FDIC Center for Financial Research Working Paper

No. 2005-01

Measuring Marginal Risk Contributions in Credit Portfolios

Paul Glasserman

December 2004

Federal Deposit Insurance Corporation • Center for Financial Research

Measuring Marginal Risk Contributions in Credit Portfolios

Paul Glasserman*
403 Uris Hall
Columbia Business School
New York, NY 10027
pg20@columbia.edu

January 2005

FDIC Center for Financial Research Working Paper No. 2005-01

Abstract

We consider the problem of decomposing the credit risk in a portfolio into a sum of risk contributions associated with individual obligors or transactions. For some standard measures of risk — including value-at-risk and expected shortfall — the total risk can be usefully decomposed into a sum of marginal risk contributions from individual obligors. Each marginal risk contribution is the conditional expected loss from that obligor, conditional on a large loss for the full portfolio. We develop methods for calculating or approximating these conditional expectations. Ordinary Monte Carlo estimation is impractical for this problem because the conditional expectations defining the marginal risk contributions are conditioned on rare events. We develop three techniques to address this difficulty. First, we develop importance sampling estimators specifically designed for conditioning on large losses. Next, we use the analysis underlying the importance sampling technique to develop a hybrid method that combines an approximation with Monte Carlo. Finally, we take this approach a step further and develop a rough but fast approximation that dispenses entirely with Monte Carlo. We develop these methods in the Gaussian copula framework and illustrate their performance in multifactor models.

Key words: Portfolio credit risk, value-at-risk, expected shortfall, Monte Carlo methods

JEL Classification: G11; G21; C15

CFR Research Program: risk measurement.

*This work is supported, in part, by the Risk Measurement Research Program of the FDIC Center for Financial Research and by NSF grants DMI0300044 and DMS0410234. I thank Rosalind Bennett and Paul Kupiec of the FDIC for their comments on an early draft of this paper and Jingyi Li for assistance with numerical experiments.

1 Introduction

For purposes of internal risk management, the measurement of portfolio credit risk is often just a first step in a more extensive process of allocating capital to transactions with counterparties. This process requires decomposing the total credit risk in a portfolio into individual risk contributions. Each risk contribution assigns part of the total risk to a particular transaction or counterparty, and these risk contributions are then used to allocate capital.

For a bank following the internal ratings based approach of the Basel Committee's New Capital Accord [2], the credit risk in a portfolio is summarized through its value-at-risk (VaR), a quantile of the loss distribution. Similar considerations apply if risk (or capital adequacy) is measured using standard deviation, expected shortfall, or several other candidate measures of risk. We focus on VaR and expected shortfall.

Many authors (including Denault [5], Garman [6], Kalkbrener [10], Kurth and Tasche [12], and Litterman [15]) have noted that several commonly used risk measures — including VaR and expected shortfall — can be usefully decomposed as a sum of sensitivities. In this decomposition, each risk contribution can be interpreted as a marginal impact on total portfolio risk. The key property underlying this decomposition is positive homogeneity of the risk measure, a property that states that the risk (or capital requirement) of a portfolio scales in proportion to the size of the portfolio.

The marginal risk contributions associated with both VaR and expected shortfall can be represented as conditional expectations of losses on subportfolios, conditioned on events in the tail of the loss distribution for the full portfolio. The rarity of these tail events presents an obstacle to practical calculation of these conditional expectations. Each contribution depends on the probability of a rare event (a default) conditional on an even rarer event (an extreme loss for the portfolio).

To address the practical difficulties of calculating marginal risk contributions, we develop efficient Monte Carlo methods and approximations. The Monte Carlo methods use importance sampling to address difficulties associated with simulating rare events. This part of the paper builds on the method developed in Glasserman and Li [8] for estimating the tail of the loss distribution in credit portfolios. The techniques underlying the importance sampling method also lead to approximations that can be used together with Monte Carlo or as an alternative. Throughout this paper, we work in the Gaussian copula model of credit risk (Gupton, Finger, and Bhatia [9] and Li [13]), widely used both for measuring portfolio credit risk and for pricing credit derivatives.

We review the Gaussian copula model in Section 2 and discuss the representation of marginal risk contributions as conditional expectations in Section 3. Section 4 analyzes the estimation of conditional expectations using ordinary Monte Carlo; we use this as a benchmark. Section 6 develops the importance sampling method based on shifting the factor means in the Gaussian copula and increasing the conditional default probabilities. Section 7 uses related ideas to shrink the factor variance (as well as shifting the factor mean) in the importance sampling procedure. Section 8 develops an asymptotic approximation for marginal risk contributions and a hybrid method that combines Monte Carlo with an approximation. Section 9 takes the approximation a step further and dispenses with Monte Carlo.

