OPERATIONS RESEARCH
Vol. 56, No. 5, September–October 2008, pp. 1200–1217
ISSN 0030-364X | EISSN 1526-5463 | 08 | 5605 | 1200

DOI 10.1287/opre.1080.0558

Fast Simulation of Multifactor Portfolio Credit Risk

Paul Glasserman
Graduate School of Business, Columbia University, New York, New York 10027,
pg20@columbia.edu

Wanmo Kang
Department of Mathematical Sciences, Korea Advanced Institute of Science and Technology (KAIST),
Daejeon 305-701, Republic of Korea, wanmo.kang@kaist.ac.kr

Perwez Shahabuddin
Deceased
Formerly at Department of Industrial Engineering and Operations Research, Columbia University, New York

This paper develops rare-event simulation methods for the estimation of portfolio credit risk—the risk of losses to a portfolio resulting from defaults of assets in the portfolio. Portfolio credit risk is measured through probabilities of large losses, which are typically due to defaults of many obligors (sources of credit risk) to which a portfolio is exposed. An essential element of a portfolio view of credit risk is a model of dependence between these sources of credit risk: large losses occur rarely and are most likely to result from systematic risk factors that affect multiple obligors. As a consequence, estimating portfolio credit risk poses a challenge both because of the rare-event property of large losses and the dependence between defaults. To address this problem, we develop an importance sampling technique within the widely used Gaussian copula model of dependence. We focus on difficulties arising in multifactor models—that is, models in which multiple factors may be common to multiple obligors, resulting in complex dependence between defaults. Our importance sampling procedure shifts the mean of the common factor to increase the frequency of large losses. In multifactor models, different combinations of factor outcomes and defaults can produce large losses, so our method combines multiple importance sampling distributions, each associated with a shift in the mean of common factors. We characterize "optimal" mean shifts. Finding these points is both a combinatorial problem and a convex optimization problem, so we address computational aspects of this step as well. We establish asymptotic optimality results for our method, showing that—unlike standard simulation—it remains efficient as the event of interest becomes rarer.

Subject classifications: simulation: efficiency, variance reduction; probability: large deviations; finance: credit risk, portfolio.
Area of review: Simulation.
History: Received August 2005; revisions received January 2007, May 2007; accepted June 2007.

1. Introduction

The risk in a portfolio depends not only on the risk in each element of the portfolio, but also on the dependence between these sources of risk. The types of dependence mechanisms used in credit-risk modeling often lead to models for which Monte Carlo methods are the only practical computational tool. However, estimating risk usually involves estimating small probabilities of rare but significant events, and crude Monte Carlo methods are generally inefficient in estimating such probabilities. Efficient simulation then requires variance reduction techniques specifically designed for rare-event simulation.

Importance sampling (IS) is commonly used to increase efficiency in rare-event simulation. It involves changing the probability distributions used to simulate a model to increase the frequency with which the event of interest is observed. To correct for the change in distribution, each replication is weighted by a likelihood ratio. The critical step in developing an IS method is finding an effective change of distribution. Indeed, a poor choice of distribution—even one that increases the probability of the rare event—may increase variance and may even produce infinite variance. Moreover, the poor performance of an IS method may be hard to detect from a small number of samples.

We address rare-event simulation problems within the Gaussian copula model of portfolio credit risk. This model originated with J. P. Morgan's CreditMetrics system (Gupton et al. 1997) and is now widely used in practice. In this model, the default of an obligor is triggered when an associated latent variable exceeds a prespecified threshold. The dependence across obligors is captured by the correlations between the latent variables, which have a multivariate normal distribution.

Applications of IS in this setting include Avranitis and Gregory (2001), Kalkbrener et al. (2004), Merino and Nyfeler (2004), and Morokoff (2004), but these are largely heuristic in the sense that they provide no theoretical support. Glasserman and Li (2005) develop and analyze a two-step IS method for this problem that utilizes the conditional

independence of defaults in the Gaussian copula model. The first step samples common factors driving the latent variables from a multivariate normal distribution with a shifted mean. Conditional on the common factors, the individual obligors become independent and the second step increases their default probabilities. Glasserman and Li (2005) establish asymptotic optimality results for this method when applied to single-factor homogeneous portfolios; i.e., models in which all obligors are identical and in which they become independent conditional on a single underlying normal random variable. In this context, the new factor mean used for IS is a scalar.

In a single-factor homogeneous model, there is a single direction in which shifting the underlying factor increases default probabilities, so the only question is how far to shift the mean. Glasserman and Li (2005) select the mean by maximizing a bound on the product of the conditional loss distribution and the density of the common factor. This step relies on the logconcavity of the bound, a property that does not extend widely to more general models. Moreover, in a multifactor heterogeneous model, there may be many combinations and directions of factor shifts that increase conditional default probabilities, and this potentially necessitates the use of a mixture of IS distributions, each associated with a different shift of mean.

In this paper, we therefore develop an alternative approach for selecting shifts in factor means, building on insights gleaned from the large deviations analysis in Glasserman, Kang, and Shahabuddin (2007) (henceforth, GKS). When applied to a single-factor homogeneous model, this method is somewhat coarser than that in Glasserman and Li (2005), though asymptotically equivalent. However, this method lends itself much more readily to the selection of multiple mean shifts for multifactor heterogeneous models. The new method divides the space of factor outcomes into sets that lead to the default of different combinations of obligors; the possible mean shifts are the minimal points in these sets. Each of these minimal points corresponds to the "most likely" factor outcome leading to the default of a particular set of obligors. Finding the appropriate sets of factor outcomes is a combinatorial problem and finding each minimal point is a quadratic programming problem. We therefore investigate methods to accelerate implementation.

From the general perspective of rare-event simulation, the main contribution of this paper lies in the handling of complex dependence between defaults. Moreover, the techniques we use for characterizing, calculating, and reducing the large number of potential mean shifts is potentially applicable to other applications.

The rest of this paper is organized as follows. Section 2 describes the portfolio credit-risk problem and introduces our IS procedure. Section 3 analyzes the IS algorithm in two limiting regimes and establishes asymptotic optimality; most proofs are deferred to the online appendix. Section 4 gives simple examples to motivate the use of a mixture of IS distributions—i.e., multiple mean shifts. Section 5 analyzes two optimization problems associated with the choice of IS distribution, where we also deferred some proofs to the online appendix. Section 6 presents numerical results. An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.

2. Portfolio Credit Risk and Importance Sampling

2.1. The Problem

We consider the distribution of losses from default over a fixed horizon. We are interested in the estimation of the probability that the credit loss of a portfolio exceeds a given threshold. The default of each obligor is triggered if a latent variable associated with the obligor exceeds a threshold determined from its marginal default probability. The latent variables are linear combinations of factor variables representing idiosyncratic risk and common risks of all obligors. We use the following notation:

$m =$ number of obligors to which the portfolio is exposed;
$Y_k =$ default indicator ($=1$ for default, $=0$ otherwise) for the kth obligor;
$p_k =$ marginal probability that the kth obligor defaults;
$\ell_k =$ loss resulting from default of the kth obligor;
$L_m = \ell_1 Y_1 + \cdots + \ell_m Y_m =$ total loss from defaults.

We are interested in the estimation of $\mathsf{P}(L_m > x)$ for a given threshold x when the event $\{L_m > x\}$ is rare. (For easy reference, we refer to the event $\{L_m > x\}$ as a *large loss event*.) The loss amount due to a default of the kth obligor, ℓ_k, can be stochastic. We assume that the ℓ_ks are bounded by a common upper bound and, for stochastic ℓ_ks, that they are independent of all other random variables. We denote $\mathsf{E}[\ell_k] = c_k$. Dependence among the default indicators Y_k is given by a *multifactor Gaussian copula model with a finite number of types*. In the Gaussian copula, the latent variables that trigger default have a multivariate normal distribution. Thus, we set $Y_k = \mathbf{1}\{X_k > \Phi^{-1}(1-p_k)\}$, with Φ the cumulative normal distribution and $X_1, X_2, \ldots$ correlated standard normal random variables; this makes $\mathsf{P}(Y_k = 1) = p_k$. Correlations between the latent variables $X_1, X_2, \ldots$ determine the dependence among the default indicators. In practice, these correlations are often derived from correlations in asset values or equity returns.

By types, we mean groups of obligors that share a dependence structure, in a sense made precise in **M1**:

M1. There are d factors and t types of obligors. $\{\mathcal{I}_1^{(m)}, \ldots, \mathcal{I}_t^{(m)}\}$ is a partition of the set of obligors $\{1, \ldots, m\}$ into types. If $k \in \mathcal{I}_j^{(m)}$, then the kth obligor is of type j and its latent variable is given by

$$X_k = \mathbf{a}_j^\top \mathbf{Z} + b_j \varepsilon_k,$$

where $\mathbf{a}_j \in \mathbb{R}^d$ with $0 < \|\mathbf{a}_j\| < 1$, $\mathbf{Z}$ is a d-dimensional standard normal random vector, $b_j = \sqrt{1 - \mathbf{a}_j^\top \mathbf{a}_j}$, and ε_k are

independent standard normal random variables. Here, $\mathbf{Z}$ represents a vector of systematic risks (i.e., common factors) and ε_k represents an idiosyncratic risk of the kth obligor. The common factors may be correlated, but we can transform them into independent random variables by a factorization of their covariance matrix. $\mathbf{a}_j$ is the vector of factor-loading coefficients of obligors belonging to type j. Let $n_j^{(m)} = |\mathcal{I}_j^{(m)}|$ denote the number of obligors of type j.

This is the Gaussian copula model of Gupton et al. (1997) with a factor structure in the correlation of the latent variables, and with the added condition that the obligors can be grouped into types. We think of the number of types as substantially smaller than the number of obligors. (Later, in fact, we will let the number of obligors increase while the number of types remains fixed.) Depending on how the factor-loading vectors are determined, the assumption that multiple obligors share the same type may involve an approximation and even a restriction. Clustering techniques may be useful in selecting factor-loading vectors for a finite number of types.

2.2. Importance Sampling

IS is a standard approach to variance reduction in Monte Carlo methods. Suppose that we have to estimate $\mathsf{E}_f[h(W)]$, where f is the density function of a random variable W and h is a measurable function with respect to the underlying probability space. Let g be another density function such that $g(w) > 0$ if $f(w) > 0$. Then, the expectation can be rewritten as

$$\begin{aligned}\mathsf{E}_f[h(W)] &= \int h(w)f(w)\,dw = \int h(w)\frac{f(w)}{g(w)}g(w)\,dw \\ &= \mathsf{E}_g\left[h(W)\frac{f(W)}{g(W)}\right].\end{aligned}$$

One can then sample W from g and use $h(W)f(W)/g(W)$ as an unbiased estimator of the original expectation. The objective is to choose g so that the variance of the new estimator is much smaller than that of the original estimator. We consider an IS method that is effective for the estimation of default probabilities when defaults are rare. One way is to develop an *asymptotically optimal method* to simulate a sequence of probabilities $\mathsf{P}(L_m > x_m)$ in which parameters change in such a manner that the default becomes rare as m approaches infinity. (More discussion of asymptotic optimality is given after Theorem 2.) Then, we apply this asymptotically optimal method to cases of finite m and expect a large reduction of variance.

We exploit the conditional independence property by adopting a two-step IS. Glasserman and Li (2005) show that IS applied only to the idiosyncratic risk, ε_k, of each obligor is not sufficient to get an asymptotic optimality unless the dependence among obligors is sufficiently weak. So, they suggest a two-step IS method combining a change of measure on the common factors, $\mathbf{Z}$, and conditional on $\mathbf{Z}$, IS on the default probabilities.

2.2.1. Importance Sampling Conditional on Common Factors. Once we condition on $\mathbf{Z}$, the loss L_m becomes a sum of independent random variables for which there is a standard way of finding an asymptotically optimal IS procedure for estimating the chance of the rare event $\{L_m > x\}$. The development below is similar to the one in Glasserman and Li (2005) except that we allow ℓ_k to be random instead of constant. Define

$$\Lambda_k(\lambda) = \log \mathsf{E}[e^{\lambda \ell_k}].$$

The change of measure for the second step of the IS is defined by a conditional change of measure given $\mathbf{Z}$. Conditional on $\mathbf{Z}$, the default events are independent Bernoulli random variables and the conditional default probability of the kth obligor (of type j) is given by

$$\begin{aligned}p_k(\mathbf{Z}) &= \mathsf{P}(Y_k = 1 \mid \mathbf{Z}) = \mathsf{P}(X_k > \Phi^{-1}(1-p_k) \mid \mathbf{Z}) \\ &= \Phi\left(\frac{\mathbf{a}_j^\top \mathbf{Z} + \Phi^{-1}(p_k)}{b_j}\right). \qquad (1)\end{aligned}$$

We change these probabilities to "exponentially twisted" probabilities, $p_{k,\theta}(\mathbf{Z})$, given by

$$p_{k,\theta}(\mathbf{Z}) = \frac{p_k(\mathbf{Z})e^{\Lambda_k(\theta)}}{1 + p_k(\mathbf{Z})(e^{\Lambda_k(\theta)} - 1)} \qquad (2)$$

for some $\theta \geqslant 0$. (For a detailed discussion on exponential twisting (2), refer to §3.1 in Glasserman and Li 2005.) The conditional likelihood ratio associated with this change of default probabilities is given by

$$\begin{aligned}&\prod_{k=1}^m \left(\frac{p_k(\mathbf{Z})}{p_{k,\theta}(\mathbf{Z})}\right)^{Y_k}\left(\frac{1-p_k(\mathbf{Z})}{1-p_{k,\theta}(\mathbf{Z})}\right)^{1-Y_k} \\ &\quad = e^{-\sum_{k=1}^m Y_k\Lambda_k(\theta) + m\psi_m(\theta,\mathbf{Z})}, \qquad (3)\end{aligned}$$

where $\psi_m(\theta, \mathbf{z})$ is the conditional cumulant generating function divided by m,

$$\begin{aligned}\psi_m(\theta,\mathbf{z}) &\triangleq \frac{1}{m}\log \mathsf{E}[e^{\theta L_m} \mid \mathbf{Z} = \mathbf{z}] \\ &= \frac{1}{m}\sum_{k=1}^m \log(1 + p_k(\mathbf{z})(e^{\Lambda_k(\theta)} - 1)). \qquad (4)\end{aligned}$$

For given $\mathbf{Z}$ and Y_ks, we apply another conditional IS for the ℓ_ks. This step is an exponential twisting of ℓ_k by θY_k, i.e., if $f_{\ell_k}(l)$ is the original density function for ℓ_k, then the new density function is given by

$$f_{\ell_k,\theta}(l) = f_{\ell_k}(l)e^{\theta Y_k l - \Lambda_k(\theta Y_k)}. \qquad (5)$$

