MANAGEMENT SCIENCE
Vol. 51, No. 11, November 2005, pp. 1643–1656
ISSN 0025-1909 | EISSN 1526-5501 | 05 | 5111 | 1643
DOI 10.1287/mnsc.1050.0415

Importance Sampling for Portfolio Credit Risk

Paul Glasserman, Jingyi Li
Columbia Business School, Columbia University, New York, New York 10027
{pg20@columbia.edu, jl976@columbia.edu}

Monte Carlo simulation is widely used to measure the credit risk in portfolios of loans, corporate bonds, and other instruments subject to possible default. The accurate measurement of credit risk is often a rare-event simulation problem because default probabilities are low for highly rated obligors and because risk management is particularly concerned with rare but significant losses resulting from a large number of defaults. This makes importance sampling (IS) potentially attractive. But the application of IS is complicated by the mechanisms used to model dependence between obligors, and capturing this dependence is essential to a portfolio view of credit risk. This paper provides an IS procedure for the widely used normal copula model of portfolio credit risk. The procedure has two parts: One applies IS conditional on a set of common factors affecting multiple obligors, the other applies IS to the factors themselves. The relative importance of the two parts of the procedure is determined by the strength of the dependence between obligors. We provide both theoretical and numerical support for the method.

Key words: Monte Carlo simulation; variance reduction; importance sampling; portfolio credit risk
History: Accepted by Wallace J. Hopp, stochastic models and simulation; received January 19, 2004. This paper was with the authors 1 month for 3 revisions.

1. Introduction

Developments in bank supervision and in markets for transferring and trading credit risk have led the financial industry to develop new tools to measure and manage this risk. An important feature of modern credit risk management is that it takes a portfolio view, meaning that it tries to capture the effect of dependence across sources of credit risk to which a bank or other financial institution is exposed. Capturing dependence adds complexity both to the models used and to the computational methods required to calculate outputs of a model.

Monte Carlo simulation is among the most widely used computational tools in risk management. As in other application areas, it has the advantage of being very general and the disadvantage of being rather slow. This motivates research on methods to accelerate simulation through variance reduction. Two features of the credit risk setting pose a particular challenge: (i) it requires accurate estimation of low-probability events of large losses; (ii) the dependence mechanisms commonly used in modeling portfolio credit risk do not immediately lend themselves to rare-event simulation techniques used in other settings.

This paper develops importance sampling (IS) procedures for rare-event simulation for credit risk measurement. We focus on the *normal copula* model originally associated with J. P. Morgan's CreditMetrics system (Gupton et al. 1997) and now widely used. In this framework, dependence between obligors (e.g., corporations to which a bank has extended credit) is captured through a multivariate normal vector of latent variables; a particular obligor defaults if its associated latent variable crosses some threshold. The Creditrisk+ system developed by Credit Suisse First Boston (Wilde 1997) has a similar structure but uses a mixed Poisson mechanism to capture dependence. We present an IS method for that model in Glasserman and Li (2003) together with some preliminary results on the model considered here.

In the normal copula framework, dependence is typically introduced through a set of underlying "factors" affecting multiple obligors. These sometimes have a tangible interpretation—for example, as industry or geographic factors—but they can also be by-products of an estimation procedure. Conditional on the factors, the obligors become independent, and this feature divides our IS procedure into two steps: We apply a change of distribution to the conditional default probabilities, given the factors, and we apply a shift in mean to the factors themselves. The relative importance of the two steps depends on the strength of the dependence across obligors, with greater dependence putting greater importance on the shift in factor mean.

The idea of shifting the factor mean to generate more scenarios with large losses has also been suggested in Avranitis and Gregory (2001), Finger (2001), and Kalkbrener et al. (2003), though with little theoretical support and primarily in single-factor models. The single-factor case is also discussed in

Glasserman (2004, §9.4.3). The main contributions of this paper lie in providing a systematic way of selecting a shift in mean for multifactor models, in integrating the shift in mean with IS for the conditional default probabilities, and in providing a rigorous analysis of the effectiveness of IS in simple models. This analysis makes precise the role played by the strength of the dependence between obligors in determining the impact of the two steps in the IS procedures.

The rest of the paper is organized as follows. Section 2 describes the normal copula model. Section 3 presents background on an IS procedure for the case of independent obligors, which we extend in §4 to the case of *conditionally* independent obligors. Section 5 combines this with a shift in factor mean. We present numerical examples in §6 and collect most proofs in the appendix.

2. Portfolio Credit Risk in the Normal Copula Model

A key element of any model of portfolio credit risk is a mechanism for capturing dependence among obligors. In this section, we describe the widely used normal copula model associated with CreditMetrics (Gupton et al. 1997) and related settings (Li 2000).

Our interest centers on the distribution of losses from default over a fixed horizon. To specify this distribution, we introduce the following notation.

$m =$ number of obligors to which portfolio is exposed
$Y_k =$ default indicator for kth obligor
$= 1$ if kth obligor defaults, 0 otherwise
$p_k =$ marginal probability that kth obligor defaults
$c_k =$ loss resulting from default of kth obligor
$L = c_1Y_1 + \cdots + c_mY_m =$ total loss from defaults

The individual default probabilities p_k are assumed known, either from credit ratings or from the market prices of corporate bonds or credit default swaps. We take the c_k to be known constants for simplicity, though it would suffice to know the distribution of c_kY_k. Our goal is to estimate tail probabilities $P(L > x)$, especially at large values of x.

To model dependence among obligors we need to introduce dependence among the default indicators $Y_1, \ldots, Y_m$. In the normal copula model, dependence is introduced through a multivariate normal vector $(X_1, \ldots, X_m)$ of latent variables. Each default indicator is represented as

$$Y_k = \mathbf{1}\{X_k > x_k\}, \quad k = 1, \ldots, m,$$

with x_k chosen to match the marginal default probability p_k. The threshold x_k is sometimes interpreted as a default boundary of the type arising in the foundational work of Merton (1974). Without loss of generality, we take each X_k to have a standard normal distribution and set $x_k = \Phi^{-1}(1-p_k)$, with Φ the cumulative normal distribution. Thus,

$$P(Y_k=1) = P(X_k > \Phi^{-1}(1-p_k)) = 1-\Phi(\Phi^{-1}(1-p_k)) = p_k.$$

Through this construction, the correlations among the X_k determine the dependence among the Y_k. The underlying correlations are often specified through a factor model of the form

$$X_k = a_{k1}Z_1 + \cdots + a_{kd}Z_d + b_k\epsilon_k, \tag{1}$$

in which

- $Z_1, \ldots, Z_d$ are systematic risk factors, each having an $N(0,1)$ (standard normal) distribution;
- ϵ_k is an idiosyncratic risk associated with the kth obligor, also $N(0,1)$ distributed;
- $a_{k1}, \ldots, a_{kd}$ are the factor loadings for the kth obligor, $a_{k1}^2 + \cdots + a_{kd}^2 \le 1$.
- $b_k = \sqrt{1-(a_{k1}^2 + \cdots + a_{kd}^2)}$ so that X_k is $N(0,1)$.

The underlying factors Z_j are sometimes given economic interpretations (as industry or regional risk factors, for example). We assume that the factor loadings a_{kj} are nonnegative. Though not essential, this condition simplifies our discussion by ensuring that larger values of the factors Z_i lead to a larger number of defaults. Nonnegativity of the a_{kj} is often imposed in practice as a conservative assumption ensuring that all default indicators are positively correlated.

Write a_k for the row vector $(a_{k1}, \ldots, a_{kd})$ of factor loadings for the kth obligor. The correlation between X_k and X_j, $j \neq k$, is given by $a_k a_j^\top$. The conditional default probability for the kth obligor given the factor loadings $Z = (Z_1, \ldots, Z_d)^\top$ is

$$\begin{aligned} p_k(Z) &= P(Y_k = 1 \mid Z) \\ &= P(X_k > x_k \mid Z) = P(a_kZ + b_k\epsilon_k > \Phi^{-1}(1-p_k) \mid Z) \\ &= \Phi\left(\frac{a_kZ + \Phi^{-1}(p_k)}{b_k}\right). \end{aligned} \tag{2}$$

3. Importance Sampling: Independent Obligors

Before discussing IS in the normal copula model, it is useful to consider the simpler problem of estimating loss probabilities when the obligors are independent. So, in this section, we take the default indicators $Y_1, \ldots, Y_m$ to be independent; equivalently, we take all $a_{kj} = 0$.

In this setting, the problem of efficient estimation of $P(L > x)$ reduces to one of applying IS to a sum of independent (but not identically distributed) random variables. For this type of problem there is a fairly well-established approach, which we now describe.

It is intuitively clear that to improve our estimate of a tail probability $P(L > x)$ we want to increase the default probabilities. If we were to replace each default probability p_k by some other default probability q_k, the basic IS identity would be

$$P(L > x) = \tilde{\mathsf{E}}\left[\mathbf{1}\{L > x\}\prod_{k=1}^{m}\left(\frac{p_k}{q_k}\right)^{Y_k}\left(\frac{1-p_k}{1-q_k}\right)^{1-Y_k}\right], \quad (3)$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator of the event in braces, $\tilde{\mathsf{E}}$ indicates that the expectation is taken using the new default probabilities $q_1, \ldots, q_m$, and the product inside the expectation is the likelihood ratio relating the original distribution of $(Y_1, \ldots, Y_m)$ to the new one. Thus, the expression inside the expectation is an unbiased estimate of $P(L > x)$ if the default indicators are sampled using the new default probabilities.

3.1. Exponential Twisting

Rather than increase the default probabilities arbitrarily, we apply an *exponential twist*: We choose a parameter θ and set

$$p_{k,\theta} = \frac{p_k e^{\theta c_k}}{1 + p_k(e^{\theta c_k} - 1)}. \quad (4)$$

If $\theta > 0$, then this does indeed increase the default probabilities; a larger exposure c_k results in a greater increase in the default probability. The original probabilities correspond to $\theta = 0$.

With this choice of probabilities, straightforward calculation shows that the likelihood ratio simplifies to

$$\prod_{k=1}^{m}\left(\frac{p_k}{p_{k,\theta}}\right)^{Y_k}\left(\frac{1-p_k}{1-p_{k,\theta}}\right)^{1-Y_k} = \exp(-\theta L + \psi(\theta)), \quad (5)$$

where

$$\psi(\theta) = \log \mathsf{E}[e^{\theta L}] = \sum_{k=1}^{m}\log(1 + p_k(e^{\theta c_k} - 1))$$

is the cumulant generating function (CGF) of L. For any θ, the estimator

$$\mathbf{1}\{L > x\}e^{-\theta L + \psi(\theta)}$$

is unbiased for $P(L > x)$ if L is generated using the probabilities $p_{k,\theta}$. Equation (5) shows that exponentially twisting the probabilities as in (4) is equivalent to applying an exponential twist to L itself.