2 Portfolio Credit Risk in the Gaussian Copula Model

A key element of any model of portfolio credit risk is a mechanism for capturing dependence among obligors. In this section, we describe the widely used Gaussian copula model for portfolio credit risk. Our description follows that in Glasserman [7].

We focus on the distribution of losses from default over a fixed horizon. The ingredients of this distribution are as follows:

$$\begin{aligned}
m &= \text{number of obligors to which portfolio is exposed;}\\
Y_k &= \text{default indicator for } k\text{th obligor}\\
&= 1 \text{ if } k\text{th obligor defaults, 0 otherwise;}\\
p_k &= \text{marginal probability that } k\text{th obligor defaults;}\\
c_k &= \text{loss given default for the } k\text{th obligor;}\\
X_k &= \text{loss from the } k\text{th obligor} = c_k Y_k;\\
L &= X_1 + \cdots + X_m = \text{total loss from defaults.}
\end{aligned}$$

We assume the marginal default probabilities p_k are known, either from credit ratings or from the market prices of corporate bonds or credit default swaps. We take the c_k to be constants for simplicity, though it would suffice to know the distribution of the kth loss X_k. This would allow for random recovery.

The Gaussian copula model provides a mechanism for specifying dependence among the default indicators $Y_1, \ldots, Y_m$. Dependence is introduced through a multivariate normal vector $(\xi_1, \ldots, \xi_m)$ of latent variables. Each default indicator is represented as

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

with x_k chosen to match the marginal default probability p_k. The threshold x_k is sometimes interpreted as a default boundary of the type arising in the foundational work of Merton [16]. Without loss of generality, we take each ξ_k to have a standard normal distribution and set $x_k = \Phi^{-1}(1-p_k)$, with Φ the cumulative normal distribution. Thus,

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

The correlations among the ξ_k determine the dependence among the Y_k. The underlying correlations are specified through a factor model of the form

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

in which

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

The underlying factors Z_j are sometimes derived from economic variables (industry or regional risk factors, for example).

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

$$p_k(Z) = P(Y_k = 1|Z) = P(\xi_k > x_k|Z) = \Phi\left(\frac{a_k Z + \Phi^{-1}(p_k)}{b_k}\right). \tag{2}$$

3 Risk Measures and Marginal Risk Contributions

The distribution of credit losses in a portfolio (i.e., the distribution of L) is typically summarized through a scalar measure of risk. Two of the most commonly used risk measures are value-at-risk and expected shortfall. The value-at-risk associated with probability $1-\alpha$ (with, e.g., $\alpha = 1\%$) is the quantile

$$VaR_\alpha = \inf\{x : P(L \geq x) \leq \alpha\}.$$

The corresponding expected shortfall is

$$ES_\alpha = \mathsf{E}[L|L \geq VaR_\alpha].$$

Value-at-risk is in more widespread use, but expected shortfall is coherent (in the sense of Artzner et al. [1]) whereas VaR is not. In particular, VaR is not in general subadditive, which means that the sum of the VaRs for two portfolios may be less than the VaR for the combined portfolio.

The calculation of a portfolio risk measure is often followed by a process of allocating the risk to elements of the portfolio based on their marginal contribution to the total risk. This type of decomposition is used for capital allocation and for measuring risk-adjusted performance.

Several authors have shown that the marginal risk contributions associated with VaR and expected shortfall can be represented as conditional expectations. In more detail, consider increasing the exposure to the kth obligor by an amount ϵ. The resulting portfolio loss random variable L^ϵ is related to the original loss L by

$$L^\epsilon = L + \epsilon X_k,$$

where X_k is the loss random variable for the kth obligor. Let VaR_α^ϵ and ES_α^ϵ denote the value-at-risk and expected shortfall for L^ϵ. Then, under appropriate conditions (see, e.g., Kalkbrener, Lotter and Overbeck [11] or Kurth and Tasche [12]) the marginal VaR contribution of the kth obligor is

$$\left.\frac{\partial\, VaR_\alpha^\epsilon}{\partial \epsilon}\right|_{\epsilon=0} = \mathsf{E}[X_k|L = VaR_\alpha], \tag{3}$$

provided $P(L = VaR_\alpha) > 0$. The marginal contribution to expected shortfall is

$$\left.\frac{\partial ES_\alpha^\epsilon}{\partial \epsilon}\right|_{\epsilon=0} = \mathsf{E}[X_k|L \geq VaR_\alpha]. \tag{4}$$

Thus, in both cases, the marginal risk contributions are conditional expectations of the individual loss random variables, conditioned on rare values of the portfolio loss L.