Hence, the likelihood ratio given $\mathbf{Z}$ and Y_ks is given by

$$\begin{aligned}&\prod_{k=1}^m \frac{f_{\ell_k}(\ell_k)}{f_{\ell_k,\theta}(\ell_k)} \\ &\quad = \prod_{k=1}^m e^{-\theta Y_k \ell_k + \Lambda_k(Y_k\theta)} = e^{-\theta\sum_{k=1}^m Y_k\ell_k + \sum_{k=1}^m \Lambda_k(Y_k\theta)}. \qquad (6)\end{aligned}$$

Note that for deterministic ℓ_k, (6) is a constant, 1. Because Y_k is binary and $\Lambda_k(0) = 0$, we have $Y_k\Lambda_k(\theta) = \Lambda_k(Y_k\theta)$, w.p. 1. So, the product of the two likelihood ratios, (3) and (6), simplifies to

$$e^{-\theta L_m + m\psi_m(\theta, \mathbf{Z})}.$$

Now we choose θ to reduce the variance of the new estimator of $\mathsf{P}(L_m > x \mid \mathbf{Z})$. An upper bound of the second moment of this estimator is

$$\begin{aligned}&\tilde{\mathsf{E}}[\mathbf{1}\{L_m > x\}e^{-2\theta L_m + 2m\psi_m(\theta, \mathbf{Z})} \mid \mathbf{Z}] \\ &\quad \leqslant \tilde{\mathsf{E}}[\mathbf{1}\{L_m > x\}e^{-2\theta x + 2m\psi_m(\theta, \mathbf{Z})} \mid \mathbf{Z}] \leqslant e^{-2\theta x + 2m\psi_m(\theta, \mathbf{Z})},\end{aligned}$$

where $\tilde{\mathsf{E}}$ is an expectation under the new measure. We minimize this (conditionally) deterministic upper bound and choose θ as

$$\theta_m(\mathbf{z}) \triangleq \arg\min_{\theta \geqslant 0}\{-\theta x + m\psi_m(\theta, \mathbf{z})\}. \tag{7}$$

So, the likelihood ratio that we use for the conditional IS is given by

$$e^{-\theta_m(\mathbf{Z})L_m + m\psi_m(\theta_m(\mathbf{Z}), \mathbf{Z})}. \tag{8}$$

2.2.2. Importance Sampling for Common Factors. Now consider IS for common factors. We limit ourselves to changes in the mean of the common factor distribution. Glasserman and Li (2005) suggest that a solution to

$$\max_{\mathbf{z}} \mathsf{P}(L_m > x \mid \mathbf{Z} = \mathbf{z})e^{-\mathbf{z}^\top \mathbf{z}/2} \tag{9}$$

should be an effective choice of mean for the new multivariate normal distribution. However, it is difficult to solve this problem exactly. So, instead, they use an upper bound $\exp(-\theta_m(\mathbf{z})x + m\psi_m(\theta_m(\mathbf{z}), \mathbf{z}))$ as a surrogate for $\mathsf{P}(L_m > x \mid \mathbf{Z} = \mathbf{z})$ and then solve

$$\max_{\mathbf{z}}\left\{-\theta_m(\mathbf{z})x + m\psi_m(\theta_m(\mathbf{z}), \mathbf{z}) - \tfrac{1}{2}\mathbf{z}^\top \mathbf{z}\right\}. \tag{10}$$

Glasserman and Li (2005) proved that this approach is asymptotically optimal for the homogeneous single-factor case. However, the result does not generalize to the multifactor case because the function $-\theta_m(\mathbf{z})x + m\psi_m(\theta_m(\mathbf{z}), \mathbf{z})$ may not have a dominating hyperplane at the maximum point of (10) for general portfolios, which is crucial in their approach. This condition is used to ensure that a single shift in the mean of the factors produces effective variance reduction; when the condition fails, we are led to consider a mixture of distributions with multiple mean shifts. See Sadowsky and Bucklew (1990) for related observations and results.

This leads to the problem of choosing the mean shifts for the mixture. As a heuristic, one might consider using all local maxima of (10) as candidates. However, in §4, we give an example in which the dominating hyperplane requirement fails, but for which (10) has a unique local (and, in fact, global) maximum. Thus, any method based on mixtures of local maxima would reduce to that of Glasserman and Li (2005) in this case. As a general IS procedure, Avramidis (2002) proposes the use of local maxima in the case of a multimodal importance function. Because he considered the unconditional importance function (without the conditional IS step in §2.2.1), the size of the relevant optimization problem (with $m + d$ variables) is too large to be handled efficiently. Moreover, the unconditional importance function is not only nondifferentiable but also discontinuous, further complicating the optimization. If instead we were to apply Avramidis's (2002) method for IS on the common factors, we would still need an approximation or a surrogate for $\mathsf{P}(L_m > x \mid \mathbf{Z} = \mathbf{z})e^{-\mathbf{z}^\top \mathbf{z}/2}$, which itself is a challenging problem. Using (10) as a surrogate would in some cases result in a single maximum, as noted above.

We therefore develop a different approach that applies directly to the multifactor setting. Our approach identifies regions in the factor space where large portfolio losses are more likely to occur. Under some limiting regimes, these regions can be identified by polyhedra in the factor space that do not include the origin. Our IS change of measure uses a mixture of normals in which each component of the mixture shifts the mean of $\mathbf{Z}$ to the closest point to the origin in one of the polyhedra. This approach separates consideration of the credit exposures from the dependence mechanism and default probabilities.

We adopt the *constant approximation* in §5.1 of Glasserman and Li (2005) for our explanation. In this approximation, L_m and $\mathsf{P}(L_m > x \mid \mathbf{Z} = \mathbf{z})$ are replaced by $\mathsf{E}[L_m \mid \mathbf{Z} = \mathbf{z}]$ and $\mathbf{1}\{\mathsf{E}[L_m \mid \mathbf{Z} = \mathbf{z}] > x\}$, respectively. Then, (9) becomes

$$\min\{\mathbf{z}^\top \mathbf{z}\colon \mathsf{E}[L_m \mid \mathbf{Z} = \mathbf{z}] > x\}. \tag{11}$$

To analyze this further, we define aggregated credit exposures,

$$C_j \triangleq \frac{1}{m}\sum_{k \in \mathcal{I}_j^{(m)}} \mathsf{E}[\ell_k] \quad \text{for } j = 1, \ldots, t \quad \text{and}$$

$$C \triangleq \frac{1}{m}\sum_{k=1}^{m} \mathsf{E}[\ell_k] = \sum_{j=1}^{t} C_j.$$

For simplicity, suppose (temporarily) that $p_k = p_j$ for all $k \in \mathcal{I}_j^{(m)}$, meaning that obligors of the same type have the same default probability. Using (1), we get

$$\begin{aligned}\frac{1}{m}\mathsf{E}[L_m \mid \mathbf{Z} = \mathbf{z}] &= \frac{1}{m}\sum_{k=1}^{m} \mathsf{E}[\ell_k]\Phi\left(\frac{\mathbf{a}_j^\top \mathbf{z} + \Phi^{-1}(p_k)}{b_j}\right) \\ &= \sum_{j=1}^{t} C_j \Phi\left(\frac{\mathbf{a}_j^\top \mathbf{z} + \Phi^{-1}(p_j)}{b_j}\right).\end{aligned}$$

If we set q as the loss-percentage threshold in which we are interested (that is, $q = x/mC$), then $\{\mathbf{z}\colon \mathsf{E}[L_m \mid \mathbf{Z} = \mathbf{z}] > x\}$ becomes

$$\left\{\mathbf{z}\colon \sum_{j=1}^{t} C_j \Phi\left(\frac{\mathbf{a}_j^\top \mathbf{z} + \Phi^{-1}(p_j)}{b_j}\right) > qC\right\}. \tag{12}$$

Our goal is to characterize this set when the default probabilities p_j are small or when the fraction q is close to one.

As a building block for the case of multiple types, observe that in the case of a single type, j, (12) reduces to

$$\overline{G}_j \triangleq \left\{\mathbf{z}\colon \mathbf{a}_j^\top \mathbf{z} > \Phi^{-1}(1-p_j) + b_j \Phi^{-1}(q)\right\}. \tag{13}$$

(In our implementation, we will tune the right side of the inequality defining $\overline{G}_j$ and set

$$G_j^{(m)} \triangleq \left\{\mathbf{z} \in \mathbb{R}^d\colon \mathbf{a}_j^\top \mathbf{z} \geqslant \alpha_1^{(m)} \Phi^{-1}(1-\bar{p}_j) + \alpha_2^{(m)} b_j \Phi^{-1}(q)\right\}, \tag{14}$$

where $\alpha_1^{(m)}$ and $\alpha_2^{(m)}$ are two sequences satisfying $0 \leqslant \alpha_1^{(m)} < 1$ and $0 \leqslant \alpha_2^{(m)} < 1$ with $\alpha_1^{(m)} \to 1$ and $\alpha_2^{(m)} \to 1$ as $m \to \infty$. We will also allow different default probabilities for obligors of the same type, so we set $\bar{p}_j = \max_{k \in \mathcal{J}_j^{(m)}} p_k$. In this subsection, for simplicity, we explain our approach using (13) instead of (14).)

In our asymptotic analysis, we introduce two limiting regimes called the *small default probabilities* (*SDP*) case and the *large loss threshold* (*LLT*) case; see §§3.1 and 3.2 for precise definitions of these regimes. In the SDP regime, the default probabilities p_j decrease to zero while the loss threshold q is fixed; this parameterization makes $\Phi^{-1}(1-p_j)$ dominate the right side of the inequality in (13). In the LLT regime, the loss threshold q increases toward one while the default probabilities are fixed, so $\Phi^{-1}(q)$ dominates the inequality in (13).

For the LLT regime, with the loss threshold q close to 100%, the event $\{\mathsf{E}[L_m \mid \mathbf{Z}] > qC\}$ requires that the common factors increase the chances of default for *all* obligor types. This occurs when $\mathbf{z}$ is in $\bigcap_{j=1}^{t} \overline{G}_j$; and the asymptotic analysis in GKS confirms that the rate of decrease of the probability of this set does indeed determine the rate of decrease of the probability that $L_m > qC$. Hence, the unique solution of $\min\{\mathbf{z}^\top \mathbf{z}\colon \mathbf{z} \in \bigcap_{j=1}^{t} \overline{G}_j\}$ gives an asymptotically optimal conditional mean shift for the factors.

Whereas in the LLT regime all types of obligors default, in the SDP regime defaults of different combinations of types may result in losses that exceed the threshold. This makes the SDP regime trickier. The boundaries of $\{\mathbf{z}\colon \mathsf{E}[L_m \mid \mathbf{Z} = \mathbf{z}] > qC\}$ can look very different, depending on the value of q, and they are quite complicated for some q values; see Figure 2, for example. We approach this case by dividing $\{\mathbf{z}\colon \mathsf{E}[L_m \mid \mathbf{Z} = \mathbf{z}] > qC\}$ into tractable regions. Each region is characterized by an intersection of $\overline{G}_j$s over a subset of types; roughly speaking, the subsets are chosen so that defaults of obligors of those types produce losses exceeding the threshold. We choose these sets to be minimal, in the sense that no subset would produce sufficiently large losses. Choosing minimal sets ensures that we shift the factor mean as little as necessary; shifting the mean too far can produce an increase in variance.

In more detail, we define $\mathcal{J}$ to be a *q-minimal index set* if $\mathcal{J} \subset \{1, \ldots, t\}$ and

$$\max_{\mathcal{J}' \subsetneq \mathcal{J}} \sum_{j \in \mathcal{J}'} C_j < qC \leqslant \sum_{j \in \mathcal{J}} C_j. \tag{15}$$

These are the minimal combinations of types whose sum of C_js are larger than the threshold qC. In the SDP limit, the union (over all q-minimal index sets $\mathcal{J}$) of the sets $\bigcap_{j \in \mathcal{J}} \overline{G}_j$ includes $\{\mathbf{z}\colon \mathsf{E}[L_m \mid \mathbf{Z} = \mathbf{z}] > qC\}$, and each $\bigcap_{j \in \mathcal{J}} \overline{G}_j$ covers some portion of the boundary of $\{\mathbf{z}\colon \mathsf{E}[L_m \mid \mathbf{Z} = \mathbf{z}] > qC\}$. Hence, a large deviation perspective on efficient IS suggests the use of a mixture of mean shifts given by $\arg\min\{\mathbf{z}^\top \mathbf{z}\colon \mathbf{z} \in \bigcap_{j \in \mathcal{J}} \overline{G}_j\}$, with $\mathcal{J}$ ranging over all q-minimal index sets. A factor outcome $\mathbf{z}$ in $\bigcap_{j \in \mathcal{J}} \overline{G}_j$ facilitates the defaults of obligors of types in $\mathcal{J}$, so this set of $\mathbf{z}$ approximates

$$\left\{\mathbf{z}\colon \sum_{j \in \mathcal{J}} C_j \Phi\left(\frac{\mathbf{a}_j^\top \mathbf{z} + \Phi^{-1}(p_j)}{b_j}\right) > qC\right\}. \tag{16}$$

The dominance of $\Phi^{-1}(1-p_j)$ on the right side of the inequality in (13) under the SDP regime makes $\overline{G}_j$ insensitive to q, resulting in the asymptotic equivalence of (16) and $\bigcap_{j \in \mathcal{J}} \overline{G}_j$. These are main ideas of the asymptotic analysis of the SDP regime.

Example 1. To build a simple visual example, we take $\mathsf{E}[\ell_k] \equiv 1$, $p_k \equiv 0.01$, and we consider four types in $\mathbb{R}^2$: $\mathbf{a}_1^\top = (0.85, 0)$, $\mathbf{a}_2^\top = (0.25, 0)$, $\mathbf{a}_3^\top = (0, 0.85)$, and $\mathbf{a}_4^\top = (0, 0.25)$. Each type has the same number of obligors. Figure 1 depicts the surface of $(1/C)\mathsf{E}[L_m \mid \mathbf{Z} = \mathbf{z}]$. Figure 2 shows the contours (level curves) of $(1/C)\mathsf{E}[L_m \mid \mathbf{Z} = \mathbf{z}]$, which are comparable to $\{\mathbf{z}\colon \mathsf{E}[L_m \mid \mathbf{Z} = \mathbf{z}] = qC\}$ for the loss thresholds q. The four $\overline{G}_j$s for each q level are halfspaces whose boundaries are two horizontal and two vertical lines with different z_1- and z_2-intercepts, respectively.

