced with any other positive random variables having reasonably well-behaved moment generating functions.

Simulation without IS is straightforward. In each replication, we first generate the common risk factors Γ_j independently from the distributions $Gamma(\alpha_j, \beta_j)$, $j = 1, \ldots, d$, with

$$\alpha_j = \frac{1}{\sigma_j^2}, \quad \beta_j = \sigma_j^2, \quad j = 1, \ldots, d.$$

This gives Γ_j mean 1 and variance σ_j^2. Then we generate Y_i from $Poisson(R_i)$ with the R_i calculated as in (4). From the Y_i we evaluate the portfolio loss (1).

3 IS FOR THE NORMAL COPULA MODEL

3.1 IS Estimator

In this section, we review an IS technique proposed in Glasserman and Li (2003) for the normal copula model. We begin by considering the simpler case of independent obligors in which $Y_1, \ldots, Y_m$ are independent Bernoulli random variables with parameters $p_1, \ldots, p_m$.

In order to generate large losses more often, it is natural to consider IS based on increasing each default probability p_i to some larger value q_i. The associated estimator of $P(L > x)$ is the product of the indicator $\mathbf{1}\{L > x\}$ and the likelihood ratio

$$\prod_{i=1}^{m}\left(\frac{p_i}{q_i}\right)^{Y_i}\left(\frac{1-p_i}{1-q_i}\right)^{1-Y_i}.$$

For the new probabilities, it turns out to be convenient to restrict attention to a one-parameter family of the form

$$p_i(\theta) = \frac{p_i e^{\theta c_i}}{p_i e^{\theta c_i} + (1-p_i)}.$$

By choosing $\theta > 0$, we increase each default probability and we do so in a way that takes account of the original p_i and also the loss magnitudes c_i.

With this choice of default probabilities, some algebra shows that the likelihood ratio can be rewritten as $\exp(-\theta L + \psi_L(\theta))$, where

$$\psi_L(\theta) = \log E[\exp(\theta L)] = \sum_{i=1}^{m}\log(1 + p_i(e^{c_i\theta} - 1))$$

is the cumulant generating function of L. Thus, using the probabilities $p_i(\theta)$ is equivalent to *exponentially twisting* L, a standard technique in IS. To sample under the twisted distribution, we simply replace the original default probability p_i with $p_i(\theta)$.

It remains to choose the parameter θ. For this we look at the second moment of the estimator, which is given by

$$\begin{aligned} M_2(x,\theta) &\equiv E_\theta[e^{-2\theta L + 2\psi_L(\theta)}\mathbf{1}\{L > x\}] \\ &\leq \exp(-2\theta x + 2\psi_L(\theta)). \end{aligned} \quad (5)$$

The subscript on the expectation indicates that it is calculated under the IS distribution for parameter θ. While finding the value of θ minimizing $M_2(x,\theta)$ is difficult, it is a simple matter to minimize the upper bound in (5). The minimizer θ_x is the unique solution to

$$\psi_L'(\theta_x) = x. \quad (6)$$

The expectation of L under this changed measure is

$$E_{\theta_x}[L] = E_{\theta_x}\left[\sum_{i=1}^{m} c_i Y_i\right] = \psi_L'(\theta_x) = x.$$

Thus, to estimate $P(L > x)$ for large values of x, we increase the individual default probabilities to make x the expected loss.

We now turn to the more interesting case in which the Y_i are dependent. We consider dependence introduced through a normal copula as discussed in Section 2.1. We apply IS as in the independent case, but we do so conditional on the common factors $Z = (Z_1, \ldots, Z_d)^\top$.

Observe that, given Z, the Y_i are indeed independent with conditional default probabilities

$$\tilde{p}_i = P(Y_i = 1|Z) = \Phi\left(\frac{a_{i1}Z_1 + \cdots + a_{id}Z_d - x_i}{a_{i0}}\right).$$

From these we can calculate the conditional cumulant generating function

$$\psi_{L|Z}(\theta) = \log E[e^{\theta L}|Z] = \sum_{i=1}^{m}\log(1 + \tilde{p}_i(e^{\theta c_i} - 1))$$

and solve for the parameter $\tilde{\theta}_x$,

$$\psi_{L|Z}'(\tilde{\theta}_x) = x.$$

We can then define new conditional default probabilities

$$\tilde{p}_i(\tilde{\theta}_x) = \frac{\tilde{p}_i e^{\tilde{\theta}_x c_i}}{\tilde{p}_i e^{\tilde{\theta}_x c_i} + 1 - \tilde{p}_i}, \quad i = 1, \ldots, m.$$

The IS procedure now generates default indicators $Y_1, \ldots, Y_m$ independently (given Z) with Y_i taking the value 1 with probability $\tilde{p}_i(\tilde{\theta}_x)$.