We will not present (or verify) conditions under which (3) and (4) are valid. Instead, we simply note that the expressions on the right do indeed decompose the total risk. In the case of VaR,

$$\sum_{k=1}^m \mathsf{E}[X_k|L = VaR_\alpha] = \mathsf{E}[\sum_{k=1}^m X_k|L = VaR_\alpha] = \mathsf{E}[L|L = VaR_\alpha] = VaR_\alpha;$$

and for expected shortfall,

$$\sum_{k=1}^m \mathsf{E}[X_k|L \geq VaR_\alpha] = \mathsf{E}[\sum_{k=1}^m X_k|L \geq VaR_\alpha] = \mathsf{E}[L|L \geq VaR_\alpha] = ES_\alpha.$$

The calculation of these marginal risk contributions presents practical difficulties precisely because of the rarity of the conditioning events and because the calculation needs to be repeated for every obligor $k = 1, \ldots, m$.

4 Monte Carlo Estimates of Conditional Expectations

Estimation of the marginal risk contributions (3) and (4) by Monte Carlo can be thought of as a two-phase procedure in which one first estimates VaR (and possibly also expected shortfall) and then estimates the risk contributions using the estimated VaR from the first phase in place of the true VaR in the conditional expectations (3) and (4).

For the Gaussian copula model, the first-phase problem is addressed by the importance sampling (IS) procedure developed in Glasserman and Li [8]. That procedure involves shifting the mean of the common factors Z and then increasing the conditional default probabilities given the outcome of Z. In the numerical examples reported in Glasserman and Li [8], IS allows precise estimation of quantiles at probabilities as small as 0.0001 (and even smaller) from as few as 1000 replications. (Other approaches to IS for credit risk are discussed in Kalkbrener et al. [11] and Morokoff [17].)

Here we focus on the second-phase problem of estimating, for a given loss level x, $\mathsf{E}[X_k|L = x]$ and $\mathsf{E}[X_k|L \geq x]$. We can treat the two cases together by considering

$$v_k = \mathsf{E}[X_k|L \in \mathcal{A}]$$

with $\mathcal{A} = \{x\}$ or $\mathcal{A} = [x, \infty)$. When we condition on $L \in \mathcal{A}$, we assume $P(L \in \mathcal{A}) > 0$. If $P(L = x) = 0$, we can replace the condition $L = x$ with $|L - x| < \epsilon$ for some $\epsilon > 0$ large enough to have $P(|L - x| < \epsilon) > 0$.

We first consider the estimation of v_k, $k = 1, \ldots, m$, using ordinary Monte Carlo simulation. Each replication of an ordinary simulation proceeds along the following steps:

1. Generate independent $N(0, 1)$ factor levels $Z_1, \ldots, Z_d$ and set $Z = (Z_1, \ldots, Z_d)^\top$;

2. For each obligor $k = 1, \ldots, m$, generate an independent $N(0, 1)$ variable ϵ_k and set $\xi_k = a_{k1}Z_1 + \cdots + a_{kd}Z_d + b_k\epsilon_k$;

3. For each obligor $k = 1, \ldots, m$, generate the default indicator $Y_k = \mathbf{1}\{\xi_k > \Phi^{-1}(1 - p_k)\}$ and the loss given default c_k. Set $X_k = Y_k c_k$.

These steps are repeated to generate multiple independent replications.

Let $(X_1^{(i)}, \ldots, X_m^{(i)})$, $i = 1, \ldots, n$, be n i.i.d. replications of the vector of individual losses X_k and let

$$L^{(i)} = X_1^{(i)} + \cdots + X_m^{(i)}$$

denote the total portfolio loss on the ith replication. To estimate the risk contributions v_k, $k = 1, \ldots, m$, we use

$$\hat{v}_k = \frac{\sum_{i=1}^n X_k^{(i)} \mathbf{1}\{L^{(i)} \in \mathcal{A}\}}{\sum_{i=1}^n \mathbf{1}\{L^{(i)} \in \mathcal{A}\}}. \tag{5}$$

We take the ratio to be zero whenever the denominator is zero.