Figure 1. 3D plot of $(1/C)\mathsf{E}[L_m \mid \mathbf{Z} = \mathbf{z}]$.

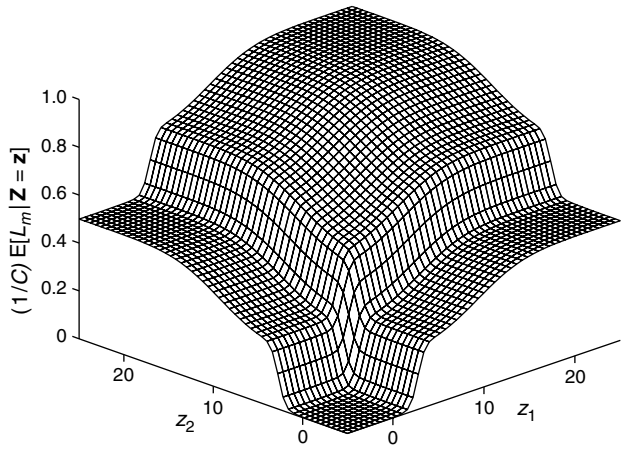
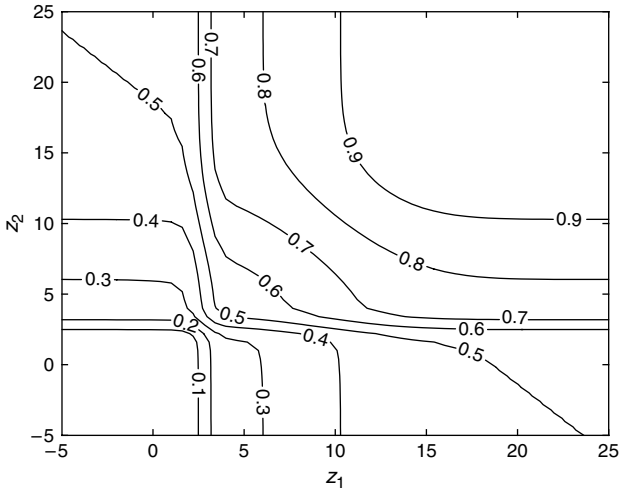


Figure 2. Contours of $(1/C)\mathsf{E}[L_m \mid \mathbf{Z}=\mathbf{z}]$.



Note. The numbers on the contours are threshold values q.

In Figure 2, the level curves corresponding to $q = 0.8$ and 0.9 seem to be covered well by intersections of half-spaces whose boundaries are one horizontal line and one vertical line, respectively, because the only q-minimal index set is $\{1, 2, 3, 4\}$. Different q values define different $\overline{G}_j$s. For $q = 0.1$ and 0.2, the corresponding level curves seem to be boundaries of unions of half-spaces associated with one horizontal line and one vertical line because q-minimal index sets are $\{1\}$, $\{2\}$, $\{3\}$, and $\{4\}$. As the above discussion suggests, the level curves corresponding to intermediate q levels show complicated shapes because of multiple q-minimal index sets. We expect that unions of intersections of $\overline{G}_j$s would cover the $\{\mathbf{z}\colon \mathsf{E}[L_m \mid \mathbf{Z}=\mathbf{z}] > qC\}$ regions.

Define $\mathcal{M}_q$ as the family of all q-minimal index sets and

$$G_{\mathcal{J}}^{(m)} \triangleq \bigcap_{j\in\mathcal{J}} G_j^{(m)} \quad \text{for } \mathcal{J}\in\mathcal{M}_q \quad \text{and} \quad G^{(m)} \triangleq \bigcup_{\mathcal{J}\in\mathcal{M}_q} G_{\mathcal{J}}^{(m)}.$$

The condition $\|\mathbf{a}_j\| > 0$ in **M1** implies $\mathbf{a}_j \neq \mathbf{0}$. If all $\mathbf{a}_j \geqslant \mathbf{0}$, then $G_{\mathcal{J}}^{(m)} \neq \varnothing$ for any $\mathcal{J} \subset \{1, \ldots, t\}$. However, if some components of $\mathbf{a}_j$ are negative, these sets may be empty. Because we need to define the new IS distribution using these minimal index sets, smaller $\mathcal{M}_q$ is desirable for efficient implementation. Hence, we introduce a *sufficient subfamily* of $\mathcal{M}_q$, which includes minimal index sets enough to define an efficient IS distribution. We denote it by $\mathcal{S}_q$. It satisfies:

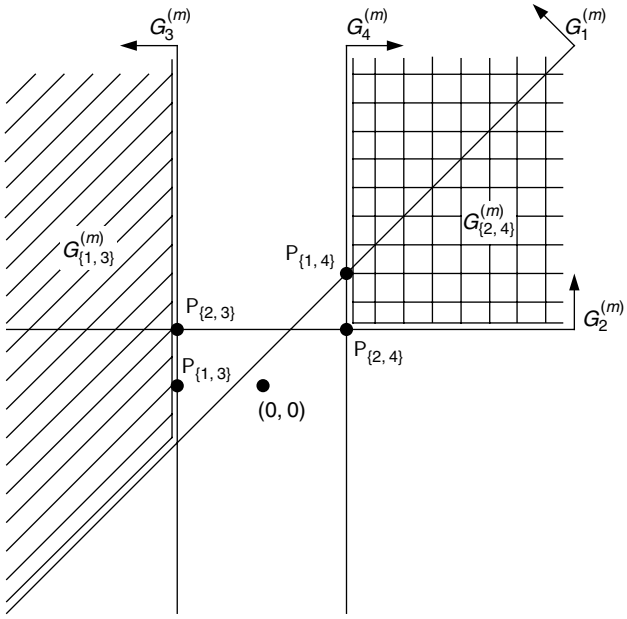
Feasibility: For each $\mathcal{J}\in\mathcal{S}_q$, $G_{\mathcal{J}}^{(m)} \neq \varnothing$ for all m;

Covering property: $\bigcup_{\mathcal{J}\in\mathcal{S}_q} G_{\mathcal{J}}^{(m)} = G^{(m)}$ for all m.

The choice of $\mathcal{S}_q$ may not be unique (see Example 2), but the asymptotic optimality of IS does not depend on this choice. For each $\mathcal{J}\in\mathcal{S}_q$, we define $\boldsymbol{\mu}_{\mathcal{J}}^{(m)}$ as the *unique solution* of the following *linearly constrained quadratic optimization problem*:

$$\boldsymbol{\mu}_{\mathcal{J}}^{(m)} \triangleq \arg\min\{\|\mathbf{z}\|\colon \mathbf{z}\in G_{\mathcal{J}}^{(m)}\}. \tag{17}$$

Figure 3. Illustration of the half-spaces associated with obligor types.



The IS approach uses a mixture of $N(\boldsymbol{\mu}_{\mathcal{J}}^{(m)}, \mathbf{I})$ for $\mathcal{J}\in\mathcal{S}_q$ as a new measure.

Example 2. To illustrate, we consider a simple example ($d = 2$ and $t = 4$ case) from GKS (2007). There are four obligor types with $C_1 = 2$, $C_2 = 2$, $C_3 = 3$, and $C_4 = 3$, so $C = 10$. Set $q = 0.45$. Then, $\mathcal{M}_q = \{\{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}\}$. Figure 3 shows the coefficient vector for each type (indicated by arrows) and the resulting sets $G_j^{(m)}$, $j = 1, 2, 3, 4$. $G_{\{1,3\}}^{(m)}$ is indicated by slanted lines and $G_{\{2,4\}}^{(m)}$ is indicated by crossed lines. $G^{(m)} = G_{\{1,3\}}^{(m)} \cup G_{\{2,4\}}^{(m)}$ because $G_{\{2,3\}}^{(m)} \subset G_{\{1,3\}}^{(m)}$, $G_{\{1,4\}}^{(m)} \subset G_{\{2,4\}}^{(m)}$, and $G_{\{3,4\}}^{(m)} = \varnothing$. The point $\mathsf{P}_{\mathcal{J}}$ denotes $\boldsymbol{\mu}_{\mathcal{J}}^{(m)}$ for each $\mathcal{J}\in\mathcal{M}_q$ and $G_{\mathcal{J}}^{(m)} \neq \varnothing$. One minimal $\mathcal{S}_q$ is $\{\{1,3\}, \{2,4\}\}$. Hence, $\mathsf{P}_{\{1,3\}}$ and $\mathsf{P}_{\{2,4\}}$ can define the change of measure on common factors. Note that any subfamily of $\mathcal{M}_q$ containing both $\{1,3\}$ and $\{2,4\}$ is valid as an $\mathcal{S}_q$.

2.2.3. Importance Sampling Procedure. Now we are in a position to state the complete IS procedure. As mentioned before, instead of choosing a single mean shift, we come up with a set of mean shifts $\boldsymbol{\mu}_{\mathcal{J}}^{(m)}$ for $\mathcal{J}\in\mathcal{S}_q$. Let $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K)$ denote these mean shifts. We then use a mixture of normal random vectors with mean vectors $\boldsymbol{\mu}_i$. We assign weight λ_i to the ith mean, with $\sum_{i=1}^K \lambda_i = 1$.

Mixed Importance Sampling (MIS)

Choosing the Means for Factor Shifting: This step is executed once in the beginning.

Step 1. Find $\mathcal{S}_q$ by solving (15). Set $K = |\mathcal{S}_q|$.

Step 2. Find $\boldsymbol{\mu}_{\mathcal{J}}^{(m)}$ by solving (17) for each $\mathcal{J}\in\mathcal{S}_q$. $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K$ denote the vectors found.

Step 3. Choose the weights $\lambda_1, \ldots, \lambda_K$ for $K \geqslant 1$. We choose the number of replications, n, and λ_is so that $\lambda_i \cdot n$ is an integer for all i.

Main Loop: Repeat for replications $r = 1, \ldots, \lambda_i \cdot n$, and for $i = 1, \ldots, K$.

Step 1. Sample $\mathbf{Z}$ from $N(\boldsymbol{\mu}_i, \mathbf{I})$.

Step 2. Find $\theta_m(\mathbf{Z})$ by solving (7).

Step 3. For $k = 1, \ldots, m$, compute the twisted conditional default probabilities $p_{k,\theta_m(\mathbf{Z})}(\mathbf{Z})$ by (2) and then sample Y_k from Bernoulli distribution with $p_{k,\theta_m(\mathbf{Z})}(\mathbf{Z})$.

Step 4. For k with $Y_k = 1$, generate the loss ℓ_k under the twisted conditional distribution given by (5). If the loss is deterministic, set $\ell_k = c_k$.

Step 5. Calculate

$$I_r^{(i)} = 1\{L_m > x\} e^{-\theta_m(\mathbf{Z})L_m + m\psi_m(\theta_m(\mathbf{Z}),\mathbf{Z})} \cdot \left(\sum_{i=1}^{K} \lambda_i \exp\left(\boldsymbol{\mu}_i^\top \mathbf{Z} - \frac{1}{2}\boldsymbol{\mu}_i^\top \boldsymbol{\mu}_i\right)\right)^{-1}.$$

Return the estimate: $(1/n)\sum_{i=1}^{K}\sum_{r=1}^{\lambda_i \cdot n} I_r^{(i)}$.

There is some flexibility in choosing the λ_is. We compared the performance of various rules including uniform weights and a rule minimizing an upper bound of $(\sum_{i=1}^{K} \lambda_i \exp(\boldsymbol{\mu}_i^\top \mathbf{Z} - \frac{1}{2}\boldsymbol{\mu}_i^\top \boldsymbol{\mu}_i))^{-1}$. However, we did not find any significant difference. So, we use the uniform weights $\lambda_i = 1/K$ in our implementation.

In **MIS**, we defined the $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K$ by setting $K = |\mathcal{S}_q|$ and $\boldsymbol{\mu}_i = \boldsymbol{\mu}_{\mathcal{J}}$ under an appropriate order on $\mathcal{J}$s in $\mathcal{S}_q$. However, if $\boldsymbol{\mu}_{\mathcal{J}}^{(m)} = \boldsymbol{\mu}_{\mathcal{J}'}^{(m)}$ for two different minimal index sets $\mathcal{J}$ and $\mathcal{J}'$, the definition of the IS mixture distribution allows the addition of the two associated weights. That is, even though $|\{\boldsymbol{\mu}_{\mathcal{J}}^{(m)}\colon\ \mathcal{J} \in \mathcal{S}_q\}| < |\mathcal{S}_q|$ in most cases, the two IS distributions defined by the $\boldsymbol{\mu}_{\mathcal{J}}^{(m)}$s for $\mathcal{J} \in \mathcal{S}_q$ and $\{\boldsymbol{\mu}_{\mathcal{J}}^{(m)}\colon\ \mathcal{J} \in \mathcal{S}_q\}$ coincide if the weights are added appropriately. By Lemma 1 in §5.1, we may focus on the characterization of $\{\boldsymbol{\mu}_{\mathcal{J}}^{(m)}\colon\ \mathcal{J} \subset \{1, \ldots, t\}, |\mathcal{J}| \leqslant d, G_{\mathcal{J}}^{(m)} \neq \varnothing\}$ instead of $\mathcal{S}_q$. This will greatly reduce the computational burden in choosing the $\boldsymbol{\mu}_i$s.

3. Analysis of Importance Sampling Estimators

In this section, we show that **MIS** remains efficient as the default event becomes rarer if we choose the parameters $\alpha_1^{(m)}$ and $\alpha_2^{(m)}$ appropriately. As in Glasserman and Li (2005), we consider two limiting parametric regimes which result in small default probabilities of portfolios:

Small default probability limit: small p_ks and moderate x values;

Large loss threshold limit: large x value and moderate p_ks.

More specifically, we consider a sequence of portfolios $\{L_m\}_{m=1}^{\infty}$ and we show that **MIS** is asymptotically optimal in the two regimes. First, we add a regularity condition on the limiting behavior of the number of obligors of each type.

M1 (Continued). We assume that for each $j = 1, \ldots, t$, $\lim_{m\to\infty} n_j^{(m)}/m = r_j > 0$.

The aggregated losses C_j and C will be redefined as limits as the size of the portfolio increases.

3.1. Small Default Probabilities Regime

We specify the small default probability regime by imposing the assumptions in **M2** in addition to those in **M1**.

M2. (1) The default loss, ℓ_k, is deterministic and equal to c_k, $0 < c_k \leqslant \bar{c} < \infty$, for $k = 1, \ldots, m$.