Setting L equal to the sum of the $Y_i c_i$ yields the one-step IS estimator

$$e^{-\tilde{\theta}_x L + \psi_{L|Z}(\tilde{\theta}_x)}\mathbf{1}\{L > x\}; \quad (7)$$

this is the conditional counterpart of the IS estimator in the independent case. Its conditional expectation is $P(L > x|Z)$ and its unconditional expectation is therefore $P(L > x)$.

To further reduce variance, we can apply a second step of importance sampling to Z, viewing $P(L > x|Z)$ as a function of Z and the calculation of $P(L > x)$ as a problem of integrating over the distribution of Z. For this we consider shifting the mean of Z from the origin to some point μ. The likelihood ratio for this change of measure is

$$\exp\left(-\mu^\top Z + \frac{1}{2}\mu^\top\mu\right).$$

When multiplied by (7) this yields the two-step IS estimator

$$\exp\left(-\mu^\top Z + \frac{1}{2}\mu^\top\mu - \tilde{\theta}_x L + \psi_{L|Z}(\tilde{\theta}_x)\right)\mathbf{1}\{L > x\}$$

in which Z is sampled from $N(\mu, 1)$ and then L is sampled from the $\tilde{\theta}_x$-twisted distribution conditional on Z.

It remains to specify the new mean μ for the common factors Z. The approach of Glasserman, Heidelberger, and Shahabuddin (1999) suggests choosing μ by solving

$$\mu = \operatorname*{argmax}_z P(L > x|Z = z)e^{-\frac{1}{2}z^\top z}.$$

The product on the right is (proportional to) the optimal IS density, so this approach chooses the new mean at the mode of the optimal density. This approach is investigated in Glasserman and Li (2003).

3.2 Asymptotic Optimality of the One-Step IS Estimator

In rare event simulation, one often tries to measure the effectiveness of an estimator of a small probability by investigating its performance as the probability of the event vanishes. An estimator is said to be asymptotically optimal if its second moment decreases at twice the rate of the probability itself. By Jensen's inequality, this is the fastest possible rate of decrease for any unbiased estimator.

To see what type of asymptotic optimality we might look for in the credit risk setting, we again consider the independent case. Because the credit portfolios of financial institutions can be very large, it is natural to consider asymptotics as $m \to \infty$. In the independent case, the key condition we need is convergence of the functions ψ_L/m to a finite, convex function ψ; this holds, for example, if the (p_i, c_i) approach a limit as i increases. In this case, asymptotic optimality can be established through the argument in Sadowsky and Bucklew (1990). In more detail, for all sufficiently large q, we have

$$\lim_{m\to\infty} \frac{1}{m} \log P(L > mq) = \gamma_q$$

and

$$\lim_{m\to\infty} \frac{1}{m} \log M_2(mq, \theta_{mq}) = 2\gamma_q,$$

for some $\gamma_q < 0$. Thus, the second moment decreases at twice the exponential rate as the first moment.

It turns out that we generally cannot hope to have a result of quite this form once we introduce dependence through either a normal copula or mixed Poisson model. Indeed, once we introduce dependence, L/m will often converge to a random limit and $P(L > mq)$ may not vanish as $m \to \infty$: there is too much dependence for this formulation to lead to asymptotic optimality.

We therefore consider a limit in which the dependence weakens as m increases. Whether or not we achieve asymptotic optimality depends on how quickly it weakens. The practical implication of this formulation is that the one-step IS estimator is effective only if the underlying correlations are not too large. At larger correlations, it becomes essential to apply IS to the common risk factors as well.

To state a precise result, we limit ourselves to the case $c_i \equiv 1$, $p_i \equiv p$, and a single common factor Z, and all X_i of the form

$$X_i = \rho Z + \sqrt{1-\rho^2}\epsilon_i.$$

We take ρ to be of the form a/m^α, for some $a, \alpha > 0$, and find different behavior depending on the value of α; i.e., depending on the speed at which ρ decreases.