It remains to choose the parameter θ. We would like to choose θ to minimize variance or, equivalently, the second moment of the estimator. The second moment is given by

$$\begin{aligned} M_2(x) = M_2(x, \theta) &= \mathsf{E}_\theta\left[\mathbf{1}\{L > x\}e^{-2\theta L + 2\psi(\theta)}\right] \\ &\le e^{-2\theta x + 2\psi(\theta)}, \end{aligned} \quad (6)$$

where E_θ denotes expectation using the θ-twisted probabilities and the upper bound holds for all $\theta \ge 0$. Minimizing $M_2(x, \theta)$ is difficult, but minimizing the upper bound is easy: we need to maximize $\theta x - \psi(\theta)$ over $\theta \ge 0$. The function ψ is strictly convex and passes through the origin, so the maximum is attained at

$$\theta_x = \begin{cases} \text{unique solution to } \psi'(\theta) = x, & x > \psi'(0); \\ 0, & x \le \psi'(0). \end{cases} \quad (7)$$

To estimate $P(L > x)$, we twist by θ_x.

A standard property of exponential twisting (e.g., as in Glasserman 2004, p. 261), easily verified by direct calculation, implies that

$$\mathsf{E}_\theta[L] = \psi'(\theta),$$

and this facilitates the interpretation of θ_x. The two cases in (7) correspond to $x > \mathsf{E}[L]$ and $x \le \mathsf{E}[L]$. In the first case, our choice of twisting parameter implies

$$\mathsf{E}_{\theta_x}[L] = \psi'(\theta_x) = x;$$

thus, we have shifted the distribution of L so that x is now its mean. In the second case, the event $\{L > x\}$ is not rare, so we do not change the probabilities.

3.2. Asymptotic Optimality

A standard way of establishing the effectiveness of a simulation estimator in a rare-event setting is to show that it is asymptotically optimal as the event of interest becomes increasingly rare, meaning that the second moment decreases at the fastest possible rate among all unbiased estimators (see, e.g., Heidelberger 1995 for background). Asymptotic optimality results are often based on making precise approximations of the form

$$P(L > x) \approx e^{-\gamma x}, \qquad M_2(x) \approx e^{-2\gamma x}, \quad (8)$$

for some $\gamma > 0$. The key point here is that the second moment decays at twice the rate of the probability itself. By Jensen's inequality $M_2(x) \ge (P(L > x))^2$, so this is the fastest possible rate of decay.

To formulate a precise result, we let the number of obligors m increase together with the threshold x. This is practically meaningful—bank portfolios can easily be exposed to thousands or even tens of thousands of obligors—as well as theoretically convenient. We therefore need an infinite sequence (p_k, c_k), $k = 1, 2, \ldots,$ of obligor parameters. Write $\psi_m(\theta)$ for the CGF of the loss in the mth portfolio. We require that

$$\psi_m(\theta) = \frac{1}{m}\sum_{k=1}^{m}\log(1 + p_k(e^{\theta c_k} - 1)) \to \bar{\psi}(\theta) \quad (9)$$

for all θ, for some strictly convex $\bar{\psi}$. This holds, for example, if the $\{p_k\}$ have a limit in $(0,1)$ and the $\{c_k\}$ have a limit in $(0,\infty)$. Write L_m for the loss in the mth portfolio, write θ_m for the value of θ_{xm} in (7) for the mth portfolio, and write $M_2(xm,\theta_m)$ for the second moment of the IS estimator of $P(L_m > xm)$ using θ_m.

Theorem 1. *Suppose* (9) *holds for all* θ *and there is a* $\bar{\theta}_x > 0$ *at which* $\bar{\psi}'(\bar{\theta}_x) = x$. *Then*

$$\lim_{m\to\infty} \frac{1}{m} \log P(L_m > xm) = -\gamma$$

$$\lim_{m\to\infty} \frac{1}{m} \log M_2(xm,\theta_m) = -2\gamma,$$

where

$$\gamma = \sup_\theta \{\theta x - \bar{\psi}(\theta)\}.$$

Thus, the IS estimator is asymptotically optimal.

Proof. The result would follow from a more general result in Sadowsky and Bucklew (1990) but for a difference in the choice of twisting parameter. (They use $\bar{\theta}_x$.) Under (9), the Gärtner-Ellis theorem (as in Dembo and Zeitouni 1998, §2.3) gives

$$\liminf_{m\to\infty} \frac{1}{m} \log P(L_m > xm) \geq -\gamma. \tag{10}$$

Using (6), we find that

$$\begin{aligned}\limsup_{m\to\infty} \frac{1}{m} \log M_2(xm,\theta_m) &\leq \limsup_{m\to\infty} -2(\theta_m x - \psi_m(\theta_m)/m)\\ &\leq \limsup_{m\to\infty} -2(\bar{\theta}_x x - \psi_m(\bar{\theta}_x)/m)\\ &= -2(\bar{\theta}_x x - \bar{\psi}(\bar{\theta}_x)) = -2\gamma.\end{aligned} \tag{11}$$

By Jensen's inequality, $M_2(xm,\theta_m) \geq P(L_m > xm)^2$ so the lim inf in (10) and the lim sup in (11) hold as limits. □

This use of Jensen's inequality in the proof of the limit applies in the proofs of all other theorems in this paper. So, it will suffice to prove the lim inf of the probability and the lim sup of the second moment as in Theorem 1 in subsequent theorems.

The limits in Theorem 1 make precise the approximations in (8). This result (and its extension in the next section) continues to hold if $\{Y_k c_k, k \geq 1\}$ is replaced with a more general sequence of independent light-tailed random variables under modest regularity conditions.

4. Dependent Obligors: Conditional Importance Sampling

As a first step in extending the method of §3 to the normal copula model of §2, we apply importance sampling conditional on the common factors Z. Conditional on $Z = z$, the default indicators are independent, the kth obligor having conditional default probability $p_k(z)$ as defined in (2). We may therefore proceed exactly as in §3.

In more detail, this entails the following steps for each replication:

1. Generate $Z \sim N(0,I)$, a d-vector of independent normal random variables.
2. Calculate the conditional default probabilities $p_k(Z)$, $k = 1,\ldots,m$, in (2). If

$$\mathsf{E}[L \mid Z] \equiv \sum_{k=1}^m p_k(Z)c_k \geq x,$$

set $\theta_x(Z) = 0$; otherwise, set $\theta_x(Z)$ equal to the unique solution of

$$\frac{\partial}{\partial\theta}\psi_m(\theta,Z) = x,$$

with

$$\psi_m(\theta,z) = \sum_{k=1}^m \log(1 + p_k(z)(e^{\theta c_k} - 1)). \tag{12}$$

3. Generate default indicators $Y_1,\ldots,Y_m$ from the twisted conditional default probabilities

$$p_{k,\theta_x(Z)}(Z) = \frac{p_k(Z)e^{\theta_x(Z)c_k}}{1 + p_k(Z)(e^{\theta_x(Z)c_k} - 1)}, \quad k = 1,\ldots,m.$$

4. Compute the loss $L = c_1Y_1 + \cdots + c_mY_m$ and return the estimator

$$\mathbf{1}\{L > x\}\exp(-\theta_x(Z)L + \psi(\theta_x(Z),Z)). \tag{13}$$

Although here we have tailored the parameter θ to a particular x, we will see in the numerical results that the same value of θ can be used to estimate $P(L > x)$ over a wide range of loss levels.

The effectiveness of this algorithm depends on the strength of the dependence among obligors. When the dependence is weak, increasing the conditional default probabilities is sufficient to achieve substantial variance reduction. But this method is much less effective at high correlations, because in that case large losses occur primarily because of large outcomes of Z, and we have not yet applied IS to the distribution of Z.

To illustrate these points more precisely, we consider the special case of a single-factor, homogeneous portfolio. This means that Z is scalar ($d = 1$), all p_k are equal to some p, all c_k are equal to a fixed value (which we take to be 1), and all obligors have the same loading ρ on the single factor Z. Thus, the latent variables have the form

$$X_k = \rho Z + \sqrt{1-\rho^2}\epsilon_k, \quad k = 1,\ldots,m, \tag{14}$$

and the conditional default probabilities are

$$\begin{aligned}p(z) &= P(Y_k = 1 \mid Z = z) = P(X_k > -\Phi^{-1}(p) \mid Z = z)\\ &= \Phi\left(\frac{\rho z + \Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right).\end{aligned} \tag{15}$$

As in Theorem 1, we consider the limit as both the number of obligors m and the loss level x increase. We now also allow ρ to depend on m through a specification of the form $\rho = a/m^{\alpha}$ for some $a, \alpha > 0$. Thus, the strength of the correlation between obligors is measured by the rate at which ρ decreases, with small values of α corresponding to stronger correlations. If we kept ρ fixed, then by the law of large numbers, L_m/m would converge to $p(Z)$ and, with $0 < q < 1$,

$$P(L_m > qm) \to P(p(Z) > q) = 1 - \Phi\left(\frac{\Phi^{-1}(q)\sqrt{1-\rho^2} - \Phi^{-1}(p)}{\rho}\right) > 0. \tag{16}$$

Thus, with ρ fixed, the event $\{L_m > qm\}$ does not become vanishingly rare as m increases. By letting ρ decrease with m we will, however, get a probability that decays exponentially fast.

Define

$$G(p) = \begin{cases} \log\left(\dfrac{1-p}{1-q}\right)^{1-q}\left(\dfrac{p}{q}\right)^{q}, & p < q, \\ 0, & p \geq q; \end{cases} \tag{17}$$

$G(p) \leq 0$, and $\exp(mG(p))$ is the likelihood ratio at $L = mq$ for the independent case with marginal individual default probability p. Also define (with $\rho = a/m^{\alpha}$)

$$F(a, z) = \lim_{m\to\infty} G(p(zm^{\alpha})) = G\big(\Phi(az + \Phi^{-1}(p))\big) \tag{18}$$

and

$$F_m(z) = -\theta_{mq}(z)mq + \psi_m(\theta_{mq}(z), z) = mG(p(z)) \tag{19}$$

using the expression for $p(z)$ in (15); $F(a, z) \leq 0$ and $F_m(z) \leq 0$ because $G(p) \leq 0$. The following lemma is verified by differentiation.

Lemma 1. *In the single-factor, homogeneous portfolio specified by $p_k \equiv p$, $c_k \equiv 1$, and* (14), *the function $z \mapsto F_m(z)$ is increasing and concave.*

Write $M_2(mq, \theta_{mq})$ for the second moment of the IS estimator (13) with $x = mq$.

Theorem 2. *Consider the single-factor, homogeneous portfolio specified by $p_k \equiv p \in (0, 1/2)$, $c_k \equiv 1$, and* (14). *For constant $q > p$, if $\rho = a/m^{\alpha}$, $a > 0$, then*

(a) *For $\alpha > 1/2$,*

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

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

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

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

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

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 conditional importance sampling 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. When the obligors are highly correlated, large losses occur primarily because of large moves in Z, but this is not (yet) reflected in our IS distribution.