Applying the strong law of large numbers to the numerator and denominator in (5), we find that, with probability 1,

$$\hat{v}_k \to v_k \quad k = 1, \ldots, m,$$

again assuming $P(L \in \mathcal{A}) > 0$. To compare the precision of this estimator with the alternatives that we propose, we need an appropriate measure of its variability. This is provided by a central limit theorem and the accompanying confidence intervals. We record these in the following result. Because $\hat{v}_k$ is a ratio estimator, we cannot simply use a sample standard deviation to measure its precision; we instead apply the method in the following proposition.

Proposition 1 *Suppose $P(L \in \mathcal{A}) > 0$ and let*

$$\hat{\sigma}_k^2 = \frac{n \sum_{i=1}^n (X_k^{(i)} - \hat{v}_k)^2 \mathbf{1}\{L^{(i)} \in \mathcal{A}\}}{\left(\sum_{i=1}^n \mathbf{1}\{L^{(i)} \in \mathcal{A}\}\right)^2}, \tag{6}$$

taking the ratio to be zero whenever the denominator is zero. Then the distribution of

$$\frac{\hat{v}_k - v_k}{\hat{\sigma}_k / \sqrt{n}}$$

converges to the standard normal and

$$\hat{v}_k \pm z_{\delta/2} \frac{\hat{\sigma}_k}{\sqrt{n}}$$

is an asymptotically valid $1 - \delta$ confidence interval for v_k, with $\Phi(z_{\delta/2}) = 1 - \delta/2$.

The limiting distribution in the proposition follows from a general result on the asymptotic normality of nonlinear functions of sample means (see, e.g., p.122 of Serfling [18]) combined with the fact that (6) is a consistent estimator of the asymptotic variance of $\hat{v}_k$. In fact, this univariate central limit theorem extends in a straightforward way to a multivariate central limit theorem for the vector $(\hat{v}_1, \ldots, \hat{v}_m)$, which could be used to form simultaneous confidence intervals for multiple estimates.

5 Importance Sampling for Large Loss Levels

The main difficulty in using the standard Monte Carlo estimators $\hat{v}_k$ to calculate marginal risk contributions is that very few replications produce portfolio losses L with $L = x$ or $L \geq x$. In importance sampling (IS), we try to increase the frequency with which rare events occur by changing the distribution from which we sample; we then weight each observation by a likelihood ratio to correct for the change in distribution.

Glasserman and Li [8] develop an importance sampling technique for the estimation of $P(L \geq x)$ (which here appears as the denominator in the estimation of expected shortfall contributions). Their method applies a change of distribution to the vector of factors Z and then changes the conditional default probabilities given Z.

The likelihood ratio that corrects for changing the distribution of Z from $N(0, I)$ to $N(\mu, \Sigma)$, with $|\Sigma| > 0$, is

$$|\Sigma|^{1/2} \exp\left(\tfrac{1}{2}\left[(Z-\mu)^\top \Sigma^{-1}(Z-\mu) - Z^\top Z\right]\right).$$

If we change only the mean of Z (so $\Sigma = I$), this simplifies to

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

The effectiveness of IS depends critically on the choice of μ (and Σ), to which we will return.

5.1 Twisting the Conditional Default Probabilities

Given Z, the conditional default probability of the kth obligor is $p_k(Z)$ in (2). We embed these conditional probabilities in a parametric family, parameterized by a scalar θ, by setting

$$p_k(\theta, Z) = \frac{p_k(Z)e^{\theta c_k}}{1 + p_k(Z)(e^{\theta c_k} - 1)}. \tag{7}$$

These "exponentially twisted" probabilities are monotone increasing in θ: at $\theta = 0$, we recover the original probability $p_k(Z)$, taking $\theta > 0$ increases the conditional default probability, and taking $\theta < 0$ decreases the conditional default probability. There are many ways one might consider increasing or decreasing the conditional default probabilities, but this particular transformation has several important features that make it effective.