Define

$$G(p) = \begin{cases} \log(\frac{1-p}{1-q})^{1-q}(\frac{p}{q})^q & p < q, \\ 0 & p \geq q; \end{cases}$$

$mG(p)$ is the likelihood ratio at $L = mq$ for the independent case with marginal individual default probability p. Also define

$$F(a, z) = G(\Phi(az + \Phi^{-1}(p))).$$

The following theorem is proved in Glasserman and Li (2003):

Theorem 1 *If $\rho = a/m^\alpha$, $a > 0$, then*
(a) For $\alpha > 1/2$,

$$\begin{aligned} \lim_{m\to\infty} m^{-1} \log P(L > mq) &= F(0,0) \\ \lim_{m\to\infty} m^{-1} \log M_2(mq, \theta_{mq}) &= 2F(0,0). \end{aligned}$$

(b) For $\alpha = 1/2$,

$$\begin{aligned} \lim_{m\to\infty} m^{-1} \log P(L > mq) &= \max_z \{F(a,z) - z^2/2\} \\ \lim_{m\to\infty} m^{-1} \log M_2(mq, \theta_{mq}) &= \max_z \{2F(a,z) - z^2/2\}. \end{aligned}$$

(c) For $0 < \alpha < 1/2$,

$$\begin{aligned} &\lim_{m\to\infty} m^{-2\alpha} \log P(L > mq) \\ &= \lim_{m\to\infty} m^{-2\alpha} \log M_2(mq, \theta_{mq}) \\ &= -z_a^2/2, \end{aligned}$$

with $z_a = (\Phi^{-1}(q) - \Phi^{-1}(p))/a$.

This result shows that we achieve asymptotic optimality only in the case $\alpha > 1/2$ (in which the correlations vanish quite quickly), because only in this case does the second moment vanish at twice the rate of the first moment. At $\alpha = 1/2$, the second moment decreases faster than the first moment, but not twice as fast, so this is an intermediate case. With $\alpha < 1/2$, the two decrease at the same rate, which implies that one-step IS is (asymptotically) no more effective than ordinary simulation in this case. The failure of asymptotic optimality in (b) and (c) results from the impact of the common risk factor Z in the occurrence of a large number of defaults: at moderate or large values of ρ, large losses occur primarily because of large moves in Z. Capturing this effect requires applying IS to Z itself, rather than just to the Y_i conditional on Z.

4 IS FOR THE MIXED POISSON MODEL

4.1 IS Estimator

We next consider the simulation problem for the mixed Poisson model. As in the normal copula model, we can think of applying IS in two steps — one step changes the default probabilities conditional on the common factors, the other applies a change of distribution to the factors themselves. Because of the special structure of the mixed Poisson model, these two steps can be combined in a convenient and effective way.

In analogy with the discussion for the normal copula model, we first assume the values of the common risk factors $\Gamma_1, \ldots, \Gamma_d$ are given, so that the Y_i are independent Poisson random variables with parameters R_i. Consider the effect of exponentially twisting $c_i Y_i$ by some $\theta \in \Re$; this defines a change of distribution through the likelihood ratio

$$\exp(-\theta c_i Y_i + R_i(e^{c_i\theta} - 1)).$$

Here, $R_i(e^{c_i\theta}-1)$ is the conditional cumulant generating function of $c_i Y_i$, given R_i. The conditional mean of Y_i under the distribution defined by θ is $R_i e^{c_i\theta}$. By choosing $\theta > 0$ we thus increase the mean of Y_i.

Now apply this exponential twist to all the $c_i Y_i$. Since L is the sum of (conditionally) independent random variables $c_i Y_i$, the likelihood ratio has the form

$$\begin{aligned}\prod_{i=1}^{m} &\exp(-\theta c_i Y_i + R_i(e^{c_i\theta} - 1)) \\ &= \exp(-\theta L + \sum_{i=1}^{m} R_i(e^{c_i\theta} - 1)), \quad (8)\end{aligned}$$

and $\sum_{i=1}^{m} R_i(e^{c_i\theta} - 1)$ is the conditional cumulant generating function of L given the risk factors $\Gamma_1, \ldots, \Gamma_d$.

To further reduce variance, we apply a second importance sampling step to the risk factors. We consider exponentially twisting each Γ_j by some τ_j. This defines a change of distribution through the likelihood ratio

$$\exp\left(-\sum_{j=1}^{d}\{\tau_j\Gamma_j + \alpha_j \log(1-\beta_j\tau_j)\}\right). \quad (9)$$

Here, $-\alpha_j \log(1-\beta_j\tau_j)$ is the cumulant generating function of Γ_j, which has a $Gamma(\alpha_j, \beta_j)$ distribution under the original measure. We see from this that τ_j must be less than $1/\beta_j$, $j = 1, \ldots d$. Under the distribution defined by τ_j, Γ_j has a $Gamma(\alpha_j, \beta_j/1-\beta_j\tau_j)$ distribution. In other words, exponentially twisting a gamma distribution produces another gamma distribution with the same shape parameter and a different scale parameter.