This is also suggested by the form of the limits in the theorem. The limits in all three cases can be put in the same form as (b) (with (a) corresponding to the limiting case $a \to 0$ and (c) corresponding to $a \to \infty$) once we note that $F(0, z)$ is independent of z in (a) and $F(a, z_a) = 0$ in (c). In each case, the two terms in the expression $F(a, z) - z^2/2$ result from two sources of randomness: the second term results from the tail of the common factor Z and the first term results from the tail of the conditional loss distribution given Z. Because our conditional IS procedure is asymptotically optimal for the conditional loss probability given Z, a factor of 2 multiplies $F(a, z)$ in each second moment. To achieve asymptotic optimality for the unconditional probability, we need to apply IS to Z as well.

5. Dependent Obligors: Two-Step Importance Sampling

5.1. Shifting the Factors

We proceed now to apply IS to the distribution of the factors $Z = (Z_1, \ldots, Z_d)^{\top}$ as well as to the conditional default probabilities. To motivate the approach we take, observe that for any estimator $\hat{p}_x$ of $P(L > x)$ we have the decomposition

$$\mathsf{Var}[\hat{p}_x] = \mathsf{E}[\mathsf{Var}[\hat{p}_x \mid Z]] + \mathsf{Var}[\mathsf{E}[\hat{p}_x \mid Z]].$$

Conditional on Z, the obligors are independent, so we know from §3 how to apply asymptotically optimal IS. This makes $\mathsf{Var}[\hat{p}_x \mid Z]$ small and suggests that in applying IS to Z we should focus on the second term in the variance decomposition. When $\hat{p}_x$ is the estimator of §4, $\mathsf{E}[\hat{p}_x \mid Z] = P(L > x \mid Z)$, so this means that we should choose an IS distribution for Z that would reduce variance in estimating the integral of $P(L > x \mid Z)$ against the density of Z.

The zero-variance IS distribution for this problem would sample Z from the density proportional to the function

$$z \mapsto P(L > x \mid Z = z) e^{-z^\top z/2}.$$

But sampling from this density is generally infeasible—the normalization constant required to make it a density is the value $P(L > x)$ we seek. Faced with a similar problem in an option-pricing context, Glasserman et al. (1999) suggest using a normal density with the same mode as the optimal density. This mode occurs at the solution to the optimization problem

$$\max_z P(L > x \mid Z = z) e^{-z^\top z/2}, \tag{20}$$

which is then also the mean of the approximating normal distribution.

Once we have selected a new mean μ for Z, the IS algorithm proceeds as follows:

1. Sample Z from $N(\mu, I)$.
2. Apply the procedure in §4 to compute $\theta_x(Z)$ and the twisted conditional default probabilities $p_{k,\theta_x(Z)}(Z)$, $k = 1, \ldots, m$, and to generate the loss under the twisted conditional distribution.
3. Return the estimator

$$\mathbf{1}\{L > x\} e^{-\theta_x(Z)L + \psi(\theta_x(Z), Z)} e^{-\mu^\top Z + \mu^\top \mu/2}. \tag{21}$$

The last factor in the estimator (the only new one) is the likelihood ratio relating the density of the $N(0, I)$ distribution to that of the $N(\mu, I)$ distribution. Thus, this two-step IS procedure is no more difficult than the one-step method of §4, once μ has been determined.

The key, then, is finding μ. Exact solution of (20) is usually difficult, but there are several ways of simplifying the problem through further approximation:

Constant Approximation. Replace L with $\mathsf{E}[L \mid Z = z]$ and $P(L > x \mid Z = z)$ with $\mathbf{1}\{\mathsf{E}[L \mid Z = z] > x\}$. The optimization problem then becomes one of minimizing $z^\top z$ subject to $\mathsf{E}[L \mid Z = z] > x$.

Normal Approximation. Note that $\mathsf{E}[L \mid Z = z] = \sum_k p_k(z) c_k$ and $\mathsf{Var}[L \mid Z = z] = \sum_k c_k^2 p_k(z)(1 - p_k(z))$ and use the normal approximation

$$P(L > x \mid Z = z) \approx 1 - \Phi\left(\frac{x - \mathsf{E}[L \mid Z = z]}{\sqrt{\mathsf{Var}[L \mid Z = z]}}\right)$$

in (20).

Saddlepoint Approximation. Given the conditional cumulant generating function $\psi(\theta, Z)$ in (12), one can approximate $P(L > x \mid Z = z)$ using any of the various saddlepoint methods in Jensen (1995), as in, for example, Martin et al. (2001). These are approximate methods for inverting the conditional characteristic function of L given Z. (In cases for which precise inversion of the characteristic function is practical, one may use the conditional Monte Carlo estimator $P(L > x \mid Z)$ and dispense with simulation of the default indicators.)

Tail Bound Approximation. Define

$$F_x(z) = -\theta_x(z)x + \psi(\theta_x(z), z); \tag{22}$$

this is the logarithm of the likelihood ratio in (13) evaluated at $L = x$. The smallest maximizer of $F_x(z)$ is the solution using the constant approximation described above because $F_m(z)$ attains its maximum value 0 only when $\mathsf{E}[L \mid Z = z] \geq x$. The inequality $\mathbf{1}\{y > x\} \leq \exp(\theta(y - x))$, $\theta \geq 0$, gives

$$P(L > x \mid Z = z) \leq \mathsf{E}\left[e^{\theta_x(Z)(L-x)} \mid Z = z\right] = e^{F_x(z)}.$$

By treating this bound as an approximation in (20) and taking logarithms, we arrive at the optimization problem

$$J(x) = \max_z \left\{F_x(z) - \tfrac{1}{2} z^\top z\right\}. \tag{23}$$

Other approximations are possible. For example, in a related problem, Kalkbrener et al. (2003) calculate the optimal mean for a single-factor approximation and then "lift" this scalar mean to a mean vector for Z. We focus on (23) because it provides the most convenient way to blend IS for Z with IS conditional on Z. The expression in (22) is also the exponent of the usual (conditional) saddlepoint approximation, which lends further support to the idea of using this upper bound as an approximation. The various methods discussed above will generally produce different values of μ. Our asymptotic optimality results are based on (23), but this does not preclude the possibility that the other methods are asymptotically optimal as well.

5.2. Asymptotic Optimality: Large Loss Threshold

We now turn to the question of asymptotic optimality of our combined IS procedure. We establish asymptotic optimality in the case of the single-factor, homogeneous portfolio used in Theorem 2 and then comment on the general case. As in §§3 and 4, we consider a limit in which the loss threshold increases together with the size of the portfolio. In contrast to Theorem 2, here we do not need to assume that the correlation parameter ρ decreases; this is the main implication of applying IS to the factors Z.

As noted in (16), when the size of the homogenous portfolio increases, the normalized loss L_m/m converges to a nonrandom limit taking values in the unit interval. In order that $P(L_m > x_m)$ vanish as

m increases, we therefore let x_m/m approach 1 from below. The following specification turns out to be appropriate:

$$x_m = m \cdot q_m, \qquad q_m = \Phi\left(c\sqrt{\log m}\right), \qquad 0 < c < \sqrt{2}. \tag{24}$$

Here q_m approaches 1 from below and $1 - q_m$ is $O(m^{-c^2/2})$. Write μ_m for the maximizer in (23) with $x = x_m$ and write $J(x_m)$ for the maximum value. That the maximizer is unique is ensured by Lemma 1. It follows that the objective in (23) is strictly concave and that the maximum is attained at just one point. Write $M_2(x_m, \mu_m)$ for the second moment of the combined IS estimator of $P(L_m > x_m)$ using μ_m.

Theorem 3. *Consider the single-factor, homogeneous portfolio specified by* $p_k \equiv p$, $c_k \equiv 1$, *and* (14). *With* x_m *as in* (24) *let*

$$\gamma = \frac{c^2}{2}\frac{1-\rho^2}{\rho^2};$$

then

$$\lim_{m\to\infty} \frac{1}{\log m} \log P(L_m > x_m) = -\gamma$$

$$\lim_{m\to\infty} \frac{1}{\log m} \log M_2(x_m, \mu_m) = -2\gamma.$$

The combined IS estimator is thus asymptotically optimal.

This result indicates that our combined IS procedure should be effective in estimating probabilities of large losses in large portfolios. Because we have normalized the limits by $\log m$, this result implies that the decay rates are polynomial rather than exponential in this case: $P(L_m > x_m) \approx m^{-\gamma}$ and $M_2(x_m, \mu_m) \approx m^{-2\gamma}$.

Although Theorem 3 is specific to the single-factor, homogeneous portfolio, we expect the same IS procedure to be effective more generally. We illustrate this through numerical examples in §6, but first we offer some observations.

The key to establishing the effectiveness of the IS procedure is obtaining an upper bound for the second moment. As a direct consequence of the concavity of $F_x(\cdot)$, the argument in the appendix shows that

$$M_2(x_m, \mu_m) \le \exp\left(2m\left\{\max_z\{F_{x_m}(z) - \tfrac{1}{2}z^\top z\}\right\}\right), \tag{25}$$

without further conditions on the portfolio, and this bound already contains much of the significance of Theorem 3. In fact, the assumption of concavity can be relaxed to the requirement that the tangent to some optimizer μ be a dominating hyperplane in the sense that

$$F_x(z) \le F_x(\mu) + \nabla F_x(\mu)(z - \mu), \tag{26}$$

with ∇F_x (a row vector) the gradient of F_x. In the proof of Theorem 3, we show that the maximizer of (23) coincides, in the limit, with the smallest maximizer of F_{x_m} (the solution using the constant approximation described in §5.1) in the sense that their ratio tends to 1. Property (26) is satisfied by any maximizer of F_x, so we expect it to be satisfied by a maximizer of (23) as m increases. This suggests that (25) should hold for large m in greater generality than that provided by Theorem 3.

We view the limiting regime in Theorem 3 as more natural than that of Theorem 2 because we are interested in the probability of large losses in large portfolios. In Theorem 2, we are forced to let the correlation parameter decrease with the size of the portfolio in order to get a meaningful limit for the second moment; this indicates that twisting the (conditional) default probabilities without shifting the factor mean is effective only when the correlation parameter is small. In practice, we face fixed loss levels for fixed portfolios, not limiting regimes; we therefore supplement our asymptotic optimality results with numerical experiments.

5.3. Asymptotic Optimality: Small Default Probability

We next show that our two-step IS procedure is also effective when large losses are rare because individual default probabilities are small. This setting is relevant to portfolios of highly rated obligors, for which one-year default probabilities are extremely small. (Historical one-year default probabilities are less than 0.1% for A-rated firms and even smaller for firms rated AA or AAA.) It is also relevant to measuring risk over short time horizons. In securities lending, for example, banks need to measure credit risk over periods as short as three days, for which default probabilities are indeed small.

For this limiting regime, it is convenient to take

$$x_m = m \cdot q, \quad 0 < q < 1, \qquad p^{(m)} = \Phi(-c\sqrt{m}), \quad c > 0. \tag{27}$$

Here $p^{(m)}$ decreases toward zero and is $O(\exp(-mc^2/2))$. We again consider a single-factor, homogeneous portfolio in which all obligors have the same default probability $p = p^{(m)}$ and all have a loss given default of 1.