A special feature of (7) is the resulting form of the likelihood ratio. The likelihood ratio that corrects for changing the conditional default probability $p_k(Z)$ to some other probability q_k is

$$\left(\frac{p_k(Z)}{q_k}\right)^{Y_k} \left(\frac{1-p_k(Z)}{1-q_k}\right)^{(1-Y_k)}.$$

In the particular case of $q_k = p_k(\theta, Z)$, this becomes

$$\left(\frac{p_k(Z)}{p_k(\theta,Z)}\right)^{Y_k}\left(\frac{1-p_k(Z)}{1-p_k(\theta,Z)}\right)^{(1-Y_k)} = e^{-\theta Y_k c_k}(1+p_k(Z)(e^{\theta c_k}-1)).$$

Because the default indicators $Y_1, \ldots, Y_m$ are conditionally independent given Z, the (conditional) likelihood ratio for changing all the default probabilities is just the product of the individual (conditional) likelihood ratios, which can be written as

$$\prod_{k=1}^{m} e^{-\theta Y_k c_k}(1+p_k(Z)(e^{\theta c_k}-1)) = \exp(-\theta L + \psi(\theta, Z)). \tag{8}$$

Here, we have used the fact that $L = Y_1 c_1 + \cdots + Y_m c_m$ and introduced the conditional cumulant generating function of L,

$$\psi(\theta, z) = \log \mathsf{E}[\exp(\theta L)|Z = z] = \sum_{k=1}^{m} \log\left(1 + p_k(Z)(e^{\theta c_k}-1)\right). \tag{9}$$

The special form of the likelihood ratio in (8) — depending on the default indicators $Y_1, \ldots, Y_m$ only through the total loss L — is a consequence of the parametric specification in (7).

Glasserman and Li [8] choose the parameter θ as a function of the factors Z and the loss threshold x. Let $\theta_x(z)$ be the unique solution to the equation

$$\frac{\partial}{\partial\theta}\psi(\theta, z) = x; \tag{10}$$

a unique solution indeed exists because, for all z, the derivative increases from $-\infty$ to ∞ as θ increases from $-\infty$ to ∞. It follows from (9) and (7) that the derivative on the left side of this equation can be rewritten using the $p_k(\theta, z)$ in (7) as

$$\frac{\partial}{\partial\theta}\psi(\theta, z) = \sum_{k=1}^{m} p_k(\theta, z) c_k.$$

Thus, setting $\theta = \theta_x(z)$ in (7) adjusts the conditional default probabilities just enough to make the conditional expected loss equal to x, in the sense that

$$\sum_{k=1}^{m} p_k(\theta_x(z), z) c_k = x. \tag{11}$$

If the threshold x is larger than the conditional expected loss $\mathsf{E}[L|Z = z]$, then $\theta_x(z) > 0$ and the twisted conditional default probabilities $p_k(\theta_x(z), z)$ are greater than the original conditional default probabilities $p_k(z)$. If $x < \mathsf{E}[L|Z = z]$, then $\theta_x(z) < 0$ and $p_k(\theta_x(z), z) < p_k(z)$. In estimating $P(L > x)$, Glasserman and Li [8] use $\theta_x^+(z) = \max\{0, \theta_x(z)\}$, because in generating large values of L there is no advantage to reducing the conditional default probabilities. Negative values of $\theta_x(z)$ are, however, useful in estimating conditional expectations given $L = x$.

5.2 Shifting the Factor Mean

Equation (7) specifies how we apply importance sampling to the default probabilities conditional on the outcome Z of the factors. The method of Glasserman and Li [8] also applies importance sampling to the factors themselves.

In estimating a tail probability $P(L > x)$, a particularly effective IS distribution for the factors Z would be the probability density proportional to the function

$$z \mapsto P(L > x|Z = z) \exp(-z^\top z/2). \tag{12}$$

If we formulate the calculation of $P(L > x)$ as a problem of integrating over the distribution of Z, then the first factor on the right is the integrand and the second factor is (up to a normalization constant) the multivariate normal density of Z. Sampling from this density is difficult, so we approximate it by a multivariate normal distribution $N(\mu, I)$. We would like to choose μ so that the mode of the multivariate normal coincides with the mode of (12). In other words, we would like to choose μ as the solution to

$$\max_z P(L > x|Z = z) \exp(-z^\top z/2). \tag{13}$$

Even this approximation is not quite feasible because $P(L > x|Z = z)$ is difficult to evaluate, so we use a further approximation. Define

$$F_x^o(z) = \max_\theta \{\psi(\theta, z) - x\theta\} = \psi(\theta_x(z), z) - x\theta_x(z) \tag{14}$$

and

$$F_x(z) = \max_{\theta \geq 0} \{\psi(\theta, z) - x\theta\} = \psi(\theta_x^+(z), z) - x\theta_x^+(z). \tag{15}$$