The likelihood ratio for this two-step change of distribution is the product of the individual likelihood ratios (8) and (9). Since the R_i are determined by $\Gamma_1, \ldots, \Gamma_d$ through (4), simple algebra shows that the likelihood ratio can be expressed as

$$\exp\left(-\theta L + \psi^{(1)}(\theta) + \psi^{(2)}(\tau) + \psi^{(3)}(\theta, \tau, \Gamma)\right), \quad (10)$$

where

$$\begin{aligned}\psi^{(1)}(\theta) &= \sum_{i=1}^{m} a_{i0}(e^{c_i\theta} - 1), \\ \psi^{(2)}(\tau) &= -\sum_{j=1}^{d} \alpha_j \log(1-\beta_j\tau_j), \\ \psi^{(3)}(\theta, \tau, \Gamma) &= \sum_{j=1}^{d}(\sum_{i=1}^{m} a_{ij}(e^{c_i\theta} - 1) - \tau_j)\Gamma_j.\end{aligned}$$

It remains to choose the twisting parameters $\tau_1, \ldots, \tau_d$ and θ. Inspection of the components of (10) reveals that by linking the choices of these parameters we can eliminate the Γ_j from the likelihood ratio, leaving only the dependence on L. Because our goal is to estimate the tail distribution of L, this will prove to be an effective choice. Suppose, then, that we choose

$$\tau_j = \sum_{i=1}^{m} a_{ij}(e^{c_i\theta} - 1), \quad j = 1, \ldots d. \quad (11)$$

For sufficiently small $\theta > 0$, this will satisfy the constraint $\tau_j < 1/\beta_j$. Substituting (11) in (10) reveals that the two-step IS likelihood ratio has the form

$$\exp(-\theta L + \psi_{L,m}(\theta)) \quad (12)$$

where

$$\psi_{L,m}(\theta) = \psi^{(1)}_{L,m}(\theta) + \psi^{(2)}_{L,m}(\theta) \quad (13)$$

with $\psi^{(1)}_{L,m} = \psi^{(1)}$ as above and

$$\psi^{(2)}_{L,m}(\theta) = -\sum_{j=1}^{d} \alpha_j \log\left(1 - \beta_j \sum_{i=1}^{m} a_{ij}(e^{c_i\theta} - 1)\right).$$

In fact, this shows that $\psi_{L,m}(\theta)$ is the cumulant generating function of L and (12) is exactly the likelihood ratio for exponentially twisting L itself by θ. (In contrast, in the normal copula model it does not seem possible to blend the two IS steps in this way or to find a simple expression for the cumulant generating function of L.)

For the choice of θ, we use the same idea as in the independent case, choosing $\theta = \theta_x$ with θ_x solving

$$\psi'_{L,m}(\theta_x) = x. \quad (14)$$

The two-step IS estimator of $P(L > x)$ for the mixed Poisson model is then

$$\exp(-\theta_x L + \psi_{L,m}(\theta_x))\mathbf{1}\{L > x\}. \qquad (15)$$

This is easily implemented through the following algorithm:

1. Define $\psi_{L,m}$ as in (13) and solve for θ_x as in (14); set $\theta_x = \max\{0, \theta\}$.
2. Compute τ_j, $j = 1, \ldots, d$, from (11).
3. Generate $\Gamma_j \sim Gamma(\alpha_j, \frac{\beta_j}{1-\beta_j\tau_j})$, $j = 1, \ldots, d$.
4. Compute the conditional means R_i, $i = 1, \ldots, m$, as in (4).
5. Generate $Y_i \sim Poisson(R_i e^{c_i\theta_x})$, $i = 1, \ldots, m$.
6. Calculate loss $L = c_1Y_1 + \cdots + c_mY_m$.
7. Return estimator (15).

The condition $\theta_x > 0$ holds whenever $x > E[L]$ and thus whenever $\{L > x\}$ is a rare event. In the less interesting case that $x \leq E[L]$ (and $\theta_x \leq 0$), IS is unnecessary. Replacing θ_x with 0 (as in Step 1) restores the original sampling distribution.

4.2 Asymptotic Optimality

We now turn to the question of asymptotic optimality for the IS estimator (15) as m increases. Our first result takes the loss threshold x to be a fixed multiple of m; we write mq instead of x. We formulate an asymptotic optimality result for settings in which the probability $P(L > mq)$ decays exponentially to 0 as m increases to ∞. As before, asymptotic optimality means that the second moment

$$\begin{aligned} &M_2(q, \theta_{q,m}) \\ &\quad = E_{\theta_{q,m}}[\mathbf{1}\{L > mq\}\exp\{-2\theta_{q,m}L + 2\psi_{L,m}(\theta_{q,m})\}] \end{aligned}$$

decays at twice the rate of the probability itself. Here, we have written the twisting parameter θ_x as $\theta_{q,m}$ and written $E_{\theta_{q,m}}$ for the expectation under the IS distribution.