Theorem 4. *For a single-factor, homogeneous portfolio with default probability* $p^{(m)}$ *and loss threshold* x_m *as in* (27), *let*

$$\gamma = \frac{c^2}{2\rho^2};$$

then

$$\lim_{m\to\infty} \frac{1}{m} \log P(L_m > x_m) = -\gamma$$

$$\lim_{m\to\infty} \frac{1}{m} \log M_2(x_m, \mu_m) = -2\gamma.$$

The combined IS estimator is thus asymptotically optimal.

6. Numerical Examples

We now illustrate the performance of our two-step IS procedure in multifactor models through numerical experiments. For these results, we shift the mean of the factors Z to the solution of (23) and then twist the conditional default probabilities by $\theta_x(Z)$, as in (21).

Our first example is a portfolio of $m = 1{,}000$ obligors in a 10-factor model. The marginal default probabilities and exposures have the following form:

$$p_k = 0.01 \cdot (1 + \sin(16\pi k/m)), \quad k = 1, \ldots, m; \tag{28}$$

$$c_k = (\lceil 5k/m \rceil)^2, \quad k = 1, \ldots, m. \tag{29}$$

Thus, the marginal default probabilities vary between 0% and 2% with a mean of 1%, and the possible exposures are 1, 4, 9, 16, and 25, with 200 obligors at each level. These parameters represent a significant departure from a homogeneous model with constant p_k and c_k.

For the factor loading, we generate the a_{kj} independently and uniformly from the interval $(0, 1/\sqrt{d})$, $d = 10$; the upper limit of this interval ensures that the sum of squared entries $a_{k1}^2 + \cdots + a_{kd}^2$ for each obligor k does not exceed 1.

Our IS procedure is designed to estimate a tail probability $P(L > x)$, so we also need to select a loss threshold x. The choice of x affects the parameter $\theta_x(z)$ and the optimal solution to (23) used as the mean of the factors Z under the IS distribution. For this example, we use $x = 1{,}000$. The components of the factor mean vector (the solution we found to (23)) are all about 0.8, reflecting the equal importance of the 10 factors in this model. Using a modified Newton method from the Numerical Algorithms Group library (function e04lbc) running on a SunFire V880 workstation, the optimization takes 0.2 seconds.

Although the IS distribution depends on our choice of x, we can use the same samples to estimate $P(L > y)$ at values of y different from x. The IS estimator of $P(L > y)$ is (see (21))

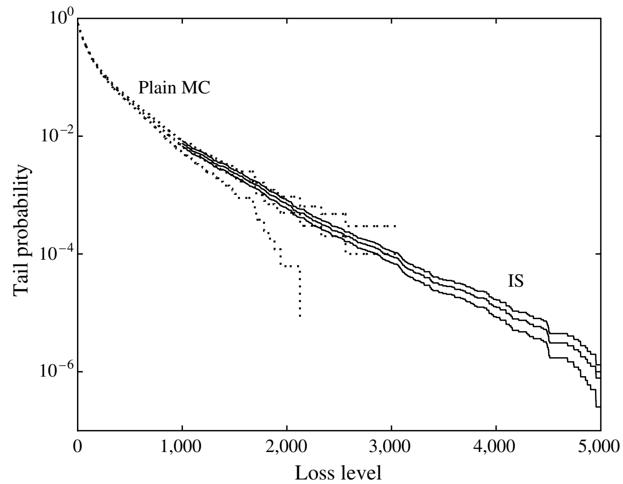
$$\mathbf{1}\{L > y\} e^{-\theta_x(Z) + \psi(\theta_x(Z), Z)} e^{-\mu^\top Z + \mu^\top \mu/2}. \tag{30}$$

In practice, we find that this works well for values of y larger than x, even much larger than x.

Figure 1 compares the performance of importance sampling (labeled IS) and ordinary Monte Carlo simulation (labeled Plain MC) in estimating the tail of the loss distribution, $P(L > y)$. The Plain MC results are based on 10,000 replications whereas the IS results use just 1,000. In each case, the three curves show the sample mean (the center line) and a 95% confidence interval (the two outer lines) computed separately at each point.

The IS estimates of $P(L > y)$ are all computed from the same samples and the same IS distribution, as in (30), with x fixed at 1,000. The figure indicates that the IS procedure gives excellent results even for values of y much larger than x. With the loss distribution of L centered near 1,000 under IS, small losses become rare events, so the IS procedure is not useful at small loss levels. To estimate the entire loss distribution, one could therefore use ordinary simulation for small loss levels and IS for the tail of the distribution. The IS methods take approximately twice as long per replication as ordinary simulation, but this is more than offset by the dramatic increase in precision in the tail.

Figure 1 Comparison of Plain Monte Carlo Simulation (Using 10,000 Replications) with IS (Using 1,000 Replications) in Estimating Loss Probabilities in a 10-Factor Model



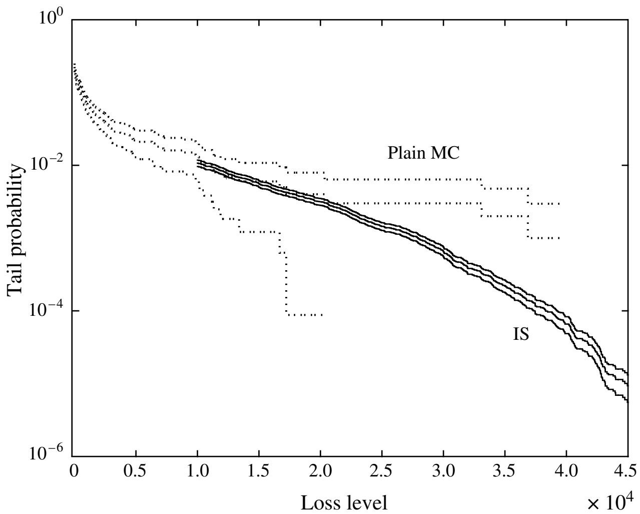
Note. In each case, the three curves show the mean and a 95% confidence interval for each point.

Our next example is a 21-factor model, again with $m = 1{,}000$ obligors. The marginal default probabilities fluctuate as in (28), and the exposures c_k increase linearly from 1 to 100 as k increases from 1 to 1,000. The matrix of factor loadings, $A = (a_{kj}, k = 1, \ldots, 1{,}000, j = 1, \ldots, 21)$, has the following block structure:

$$A = \left(R \left| \begin{matrix} F & & \\ & \ddots & \\ & & F \end{matrix} \right| \begin{matrix} G \\ \vdots \\ G \end{matrix} \right), \qquad G = \begin{pmatrix} g & & \\ & \ddots & \\ & & g \end{pmatrix}, \tag{31}$$

with R a column vector of 1,000 entries, all equal to 0.8; F, a column vector of 100 entries, all equal to 0.4; G a 100×10 matrix; and g, a column vector of 10 entries, all equal to 0.4. This structure is suggestive of the following setting: The first factor is a marketwide factor affecting all obligors, the next 10 factors (the F block) are industry factors, and the last 10 (the G block) are geographical factors. Each obligor then has a loading of 0.4 on one industry factor and on one geographical factor, as well as a loading of 0.8

Figure 2 Comparison of Plain Monte Carlo Simulation (Using 10,000 Replications) with IS (Using 1,000 Replications) in Estimating Loss Probabilities in a 21-Factor Model



Note. In each case, the three curves show the mean and a 95% confidence interval for each point.

on the marketwide factor. There are 100 combinations of industries and regions and exactly 10 obligors fall in each combination. This highly structured dependence on the factors differentiates this example from the previous one. The optimal mean vector (selected through (23)) for this example has 2.46 as its first component and much smaller values (around 0.20) for the other components, reflecting the greater importance of the first factor.

Results for this model are shown in Figure 2, which should be read in the same way as Figure 1. The parameters of the IS distribution are chosen with $x = 10{,}000$, but we again see that IS remains extremely effective even at much larger loss levels.

The effectiveness of IS for this example is quantified in Table 1. The table shows estimated variance reduction factors in estimating $P(L > y)$ at various levels of y, all using the IS parameters corresponding to $x = 10{,}000$. The variance reduction factor is the variance per replication using ordinary simulation divided by the variance per replication using IS. As expected, the variance reduction factor grows as we move farther into the tail of the distribution. The IS method takes approximately twice as long per replication as ordinary simulation, so the (large sample) efficiency ratios are about half as large as the variance ratios. This comparison does not include the fixed cost involved in finding the optimal μ. In our experiments, the time required for the optimization is equal to the time to run 100–1,000 replications.

Table 1 Variance Reduction Factors Using IS at Various Loss Levels

y	$P(L > y)$	V.R. factor
10,000	0.0114	33
14,000	0.0065	53
18,000	0.0037	83
22,000	0.0021	125
30,000	0.0006	278
40,000	0.0001	977

There is no guarantee that a local optimum in (23) is a unique global optimum. If $P(L > x \mid Z = z)$ is multimodal as a function of z, the optimal IS distribution for Z may also be multimodal. A natural extension of the procedure we have described would combine IS estimates using several local optima for (23), as in, e.g., Avramidis (2001). However, we have not found this to be necessary in the examples we have tested.

7. Concluding Remarks

We have developed and analyzed an IS procedure for estimating loss probabilities in a standard model of portfolio credit risk. Our IS procedure is specifically tailored to address the complex dependence between defaults of multiple obligors that results from a normal copula. The procedure shifts the mean of underlying factors and then applies exponential twisting to default probabilities conditional on the factors. The first step is essential if the obligors are strongly dependent; the second step is essential (and sufficient) if the obligors are weakly dependent. We have established the asymptotic optimality of our procedure in single-factor, homogeneous models and illustrated its effectiveness in more complex models through numerical results.

Acknowledgments
The authors thank the referees for their careful reading of the paper and detailed comments. This work is supported in part by NSF grants DMS007463 and DMI0300044.

Appendix. Proofs

Proof of Theorem 2. We first prove (b) and then use similar arguments for (a) and (c). In all three cases, we prove an upper bound for the second moment and a lower bound for the probability, as these are the most important steps. We omit the proof of the lower bound for the second moment and the upper bound for the probability because they follow by very similar arguments.