These functions are illustrated for a homogeneous, single-factor model (with $p_k \equiv 0.02$, $c_k \equiv 1$, $a_{k1} \equiv 0.3$, $m = 100$, and $x = 10$) in Figure 1. As shown in [8], F_x provides an upper bound on the conditional tail of L in the sense that

$$P(L > x|Z = z) \leq \exp(F_x(z)).$$

This upper bound also serves as a rough approximation. Using this approximation in (13) leads us to choose μ as the solution to

$$\max_z \{F_x(z) - \tfrac{1}{2} z^\top z\}. \tag{16}$$

Glasserman and Li [8] use the solution μ_* as the new mean of the factors in their importance sampling procedure. There is just one maximizer in a single-factor model. In multifactor

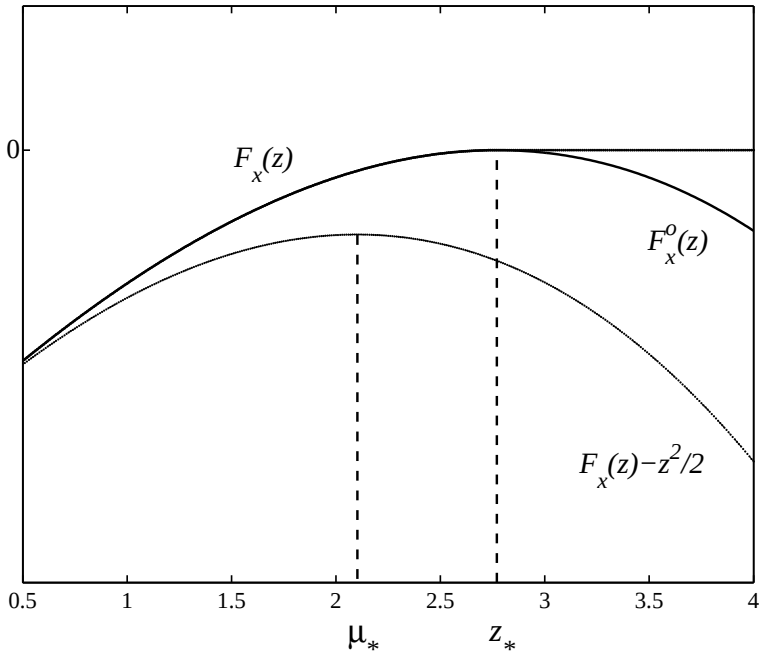

Figure 1: Graphs of the functions $F_x(z)$, $F_x^o(z)$, and $F_x(z) - z^2/2$ for a single-factor, homogeneous portfolio with $p_k \equiv 0.02$, $c_k \equiv 1$, $a_{k1} \equiv 0.3$, $m = 100$, and $x = 10$. The point μ_* maximizes $F_x(z) - z^2/2$, and z_* is the smallest point at $F_x(z) = 0$.

models, it is possible to have multiple solutions. For simplicity, we will assume a unique μ_* in our discussion.

A second point of interest for importance sampling is the solution to

$$\min_z z^\top z \quad \text{subject to} \quad F_x(z) = 0. \tag{17}$$

In Figure 1, this is the smallest z at which F_x reaches zero and also the only point at which $F_x^o(z) = 0$. This point is labeled z_*. Because $F_x^o(z_*) = 0$, (14) implies that $\theta_x(z_*) = 0$, which means that $\mathsf{E}[L|Z = z_*] = x$. Thus, shifting the mean of Z to z_* is potentially attractive for importance sampling. In fact, Glasserman and Li [8] show that in a homogeneous, single-factor model, $\mu_*/z_* \to 1$ as the size of the portfolio increases. In an asymptotic sense, all candidate means between μ_* and z_* are equally effective and, indeed, optimal.