Asymptotic optimality would follow from the existence of a limiting cumulant generating function

$$\psi_L(\theta) = \lim_{m\to\infty} \frac{1}{m}\psi_{L,m}(\theta). \qquad (16)$$

Inspection of $\psi^{(2)}_{L,m}$ reveals, however, that this limit will ordinarily be infinite for all $\theta > 0$ if the parameters of the problem are all $O(1)$. In particular, if the a_{ij} and c_i are all $O(1)$ and positive, then

$$\sum_{i=1}^{m} a_{ij}(e^{c_i\theta} - 1)$$

will be larger than $1/\beta_j$ for sufficiently large m, making $\psi^{(2)}_{L,m}(\theta)$ infinite for all sufficiently large m.

The source of the problem here (as in the normal copula setting) is that there is too much dependence among the Y_i. As a result, $P(L > mq)$ may even have a nonzero limit as m increases. To formulate an asymptotic optimality result, we consider limiting regimes in which either the impact of the common factors $\Gamma_1, \ldots, \Gamma_d$ diminishes as m increases, or in which q itself increases with m. From a practical perspective, this suggests that the IS estimator may not be very effective in the presence of strong correlations between the Y_i. Alternatively, this says that q needs to be large relative to the strength of the dependence on the common risk factors.

There are many ways one could make the parameters of the model vary with m that would lead to the convergence required in (16). We give three specifications. In the first, the coefficients a_{ij} are decreasing, so the Y_i become less sensitive to the risk factors; in the second, the number of risk factors increases and each becomes less important; in the third, the variability of each risk factor decreases.

Case (a): $c_i \to c$, $a_{i0} \to a_0$, $ma_{ij} \to a_j$, for some constants c, $a_0, a_1, \ldots, a_d$, and the limit is

$$\psi_L(\theta) = a_0(e^{c\theta} - 1).$$

Case (b): $c_i \to c$, $a_{i0} \to a_0$, $ma_{ij} \to a$, $d/m \to d_0$, and the limit is

$$\psi_L(\theta) = a_0(ce^{\theta} - 1) - \frac{d_0}{\sigma_j^2}\log(1 - \sigma_j^2 a(e^{c\theta} - 1))$$

Case (c): $c_i \to c$, $a_{i0} \to a_0$, $a_{ij} \to a_j$, $m/\alpha_j \to \sigma_j^2 > 0$ and $m\beta_j \to \sigma_j^2$, and the limit is

$$\psi_L(\theta) = a_0(e^{c\theta} - 1) - \sum_{j=1}^{d} \frac{1}{\sigma_j^2}\log(1 - \sigma_j^2 a_j(e^{c\theta} - 1)).$$

For any of these possible limits, let θ_q satisfy

$$\psi'_L(\theta_q) = q. \qquad (17)$$

Theorem 2 *For Cases (a)–(c), we have*

$$\lim_{m\to\infty} m^{-1}\log P(L > mq) = -\theta_q q + \psi_L(\theta_q). \qquad (18)$$

and

$$\lim_{m\to\infty} m^{-1}\log M_2(q, \theta_{q,m}) = -2\theta_q q + 2\psi_L(\theta_q). \qquad (19)$$

Thus, IS using exponential twisting with parameter $\theta_{q,m}$ is asymptotically optimal.

Proof. Given the existence of ψ_L, (18) is a direct consequence of the Gärtner-Ellis Theorem (as in, e.g., Dembo and Zeitouni (1998)).

From the definition of $M_2(q, \theta_{q,m})$, we have

$$\begin{aligned} M_2(q,\theta_{q,m}) &\leq \exp\{-2m(\theta_{q,m}q + m^{-1}\psi_{L,m}(\theta_{q,m}))\} \\ &\leq \exp\{-2m(\theta_q q + m^{-1}\psi_{L,m}(\theta_q))\} \\ &= \exp\{-2m(\theta_q q + \psi_L(\theta_q)) + o(m)\}. \end{aligned}$$

The second inequality holds because $\theta_{q,m}$ minimizes $-\theta q + m^{-1}\psi_{L,m}(\theta)$. And since $m^{-1}\psi_{L,m}(\theta_q) \to \psi_L(\theta_q)$, we obtain the last equality. Thus,

$$\limsup_{m\to\infty} m^{-1}\log M_2(q,\theta_{q,m}) \leq -2\theta_q q + 2\psi_L(\theta_q).$$

Since the second moment must be at least as large as the square of the first moment, using (18) we get (19). and asymptotic optimality holds. ∎

Through the argument in Sadowsky and Bucklew (1990), it follows that we also have asymptotic optimality if in the IS algorithm we replace $\theta_{q,m}$ with the fixed value θ_q solving $\psi_L'(\theta_q) = q$. This has some potential advantage in the sense that ψ_L may have a simpler form than $\psi_{L,m}$. In numerical experiments we have found, however, that using $\theta_{q,m}$ results in greater variance reduction — sometimes much greater.