(b) First we show that

$$\limsup_{m\to\infty} m^{-1} \log M_2(mq, \theta_{mq}) \le \max_z \{2F(a,z) - \tfrac{1}{2}z^2\}. \quad (32)$$

Define $\tilde{P}$ as

$$\tilde{P}(L=y, Z\in dz) = \begin{cases} \mathrm{Bin}(y; m, q)\phi(z)dz, & p(z) \le q, \\ \mathrm{Bin}(y; m, p(z))\phi(z)dz, & p(z) > q, \end{cases} \quad (33)$$

whereas under the original distribution, $P(L = y, Z \in dz) = \mathrm{Bin}(y; m, p(z))\phi(z)dz$. Here $\mathrm{Bin}(\cdot; m, p)$ denotes the probability mass function of the binomial distribution with parameters m and p. As in (5), it follows that

$$\frac{dP(L,Z)}{d\tilde{P}(L,Z)} = \begin{cases} \left(\dfrac{1-p(Z)}{1-q}\right)^{m-L}\left(\dfrac{p(Z)}{q}\right)^{L}, & p(Z)\le q,\\ 1, & p(Z)>q \end{cases}$$
$$= \exp(-\theta_{mq}(Z)L+\psi_m(\theta_{mq}(Z),Z)),$$

with

$$\theta_{mq}(Z) = \left[\log\frac{(1-p(Z))q}{(1-q)p(Z)}\right]^{+}.$$

Writing $M_2(mq,\theta_{mq})$ as an expectation $\tilde{\mathsf{E}}$ under $\tilde{P}$, we get

$$\begin{aligned} M_2(mq,\theta_{mq}) &= \tilde{\mathsf{E}}[\mathbf{1}\{L>mq\}\exp(-2\theta_{mq}(Z)L+2\psi_m(\theta_{mq}(Z),Z))]\\ &\le \tilde{\mathsf{E}}[\exp(-2\theta_{mq}(Z)mq+2\psi_m(\theta_{mq}(Z),Z))]\\ &= \tilde{\mathsf{E}}[\exp(2F_m(Z))]. \end{aligned} \quad (34)$$

Here $F_m(z)$ is as defined in (19). The function $F_m(z)$ is concave from Lemma 1 and $F_m(z)$ attains its maximum value 0 for all $z\ge z_m\sqrt{m}$ where

$$z_m = \frac{\Phi^{-1}(q)\sqrt{1-a^2/m}-\Phi^{-1}(p)}{a} \quad \text{and} \quad p(z_m\sqrt{m})=q.$$

This is obtained from the fact that $G(p(z))$ attains its maximum value 0 for all $p(z)\ge q$ and $p(z)$ is an increasing function in z. Because $q>p$, we have $z_m>0$ for all sufficiently large m.

For any z^0, the concavity of $F_m(z)$ implies $F_m(z)\le F_m(z^0)+F_m'(z^0)(z-z^0)$. Then

$$\begin{aligned} \tilde{\mathsf{E}}[\exp(2F_m(Z))] &\le \tilde{\mathsf{E}}[\exp(2F_m(z^0)+2F_m'(z^0)(Z-z^0))]\\ &= \exp(2F_m(z^0)-2F_m'(z^0)z^0+2(F_m'(z^0))^2). \end{aligned}$$

The term $2(F_m'(z^0))^2$ appears because $\tilde{\mathsf{E}}[\exp(aZ)]=\exp(a^2/2)$ for any constant a. Choose z^0 to be the point at which $2F_m(z)-\frac{1}{2}z^2$ is maximized. Then $2F_m'(z^0)=z^0$ and

$$\begin{aligned} \tilde{\mathsf{E}}[\exp(2F_m(Z))] &\le \exp\left(2F_m(z^0)-\frac{1}{2}(z^0)^2\right)\\ &= \exp\left(\max_z\left\{2F_m(z)-\frac{1}{2}z^2\right\}\right)\\ &= \exp\left(m\max_z\left\{\frac{2F_m(z\sqrt{m})}{m}-\frac{1}{2}z^2\right\}\right). \end{aligned} \quad (35)$$

Because $p\in(0,1/2)$, $\Phi^{-1}(p)<0$, comparison of (18) and (19) shows that

$$\begin{aligned} \frac{F_m(z\sqrt{m})}{m} &= G\left(\Phi\left(\frac{az+\Phi^{-1}(p)}{\sqrt{1-a^2/m}}\right)\right)\\ &\le G\left(\Phi\left(\frac{az}{\sqrt{1-a^2/m}}+\Phi^{-1}(p)\right)\right)\\ &= F\left(a,\frac{z}{\sqrt{1-a^2/m}}\right) \end{aligned}$$

and therefore

$$\begin{aligned} &\max_z\left\{\frac{2F_m(z\sqrt{m})}{m}-\frac{1}{2}z^2\right\}\\ &\quad\le \max_z\left\{2F\left(a,\frac{z}{\sqrt{1-a^2/m}}\right)-\frac{1}{2}z^2\right\}\\ &\quad= \max_z\left\{2F(a,z)-\frac{1}{2}z^2(1-a^2/m)\right\}\\ &\quad\le (1-a^2/m)\max_z\left\{2F(a,z)-\frac{1}{2}z^2\right\}. \end{aligned} \quad (36)$$

Combining this with (34) and (35) proves (32).

Next we show that

$$\liminf_{m\to\infty} m^{-1}\log P(L>mq) \ge \max_z\{F(a,z)-\tfrac{1}{2}z^2\}. \quad (37)$$

For any $\delta>0$, define $\tilde{P}_\delta$ as in (33) but with q replaced by $q+\delta$. It follows that

$$\frac{dP(L,Z)}{d\tilde{P}_\delta(L,Z)} = \exp(-\theta_\delta(Z)L+\psi_m(\theta_\delta(Z),Z)),$$

with

$$\theta_\delta(Z) = \left[\log\frac{(1-p(Z))(q+\delta)}{(1-q-\delta)p(Z)}\right]^{+}.$$

Writing $P(L>mq)$ as an expectation $\tilde{\mathsf{E}}_\delta$ under $\tilde{P}_\delta$, we get

$$\begin{aligned} P(L>mq) &= \tilde{\mathsf{E}}_\delta[\mathbf{1}\{L>mq\}\exp(-\theta_\delta(Z)L+\psi_m(\theta_\delta(Z),Z))]\\ &\ge \tilde{\mathsf{E}}_\delta[\mathbf{1}\{mq<L\le m(q+\delta)\}\\ &\qquad\cdot\exp(-\theta_\delta(Z)m(q+\delta)+\psi_m(\theta_\delta(Z),Z))]\\ &= \tilde{\mathsf{E}}_\delta[\mathbf{1}\{mq<L\le m(q+\delta)\}\exp(mG_\delta(p(Z)))]\\ &\ge \tilde{\mathsf{E}}_\delta[\mathbf{1}\{mq<L\le m(q+\delta)\}\\ &\qquad\cdot\exp(mG_\delta(p(Z)))\mathbf{1}\{p(Z)\le q+\delta\}], \end{aligned} \quad (38)$$

with

$$G_\delta(p) = \begin{cases} \log\left(\dfrac{1-p}{1-q-\delta}\right)^{1-q-\delta}\left(\dfrac{p}{q+\delta}\right)^{q+\delta}, & p\le q+\delta;\\ 0, & p>q+\delta. \end{cases}$$

Under $\tilde{P}_\delta$, L and Z are independent given $p(Z)<q+\delta$. So

$$\begin{aligned} &\tilde{\mathsf{E}}_\delta[\mathbf{1}\{mq<L\le m(q+\delta)\}\exp(mG_\delta(p(Z)))\mathbf{1}\{p(Z)\le q+\delta\}]\\ &\quad= \tilde{\mathsf{E}}_\delta[\mathbf{1}\{mq<L\le m(q+\delta)\}\mid p(Z)<q+\delta]\\ &\qquad\cdot\tilde{\mathsf{E}}_\delta[\exp(mG_\delta(p(Z)))\mathbf{1}\{p(Z)\le q+\delta\}]. \end{aligned} \quad (39)$$

Also, given $p(Z)\le q+\delta$, the loss L has a binomial distribution with parameters m and $q+\delta$ under $\tilde{P}_\delta$. By the central limit theorem, for m large enough,

$$\begin{aligned} &\tilde{P}_\delta(mq<L\le m(q+\delta)\mid p(Z)\le q+\delta)\\ &\quad= \tilde{P}_\delta\left(-\sqrt{\frac{m}{(q+\delta)(1-q-\delta)}}\,\delta<\frac{L-m(q+\delta)}{\sqrt{m(q+\delta)(1-q-\delta)}}\right.\\ &\qquad\left.\le 0\mid p(Z)\le q+\delta\right)\ge\frac{1}{4}. \end{aligned} \quad (40)$$

To bound the second factor in (39), we separate the cases $q<\frac{1}{2}$ and $q\ge\frac{1}{2}$. If $q<\frac{1}{2}$, we can choose δ small enough so that $q+\delta<\frac{1}{2}$. On the event $\{p(Z)\le q+\delta\}$, we have $(a/\sqrt{m})Z+\Phi^{-1}(p)<0$ from the expression for $p(Z)$ in (15). Then for arbitrarily small $\epsilon>0$ and $m>a^2/\epsilon$,

$$p(Z)\ge\Phi\left(\frac{(a/\sqrt{m})Z+\Phi^{-1}(p)}{\sqrt{1-\epsilon}}\right).$$

Define $z_\delta=(\Phi^{-1}(q+\delta)-\Phi^{-1}(p))/a$, combining the fact that $\mathbf{1}\{p(Z)\le q+\delta\}\ge\mathbf{1}\{(Z/\sqrt{m})\le z_\delta\}$ with (38)–(40),

$$\begin{aligned} &P(L>mq)\\ &\ge\frac{1}{4}\tilde{\mathsf{E}}_\delta\left[\exp\left(mG_\delta\left(\Phi\left(\frac{(a/\sqrt{m})Z+\Phi^{-1}(p)}{\sqrt{1-\epsilon}}\right)\right)\right)\mathbf{1}\left\{\frac{Z}{\sqrt{m}}\le z_\delta\right\}\right]\\ &=\frac{1}{4}\tilde{\mathsf{E}}_\delta\left[\exp\left(mF_{\delta,\epsilon}\left(a,\frac{Z}{\sqrt{m}}\right)\right)\mathbf{1}\left\{\frac{Z}{\sqrt{m}}\le z_\delta\right\}\right]. \end{aligned} \quad (41)$$

Here

$$F_{\delta,\epsilon}(a,z)=G_\delta\left(\Phi\left(\frac{az+\Phi^{-1}(p)}{\sqrt{1-\epsilon}}\right)\right).$$

Applying an extension of Varadhan's Lemma (as on p. 140 of Dembo and Zeitouni 1998), we get

$$\begin{aligned}\liminf_{m\to\infty} m^{-1}\log P(L>mq)&\\ \geq \liminf_{m\to\infty} m^{-1}\log\frac{1}{4}\widetilde{\mathsf{E}}_\delta\left[\exp\left(mF_{\delta,\epsilon}\left(a,\frac{Z}{\sqrt{m}}\right)\right)\mathbf{1}\left\{\frac{Z}{\sqrt{m}}\leq z_\delta\right\}\right]&\\ \geq \sup_{z\leq z_\delta}\left\{F_{\delta,\epsilon}(a,z)-\frac{1}{2}z^2\right\}=\max_{z\leq z_\delta}\left\{F_{\delta,\epsilon}(a,z)-\frac{1}{2}z^2\right\}.&\quad (42)\end{aligned}$$

The equality is obtained from the continuity of the function $F_{\delta,\epsilon}(a,z)$.