(2) If the kth obligor is of type j, then its default probability is given by $p_k = p_j^{(m)} \triangleq \Phi(-s_j\sqrt{m})$, where $s_j > 0$. Hence, the conditional default probability (given the factors $\mathbf{Z}$) of the same obligor is given by

$$p_k(\mathbf{Z}) = p_j^{(m)}(\mathbf{Z}) = \Phi\left(\frac{\mathbf{a}_j^\top \mathbf{Z} + \Phi^{-1}(p_j^{(m)})}{b_j}\right) = \Phi\left(\frac{\mathbf{a}_j^\top \mathbf{Z} - s_j\sqrt{m}}{b_j}\right).$$

(3) For each type $j = 1, \ldots, t$,

$$C_j \triangleq \lim_{m\to\infty} \frac{1}{m} \sum_{k \in \mathcal{J}_j^{(m)}} c_k < \infty \quad \text{and}$$

$$C \triangleq \lim_{m\to\infty} \frac{1}{m}\sum_{k=1}^{m} c_k = \sum_{j=1}^{t} \lim_{m\to\infty}\left(\frac{1}{m}\sum_{k\in\mathcal{J}_j^{(m)}} c_k\right) = \sum_{j=1}^{t} C_j.$$

(4) The total loss from defaults and the portfolio default threshold are

$$L_m = \sum_{k=1}^{m} c_k Y_k^{(m)} \quad \text{and} \quad x \equiv x_m = q\sum_{k=1}^{m} c_k,$$

where $Y_k^{(m)} = \mathbf{1}_{\{X_k > \Phi^{-1}(1-p_j^{(m)})\}}$ and $0 < q < 1$. Note that $\sum_{k=1}^{m} c_k$ is the maximum possible loss, and thus we are interested in the loss exceeding a fraction q of this. We impose a mild restriction on the possible values of q; q is not a value in the finite set, $\{(1/C)\sum_{j\in\mathcal{J}} C_j\colon\ \mathcal{J} \subset \{1, \ldots, t\}\}$.

We apply the original definition of the q-minimal index set with these C_j and C. For each $\mathcal{J} \in \mathcal{M}_q$, we define a polyhedron,

$$G_{\mathcal{J}} \triangleq \left\{\mathbf{z} \in \mathbb{R}^d\colon\ \mathbf{a}_j^\top \mathbf{z} \geqslant s_j \text{ for all } j \in \mathcal{J}\right\},$$

and define $\boldsymbol{\gamma}_{\mathcal{J}}$ as the *unique solution* of the following *linearly constrained problem*:

$$\boldsymbol{\gamma}_{\mathcal{J}} \triangleq \begin{cases} \arg\min\{\|\mathbf{z}\|\colon\ \mathbf{z} \in G_{\mathcal{J}}\} & \text{if } G_{\mathcal{J}} \neq \varnothing, \\ (\infty, \ldots, \infty)^\top & \text{if } G_{\mathcal{J}} = \varnothing. \end{cases} \tag{18}$$

Define

$$\boldsymbol{\gamma}_* \triangleq \arg\min\{\|\boldsymbol{\gamma}_{\mathcal{J}}\|\colon\ \mathcal{J} \in \mathcal{M}_q\}, \tag{19}$$

breaking ties arbitrarily, if necessary. Note that $\boldsymbol{\gamma}_* = (\infty, \ldots, \infty)^\top$ and $\|\boldsymbol{\gamma}_*\| = \infty$ if $G_{\mathcal{J}} = \varnothing$ for all $\mathcal{J} \in \mathcal{M}_q$ by definition. The following large deviations result was proved in GKS (2007).

THEOREM 1. *If the assumptions* **M1** *and* **M2** *are satisfied, then*

$$\lim_{m\to\infty}\frac{1}{m}\log \mathsf{P}(L_m > x_m) = -\frac{1}{2}\|\boldsymbol{\gamma}_*\|^2.$$

In this paper, we consider only the case $\|\boldsymbol{\gamma}_*\| < \infty$ and make this explicit by imposing:

ASSUMPTION 1. $\|\boldsymbol{\gamma}_*\| < \infty$.

Note that Assumption 1 is equivalent to requiring that there exist at least one minimal index set $\mathcal{J} \in \mathcal{M}_q$ such that $G_{\mathcal{J}} \neq \varnothing$. If the assumption is violated, it may be possible to consider a faster-growing parameterization to get a finite limit and then devise a similar IS estimator to the current one. We do not study this case further in this paper. Note also that s_j decides the vanishing rate of the default probability of type j and $p_j^{(m)} = \exp(-s_j^2 m/2 + o(m))$. Furthermore $\boldsymbol{\gamma}_*$ depends on all s_j which parameterize the regime as well as on the factor-loading vectors $\mathbf{a}_j$.

We denote the likelihood ratio used in **MIS**, corresponding only to the change of measure on the common factors, by $(M_{\mathbf{Z}}^{(m)})^{-1}$, where

$$M_{\mathbf{Z}}^{(m)} \triangleq \sum_{\mathcal{J}\in\mathcal{S}_q} \lambda_{\mathcal{J}} \exp\left(\boldsymbol{\mu}_{\mathcal{J}}^{(m)\top}\mathbf{Z} - \tfrac{1}{2}\boldsymbol{\mu}_{\mathcal{J}}^{(m)\top}\boldsymbol{\mu}_{\mathcal{J}}^{(m)}\right)$$

and $\sum_{\mathcal{J}\in\mathcal{S}_q}\lambda_{\mathcal{J}} = 1$. Note that the sum is over the sufficient subfamily $\mathcal{S}_q$. However, as we noted before, any other IS method having the same distribution as $M_{\mathbf{Z}}^{(m)}$ will be asymptotically optimal. Each summand in $M_{\mathbf{Z}}^{(m)}$ corresponds to a group of types whose default suffices to produce a portfolio loss exceeding the threshold.

Note that $\Lambda_k(\lambda) = c_k\lambda$ because we are assuming ℓ_k to be deterministic. The second part of **MIS** is the conditional IS exploiting the conditional independence structure. The cumulant generating function given by (4) can be written as

$$\psi_m(\theta, \mathbf{z}) = \frac{1}{m}\sum_{j=1}^{t}\sum_{k\in\mathcal{I}_j^{(m)}} \log(1 + p_j^{(m)}(\mathbf{z})(e^{\theta c_k} - 1)).$$

The likelihood ratio for the combined change of measure is given by the product of two changes of measures

$$\frac{d\mathsf{P}}{d\mathsf{P}_m} = e^{-\theta_m(\mathbf{Z})L_m + m\psi_m(\theta_m(\mathbf{Z}),\mathbf{Z})}(M_{\mathbf{Z}}^{(m)})^{-1}.$$

We write E_m for expectation under the probability measure P_m (i.e., under IS distribution), under which $\mathbf{Z}$ is distributed by the mixture defined above and Y_k is a default indicator with conditional default probability $p_{k,\theta_m(\mathbf{Z})}(\mathbf{Z})$. We also write E for expectation under the original probability measure P, under which $\mathbf{Z}$ is a d-dimensional standard normal random variable and Y_k is a default indicator with default probability p_k.

Note that, in this analysis, we restrict **MIS** to instances with homogeneous marginal default probabilities of obligors belonging to the same type.

We have the following result on the second moment of the IS estimator of $\mathsf{P}(L_m > x_m)$, which is denoted by $M_2(x_m, \theta_m(\mathbf{Z}))$. The proof can be found in the online appendix. Combining this bound with the lower bound from the large deviations analysis (i.e., Theorem 1), we conclude the asymptotic optimality of our IS estimator.

THEOREM 2. *Suppose that assumptions* **M1** *and* **M2** *hold and* $\mathcal{S}_q \neq \varnothing$. *Define* $\epsilon_m > 0$ *such that* $\epsilon_m \to 0$ *and* $\epsilon_m\sqrt{m} \to \infty$ *as* $m \to \infty$. *If we apply* **MIS** *with* $\alpha_1^{(m)} = 1 - \epsilon_m$ *and* $0 \leqslant \alpha_2^{(m)} < 1$, *then*

$$\limsup_{m\to\infty}\frac{1}{m}\log M_2(x_m, \theta_m(\mathbf{Z})) \leqslant -\|\boldsymbol{\gamma}_*\|^2.$$

Hence,

$$2\lim_{m\to\infty}\frac{1}{m}\log \mathsf{P}(L_m > x_m) = \lim_{m\to\infty}\frac{1}{m}\log M_2(x_m, \theta_m(\mathbf{Z})),$$

and thus we have asymptotic optimality of the two-step IS estimator obtained by **MIS**.

Note that one possible choice for $\alpha_2^{(m)}$ is zero. The "asymptotic optimality" can be interpreted as the following: there is a positive constant c (in fact, $c = \frac{1}{2}\|\boldsymbol{\gamma}_*\|^2$) for which $\mathsf{P}(L_m > x_m) = \exp(-c\cdot m + o(m))$ and $M_2(x_m, \theta_m(\mathbf{Z})) = \exp(-2c\cdot m + o(m))$. This means that the second moment of the estimator decreases at twice the exponential rate of the loss probability itself. This is the fastest possible rate for any unbiased estimator because of Jensen's inequality. For naive simulation, the second moment decreases as $\exp(-c\cdot m + o(m))$. In the rare-event simulation literature, estimators achieving this rate are called *asymptotically optimal*.

3.2. Large Loss Threshold Regime

Next, we consider the case of increasing loss threshold while the default probability of each obligor remains fixed. We also allow for random recovery and hence the loss resulting from the default of an obligor to be random. We add **M3** to **M1** to specify the parameters.

M3. (1) The marginal default probabilities satisfy $0 < \underline{p} \leqslant p_k \leqslant \bar{p} < 1$ for $k = 1, \ldots, m$.

(2) The maximum loss for obligor k is l_k and $0 < \underline{l} \leqslant l_k \leqslant \bar{l} < \infty$ for $k = 1, \ldots, m$. Set $\ell_k = l_k U_k$, where U_k is a $[\underline{u}, 1]$-valued random variable and $\underline{u}$ is a constant satisfying $0 < \underline{u} \leqslant 1$. For each obligor type j, $\{U_k\}_{k\in\mathcal{I}_j^{(m)}}$ is an i.i.d. sequence from a distribution with mean u_j^*. We use u_k to denote the mean of U_k; this is u_j^* if the kth obligor is of type j. Hence, $c_k = l_k u_k$. These loss random variables are independent of $\mathbf{Z}$ and $\{\varepsilon_k\}$. We assume that $(1/m)\sum_{k\in\mathcal{I}_j^{(m)}} l_k$ converges for all j as $m \to \infty$.

(3) The portfolio loss threshold is given by

$$x \equiv x_m = q_m \sum_{k=1}^{m} c_k,$$

where $q_m = \Phi(s\sqrt{\log m})$ for some $0 < s < 1$. Note that unlike the previous section, $q_m \to 1$ as $m \to \infty$.

Let

$$G \triangleq \bigcap_{j=1}^{t} \{\mathbf{z} \in \mathbb{R}^d : \mathbf{a}_j^\top \mathbf{z} \geqslant s b_j\}$$

and let $\boldsymbol{\gamma}$ be the *unique solution* of the following linearly constrained problem:

$$\boldsymbol{\gamma} \triangleq \begin{cases} \arg\min\{\|\mathbf{z}\| : \mathbf{z} \in G\} & \text{if } G \neq \varnothing, \\ (\infty, \ldots, \infty)^\top & \text{if } G = \varnothing. \end{cases} \tag{20}$$

The following large deviations result in GKS (2007) provides a lower bound on the asymptotic efficiency of any IS estimator.

THEOREM 3. *If assumptions* **M1** *and* **M3** *are satisfied, then*

$$\lim_{m\to\infty} \frac{1}{\log m} \log \mathsf{P}(L_m > x_m) = -\frac{1}{2}\|\boldsymbol{\gamma}\|^2.$$

Note the difference between the denominators in Theorem 1 and Theorem 3. The probability in Theorem 1 decreases exponentially fast in m, while the one in Theorem 3 decreases at a polynomial rate. To guarantee the existence of meaningful factor shifting, we assume, as in §3.1, that:

ASSUMPTION 2. $\|\boldsymbol{\gamma}\| < \infty$, *or equivalently,* $G \neq \varnothing$.

Note that the vector $\boldsymbol{\gamma}$ depends on s, b_j, and the factor-loading vectors.

We use the symbol $\Lambda(\cdot)$ to denote the cumulant generating function of U_k as well as ℓ_k. Because we use U_k only in this section, there should be no confusion. Denote the cumulant generating function of U_k of type j (whose mean $u_k = u_j^*$) by

$$\Lambda_j(\lambda) \triangleq \log \mathsf{E}[e^{\lambda U_k}]. \tag{21}$$

Thus, $\mathsf{E}[e^{\theta \ell_k}] = \Lambda_j(\theta l_k)$. $\Lambda_j(\cdot)$ is twice continuously differentiable on $\mathbb{R}$ because U_k is bounded. (See, e.g., pp. 72–73 of Durrett 1996. $\Lambda_j(\cdot)$ is, in fact, an analytic function.) Also, $\Lambda_j(0) = 0$, $\Lambda_j'(0) = \mathsf{E}[U_k] = u_k = u_j^* > 0$.

We construct a sequence of two-step IS distributions. Because $q_m \to 1$ as $m \to \infty$, $q_m > 1 - (\min_j C_j/C)$ for all sufficiently large m. Then, $\mathcal{S}_{q_m} = \{\{1, \ldots, t\}\}$ for these large m. Hence, to simplify the notation, if we define

$$\boldsymbol{\mu}^{(m)} \triangleq \arg\min\{\|\mathbf{z}\| : \mathbf{z} \in G^{(m)}_{\{1,\ldots,t\}}\} = \boldsymbol{\mu}^{(m)}_{\{1,\ldots,t\}}, \tag{22}$$

then we can completely specify the mixture IS distribution. (In fact, the IS distribution is defined by a single shift under this LLT regime.) The likelihood ratio of the common factor shift is given by $(M_{\mathbf{Z}}^{(m)})^{-1}$, where

$$M_{\mathbf{Z}}^{(m)} = \exp\left(\boldsymbol{\mu}^{(m)\top}\mathbf{Z} - \tfrac{1}{2}\boldsymbol{\mu}^{(m)\top}\boldsymbol{\mu}^{(m)}\right).$$

From the definition, the shifting of the factor mean to $\boldsymbol{\mu}^{(m)}$ increases the default chance of every obligor. The conditional IS part is given by (8), where $\psi_m(\theta, \mathbf{z})$ is

$$\psi_m(\theta, \mathbf{z}) = \frac{1}{m}\sum_{j=1}^{t}\sum_{k\in\mathcal{I}_j^{(m)}} \log(1 + p_k(\mathbf{z})(e^{\Lambda_j(\theta l_k)} - 1)). \tag{23}$$

Then, the combined likelihood ratio is given by

$$\frac{d\mathsf{P}}{d\mathsf{P}_m} = e^{-\theta_m(\mathbf{Z})L_m + m\psi_m(\theta_m(\mathbf{Z}),\mathbf{Z})}\left(M_{\mathbf{Z}}^{(m)}\right)^{-1}.$$

Under the probability measure P_m (i.e., under the IS distribution), $\mathbf{Z}$ is distributed by $N(\boldsymbol{\mu}^{(m)}, \mathbf{I}_d)$, Y_k is a default indicator with a conditional default probability $p_{k,\theta_m(\mathbf{Z})}$, and ℓ_k $(= l_k U_k)$ is sampled from the exponentially tilted distribution given by (5).