Now we consider another type of asymptotic optimality in which q itself increases with m. Suppose $q = xm^\alpha$ for positive constants x and α. Write θ_m for the solution to

$$\psi'_{L,m}(\theta_m) = xm^{1+\alpha}.$$

The second moment under the IS distribution with parameter θ_m is

$$\begin{aligned} &M_2(m,\theta_m) \\ &\quad = E_{\theta_m}[\mathbf{1}\{L > xm^{1+\alpha}\}\exp\{-2\theta_m L + 2\psi_{L,m}(\theta_m)\}] \end{aligned}$$

with E_{θ_m} the expectation under the IS distribution. We suppose that for each $j = 0,1,\ldots,d$, $a_{ij} \to a_j$ for some constants a_j. We assume the indices of $\Gamma_1, \ldots \Gamma_d$ are ordered so that $a_1\beta_1 \leq \cdots \leq a_d\beta_d$.

Theorem 3 *If $c_i \to c$, $a_{ij} \to a_j$, $j = 0,1,\ldots d$, then*

$$\begin{aligned} \lim_{m\to\infty} m^{-\alpha}\log P(L > xm^{1+\alpha}) &= -x/ca_1\beta_1 \\ \lim_{m\to\infty} m^{-\alpha}\log M_2(m,\theta_m) &= -2x/ca_1\beta_1. \end{aligned}$$

Thus, IS using exponential twsisting with parameter θ_m is asymptotically optimal.

Proof. First we show that

$$\liminf_{m\to\infty} m^{-\alpha}\log P(L > xm^{1+\alpha}) \geq -\frac{x}{ca_1\beta_1}. \qquad (20)$$

Since $c_i \to c$, $a_{ij} \to a_j$, for arbitrary $\epsilon > 0$ there exists an m_1 such that for any $m \geq m_1$,

$$c - \epsilon \leq c_i \leq c + \epsilon, \text{ and } a_j - \epsilon \leq a_{ij} \leq a_j + \epsilon.$$

Given Γ_1, let $N_1, N_2, \ldots$ be i.i.d. Poisson random variables with mean $(a_1 - \epsilon)\Gamma_1$. For sufficiently large m,

$$\begin{aligned} &P(L > xm^{1+\alpha}) \qquad (21) \\ &\geq P((c-\epsilon)\sum_{i=1}^m N_i > xm^{1+\alpha}) \\ &\geq P(\Gamma > \gamma_{m,\epsilon})P((c-\epsilon)\sum_{i=1}^m N_i > xm^{1+\alpha}\,|\Gamma = \gamma_{m,\epsilon}) \end{aligned}$$

where $\gamma_{m,\epsilon} = \frac{xm^\alpha}{(c-\epsilon)(a_1-\epsilon)} + \epsilon$. Using the fact that $\sum_i N_i$ has a Poisson distribution, given Γ_1, and applying the bound (4.49) of Johnson et al. (1993), we find that the second factor is greater than 1/2 for all sufficiently large m. Combining this with the fact that the tail of Γ_1 decays exponentially at rate $1/\beta_1$, for large m we get

$$\begin{aligned} P(L > xm^{1+\alpha}) &\geq \frac{1}{2}P(\Gamma_1 > \gamma_{m,\epsilon}) \\ &= \exp(-\frac{\gamma_{m,\epsilon}}{\beta_1} + o(m^\alpha)). \end{aligned}$$

Since $\epsilon > 0$ can be arbitrarily small, (20) follows.

Next we show that

$$\limsup_{m\to\infty} m^{-\alpha}\log M_2(m,\theta_m) \leq -\frac{2x}{ca_1\beta_1}. \qquad (22)$$

Define

$$\psi_{L,m,\epsilon}(\theta) = \psi^{(1)}_{L,m,\epsilon}(\theta) + \psi^{(2)}_{L,m,\epsilon}(\theta)$$

where

$$\begin{aligned} \psi^{(1)}_{L,m,\epsilon}(\theta) &= m(a_1+\epsilon)(e^{(c+\epsilon)\theta} - 1) \\ \psi^{(2)}_{L,m,\epsilon}(\theta) &= -\sum_{j=1}^d \alpha_j \log(1 - \beta_j m(a_1+\epsilon)(e^{(c+\epsilon)\theta} - 1)). \end{aligned}$$