Define $F_\delta(a,z)=G_\delta(\Phi(az+\Phi^{-1}(p)))$ and let μ_δ be the solution to the optimization problem $\max_{z\leq z_\delta}\{F_\delta(a,z)-\frac{1}{2}z^2\}$. As ϵ decreases to 0, the function $F_{\delta,\epsilon}(a,\mu_\delta)-\frac{1}{2}\mu_\delta^2$ increases to $F_\delta(a,\mu_\delta)-\frac{1}{2}\mu_\delta^2$. With (42), this gives

$$\begin{aligned}\liminf_{m\to\infty} m^{-1}\log P(L>mq)&\geq \max_{z\leq z_\delta}\left\{F_\delta(a,z)-\tfrac{1}{2}z^2\right\}\\ &=\max_z\left\{F_\delta(a,z)-\tfrac{1}{2}z^2\right\}.\quad (43)\end{aligned}$$

The equality comes from the fact that z_δ maximizes $F_\delta(a,z)$ and $z_\delta>0$ because $q>p$ and δ small enough.

Similarly, define μ to be the solution to the optimization problem $\max_z\{F(a,z)-\frac{1}{2}z^2\}$. As δ decreases to 0, $F_\delta(a,\mu)-\frac{1}{2}\mu^2$ increases to $F(a,\mu)-\frac{1}{2}\mu^2$. Combining with (43), we obtain the lim inf in (37) for $q<\frac{1}{2}$.

If $q\geq\frac{1}{2}$, for $m>a^2/\epsilon$,

$$\left(\Phi^{-1}(q+\delta)\sqrt{1-a^2/m}-\Phi^{-1}(p)\right)/a>z_{\delta,\epsilon}.$$

Instead of (41), we get

$$\begin{aligned}P(L>mq)&\\ \geq\frac{1}{4}\widetilde{\mathsf{E}}_\delta\Bigg[\exp\left(mG_\delta\left(\Phi\left(\frac{(a/\sqrt{m})Z+\Phi^{-1}(p)}{\sqrt{1-a^2/m}}\right)\right)\right)&\\ \cdot\mathbf{1}\left\{-\frac{\Phi^{-1}(p)}{a}\leq\frac{Z}{\sqrt{m}}\leq z_{\delta,\epsilon}\right\}\Bigg]&\\ \geq\frac{1}{4}\widetilde{\mathsf{E}}_\delta\left[\exp\left(mF_\delta\left(a,\frac{z}{\sqrt{m}}\right)\right)\mathbf{1}\left\{-\frac{\Phi^{-1}(p)}{a}\leq\frac{Z}{\sqrt{m}}\leq z_{\delta,\epsilon}\right\}\right].&\end{aligned}$$

Applying the extension of Varadhan's Lemma, we get

$$\liminf_{m\to\infty} m^{-1}\log P(L>mq)\geq \max_{-\Phi^{-1}(p)/a\leq z\leq z_{\delta,\epsilon}}\left\{F_\delta(a,z)-\tfrac{1}{2}z^2\right\}.$$

Because δ and ϵ are arbitrary, the argument used for $q<\frac{1}{2}$ now shows that

$$\liminf_{m\to\infty} m^{-1}\log P(L>mq)\geq \max_{-\Phi^{-1}(p)/a\leq z\leq z_\delta}\left\{F(a,z)-\tfrac{1}{2}z^2\right\}.\quad (44)$$

Also, for $m>a^2/\epsilon$ and $Z\leq-\sqrt{m}\Phi^{-1}(p)/a$, we have

$$p(Z)\geq\Phi\left(\frac{(a/\sqrt{m})Z+\Phi^{-1}(p)}{\sqrt{1-\epsilon}}\right)$$

for $p(Z)$ in (15). Because $G_\delta(\cdot)$ is an increasing function,

$$\begin{aligned}P(L>mq)&\geq\frac{1}{4}\widetilde{\mathsf{E}}_\delta\Bigg[\exp\left(mG_\delta\left(\Phi\left(\frac{(a/\sqrt{m})Z+\Phi^{-1}(p)}{\sqrt{1-\epsilon}}\right)\right)\right)\\ &\qquad\cdot\mathbf{1}\left\{\frac{Z}{\sqrt{m}}\leq-\frac{\Phi^{-1}(p)}{a}\right\}\Bigg]\\ &=\frac{1}{4}\widetilde{\mathsf{E}}_\delta\left[\exp\left(mF_{\delta,\epsilon}\left(a,\frac{z}{\sqrt{m}}\right)\right)\mathbf{1}\left\{\frac{Z}{\sqrt{m}}\leq-\frac{\Phi^{-1}(p)}{a}\right\}\right].\end{aligned}$$

Again applying the extension of Varadhan's Lemma, we get

$$\liminf_{m\to\infty} m^{-1}\log P(L>mq)\geq \max_{z\leq-\Phi^{-1}(p)/a}\left\{F_{\delta,\epsilon}(a,z)-\tfrac{1}{2}z^2\right\}.\quad (45)$$

Because ϵ is arbitrary, (45) holds with $F_{\delta,\epsilon}$ replaced by F_δ. Combining this with (44) and using the fact that δ is arbitrary, we obtain the lim inf in (37) for $q\geq\frac{1}{2}$.

(a) Define $a_m=a/m^{\alpha-1/2}$, then $\rho=a_m/\sqrt{m}$. First we show that

$$\limsup_{m\to\infty} m^{-1}\log M_2(mq,\theta_{mq})\leq 2F(0,0).\quad (46)$$

By following the same steps as in (34), (35), and (36), we get

$$M_2(mq,\theta_{mq})\leq\exp\left(m(1-a^2/m^{2\alpha})\max_z\left\{2F(a_m,z)-\tfrac{1}{2}z^2\right\}\right).\quad (47)$$

For arbitrarily small $\zeta>0$, we can find some $m_1=(\alpha/\zeta)^{2/(2\alpha-1)}$ so that $a_m\leq\zeta$ for $m\geq m_1$. Define μ_{a_m} to be the solution to the optimization problem $\max_z\{2F(a_m,z)-\frac{1}{2}z^2\}$ and μ_ζ to be the solution at $a_m=\zeta$. It is obvious that $\mu_{a_m}>0$ and $\mu_\zeta>0$ because for any $a>0$, the function $F(a,z)-\frac{1}{2}z^2$ is increasing in $z\leq 0$. We have $F(a_m,z)\leq F(\zeta,z)$ for any $z>0$ and $m>m_1$ since the function $F(a,z)$ is also increasing in a for any $z>0$. Then, with $\mu_{a_m}>0$ and $\mu_\zeta>0$,

$$\max_z\left\{2F(a_m,z)-\tfrac{1}{2}z^2\right\}\leq\max_z\left\{2F(\zeta,z)-\tfrac{1}{2}z^2\right\}.\quad (48)$$

Combining (47) and (48), we get

$$\limsup_{m\to\infty} m^{-1}\log M_2(mq,\theta_{mq})\leq\max_z\left\{2F(\zeta,z)-\tfrac{1}{2}z^2\right\}.\quad (49)$$

$\zeta>0$ is arbitrary, so (49) holds at $\zeta=0$. The function $2F(0,z)-\frac{1}{2}z^2$ has maximum $2F(0,0)$.

Next we show that

$$\liminf_{m\to\infty} m^{-1}\log P(L>mq)\geq F(0,0).\quad (50)$$

We also need to separate the cases $q<\frac{1}{2}$ and $q\geq\frac{1}{2}$ as in (b). We give out only the proof for $q<\frac{1}{2}$. Similar proof follows for $q\geq\frac{1}{2}$. The argument used to show (41) now yields

$$\begin{aligned}P(L>mq)\geq\frac{1}{4}\widetilde{\mathsf{E}}_\delta\Bigg[\exp\left(mF_{\delta,\epsilon}\left(a_m,\frac{Z}{\sqrt{m}}\right)\right)&\\ \cdot\mathbf{1}\left\{\frac{Z}{\sqrt{m}}\leq\frac{\Phi^{-1}(q+\delta)-\Phi^{-1}(p)}{a_m}\right\}\Bigg].&\quad (51)\end{aligned}$$

For $m>m_1$, we have $a_m\leq\zeta$, so

$$\begin{aligned}\frac{1}{4}\widetilde{\mathsf{E}}_\delta\left[\exp\left(mF_{\delta,\epsilon}\left(a_m,\frac{Z}{\sqrt{m}}\right)\right)\mathbf{1}\left\{\frac{Z}{\sqrt{m}}\leq\frac{\Phi^{-1}(q+\delta)-\Phi^{-1}(p)}{a_m}\right\}\right]&\\ \geq\frac{1}{4}\widetilde{\mathsf{E}}_\delta\Bigg[\exp\left(mF_{\delta,\epsilon}\left(a_m,\frac{Z}{\sqrt{m}}\right)\right)&\\ \cdot\mathbf{1}\left\{\frac{Z}{\sqrt{m}}\leq\frac{\Phi^{-1}(q+\delta)-\Phi^{-1}(p)}{\zeta}\right\}\Bigg]&\\ \geq\frac{1}{4}\widetilde{\mathsf{E}}_\delta\Bigg[\exp\left(mF_{\delta,\epsilon}\left(0,\frac{Z}{\sqrt{m}}\right)\right)&\\ \cdot\mathbf{1}\left\{0\leq\frac{Z}{\sqrt{m}}\leq\frac{\Phi^{-1}(q+\delta)-\Phi^{-1}(p)}{\zeta}\right\}\Bigg].&\quad (52)\end{aligned}$$

The second inequality comes from the fact that for any $z>0$, the function $F_{\delta,\epsilon}(a,z)$ is increasing in a. Combining (51) and (52), and applying the extension of Varadhan's Lemma, we get

$$\liminf_{m\to\infty} m^{-1}\log P(L>mq) \geq \max_{0\le z\le[\Phi^{-1}(q+\delta)-\Phi^{-1}(p)]/\zeta}\left\{F_{\delta,\epsilon}(0,z)-\tfrac{1}{2}z^2\right\}. \tag{53}$$

Because ϵ and δ are arbitrary, the argument used in (b) now shows that

$$\liminf_{m\to\infty} m^{-1}\log P(L>mq) \geq \max_{0\le z\le[\Phi^{-1}(q)-\Phi^{-1}(p)]/\zeta}\left\{F(0,z)-\tfrac{1}{2}z^2\right\} = F(0,0). \tag{54}$$

(c) Recall $a_m = a/m^{\alpha-1/2}$, $\rho = a_m/\sqrt{m}$. First we show that

$$\limsup_{m\to\infty} m^{-2\alpha}\log M_2(mq,\theta_{mq}) \le -\tfrac{1}{2}z_a^2. \tag{55}$$