6 Importance Sampling for Conditional Expectations

We now return to the problem of estimating marginal risk contributions of the form $\mathsf{E}[X_k|L \in \mathcal{A}]$, with $\mathcal{A} = \{x\}$ or $\mathcal{A} = [x, \infty)$. In both cases, we assume $P(L \in \mathcal{A}) > 0$ and use the

representation

$$r_k \equiv \mathsf{E}[X_k|L \in \mathcal{A}] = \frac{\mathsf{E}[X_k \mathbf{1}\{L \in \mathcal{A}\}]}{P(L \in \mathcal{A})}.$$

We apply importance sampling to estimate the numerator and the denominator. To keep the discussion generic, let $\tilde{P}$ and $\tilde{\mathsf{E}}$ denote probability and expectation under a different probability measure with the same null sets as P. Write ℓ for the likelihood ratio $dP/d\tilde{P}$. Then

$$r_k = \frac{\tilde{\mathsf{E}}[X_k \ell \mathbf{1}\{L \in \mathcal{A}\}]}{\tilde{\mathsf{E}}[\ell \mathbf{1}\{L \in \mathcal{A}\}]}.$$

Let $(X_1^{(i)}, \ldots, X_m^{(i)})$, $i = 1, \ldots, n$, denote independent replications generated under $\tilde{P}$ and let $L^{(i)} = X_1^{(i)} + \cdots + X_m^{(i)}$. Let $\ell^{(i)}$ denote the likelihood ratio on the ith replication. The estimator

$$\hat{r}_k^{IS} = \frac{\sum_{i=1}^n X_k^{(i)} \ell^{(i)} \mathbf{1}\{L^{(i)} \in \mathcal{A}\}}{\sum_{i=1}^n \ell^{(i)} \mathbf{1}\{L^{(i)} \in \mathcal{A}\}} \tag{18}$$

converges to r_k with probability 1. We can assess the precision of this estimator using confidence intervals calculated in accordance with the following result:

Proposition 2 *Suppose* $P(L \in \mathcal{A}) > 0$ *and*

$$\tilde{\mathsf{E}}[X_k^2 \ell^2 \mathbf{1}\{L \in \mathcal{A}\}] < \infty \quad \textit{and} \quad \tilde{\mathsf{E}}[\ell^2 \mathbf{1}\{L \in \mathcal{A}\}] < \infty.$$

Let

$$\hat{\sigma}_k^{IS} = \left(\frac{n \sum_{i=1}^n (X_k^{(i)} \ell^{(i)} - \hat{r}_k^{IS} \ell^{(i)})^2 \mathbf{1}\{L^{(i)} \in \mathcal{A}\}}{\left(\sum_{i=1}^n \ell^{(i)} \mathbf{1}\{L^{(i)} \in \mathcal{A}\} \right)^2} \right)^{1/2}, \tag{19}$$

taking the ratio to be zero whenever the denominator is zero. Then with the $X_k^{(i)}$ *sampled under* $\tilde{P}$*, the distribution of*

$$\frac{\hat{r}_k^{IS} - r_k}{\hat{\sigma}_k^{IS}/\sqrt{n}}$$

converges to the standard normal and

$$\hat{r}_k^{IS} \pm z_{\delta/2} \frac{\hat{\sigma}_k^{IS}}{\sqrt{n}}$$

is an asymptotically valid $1 - \delta$ *confidence interval for* r_k*.*

6.1 Value-at-Risk Contributions

We now specialize to conditional expectations of the form $\mathsf{E}[X_k|L = x]$, as required for VaR contributions. We apply importance sampling to the estimation of both the numerator $\mathsf{E}[X_k \mathbf{1}\{L =$

$x\}]$ and the denominator $P(L = x)$. We use the general class of IS procedures described in Section 5 in which we shift the mean of the factors Z and then exponentially twist the conditional default probabilities, given Z. To complete the specification of the method, we need to select a new mean for Z and a conditional twisting parameter θ.

For the conditional twisting parameter we use $\theta_x(z)$ rather than its positive part $\theta_x^+(z)$, used in [8]. Recall that $\theta_x(z)$ is positive whenever $\mathsf{E}[L|Z = z]$ is less than x. Thus, $\theta_x(z)$ and $\theta_x^+(z)$ differ only when $\mathsf{E}[L|Z = z]$ exceeds x. In estimating $P(L > x)$, large values of L are useful, so there is no need to shrink the conditional expected loss back to x if $\mathsf{E}[L|Z = z] > x$. But in estimating $P(L = x)$ or expectations conditioned on $L = x$, we want to force the sampling procedure to generate more samples in which the loss L exactly equals x. By using $\theta_x(z)$ rather than $\theta_x^+(z)$, we ensure that the expected loss, given Z, is equal to x *regardless of the outcome of* Z. This is simply a restatement of (11).