For $m \geq m_1$ large enough and $\theta \geq 0$, $\psi_{L,m,\epsilon}(\theta) \geq \psi_{L,m}(\theta)$. Define $\theta_{m,\epsilon}$ to be the value that solves

$$\psi'_{L,m,\epsilon}(\theta_{m,\epsilon}) = xm^{1+\alpha}$$

and observe that

$$\begin{aligned} \psi'^{(1)}_{L,m,\epsilon}(\theta) &= m(a_0+\epsilon)(c+\epsilon)e^{(c+\epsilon)\theta} \\ \psi'^{(2)}_{L,m,\epsilon}(\theta) &= \sum_{j=1}^d \frac{m(a_j+\epsilon)(c+\epsilon)e^{(c+\epsilon)\theta}}{1 - \beta_j m(a_j+\epsilon)(e^{(c+\epsilon)\theta} - 1)}. \end{aligned}$$

From the definition of $M_2(m,\theta_m)$, we know that

$$\begin{aligned} M_2(m,\theta_m) &\le \exp\left\{-2\theta_m x m^{1+\alpha} + 2\psi_{L,m}(\theta_m)\right\} \quad (23) \\ &\le \exp\left\{-2\theta_{m,\epsilon} x m^{1+\alpha} + 2\psi_{L,m}(\theta_{m,\epsilon})\right\} \\ &\le \exp\left\{-2\theta_{m,\epsilon} x m^{1+\alpha} + 2\psi_{L,m,\epsilon}(\theta_{m,\epsilon})\right\}. \end{aligned}$$

The second inequality comes from the fact that θ_m minimize the upper bound.

As $m \to \infty$, we have $\psi'_{L,m,\epsilon}(\theta_{m,\epsilon}) \to \infty$ and this requires that $\theta_{m,\epsilon}/z_m \to 1$ where

$$z_m = \log\left(\frac{1}{\beta_1 m(a_1+\epsilon)} + 1\right)/(c+\epsilon)$$

is the smallest root of the denominator $\psi'^{(2)}_{L,m,\epsilon}(\theta)$. But then $\theta_{m,\epsilon} m \to 1/(c+\epsilon)(a_1+\epsilon)\beta_1$. So

$$\begin{aligned} &\limsup_{m\to\infty} \frac{-2\theta_{m,\epsilon} x m^{1+\alpha} + 2\psi_{L,m,\epsilon}(\theta_{m,\epsilon})}{m^\alpha} \quad (24) \\ =\ &\limsup_{m\to\infty} -2\theta_{m,\epsilon} x m + \frac{2x\psi_{L,m,\epsilon}(\theta_{m,\epsilon})}{\psi'_{L,m,\epsilon}(\theta_{m,\epsilon})} \\ =\ &-\frac{2x}{(c+\epsilon)(a_1+\epsilon)\beta_1} + \limsup_{m\to\infty} \frac{2x\psi_{L,m,\epsilon}(\theta_{m,\epsilon})}{\psi'_{L,m,\epsilon}(\theta_{m,\epsilon})}. \end{aligned}$$

Because $\psi_{L,m,\epsilon}$ is a convex function passing through the origin,

$$\frac{2x\psi_{L,m,\epsilon}(\theta_{m,\epsilon})}{\psi'_{L,m,\epsilon}(\theta_{m,\epsilon})} \le \frac{2x\psi'_{L,m,\epsilon}(\theta_{m,\epsilon})\theta_{m,\epsilon}}{\psi'_{L,m,\epsilon}(\theta_{m,\epsilon})} = 2x\theta_{m,\epsilon} \to 0.$$

Since $\epsilon > 0$ can be arbitrarily small, (22) holds. By Jensen's inequality, (20) and (22) together imply the two limits in the statement of the theorem. ■

4.3 Numerical Examples

We now illustrate the effectiveness of the IS algorithm through some numerical examples. For our first example, we consider a portfolio with $m = 1000$ obligors and exposures $c_i = 0.04 + 0.00196i$ increasing linearly from 0.042 to 2. We set $a_{i0} \equiv 0.002$ and $a_{ij} \equiv 0.0002$, $j = 1,\ldots,d$. There are $d = 10$ risk factors, each with variance 9. With these parameters, $E[Y_i] = 0.004$ (think of this as the marginal probability of default over 1 year) and the standard deviation of Y_i is 0.002; these values reflect a high degree of variability in the conditional default probabilities.

Table 1 reports variance ratios (variance reduction factors) for several values of q in estimating $P(L > mq)$. Each variance ratio is calculated by estimating the variance per replication using standard simulation and dividing it by the variance per replication using IS. Each estimate in the table is based on 100,000 replications. At larger values of q, the variance ratio becomes very large. The improvement is substantial for probabilities in the range of 1% to 0.1% which are of particular interest in risk management applications.