We use the upper bound on $M_2(mq,\theta_{mq})$ in (47), which continues to apply. Define

$$z_{a_m} = \frac{\Phi^{-1}(q)-\Phi^{-1}(p)}{a_m},$$

then $\lim_{m\to\infty} z_{a_m}=0$. The function $F(a_m,\cdot)$ is concave (as can be verified by differentiation) and takes its maximum value 0 for $z\ge z_{a_m}$. Let μ_{a_m} be the point at which $2F(a_m,z)-\frac{1}{2}z^2$ is maximized. We will show that $\mu_{a_m}/z_{a_m}\to 1$; i.e., we will show that for all small $\xi>0$, we can find m_1 large enough that $\mu_{a_m}\in(z_{a_m}(1-\xi),z_{a_m})$ for all $m>m_1$. For this, it suffices to show that

$$2F'(a_m,z_{a_m}(1-\xi))-z_{a_m}(1-\xi)>0 \quad\text{and}\quad 2F'(a_m,z_{a_m})-z_{a_m}<0,$$

where F' denotes the derivative of F with respect to its second argument. The second inequality holds because $F'(a_m,z_{a_m})=0$ and $z_{a_m}>0$ when $q>p$. For the first inequality, the mean value theorem yields,

$$F'(a_m,z_{a_m}(1-\xi)) = F'(a_m,z_{a_m}) - F''(a_m,z_{a_m}^{\eta})z_{a_m}\xi, \tag{56}$$

with $z_{a_m}^{\eta}$ a point in $[z_{a_m}(1-\xi),z_{a_m}]$. For $z<z_{a_m}$, differentiation yields

$$F''(a_m,z) = -a_m^2 H(a_m,z), \tag{57}$$

with

$$H(a_m,z) = \frac{\varphi(a_m z+\Phi^{-1}(p))}{(1-\Phi(a_m z+\Phi^{-1}(p)))\Phi(a_m z+\Phi^{-1}(p))}H_1(a_m,z)$$

and

$$H_1(a_m,z) = \frac{\varphi(a_m z+\Phi^{-1}(p))((\Phi(a_m z+\Phi^{-1}(p))-q)^2+q(1-q))}{(1-\Phi(a_m z+\Phi^{-1}(p)))\Phi(a_m z+\Phi^{-1}(p))} - (a_m z+\Phi^{-1}(p))(\Phi(a_m z+\Phi^{-1}(p))-q).$$

Here $\varphi(\cdot)$ is the probability density function of the standard normal distribution. If $a_m z+\Phi^{-1}(p)\le 0$, then

$$\begin{aligned} H_1(a_m,z) &\geq \frac{(-\Phi^{-1}(p)-a_m z)((\Phi(a_m z+\Phi^{-1}(p))-q)^2+q(1-q))}{1-\Phi(a_m z+\Phi^{-1}(p))} \\ &\quad + (-\Phi^{-1}(p)-a_m z)(\Phi(a_m z+\Phi^{-1}(p))-q) \\ &= \frac{\Phi(a_m z+\Phi^{-1}(p))(-\Phi^{-1}(p)-a_m z)(1-q)}{1-\Phi(a_m z+\Phi^{-1}(p))} > 0. \end{aligned}$$

The inequality comes from the fact that $\varphi(x)/x\ge\Phi(-x)$ for $x>0$ (Feller 1968, p. 175).

Similarly, if $a_m z+\Phi^{-1}(p)>0$, then $H_1(a_m,z)>0$. Also, $H(a_m,z)>0$ because $H_1(a_m,z)>0$. Simple algebra also shows that $H(a_m,z)$ is bounded for $z\in[z_{a_m}(1-\xi),z_{a_m}]$. Thus, (57) implies that $F''(a_m,z_{a_m}^{\eta})\le -a_m^2 C$ for some constant $C>0$. We can now choose $m_1=((1-\xi)/(a^2C\xi))^{1/(1-2\alpha)}$, so that $F''(a_m,z_{a_m}^{\eta})<(\xi-1)/\xi$ for $m>m_1$. Combining this with (56), we have

$$\begin{aligned} &2F'(a_m,z_{a_m}(1-\xi))-z_{a_m}(1-\xi) \\ &\quad = 2F'(a_m,z_{a_m})-F''(a_m,z_{a_m}^{\eta})z_{a_m}\xi - z_{a_m}(1-\xi)>0. \end{aligned}$$

Therefore we have shown that for small $\xi>0$, we can find m_1 large enough that for $m>m_1$, we have $\mu_{a_m}\in(z_{a_m}(1-\xi),z_{a_m})$, i.e., $\mu_{a_m}/z_{a_m}\to 1$.

The definitions of μ_{a_m} and z_{a_m} give

$$-\tfrac{1}{2}z_{a_m}^2 = 2F(a_m,z_{a_m})-\tfrac{1}{2}z_{a_m}^2 \le 2F(a_m,\mu_{a_m})-\tfrac{1}{2}\mu_{a_m}^2 \le -\tfrac{1}{2}\mu_{a_m}^2$$

so we also have $\max_z\{2F(a_m,z)-\frac{1}{2}z^2\}/(-\frac{1}{2}z_{a_m}^2)\to 1$, and then, since $a_m=a/m^{\alpha-1/2}$ and $z_a=z_m/m^{\alpha-1/2}$,

$$m^{1-2\alpha}\max_x\left\{2F(a_m,z)-\tfrac{1}{2}z^2\right\}\to -\tfrac{1}{2}z_a^2. \tag{58}$$

Applying this to (47) proves (55).

Next we show that

$$\liminf_{m\to\infty} m^{-2\alpha}\log P(L>mq)\ge -\tfrac{1}{2}z_a^2. \tag{59}$$

Similarly we also need to separate the cases $q<\frac{1}{2}$ and $q\ge\frac{1}{2}$ as in (b). We give out only the proof for $q<\frac{1}{2}$. Similar proof follows for $q\ge\frac{1}{2}$. The argument leading to (41) gives

$$P(L>mq)\ge \frac{1}{4}\tilde{\mathsf{E}}_\delta\left[\exp\left(mF_{\delta,\epsilon}\left(a,\frac{Z}{m^\alpha}\right)\right)\mathbf{1}\left\{\frac{Z}{m^\alpha}\le z_\delta\right\}\right]. \tag{60}$$

Recall that $F_{\delta,\epsilon}(a,z_{\delta,\epsilon})=0$, $F'_{\delta,\epsilon}(a,z_{\delta,\epsilon})=0$, and arguing as in the proof of the upper bound, $\frac{1}{2}F''_{\delta,\epsilon}(a,z_{\delta,\epsilon})=-c<0$ for c some positive constant. Thus we have

$$\lim_{\eta\to 0}\frac{F_{\delta,\epsilon}(a,z_{\delta,\epsilon}-\eta)}{\eta^2}=-c<0,$$

and for sufficiently small η, $F_{\delta,\epsilon}(a,z_{\delta,\epsilon}-\eta)>-(c+\epsilon)\eta^2$. So if we choose $\eta=\sqrt{\epsilon/(c+\epsilon)}m^{\alpha-1/2}$, then for m sufficiently large, $F_{\delta,\epsilon}(a,z_{\delta,\epsilon}-\eta)>-\epsilon m^{2\alpha-1}$. Therefore we have

$$\begin{aligned} &\frac{1}{4}\tilde{\mathsf{E}}_\delta\left[\exp\left(mF_{\delta,\epsilon}\left(a,\frac{Z}{m^\alpha}\right)\right)\mathbf{1}\left\{\frac{Z}{m^\alpha}\le z_\delta\right\}\right] \\ &\quad\ge \frac{1}{4}\exp(-\epsilon m^{2\alpha})P\left(z_\delta-\eta_m<\frac{Z}{m^\alpha}\le z_\delta\right) \\ &\quad\ge \frac{1}{4}\exp(-\epsilon m^{2\alpha})\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z_\delta^2 m^{2\alpha}\right)\eta_m m^\alpha \\ &\quad= \frac{1}{4}\exp(-\epsilon m^{2\alpha})\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z_\delta^2 m^{2\alpha}\right)\sqrt{\frac{\epsilon}{c+\epsilon}}m^{2\alpha-1/2}. \end{aligned} \tag{61}$$

The second inequality comes from the fact that $P(x-\delta\le Z<x)\ge\varphi(x)\delta$ if $x>\delta$.

Combining (60) and (61), we have

$$\begin{aligned} &\liminf_{m\to\infty} m^{-2\alpha}\log P(L>mq) \\ &\quad\ge -\epsilon-\frac{1}{2}z_\delta^2+\liminf_{m\to\infty} m^{-2\alpha}\log\left(\sqrt{\frac{\epsilon}{c+\epsilon}}m^{2\alpha-1/2}\right) \\ &\quad= -\epsilon-\frac{1}{2}z_\delta^2, \end{aligned}$$

and because ϵ and δ are arbitrary, the lim inf in (59) follows.

Proof of Theorem 3. First we show that

$$\limsup_{m\to\infty} \frac{1}{\log m} \log M_2(x_m, \mu_m) \le -2\gamma. \tag{62}$$

Let $\bar{\mathsf{E}}$ denote expectation under the IS distribution. Then

$$\begin{aligned} M_2(x_m, \mu_m) &= \bar{\mathsf{E}}\big[\mathbf{1}\{L_m > x_m\}\exp(-2\theta_{x_m}(Z)L + 2\psi_m(\theta_{x_m}(Z), Z)) \\ &\qquad \cdot \exp(-2\mu_m Z + \mu_m^2)\big] \\ &\le \bar{\mathsf{E}}\big[\exp(-2\theta_{x_m}(Z)x_m + 2\psi_m(\theta_{x_m}(Z), Z)) \\ &\qquad \cdot \exp(-2\mu_m Z + \mu_m^2)\big] \\ &= \bar{\mathsf{E}}\big[\exp(2F_m(Z))\exp(-2\mu_m Z + \mu_m^2)\big], \end{aligned} \tag{63}$$

with $F_m(z)$ as in (19) but with q replaced by q_m.

For any z, the concavity of $F_m(z)$ implies

$$F_m(z) \le F_m(\mu_m) + F_m'(\mu_m)(z - \mu_m) = F_m(\mu_m) + \mu_m z - \mu_m^2 \tag{64}$$

The equality comes from the fact that $F_m'(\mu_m) = \mu_m$ because μ_m is the unique maximizer in (23) with $x = x_m$. Substituting (64) into the upper bound in (63), we have

$$\begin{aligned} &\bar{\mathsf{E}}\big[\exp(2F_m(Z))\exp(-2\mu_m Z + \mu_m^2)\big] \\ &\quad \le \bar{\mathsf{E}}\big[\exp(2F_m(\mu_m) + 2\mu_m Z - 2\mu_m^2 - 2\mu_m Z + \mu_m^2)\big] \\ &\quad = \exp(2F_m(\mu_m) - \mu_m^2). \end{aligned} \tag{65}$$

Define

$$z_m = \frac{\sqrt{1-\rho^2}\,c\sqrt{\log m} - \Phi^{-1}(p)}{\rho}.$$

We have $p(z_m) = q_m$ and $F_m(z)$ takes its maximum value 0 for all $z \ge z_m$. We show that for all small $\xi > 0$, we can find m_1 large enough that $\mu_m \in (z_m(1-\xi), z_m)$ for all $m > m_1$. It suffices to show that

$$F_m'(z_m(1-\xi)) - z_m(1-\xi) > 0 \quad \text{and} \quad F_m'(z_m) - z_m < 0. \tag{66}$$

The second inequality holds because $F_m'(z_m) = 0$ and $z_m > 0$ when $p < q_m$.