REMARKS. The valid range of x is $\frac{1}{2}\sum_{k=1}^m l_k u_k < x < \Phi(\sqrt{\log m})\sum_{k=1}^m l_k u_k$ from the restriction of $0 < s < 1$. However, we apply **MIS** to instances having thresholds less than $\frac{1}{2}\sum_{k=1}^m l_k u_k$ by allowing negative s values when we solve (22). If the threshold value x is small enough, then s will be a largeative value so that $\mathbf{0} \in G^{(m)}$ and $\boldsymbol{\mu}^{(m)} = \mathbf{0}$. This coincides with the intuition that we do not need to shift the common factors if x is small.

The following theorem establishes the asymptotic optimality of the IS estimator of $\mathsf{P}(L_m > x_m)$. The proof is in the online appendix.

THEOREM 4. *Suppose that assumptions* **M1** *and* **M3** *hold and we apply* **MIS** *with* $\alpha_2^{(m)} = 1 - (1/\sqrt{\log m})$ *and* $\alpha_1^{(m)} \to 1$ *as* $m \to \infty$. *If* $G \neq \varnothing$, *then*

$$\limsup_{m\to\infty} \frac{1}{\log m} \log M_2(x_m, \theta_m(\mathbf{Z})) \leqslant -\|\boldsymbol{\gamma}\|^2.$$

Hence,

$$2\lim_{m\to\infty}\frac{1}{\log m}\log \mathsf{P}(L_m > x_m) = \lim_{m\to\infty}\frac{1}{\log m}\log M_2(x_m, \theta_m(\mathbf{Z})),$$

which proves the asymptotic optimality of the two-step IS estimator obtained by **MIS**.

3.3. Choices of $\alpha_1^{(m)}$ and $\alpha_2^{(m)}$

Finally, note that **MIS** satisfies both Theorems 2 and 4 if we set $\alpha_1^{(m)} = 1 - \epsilon_m$ and $\alpha_2^{(m)} = 1 - (1/\sqrt{\log m})$, where ϵ_m is defined as in Theorem 2. However, we still have some flexibility in choosing ϵ_m. We use one of the following two parameterizations of (14) in our numerical examples:

$$\alpha_1^{(m)} = 1 - m^{-1/3} \quad \text{and} \quad \alpha_2^{(m)} = 1 - \frac{1}{\sqrt{\log m}} \tag{24}$$

or

$$\alpha_1^{(m)} = 1 - m^{-1/\beta(\mathbf{b})} \quad \text{and} \quad \alpha_2^{(m)} = 0, \tag{25}$$

where $\beta(\mathbf{b}) \triangleq \max\{3, 15 \cdot (1/t)\sum_{j=1}^{t} b_j\}$. The dependence on the b_js for (25) prevents $\boldsymbol{\mu}_{\mathcal{J}}^{(m)}$ from increasing too much for the types having small $\|\mathbf{a}_j\|$ values. (24) can be used for both limiting regimes, while (25) guarantees the asymptotic optimality only for the small default probability regime.

Even though both parameterizations are asymptotically optimal, note the different advantages of two parameterizations: (24) utilizes more information on q; by using (25), we can exploit the independence from q values, which reduces the amount of computation if we estimate the credit risk at several q levels or if we estimate newly structured portfolios under the same dependence structure (see §5.1). However, we do not have a general rule for how best to choose them. The structure of dependence, the range of marginal default probabilities, and the loss threshold affect the performance of these parameterizations. We restrict our choices to the two given above. For any class of problems, we try both parameterizations and choose the better one.

4. The Necessity of Mixture Distributions for Factor Shifting: A Simple Example

To illustrate the importance of using a mixture of factor shifts, we consider a simple, if rather extreme, example. We consider the following Gaussian copula model with 1,000 obligors, two factors, two types, and nonnegative, orthogonal factor-loading coefficient vectors:

$$\begin{aligned} X_{2k-1} &= 0.7Z_1 + \sqrt{0.51}\varepsilon_{2k-1}, \\ X_{2k} &= 0.65Z_2 + \sqrt{0.5775}\varepsilon_{2k}, \end{aligned} \tag{26}$$

for $k = 1, \ldots, 1{,}000$. Here, Z_1, Z_2, and the ε_ks are independent $N(0,1)$ random variables. We assume that the marginal default probabilities and potential loss amounts are at the same level of $p_k \equiv 5\%$ and $\ell_k \equiv 1$. We restrict the dimension of common factors to two to make the visualization easy, and we choose orthogonal factor loadings to exploit the effects of multiple factors. Furthermore, we perturb the data to make the example asymmetric because some algorithms based on a single-factor model may have simple remedies for symmetric multifactor problems.

We are interested in the computation of the probability of 30% loss of the total credit exposure:

$$\mathrm{P}\left(\sum_{1\leqslant k\leqslant 1{,}000} \mathbf{1}\{X_k > \Phi^{-1}(0.95)\} > 0.3 \cdot 1{,}000\right),$$

i.e., $q = 0.3$ in the notation of §2. Note that $q < 0.5 = C_1 = C_2$ and recall (14). We get $G_1^{(m)} = \{(z_1, z_2) \in \mathbb{R}^2 : z_1 \geqslant 1.7834\}$ and $G_2^{(m)} = \{(z_1, z_2) \in \mathbb{R}^2 : z_2 \geqslant 1.8977\}$, which are computed by (14) with (24). In this case, $\mathcal{S}_q = \{\{1\}, \{2\}\}$, $\boldsymbol{\mu}_{\{1\}}^{(m)} = (1.7834, 0)$, and $\boldsymbol{\mu}_{\{2\}}^{(m)} = (0, 1.8977)$. **MIS** suggests a mixture of $\boldsymbol{\mu}_{\{1\}}^{(m)}$ and $\boldsymbol{\mu}_{\{2\}}^{(m)}$.

As a candidate for a single mean shift, we choose a point minimizing the upper bound (10) of a zero-variance IS distribution as done in Glasserman and Li (2005). Figures 4 and 6 show the level curves of $F_x(\mathbf{z})$ and $-\log(-F_x(\mathbf{z}) + \frac{1}{2}\mathbf{z}^\top\mathbf{z})$, respectively, where $F_x(\mathbf{z}) = -\theta_m(\mathbf{z})x + m\psi_m(\theta_m(\mathbf{z}), \mathbf{z})$. Figures 5 and 7 are their 3D counterparts. (Figure 6 is drawn to identify locations of local maxima. To magnify the shape of the function near maximum, we apply the logarithmic transformation.) Because of the small difference in the factor-loading coefficients, there exists a unique maximum of (10). It is given by $\boldsymbol{\mu} = (2.5051, 0.4343)$. Note that, from these figures and Figure 8 as well, we observe that there is no tangent plane at $\boldsymbol{\mu}$ dominating $F_x(\mathbf{z})$, which is a key condition to get a good bound on second moments of IS estimators in Glasserman and Li (2005). Furthermore, if $\exp(-F_x(\mathbf{z}) + \frac{1}{2}\mathbf{z}^\top\mathbf{z})$ is used as a proxy of the zero-variance IS distribution, any heuristic mixture approach using all modal points of the proxy suggests the same single shift as above because there is no other local maximum except the global maximum point. In Figures 4 and 6, the differences in heights between any two neighboring level curves are equal. Along any radial direction from the origin, $-\log(-F_x(\mathbf{z}) + \frac{1}{2}\mathbf{z}^\top\mathbf{z})$ in Figure 6 increases initially and then decreases although the downward slope is

Figure 4. Level curves of $F_x(\mathbf{z})$.

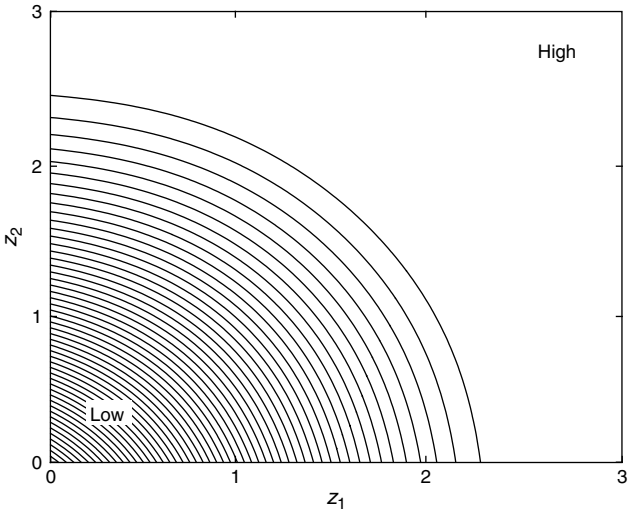


Figure 5. 3D plot of $F_x(\mathbf{z})$.

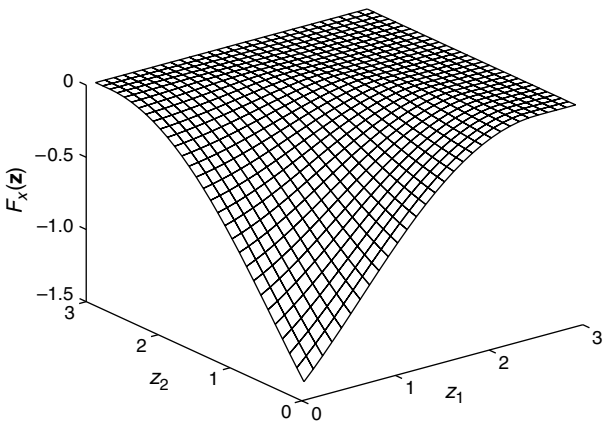
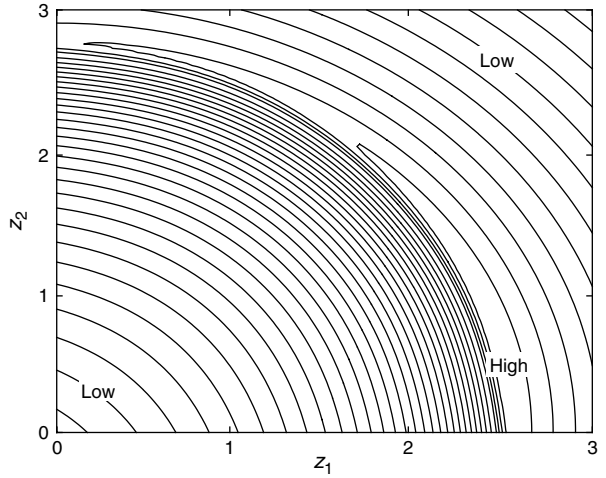


Figure 6. Level curves of $-\log(-F_x(\mathbf{z}) + \frac{1}{2}\mathbf{z}^\top\mathbf{z})$.



Notes. The unique maximum points are located near $(2.5, 0.5)$. There is no other local maximum point.

very small (see also Figure 7). Therefore, a unique maximum exists along each radial direction. Furthermore, the profile curve in Figure 8 shows that there is a unique maximum of $-\log(-F_x(\mathbf{z}) + \frac{1}{2}\mathbf{z}^\top\mathbf{z})$ among the radial maxima, given by $\boldsymbol{\mu}$ mentioned above. Figures 4 and 5 also show that the surface of $F_x(\mathbf{z})$ resembles a quarter of an *inverted bell*. Hence, the tangent plane of $F_x(\mathbf{z})$ at $\boldsymbol{\mu}$ does not dominate a part of $F_x(\mathbf{z})$ near $(0, 2)$.

We compare three methods, all of which provide unbiased estimators:

MIS: Shift the factor mean to $\boldsymbol{\mu}_{\{1\}}^{(m)}$ and $\boldsymbol{\mu}_{\{2\}}^{(m)}$ with equal chance.

MC1: Two-step IS algorithm of Glasserman and Li (2005) with a single mean shifting to $\boldsymbol{\mu}$.

CMC: Crude Monte Carlo simulation.

Figure 9 summarizes the results. The variance of **MC1** does not stabilize as the number of replications increases. This observation can be explained by the following reasoning: if we apply **MC1**, then (i) the obligors with $(0, 0.65)$-factor loadings hardly default; (ii) when any obligor with

Figure 7. 3D plot of $-\log_{10}(-F_x(\mathbf{z}) + \frac{1}{2}\mathbf{z}^\top\mathbf{z})$.