We have carried out the same experiments using $c_i \equiv 1$ and obtained very similar results. We also obtained very similar results in estimating

$$P(\sum_{i=1}^m \min\{Y_i, 1\}c_i > qm)$$

(i.e., dropping the Poisson approximation in the original model) using the same IS distribution. We obtained greater variance reduction in models with smaller values of σ_j. However, the main determinant of the variance ratio seems to be the magnitude of the probability $P(L > mq)$.

Table 1: Variance Reduction for Increasing q

q	$P(L > mq)$	Var Ratio
0.0080	10.27%	3.27
0.0099	4.94%	5.46
0.0138	1.05%	17.30
0.0156	0.51%	30.81
0.0197	0.10%	120.57

Next we illustrate the effect of increasing m while holding q fixed. For this example, we take $c_i \equiv 1$ and $q = 0.009$. The resulting probabilities $P(L > mq)$ and variance ratios for increasing m are reported in Table 2, each based on 20,000 replications. As expected, the probabilities approach a limit as m increases and the variance ratio also appears to reach a limit. This contrasts with Table 1 where the variance ratio grows quickly with q. The results are consistent with Theorems 2 and 3, which indicate that for large variance reduction we either need q to grow or the effect of the underlying gamma risk factors to weaken.

Table 2: Variance Reduction for Increasing m

m	$P(L > mq)$	Var Ratio
500	8.28%	4.00
1000	4.45%	5.74
2000	3.52%	6.77
5000	2.95%	7.36
10000	2.83%	7.61
20000	2.68%	7.94
100000	2.57%	8.18

5 CONCLUDING REMARKS

This paper has proposed, analyzed, and tested a two-step IS method for estimating loss probabilities in a mixed Poisson model of credit risk. The method applies an exponential twist to the default random variables conditional on the values of common risk factors, and it applies a second exponential twist to the risk factors themselves. We have identified limiting regimes under which this method is asymptotically optimal and illustrated its effectiveness through numerical examples.

The loss distribution in the mixed Poisson model can be calculated numerically through transform inversion (as in CSFP 1997), essentially by using the same cumulant generating function we use in importance sampling. Nevertheless, the strategy we have used here is applicable more generally. It applies, for example, to the normal copula model, for which simulation is the most practical computational tool. In the normal copula model, applying just a single IS step leads to asymptotic optimality only if the model's correlations decrease as the portfolio size increases; effective variance reduction usually requires applying IS to the risk factors as well. The special structure of the mixed Poisson model allows the two steps to be combined into a single exponential change of distribution.

ACKNOWLEDGMENTS

This research is supported in part by NSF grants DMS007463 and DMI0300044.

REFERENCES

Cario, M.C., and B.L. Nelson. 1997. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. IEMS Department, Northwestern University, Evanston, IL.

Credit Suisse Financial Products. 1997. *CreditRisk$^+$: A CreditRisk Management Framework.* London.

Crouhy, M., D. Galai, and R. Mark. 2001. *Risk Management.* New York: McGraw-Hill.

Dembo A. and O. Zeitouni. 1998. *Large Deviations Techniques and Applications*, 2nd ed. New York: Springer.

Glasserman P., P. Heidelberger, P. Shahabuddin. 1999. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Mathematical Finance* 9:117–152.

Glasserman, P., and J. Li. (2003). In preparation.

Gupton, G., C. Finger, M. Bhatia. 1997. CreditMetrics Technical Document. J.P. Morgan & Co., New York. Available at `<www.riskmetrics.com>`.

Johnson, N.L., S. Kotz, and A.W. Kemp. 1993. *Univariate Discrete Distributions,* 2nd ed. New York: Wiley.

Li, D. 2000 On default correlation: a copula function approach. *Journal of Fixed Income* 9:43–54.

Sadowsky J. and J. Bucklew. 1990. On large deviations theory and asymptotically efficient Monte Carlo estimation. *IEEE transactions on information theory* 36:579–588.

AUTHOR BIOGRAPHIES

PAUL GLASSERMAN is the Jack R. Anderson Professor in the Decision, Risk, and Operations Division of Columbia Business School. His research interests include variance reduction techniques and modeling and computational issues in risk management and the pricing of derivative securities. He is author of *Monte Carlo Methods in Financial Engineering*, published by Springer-Verlag in 2003. His e-mail address is `<pg20@columbia.edu>` and his web page is `<www.paulglasserman.net>`.

JINGYI LI is a Ph.D. candidate in the Division of Decision, Risk, and Operations in Columbia Business School, New York, NY. Her main focus of interest has been the application of simulation methods in finance. She received her B.E. and M.E. in electrical engineering from Tsinghua University in 1996 and 1999. Her e-mail address is `<jl976@columbia.edu>`.