For the first inequality in (66), we have

$$\begin{aligned} F_m'(z_m(1-\xi)) &= m\left(\frac{q_m}{p(z_m(1-\xi))} - \frac{1-q_m}{1-p(z_m(1-\xi))}\right) \\ &\qquad \cdot \varphi\left(\frac{\rho z_m(1-\xi) + \Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right)\frac{\rho}{\sqrt{1-\rho^2}}. \end{aligned}$$

Using the property that $1 - \Phi(x) \sim \varphi(x)/x$ as $x \to \infty$ (Feller 1968, p. 175), we conclude that

$$\frac{q_m}{p(z_m(1-\xi))} = O(1), \qquad \frac{1-q_m}{1-p(z_m(1-\xi))} = o(1),$$

and

$$\varphi\left(\frac{\rho z_m(1-\xi) + \Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right) = O\big(m^{-(c^2/2)(1-\xi)^2}\big).$$

Therefore we have

$$F_m'(z_m(1-\xi)) = O\big(m^{1-(c^2/2)(1-\xi)^2}\big),$$

and as $0 < c < \sqrt{2}$,

$$z_m = O\big(c\sqrt{\log m}\big) = o\big(m^{1-(c^2/2)(1-\xi)^2}\big).$$

Then for m large enough, we obtain (66).

We have shown that for all $\xi > 0$, we have $\mu_m / z_m \in (1-\xi, 1)$ for sufficiently large m. It follows that for all $\upsilon_\xi > 0$ and all sufficiently large m we have $-\frac{1}{2}z_m^2 < F_m(\mu_m) - \frac{1}{2}\mu_m^2 < -\frac{1}{2}z_m^2(1-\upsilon_\xi)$. Combining this with (63) and (65), we get $M_2(x_m, \mu_m) < \exp(-z_m^2(1-\upsilon_\xi))$, and as υ_ξ is arbitrary the lim sup in (62) follows.

Next we show that

$$\liminf_{m\to\infty} \frac{1}{\log m} \log P(L_m > x_m) \ge -\gamma. \tag{67}$$

For any $\delta_m > 0$,

$$P(L_m > x_m) \ge P(L_m > x_m \mid p(Z) = q_m + \delta_m) P(p(Z) \ge q_m + \delta_m). \tag{68}$$

Given $p(Z) = q_m + \delta_m > q_m$, from the definition of our change of measure, L_m is binomially distributed with parameters m and $q_m + \delta_m$. Applying the bound (3.62) of Johnson et al. (1993), we have

$$\begin{aligned} &P(L_m > x_m \mid p(Z) = q_m + \delta_m) \\ &\qquad \ge 1 - \Phi\left(-\sqrt{\frac{m}{(q_m+\delta_m)(1-q_m-\delta_m)}}\,\delta_m\right) \ge \frac{1}{2}. \end{aligned}$$

It therefore suffices to consider the second factor in (68). If we take

$$\delta_m = \Phi\big(c\sqrt{\log m} + \epsilon\big) - \Phi\big(c\sqrt{\log m}\big),$$

for some $\epsilon > 0$, then $\Phi^{-1}(q_m + \delta_m) = c\sqrt{\log m} + \epsilon$. Thus,

$$\begin{aligned} &P(p(Z) \ge q_m + \delta_m) \\ &\quad = P\left(Z \ge \frac{\sqrt{1-\rho^2}\,\Phi^{-1}(q_m+\delta_m) - \Phi^{-1}(p)}{\rho}\right) \\ &\quad \ge P\left(\frac{\sqrt{1-\rho^2}\,\Phi^{-1}(q_m+\delta_m) - \Phi^{-1}(p)}{\rho} \le Z \right. \\ &\qquad\qquad \left. \le \frac{\sqrt{1-\rho^2}\,\Phi^{-1}(q_m+\delta_m)}{\rho}\right) \\ &\quad \ge \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(1-\rho^2)(\Phi^{-1}(q_m+\delta_m))^2}{\rho^2}\right)\frac{\Phi^{-1}(p)}{\rho} \\ &\quad = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(1-\rho^2)(c\sqrt{\log m}+\epsilon)^2}{\rho^2}\right)\frac{\Phi^{-1}(p)}{\rho} \end{aligned} \tag{69}$$

and

$$\begin{aligned} &\liminf_{m\to\infty} \frac{1}{\log m} \log P(p(Z) \ge q_m + \delta_m) \\ &\qquad \ge -\frac{1-\rho^2}{2\rho^2} \liminf_{m\to\infty} \frac{(c\sqrt{\log m}+\epsilon)^2}{\log m} = -\gamma, \end{aligned}$$

which proves (67).

Proof of Theorem 4. The proof is similar to that of Theorem 3. First we show that

$$\limsup_{m\to\infty} \frac{1}{m} \log M_2(x_m, \mu_m) \le -2\gamma. \tag{70}$$

For this, set $F_m(z) = G(p^{(m)}(z))$ where

$$p^{(m)}(z) = \Phi\left(\frac{\rho z + \Phi^{-1}(p^{(m)})}{\sqrt{1-\rho^2}}\right).$$

Define

$$z_m = \frac{\sqrt{1-\rho^2}\,\Phi^{-1}(q) + c\sqrt{m}}{\rho}.$$

Here $p(z_m) = q$ and $F_m(z)$ takes its maximum value 0 for all $z \geq z_m$. We show that for small $\xi > 0$, we can find m_1 large enough that $\mu_m \in (z_m(1-\xi), z_m)$ for all $m > m_1$. It suffices to show (66). The second inequality in (66) holds because $F_m'(z_m) = 0$ and $z_m > 0$ when $p^{(m)} < q$. For the first inequality, we have

$$F_m'(z_m(1-\xi)) = m\left(\frac{q}{p^{(m)}(z_m(1-\xi))} - \frac{1-q}{1-p^{(m)}(z_m(1-\xi))}\right) \cdot \varphi\left(\frac{\rho z_m(1-\xi) - c\sqrt{m}}{\sqrt{1-\rho^2}}\right)\frac{\rho}{\sqrt{1-\rho^2}}. \tag{71}$$

Since

$$\frac{q}{p^{(m)}(z_m(1-\xi))} = O\left(1\bigg/\Phi\left(\frac{\rho z_m(1-\xi) - c\sqrt{m}}{\sqrt{1-\rho^2}}\right)\right) \quad \text{and} \quad \frac{1-q}{1-p^{(m)}(z_m(1-\xi))} = O(1),$$

by applying the property that $\varphi(x)/\Phi(-x) \sim x$ as $x \to \infty$ to (71), we conclude that

$$F_m'(z_m(1-\xi)) = O\left(\frac{\xi c m^{3/2}}{\sqrt{1-\rho^2}}\right),$$

whereas $z_m = O(c\sqrt{m})$. We thus obtain (66) for all sufficiently large m.

It follows that for $m > m_1$, $\mu_m/z_m \in (1-\xi, 1)$, and correspondingly $-\frac{1}{2}z_m^2 < F_m(\mu_m) - \frac{1}{2}\mu_m^2 < -\frac{1}{2}z_m^2(1-\xi)^2$. Combining this with (63) and (65), we have $M_2(x_m, \mu_m) < \exp(-z_m^2(1-\xi)^2)$. Since ξ is arbitrary, we obtain the lim sup in (70).

Next we show that

$$\liminf_{m\to\infty} \frac{1}{m} \log P(L_m > x_m) \geq -\gamma. \tag{72}$$

For arbitrary $\delta > 0$,

$$P(L_m > x_m) \geq P(L_m > x_m \mid p^{(m)}(Z) = q+\delta) P(p^{(m)}(Z) \geq q+\delta). \tag{73}$$

Given $p^{(m)}(Z) = q+\delta > q$, from the definition of our change of measure, L_m is binomial distributed with parameter m and $q+\delta$. Applying the bound (3.62) of Johnson et al. (1993), we have

$$P(L_m > x_m \mid p^{(m)}(Z) = q+\delta) \geq 1 - \Phi\left(-\sqrt{\frac{m}{(q+\delta)(1-q-\delta)}}\delta\right) \geq \frac{1}{2}. \tag{74}$$

Also notice that for any constant $c_0 > 0$,

$$\begin{aligned} &P(p^{(m)}(Z) \geq q+\delta) \\ &= P\left(Z \geq \frac{\sqrt{1-\rho^2}\Phi^{-1}(q+\delta) - \Phi^{-1}(p^{(m)})}{\rho}\right) \\ &\geq P\left(\frac{\sqrt{1-\rho^2}\Phi^{-1}(q+\delta) - \Phi^{-1}(p^{(m)})}{\rho} \leq Z \leq \frac{\sqrt{1-\rho^2}\Phi^{-1}(q+\delta) - \Phi^{-1}(p^{(m)})}{\rho} + c_0\right) \\ &\geq \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\sqrt{1-\rho^2}\Phi^{-1}(q+\delta) - \Phi^{-1}(p^{(m)})}{\rho} + c_0\right)^2\right)c_0 \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\sqrt{1-\rho^2}\Phi^{-1}(q+\delta) + c\sqrt{m}}{\rho} + c_0\right)^2\right)c_0. \end{aligned} \tag{75}$$

Combining (73), (74), and (75), we obtain the lim inf in (72). By Jensen's inequality, $M_2(x_m, \mu_m) \geq P(L_m > x_m)^2$, so the lim sup in (70) and the lim inf in (72) hold as limits.

References

Avramidis, A. 2001. Importance sampling for multimodal functions and application to pricing exotic options. *Proc. Winter Simulation Conf.*, IEEE Press, Piscataway, NJ, 1493–1501.

Avranitis, A., J. Gregory. 2001. *Credit: The Complete Guide to Pricing, Hedging and Risk Management*. Risk Books, London, UK.

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

Feller, W. 1968. *An Introduction to Probability Theory and Its Applications*, 3rd ed. John Wiley and Sons, New York.

Finger, C. 2001. Enhancing Monte Carlo for credit portfolio models. Presentation at The International Centre for Business Information Global Derivatives Conference, Juan-les-Pins, France.

Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*. Springer, New York.

Glasserman, P., J. Li. 2003. Importance sampling for a mixed Poisson model of portfolio credit risk. *Proc. Winter Simulation Conf.*, IEEE Press, Piscataway, NJ.

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

Gupton, G., C. Finger, M. Bhatia. 1997. CreditMetrics technical document. J. P. Morgan & Co., New York.

Heidelberger, P. 1995. Fast simulation of rare events in queueing and reliability models. *ACM Trans. Model. Comput. Simulation* **5** 43–85.

Jensen, J. L. 1995. *Saddlepoint Approximations*. Oxford University Press, Oxford, UK.

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

Kalkbrener, M., H. Lotter, L. Overbeck. 2003. Sensible and efficient capital allocation for credit portfolios. Deutsche Bank, AG.

Li, D. 2000. On default correlation: A copula function approach. *J. Fixed Income* **9** 43–54.

Martin, R., K. Thompson, C. Browne. 2001. Taking to the saddle. *Risk* **14**(6) 91–94.

Merton, R. C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *J. Finance* **29** 449–470.

Sadowsky, J. S., J. A. Bucklew. 1990. On large deviations theory and asymptotically efficient Monte Carlo estimation. *IEEE Trans. Inform. Theory* **36** 579–588.

Wilde, T. 1997. *CreditRisk⁺: A CreditRisk Management Framework*. Credit Suisse First Boston, London, UK.