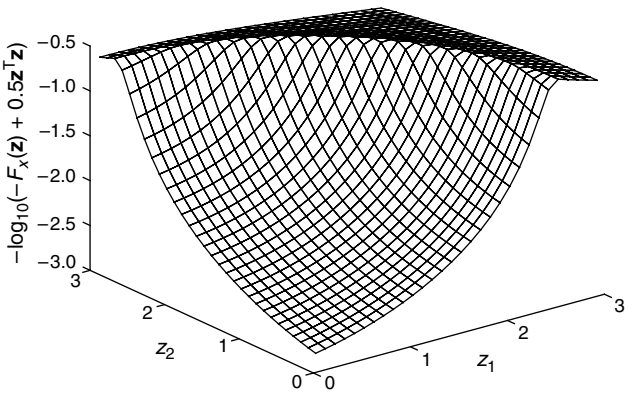
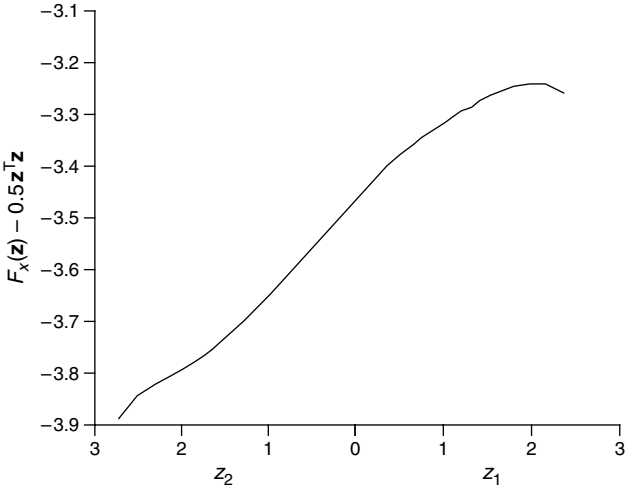


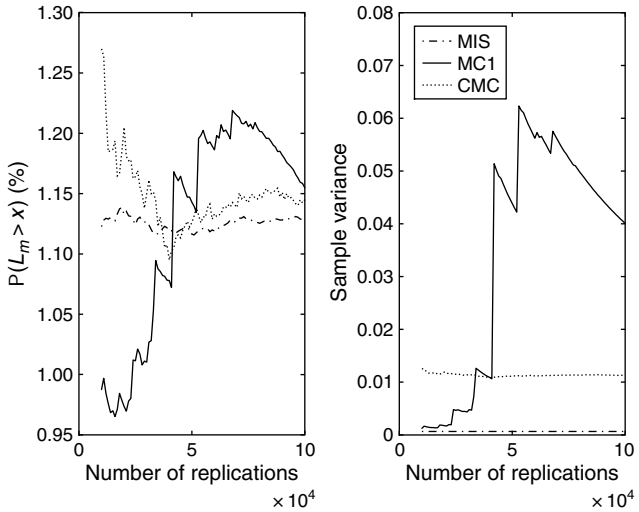
Figure 8. A profile curve of $F_x(\mathbf{z}) - \frac{1}{2}\mathbf{z}^\top\mathbf{z}$ from a view-direction $(1, 1, 0)$.



Note. This profile combined with Figure 6 shows the uniqueness of the maximum point.

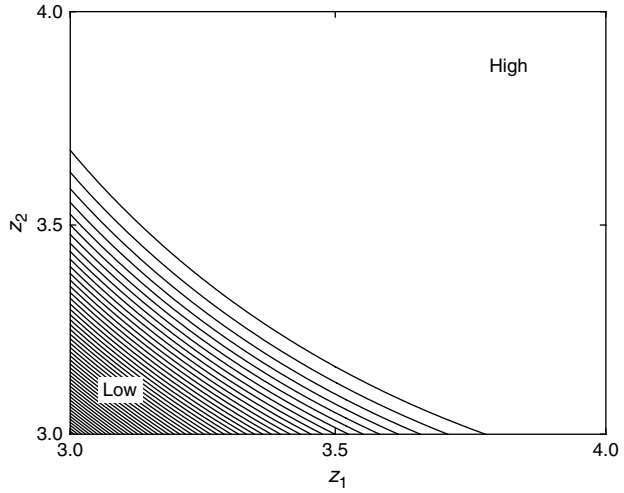
$(0, 0.65)$-factor loadings does default (which rarely happens), it adds a large term to the estimates of both the probability and the variance. This explains the sudden jumps in the sample variance and the estimate of the default probability as the number of replications increases; the jumps are caused by the rare-events large likelihood ratio values. In general, rare events with large likelihood ratio values are rarely generated in the actual simulation runs. Therefore, the theoretically large variance is prone to being underestimated in the simulation. By the same reasoning, estimates of the probability will tend to be too low without a very large number of replications.

Figure 9. Accumulated estimates of portfolio default probability and sample variance from a single run of 10^5 replications for three methods, with $q = 0.3$.



Notes. We use common random numbers across the three methods for comparison. The average sample variance of **MIS** is 6.5×10^{-4}.

Figure 10. Level curves of $F_x(\mathbf{z})$.



This example shows the importance of mixture IS distributions characterized in §3. Similar results were observed in Glasserman and Wang (1997) for the sample mean of i.i.d. random variables. Dupuis and Wang (2004) develop an adaptive IS method to address this type of problem.

For large q (in the sense of $q > 1 - (\min_j C_j/C)$), it seems that the tangent plane at $\boldsymbol{\mu}$ *almost* dominates $F_x(\mathbf{z})$ and a single shift of common factor mean could work well. We include Figures 10–13 to show the dramatic changes of the function shapes as q increases to 0.8.

In this case, using (14) and (24), $G_1^{(m)} = \{(z_1, z_2) \in \mathbb{R}^2\colon z_1 \geqslant 2.6467\}$ and $G_2^{(m)} = \{(z_1, z_2) \in \mathbb{R}^2\colon z_2 \geqslant 2.8871\}$. Also, $\mathcal{S}_q = \{\{1, 2\}\}$ and $G_{\{1,2\}}^{(m)} = \{(z_1, z_2) \in \mathbb{R}^2\colon z_1 \geqslant 2.6467, z_2 \geqslant 2.8871\}$. So, $\boldsymbol{\mu}_{\{1,2\}}^{(m)} = \arg\min_{\mathbf{z} \in G_{\{1,2\}}^{(m)}} \|\mathbf{z}\| = (2.6467, 2.8871)$ by (17). **MIS** now involves a single mean shift to $\boldsymbol{\mu}_{\{1,2\}}^{(m)}$. **MC1** shifts the mean of common factors to (3.3030, 3.3838). Figure 14 summarizes the results. As expected, the two methods, **MC1** and **MIS**, are well behaved and **MC1** produces smaller variance. Crude Monte Carlo could not generate any sample whose loss exceeds 80% of the total credit exposure.

Figure 11. 3D plot of $F_x(\mathbf{z})$.

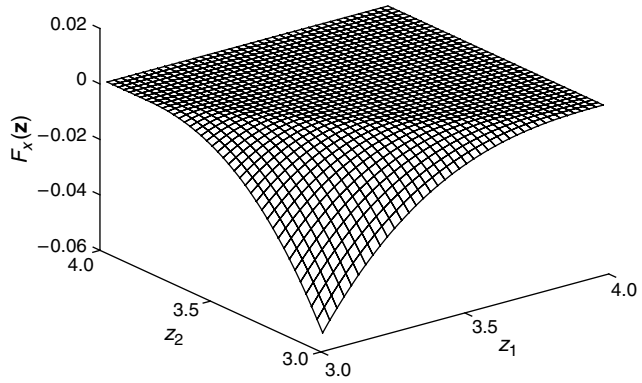
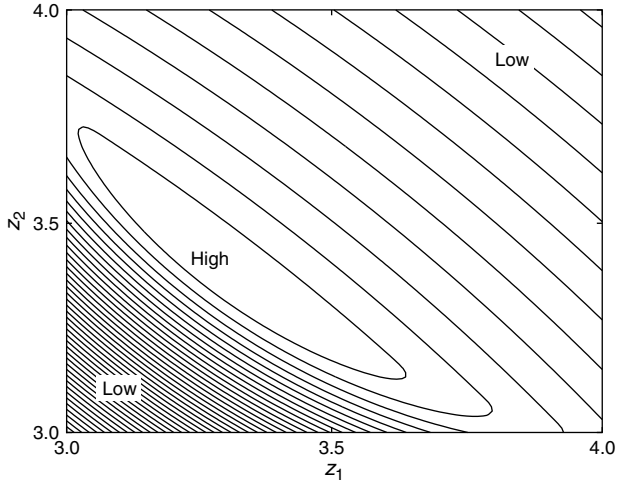


Figure 12. Level curves of $-\log(-F_x(\mathbf{z}) + \frac{1}{2}\mathbf{z}^\top\mathbf{z})$.



Note. The unique maximum point is (3.3030, 3.3838).

5. Computational Issues in the Implementation

Choosing the factor mean vectors involves optimization procedures. In this section, we address algorithms used to solve these problems. Furthermore, we focus on the enumeration of the set $\{\boldsymbol{\mu}_{\mathcal{J}}^{(m)}\colon \mathcal{J} \in \mathcal{S}_q\}$ instead of $\mathcal{S}_q$; the latter possibly consists of exponentially many elements. The issue here is how to find the candidate IS distributions as fast as possible when the number of types, t, and the dimension of factors, d, are fixed. We also characterize the single-factor case completely. In this section, we use the notation $d_j = \alpha_1^{(m)}\Phi^{-1}(1 - \bar{p}_j) + \alpha_2^{(m)} b_j \Phi^{-1}(q)$, which is a constant for each j. Recall that $G_j^{(m)} = \{\mathbf{z}\colon \mathbf{a}_j^\top \mathbf{z} \geqslant d_j\}$.

5.1. Reduction of the Number of Candidate Importance Sampling Distributions

When implementing the IS algorithm, finding $\mathcal{S}_q$ could take a long time or $|\mathcal{S}_q|$ could be prohibitively large to run the simulation efficiently. For a given problem instance, the tractability (the size of $\mathcal{S}_q$) depends on the value q. Furthermore, using a simple enumeration, a small q such

Figure 13. 3D plot of $-\log_{10}(-F_x(\mathbf{z}) + \frac{1}{2}\mathbf{z}^\top\mathbf{z})$.

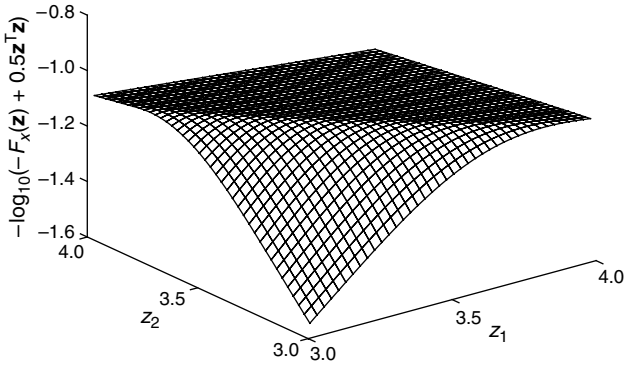
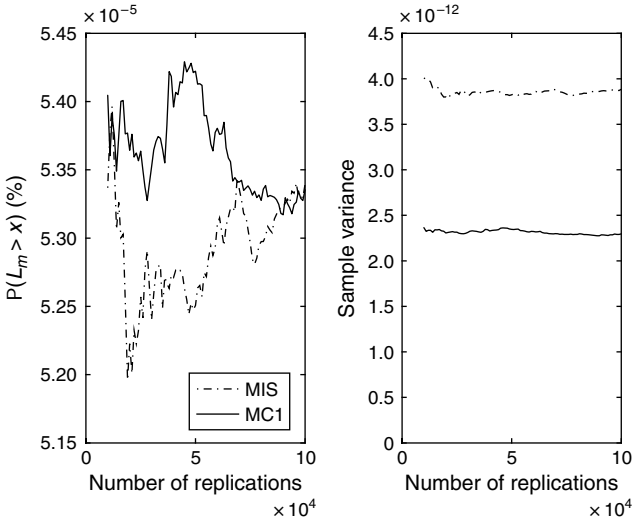


Figure 14. Accumulated estimates of portfolio default probability and sample variance from a single run of 10^5 replications for two methods, with $q = 0.8$.



Note. We use common random numbers across the two methods for comparison.

as $q < 2 \cdot \min_j C_j/C$ allows an efficient implementation because $|\mathcal{S}_q| \leqslant \binom{t}{2}$. The other easy cases are for large q values such as $q > 1 - 2 \cdot \min_j C_j/C$, which has $|\mathcal{S}_q| \leqslant t$. However, in the worst case, the size of $\mathcal{S}_q$ will be $\binom{t}{[t/2]}$, where the application of **MIS** is intractable for instances with a large number of types. (This worst case can be obtained by setting all C_js to be the same, $\mathbf{a}_j > \mathbf{0}$, and q just above 0.5.) To avoid this difficulty, we need to devise a method that does not involve an enumeration of the index sets in $\mathcal{S}_q$. The key observation is the following lemma, proved in the online appendix.

LEMMA 1. *For any $\mathcal{J} \in \mathcal{S}_q$ satisfying $G_{\mathcal{J}} \neq \varnothing$, there exists a $\mathcal{J}' \subset \mathcal{J}$ with $|\mathcal{J}'| \leqslant d$ such that*

$$\boldsymbol{\mu}_{\mathcal{J}}^{(m)} = \boldsymbol{\mu}_{\mathcal{J}'}^{(m)}.$$

From this lemma, we have the following upper bound:

LEMMA 2. *For an instance with d factors and t ($\geqslant d$) types,*

$$\left|\left\{\boldsymbol{\mu}_{\mathcal{J}}^{(m)}\colon\ \mathcal{J} \in \mathcal{S}_q\right\}\right| \leqslant \binom{t}{d} + \binom{t}{d-1} + \cdots + t < t^d$$

holds for all m.

PROOF. The right side of the inequality is the number of ways choosing d or fewer constraints out of t possible constraints. Combining this with Lemma 1, we complete the proof. □

Define $\mathcal{V} \triangleq \{\boldsymbol{\mu}_{\mathcal{J}}^{(m)}\colon\ \mathcal{J} \subset \{1, \ldots, t\}, |\mathcal{J}| \leqslant d, G_{\mathcal{J}} \neq \varnothing\}$. Clearly, we have $\{\boldsymbol{\mu}_{\mathcal{J}}^{(m)}\colon\ \mathcal{J} \in \mathcal{S}_q\} \subset \mathcal{V}$ by Lemma 1. Note that the upper bound in Lemma 2 is also an upper bound on $|\mathcal{V}|$. By adjusting the associated weights, the polynomially bounded number of mean vectors for the mixture IS distribution may allow an efficient implementation for the factor models described by a moderate number of factors and types.

Our approach will be to first restrict the candidate mean vectors to $\mathcal{V}$, and to then use that set to find $\{\boldsymbol{\mu}_{\mathcal{J}}^{(m)}\colon\ \mathcal{J} \in \mathcal{S}_q\}$. Write $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and define $\mathcal{H}(\mathbf{v}) \triangleq \{j \in \{1, \ldots, t\}\colon\ \mathbf{a}_j^\top \mathbf{v} \geqslant d_j\}$ for $\mathbf{v} \in \mathcal{V}$. $\mathcal{H}(\mathbf{v})$ is the maximal index set satisfying $\mathbf{v} = \boldsymbol{\mu}_{\mathcal{H}(\mathbf{v})}^{(m)}$. For each $\mathbf{v} \in \mathcal{V}$, consider all the minimal constraint sets producing $\mathbf{v}$ as a unique optimal solution; denote this family by $\mathcal{F}(\mathbf{v}) = \{F\colon\ F \subset \mathcal{H}(\mathbf{v}), \mathbf{v} = \boldsymbol{\mu}_F^{(m)}, \mathbf{v} \neq \boldsymbol{\mu}_{F\setminus\{j\}}^{(m)} \text{ for all } j \in F\}$. Note that $|F| \leqslant d$ for each $F \in \mathcal{F}(\mathbf{v})$ by Lemma 1, and hence the cardinality of $\bigcup_{\mathbf{v}\in\mathcal{V}} \mathcal{F}(\mathbf{v})$ has the same upper bound as the one in Lemma 2. Because we search $\mathcal{V}$ by probing all index sets of cardinality less than or equal to d, we get $\mathcal{F}(\mathbf{v})$s as by-products of the search.

For any index set A and B with $B \subset A$, define the following knapsack problem:

(SSP):

$$f^*(\{C_j\}_{j=1}^t, q, A, B) = \max \sum_{j\in A\setminus B} C_j x_j$$
$$\text{s.t.} \sum_{j\in A\setminus B} C_j x_j \leqslant \sum_{j\in A} C_j - qC,$$
$$x_j \in \{0, 1\} \quad \text{for all } j \in A\setminus B.$$

The procedure identifying $\{\boldsymbol{\mu}_{\mathcal{J}}^{(m)}\colon\ \mathcal{J} \in \mathcal{S}_q\}$ is as follows:

Step 1. Identify $\mathcal{V}$ by solving the norm minimization problems (17) associated with all possible combinations of type indices, $\mathcal{J} \subset \{1, \ldots, t\}$, $|\mathcal{J}| \leqslant d$.

Step 2. For each $\mathbf{v} \in \mathcal{V}$,
  For each $F \in \mathcal{F}(\mathbf{v})$,
  - Solve (SSP) with $A = \mathcal{H}(\mathbf{v})$ and $B = F$;
  - If $f^*(\{C_j\}_{j=1}^t, q, \mathcal{H}(\mathbf{v}), F) > \sum_{j\in\mathcal{H}(\mathbf{v})} C_j - qC - \min_{i\in F} C_i$, then include $\mathbf{v}$ in among the mean shift vectors for IS.

  End
End

The validity of this procedure is shown in the online appendix. Additionally, we include a more explicit procedure for the single-factor case in the online appendix. We assume that all C_js are positive integers, which is a necessary assumption for knapsack problems. This integrality assumption would not cause any difficulty in applications because, in practice, exposures in a large portfolio are denominated in multiples of a base amount (e.g., one million dollars) for risk-management purposes.

The knapsack problem (SSP) has a special structure and is called a *subset sum problem*, which is NP-complete, and we can show that (SSP) is also NP-complete (see the online appendix). However, subset sum problems have fully

polynomial time approximation algorithms (running in time $O(\min\{n\cdot 1/\epsilon, n+(1/\epsilon^2)\log(1/\epsilon)\})$ and space $O(n+1/\epsilon)$ to achieve an accuracy ϵ $(0<\epsilon<1)$, where n is the number of variables in the subset sum problem). Furthermore, knapsack problems arising in practice can usually be solved very quickly. (See, e.g., Chapter 4 of Kellerer et al. 2004.) For example, using the code available at http://www.diku.dk/∼pisinger/subsum.c, we measured the time spent to solve 10^6 subset sum problems. Each instance consists of 100 randomly generated weights (i.e., $|A\backslash B| = 100$ in (SSP)) with knapsack weights $1 \leqslant C_j \leqslant 10^4$. All 10^6 problems were solved in 21.88 seconds. (All experiments in this paper were executed using a notebook computer with a 1.7 GHz Intel Pentium M CPU and 512 MB of RAM.) This number of problems, 10^6, is roughly the upper bound of the cardinality of $\mathscr{V}$ for a factor model having 100 types and three factors. In solving a subset sum problem, the range of knapsack weights are most crucial for the running time of the algorithm. The ranges in our test cases imply that the potential loss amount of each obligor will take its value among multiples of up to 10^4 of some base amount. In practice, we expect the relevant ranges to be narrower than this.

If we choose $\alpha_2^{(m)} = 0$ in (14), then $\mathscr{V}$, $\mathscr{H}(\mathbf{v})$, and $\mathscr{F}(\mathbf{v})$ do not depend on the value of q. Hence, they do not change unless the dependence structure changes. This implies that if we keep this information between the changes of the dependence structure, then we just need to solve multiple subset sum problems to find the mixture IS distribution for a newly structured portfolio. Table 1 shows the average cardinalities of $\{\boldsymbol{\mu}_{\mathcal{J}}^{(m)}\colon \mathcal{J}\in\mathscr{S}_q\}$ and $\mathscr{V}$ for 30 randomly generated 20- or 25-type instances with factor spaces in $\mathbb{R}^4$ or $\mathbb{R}^5$ (i.e., factor dimension 4 or 5). (See §6.1 for more detailed specifications of this random generation of instances.) The values of the upper bound on $|\mathscr{V}|$ in Lemma 2 are 6,195, 21,699, 15,275, and 68,405, respectively. However, we just need to keep a smaller size (at most 2,000 on average) of $\mathscr{V}$ to get $\{\boldsymbol{\mu}_{\mathcal{J}}^{(m)}\colon \mathcal{J}\in\mathscr{S}_q\}$. The computing time of $\mathscr{V}$ takes about 28, 100, 65, and 300 seconds for each instance, respectively, if we use the MATLAB function `quadprog` for the norm minimization (17). (By a specialized algorithm in §5.2, the time can be reduced to 0.3, 1, 1, and 6 seconds for each instance, respectively.) The total times in solving nine subset sum problems to find $\{\boldsymbol{\mu}_{\mathcal{J}}^{(m)}\colon \mathcal{J}\in\mathscr{S}_q\}$s for $q = 0.1, 0.2, \ldots, 0.9$ from $\mathscr{V}$ are at most 0.2, 0.5, 0.4, and 1.5 seconds, respectively, and this short solving time makes our approach more attractive. Furthermore, the cardinalities of $\{\boldsymbol{\mu}_{\mathcal{J}}^{(m)}\colon \mathcal{J}\in\mathscr{S}_q\}$ are much smaller than the theoretical upper bound. This fact implies that we can implement the IS efficiently.

5.2. Quadratic Optimization

To implement the IS algorithms, we need to solve (17). These problems are norm minimizations over a polyhedron, $\min\{\|\mathbf{z}\|\colon \mathbf{a}_j^\top\mathbf{z}\geqslant d_j, j\in\mathcal{J}\}$. We can apply general quadratic programming (QP) algorithms to these problems. They have strictly convex quadratic objective functions and linear inequality constraints. These features allow very fast and robust convergence of algorithms when we apply general QP algorithms (see, e.g., Nocedal and Wright 1999). In implementation, we can use any available code (e.g., the MATLAB function `quadprog`). However, we can exploit the hierarchy of QP problems further. That is, we characterize $\mathscr{V}$ by solving a QP for each $\mathcal{J}\subset\{1,\ldots,t\}$, $|\mathcal{J}|\leqslant d$. This strategy allows us to solve $\boldsymbol{\nu}_{\mathcal{J}} = \arg\min\{\|\mathbf{z}\|\colon \mathbf{a}_j^\top\mathbf{z} = d_j \text{ for all } j\in\mathcal{J}\}$ instead of the original inequality constrained QP. This equality constrained problem can be solved by simple Gaussian eliminations.

In general, $\|\boldsymbol{\nu}_{\mathcal{J}}\|\geqslant\|\boldsymbol{\mu}_{\mathcal{J}}\|$, so we have to detect the case of $\|\boldsymbol{\nu}_{\mathcal{J}}\| > \|\boldsymbol{\mu}_{\mathcal{J}}\|$. Consider the following procedure:

Set $L = \varnothing$.
For $i = 1$ to d,
  For all $\mathcal{J}\subset\{1,\ldots,t\}$ of $|\mathcal{J}| = i$,
    • Find $\boldsymbol{\nu}_{\mathcal{J}}$;
    • Check the existence of $\mathcal{J}'\in L$ for which $\mathcal{J}'\subset\mathcal{J}$ and $\boldsymbol{\nu}_{\mathcal{J}'}\leqslant\boldsymbol{\nu}_{\mathcal{J}}$;
    • If no such $\mathcal{J}'$ exists, then add $\mathcal{J}$ to L.
  End
End

Note that there exists a $\mathcal{J}'\subset\mathcal{J}$ such that $\boldsymbol{\nu}_{\mathcal{J}'} = \boldsymbol{\mu}_{\mathcal{J}}$ if $\|\boldsymbol{\nu}_{\mathcal{J}}\| > \|\boldsymbol{\mu}_{\mathcal{J}}\|$. Furthermore, $\boldsymbol{\nu}_{\mathcal{J}'} = \boldsymbol{\mu}_{\mathcal{J}'}$. Because the enumeration is done in increasing order of $|\mathcal{J}|$, $\boldsymbol{\nu}_{\mathcal{J}'}$ will be found in the list L (because $|\mathcal{J}'| < |\mathcal{J}|$). Hence, $\mathcal{J}$ is discarded correctly. This procedures substantially reduces the amount of time required to identify $\mathscr{V}$.

5.3. Approximate Importance Sampling on Common Factors

For an instance with a large number of common factors, **MIS** may be computationally intractable. We now present an approximate approach for handling such cases.

Table 1. The average number of minimum norm points in $\mathbb{R}^d$.

Types	d	Bound	$\lvert\mathscr{V}\rvert$	$n_{0.1}$	$n_{0.2}$	$n_{0.3}$	$n_{0.4}$	$n_{0.5}$	$n_{0.6}$	$n_{0.7}$	$n_{0.8}$	$n_{0.9}$
20	4	6,195	574.6	16.9	36.1	48.5	52.2	44.5	29.6	14.3	3.9	0.2
20	5	21,699	932.2	25.0	57.0	78.8	84.4	69.0	44.2	19.5	4.9	0.4
25	4	15,275	1,224.9	33.5	65.7	90.5	91.7	74.6	44.1	16.0	2.4	0.2
25	5	68,405	2,036.5	39.7	96.3	138.4	157.1	137.7	79.8	28.2	3.1	0.0

Note. n_q denotes the average of $|\{\boldsymbol{\mu}_{\mathcal{J}}^{(m)}\colon \mathcal{J}\in\mathscr{S}_q\}|$.

Until now, we considered the issue of finding the *exact* IS distributions induced by the given t types in $\mathbb{R}^d$. In §5.1, we reformulated this original problem as a problem of identifying $\mathscr{V}$ by additional subset sum problems with negligible computing effort. This uses a crude procedure to find $\mathscr{V}$—enumerating all possible combinations of d or fewer types at a time. The number of such combinations is polynomial of order t^d. For large t and d, this enumeration could be prohibitive. (Note, however, that the characterization of $\mathscr{V}$ is required only once unless the dependence structure changes by setting $\alpha_2^{(m)} = 0$. Hence, even a rather long computation time in this step could be acceptable.) To reduce the computing burden, we have to reduce t or d. To reduce the number of types, one may consider clustering ideas as in Hastie et al. (2001). We focus on the reduction of the number of factors.

Assume that $\{\mathbf{a}_j\}_{j=1}^t \subset \mathbb{R}^D$. We want to reduce the factor dimension from D to d ($<D$). We suggest the use of *Principal Components Analysis* (PCA). By applying PCA (without mean adjusting—i.e. by setting the mean (location) vector null in PCA to find an appropriate basis spanning the best *subspace* rather than the best *affine space*) to $[\mathbf{a}_1 \dots \mathbf{a}_t]^\top$, we can choose the best subspace of $\mathbb{R}^D$ to explain the variations among factor-loading vectors under the restriction of subspace dimension d. Then, using the projections of the factor-loading vectors onto this subspace, we compute $\{\boldsymbol{\mu}'_{\mathcal{J}}\colon \mathcal{J} \in \mathscr{S}'_q\} \subset \mathbb{R}^d$ by solving the convex quadratic optimization (17) and the more tractable subset sum problems in $\mathbb{R}^d$ with t types. Here, we use $\mathscr{S}'_q$ to emphasize that these factor-shifting mean vectors come from the approximation. We can also reduce the number of types by aggregating two types if their projected factor-loading vectors are close to each other because the marginal default probabilities are allowed to vary within a type. Using the orthonormal basis on the subspace constructed by PCA, we can recover $\{\boldsymbol{\mu}_{\mathcal{J}}\colon \mathcal{J} \in \mathscr{S}'_q\} \subset \mathbb{R}^D$ corresponding to $\{\boldsymbol{\mu}'_{\mathcal{J}}\colon \mathcal{J} \in \mathscr{S}'_q\}$.

Because we get a set of factor-shifting mean vectors, we can construct the IS distribution based on these *approximate* mean vectors. Thus, we use $\{\boldsymbol{\mu}_{\mathcal{J}}\colon \mathcal{J} \in \mathscr{S}'_q\}$ to shift the common factors and to compute the likelihood ratios, but we use the *exact* factor loadings, $\{\mathbf{a}_j\}_{j=1}^T$, in the evaluation of latent variables for each obligor. This makes the IS procedure with approximate mean-shifting vectors unbiased. We expect variance reduction because we consider the most important d dimensions of factor loadings.

6. Numerical Examples

We do experiments on two settings of parameters.

- We apply **MIS** to a factor model of small size—25 types and five factors. As the tuning parameters for (14), we use (25), which does not depend on q. The data for the test cases are generated uniformly over specified ranges and the factor-loading matrices are highly dense. This example is not particularly realistic from a financial viewpoint because in practice we would expect the factor loadings to be fairly sparse. However, we test this case as a challenging example to validate the efficiency of our method.
- We apply **MIS** to structured factor models with sparse factor loadings: 100 types and 21 or 22 factors. For these problems, because t and d are too large to apply the original **MIS**, we apply the approximate IS method in §5.3 with the choice of (24) as the parameters for (14). This example seems more realistic than the previous one. The leading columns in the loading matrix can be regarded as marketwide factors, or geographic factors associated with the economic environment in a country or region.

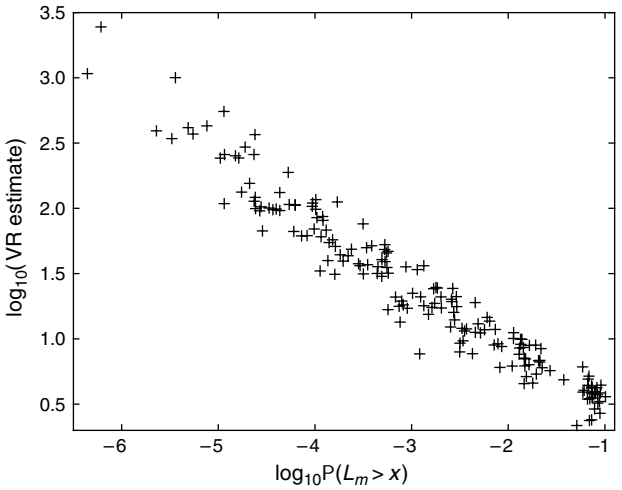
The number of obligors is 1,000 and the number of replications is 10,000 in both settings. We relax the restrictions imposed in **M2**-2 (the same default probabilities within each group) and **M3**-3 (the range of s) in these experiments. For the results, we report two efficiency measures: variance ratio (VR) and efficiency ratio (ER). The VR is measured by dividing the variance of the crude Monte Carlo estimator by that of the **MIS** estimator. Because the sample variance estimate of the crude Monte Carlo estimator is not always stable for large loss events, we replace it with $\hat{p}(1-\hat{p})$, where $\hat{p}$ is the estimate of the probability of the large losses produced by **MIS**. The ER is calculated in a similar way to VR except that each variance estimate is multiplied by the CPU time spent computing the corresponding probability estimate. By including the computing time, ER gives a more applicable measure than VR in general. However, ER depends significantly on particular implementations of the algorithm and can vary widely across different implementations. Hence, we also consider VR a meaningful measure of the performance of the IS estimators.

6.1. MIS for Instances with a Small Number of Factors

We tested 30 randomly generated instances of 1,000 obligors belonging to one of 25 types. 60% of the coefficients of the factor-loading vectors are nonzero. Each nonzero component is drawn uniformly from $[-0.2, 1]$. They are then scaled so that $\|\mathbf{a}_j\|$, $j = 1, \dots, t$, are distributed uniformly on $[0.1, 0.7]$. The potential maximum loss amount of each obligor is deterministic and chosen from a discrete uniform distribution on $\{1, 2, \dots, 30\}$. The marginal default probability for the kth obligor is $0.0255 + 0.0245 \times \sin(16\pi k/m)$, so that it lies within $(0.1\%, 5\%)$. We also randomize the number of obligors in each type such that they are uniformly distributed, and so that the number of obligors in one type does not exceed 150% of that of any other type. Because, even for $q = 0.2$, the probability that $L_m > x$ vanishes to 10^{-7} or 10^{-8}, we simulate the portfolio losses for $q = 0.05, 0.075, \dots, 0.175$.

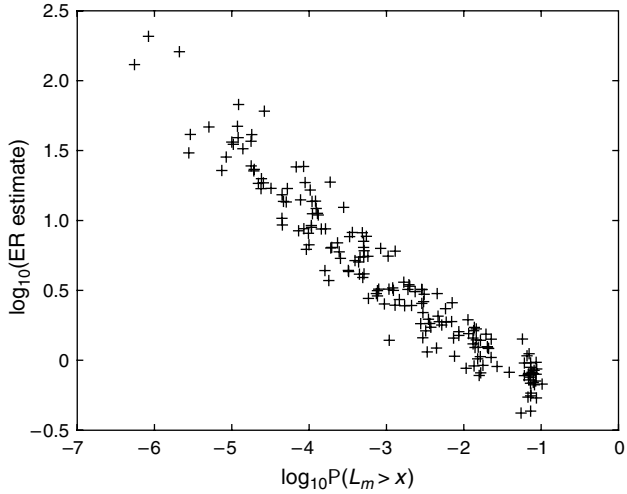
Figure 15 depicts the observed pairs—the portfolio loss probability and the estimated VRs—for each parameter on a logarithmic scale. As the theory predicts, we observe that larger VRs are achieved by **MIS** as the defaults become rarer in Figure 15. In Figure 16, we observe that the ERs

Figure 15. Each point shows a portfolio loss probability and corresponding variance ratio on log scales.



Note. The average number of mean vectors for each loss level is 17.9, 26.7, 39.7, 52.1, 68.3, and 79.4, respectively.

Figure 16. Each point shows a portfolio loss probability and corresponding efficiency ratio on log scales.



are smaller than the VRs because of the additional computations required for IS. However, we again see a trend of bigger improvement for smaller probabilities.

6.2. MIS for Structured Factor Models with Sparse Factor Loadings

This example is generalized from one given by Glasserman and Li (2005). They tested one 21-factor case among eight cases in this extended model. The parameters are given by

$$p_k = 0.01 \cdot (1 + \sin(16\pi k/1{,}000)), \quad k = 1, \ldots, 1{,}000,$$

$$\ell_k = 1 + \frac{99}{999}(k-1), \quad k = 1, \ldots, 1{,}000,$$

and the 100 types have the factor loadings given by the rows of A,

$$A = \left(\begin{array}{c|ccc|c} & F & & & G \\ R & & \ddots & & \vdots \\ & & & F & G \end{array}\right), \quad G = \begin{pmatrix} c_G & & \\ & \ddots & \\ & & c_G \end{pmatrix}.$$

Here, F is a column vector of 10 entries, all equal to c_F, and G is a 10×10 matrix. In the 21-factor model, R is a column vector of 100 entries, all equal to c_R; in the 22-factor model, R is a 100×2 matrix in which the first 50 entries of the first column and the last 50 entries of the second column are equal to c_R, and all other entries are zero. The 21-factor model thus has a single "market factor" (affecting all obligors) and the 22-factor model has two orthogonal "market factors." Consecutive groups of 10 obligors (1–10, 11–20, ...) are of the same type. We consider different factor loadings: $(c_R, c_F, c_G) = (0.8, 0.4, 0.4)$ is associated with a large market factor case, $(c_R, c_F, c_G) = (0.5, 0.4, 0.4)$ with a medium market factor, and $(c_R, c_F, c_G) = (0.2, 0.4, 0.4)$ with a small market factor. We also consider $(c_R, c_F, c_G) = (0.25, 0.15, 0.05)$, suggested by Morokoff (2004). Note that all these models satisfy $\mathbf{a}_j \geqslant 0$ for all j, which implies that the intersection of all the half-spaces associated with obligor types is nonempty. Hence, these models satisfy Assumption 2.

To apply approximate IS, we apply PCA to $A^\top A$. We measure the effectiveness of the PCA approximation to the full matrix through the usual PCA measure of explained variability. These values are summarized in Table 2.

Approximate IS works well for both models. Tables 3–6 summarize the VR and ER estimates for the 21-factor model, and Tables 7–10 for the 22-factor model. It seems

Table 2. The explained variability (ratio of squared variations) applying PCA to the factor-loading coefficients.

	(c_R, c_F, c_G)			
	(0.8, 0.4, 0.4) (%)	(0.5, 0.4, 0.4) (%)	(0.2, 0.4, 0.4) (%)	(0.25, 0.15, 0.05) (%)
Single dominating factor in $\mathbb{R}^{21}$	79	60	25	74
Two dominating factors in $\mathbb{R}^{22}$	80	64	31	77

Table 3. VR and ER estimates using approximate IS at various loss levels in the 21-factor model with 100 types.

Loss $x(q)$	$P(L_m > x)$	VR est.	ER est.
10,000 (20%)	0.0116	25	5
15,000 (30%)	0.0053	44	9
20,000 (40%)	0.0027	74	14
25,000 (50%)	0.0013	126	23
30,000 (60%)	0.0006	223	39
35,000 (70%)	0.0002	443	74
40,000 (80%)	0.0001	1,043	167

Note. The factor-loading coefficients are (0.8, 0.4, 0.4).

Table 4. VR and ER estimates using approximate IS at various loss levels in the 21-factor model with 100 types.

Loss $x(q)$	$P(L_m > x)$	VR est.	ER est.
5,000 (10%)	0.0084	34	8
7,500 (15%)	0.0025	88	19
10,000 (20%)	0.0008	217	44
12,500 (25%)	0.0003	494	97
15,000 (30%)	0.0001	1,133	213

Note. The factor-loading coefficients are (0.5, 0.4, 0.4).

Table 5. VR and ER estimates using approximate IS at various loss levels in the 21-factor model with 100 types.

Loss $x(q)$	$P(L_m > x)$	VR est.	ER est.
2,000 (4%)	0.0132	3	1
2,500 (5%)	0.0042	8	2
3,000 (6%)	0.0016	18	5
3,500 (7%)	0.0006	48	14
4,000 (8%)	0.0002	166	44
4,500 (9%)	0.0001	74	19

Note. The factor-loading coefficients are (0.2, 0.4, 0.4).

Table 6. VR and ER estimates using approximate IS at various loss levels in the 21-factor model with 100 types.

Loss $x(q)$	$P(L_m > x)$	VR est.	ER est.
1,000 (2%)	0.0941	3	2
1,500 (3%)	0.0237	12	5
2,000 (4%)	0.0061	45	17
2,500 (5%)	0.0018	145	48
3,000 (6%)	0.0005	444	136
3,500 (7%)	0.0002	1,390	397

Note. The factor-loading coefficients are (0.25, 0.15, 0.05).

Table 7. VR and ER estimates using approximate IS at various loss levels in the 22-factor model with 100 types.

Loss $x(q)$	$P(L_m > x)$	VR est.	ER est.
10,000 (20%)	0.0077	16	3
15,000 (30%)	0.0030	61	11
20,000 (40%)	0.0012	118	21
25,000 (50%)	0.0004	231	39
30,000 (60%)	0.0001	600	95

Note. The factor-loading coefficients are (0.8, 0.4, 0.4).

Table 8. VR and ER estimates using approximate IS at various loss levels in the 22-factor model with 100 types.

Loss $x(q)$	$P(L_m > x)$	VR est.	ER est.
5,000 (10%)	0.0050	24	4
7,500 (15%)	0.0011	76	13
10,000 (20%)	0.0003	223	36
12,500 (25%)	0.0001	612	96

Note. The factor-loading coefficients are (0.5, 0.4, 0.4).

Table 9. VR and ER estimates using approximate IS at various loss levels in the 22-factor model with 100 types.

Loss $x(q)$	$P(L_m > x)$	VR est.	ER est.
2,000 (4%)	0.0095	6	1
2,500 (5%)	0.0030	16	3
3,000 (6%)	0.0010	35	7
3,500 (7%)	0.0004	70	13
4,000 (8%)	0.0001	290	52

Note. The factor-loading coefficients are (0.2, 0.4, 0.4).

Table 10. VR and ER estimates using approximate IS at various loss levels in the 22-factor model with 100 types.

Loss $x(q)$	$P(L_m > x)$	VR est.	ER est.
1,000 (2%)	0.0748	3	1
1,500 (3%)	0.0134	13	3
2,000 (4%)	0.0027	53	11
2,500 (5%)	0.0006	214	41
3,000 (6%)	0.0001	852	158

Note. The factor-loading coefficients are (0.25, 0.15, 0.05).

that the relatively small improvements for $(0.2, 0.4, 0.4)$ come from the small ratios of explained squared variations of factor-loading coefficients in Table 2. This may also explain the lack of monotonicity of VR in Table 5.

7. Concluding Remarks

We have proposed an importance sampling procedure for the estimation of portfolio credit risk in the multifactor Gaussian copula model. Our procedure shifts the mean in the underlying factor structure to increase the probability of large losses. We established the asymptotic optimality of the procedure under two limiting parameter regimes, assuming a finite number of types of obligors. For practical implementation, we considered two relevant optimization problems for choosing the mean shifts used for the common factors. We also reported numerical examples showing the variance reductions and efficiency improvements achieved by applying the IS method. We developed approximations to facilitate the application of the procedure to problems with a large number of types and common factors, for which the exact procedure may be computationally intractable.

While the focus of this paper is on a specific problem in the measurement of portfolio credit risk, the main issue we address also arises in other problems of rare-event simulation. The key challenge in the multifactor setting we consider is that there are multiple ways the rare event of interest (large portfolio losses) can occur. In this type of setting, an effective IS procedure often requires using a mixture of distributions, with each component of the mixture associated with one of the "ways" the rare event occurs. Implementing this general approach in specific settings requires identifying a potentially large number of such paths to a rare event, and then characterizing which of these are sufficiently important to get extra weight under the IS distribution. The techniques developed in this paper to address these two problems in the credit-risk setting are potentially applicable to other problems as well.

8. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://or.journal.informs.org/.

Acknowledgments

The first two authors dedicate this work to the memory of Perwez Shahabuddin, who died after the original submission of the paper. They thank the reviewers for their careful reading of the paper and many constructive comments. The second author thanks Kyungsik Lee for a discussion on the subset sum problem. This research was partially supported by National Science Foundation grant DMI 03-00044.

References

Avramidis, A. N. 2002. Importance sampling for multimodal functions and applications to pricing exotic options. E. Yucesan, C.-H. Chen, J. L. Snowdon, J. M. Charnes, eds. *Proc. 2002 Winter Simulation Conf.*, 1493–1501.

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

Dupuis, P., H. Wang. 2004. Important sampling, large deviations, and differential games. *Stochastics and Stochastics Rep.* **76**(6) 481–508.

Durrett, R. 1996. *Probability: Theory and Examples*, 2nd ed. Duxbury Press, Belmont, CA.

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

Glasserman, P., Y. Wang. 1997. Counterexamples in importance sampling for large deviations probabilities. *Ann. Appl. Probab.* **7**(3) 731–746.

Glasserman, P., W. Kang, P. Shahabuddin. 2007. Large deviations in multifactor portfolio credit risk. *Math. Finance* **17**(3) 345–379.

Gupton, G., C. Finger, M. Bhatia. 1997. *CreditMetrics Technical Document*. J. P. Morgan & Co., New York.

Hastie, T., R. Tibshirani, J. Friedman. 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, New York.

Kalkbrener, M., H. Lotter, L. Overbeck. 2004. Sensible and efficient capital allocation for credit portfolios. *RISK* **17**(1) S19–S24.

Kellerer, H., U. Pferschy, D. Pisinger. 2004. *Knapsack Problems*. Springer-Verlag, Berlin/Heidelberg, Germany.

Merino, S., M. A. Nyfeler. 2004. Applying importance sampling for estimating coherent credit risk contributions. *Quant. Finance* **4** 199–207.

Morokoff, W. J. 2004. An importance sampling method for portfolios of credit risky assets. R. G. Ingalls, M. D. Rossetti, J. S. Smith, B. A. Peters, eds. *Proc. 2004 Winter Simulation Conf.*, 1668–1676.

Nocedal, J., S. J. Wright. 1999. *Numerical Optimization*. Springer-Verlag, New York.

Sadowsky, J. S., J. A. Bucklew. 1990. On large deviations theory and asymptotically efficient Monte Carlo estimation. *IEEE Trans. Inform. Theory* **36**(3) 579–588.