For the new factor mean, we use μ_*, the optimizer in (16). The resulting estimator of $\mathsf{E}[X_k\mathbf{1}\{L = x\}]$ is the sample mean of independent replications of

$$X_k\ell\mathbf{1}\{L = x\} = X_k e^{-\theta_x(Z)L+\psi(\theta_x(Z),Z)}e^{-\mu_*^\top Z-\mu_*^\top\mu_*/2}\mathbf{1}\{L = x\};$$

we estimate the denominator using independent replications of

$$\ell\mathbf{1}\{L = x\} = e^{-\theta_x(Z)L+\psi(\theta_x(Z),Z)}e^{-\mu_*^\top Z-\mu_*^\top\mu_*/2}\mathbf{1}\{L = x\};$$

and, as in (18), the IS estimator of $\mathsf{E}[X_k|L = x]$ is the ratio of the averages of the two expressions over multiple replications. We detail the steps in the following:

Algorithm 6.1

1. Find μ_* by solving (16)
2. Repeat for replications $i = 1, \ldots, n$
   (a) Generate Z from $N(\mu_*, I)$
   (b) Calculate $p_k(Z)$ as in (2), $k = 1, \ldots, m$
   (c) Solve for $\theta_x(Z)$ in (11) and calculate $p_k(\theta_x(Z), Z)$, $k = 1, \ldots, m$
   (d) Set $Y_k = 1$ with probability $p_k(\theta_x(Z), Z)$ and $Y_k = 0$ otherwise, $k = 1, \ldots, m$
   (e) Calculate total loss $L = Y_1c_1 + \cdots + Y_mc_m$
   (f) Calculate likelihood ratio

$$\ell = \exp(-\theta_x(Z)L + \psi(\theta_x(Z), Z) - \mu_*^\top Z + \mu_*^\top\mu_*/2)$$

(g) Set $Num_k = Num_k + Y_k c_k \ell \mathbf{1}\{L = x\}$ and $Den_k = Den_k + \ell \mathbf{1}\{L = x\}$

3. Return $\hat{r}_k = Num_k / Den_k$

6.2 Expected Shortfall Contributions

We now modify the method of Section 6.1 to estimate $\mathsf{E}[X_k | L \geq x]$. The structure of the algorithm is the same, except that now we use $\theta_x^+(Z)$ rather than $\theta_x(Z)$: if $\mathsf{E}[L|Z] < x$, we adjust the default probabilities to increase the conditional expected loss to x, but if $\mathsf{E}[L|Z] \geq x$, we do not change the default probabilities. The algorithm is as follows:

Algorithm 6.2

1. Find μ_* by solving (16)

2. Repeat for replications $i = 1, \ldots, n$

   (a) Generate Z from $N(\mu_*, I)$

   (b) Calculate $p_k(Z)$ as in (2), $k = 1, \ldots, m$

   (c) Solve for $\theta_x(Z)$ in (11) and calculate $p_k(\theta_x^+(Z), Z)$, $k = 1, \ldots, m$

   (d) Set $Y_k = 1$ with probability $p_k(\theta_x^+(Z), Z)$ and $Y_k = 0$ otherwise, $k = 1, \ldots, m$

   (e) Calculate total loss $L = Y_1 c_1 + \cdots + Y_m c_m$

   (f) Calculate likelihood ratio

$$\ell = \exp(-\theta_x^+(Z) L + \psi(\theta_x^+(Z), Z) - \mu_*^\top Z + \mu_*^\top \mu_*/2)$$

   (g) Set $Num_k = Num_k + Y_k c_k \ell \mathbf{1}\{L \geq x\}$ and $Den_k = Den_k + \ell \mathbf{1}\{L \geq x\}$

3. Return $\hat{r}_k = Num_k / Den_k$

6.3 Numerical Examples

We now illustrate the performance of Algorithms 6.1 and 6.2 through examples. Each example is a portfolio of $m = 100$ obligors; we have obtained nearly identical results with $m = 1000$, so the size of the portfolio does not appear to have much effect on the quality of the estimators. (Indeed, the asymptotic optimality results in Glasserman and Li [8] assume $m \to \infty$, suggesting that the method is even more effective in large portfolios.) The effect of varying the default probabilities p_k is similar to the effect of varying the exposures c_k, so we keep the default probabilities fixed at 1%. Generally speaking, the benefit of the IS estimators is greater for

rarer events; increasing or decreasing the overall level of the default probabilities has roughly the same effect as varying the loss level x in the conditional expectations.

Example 1. In our first example, the obligors are independent. The exposures are

$$c_k = \begin{cases} 1, & k = 1, \ldots, 20, \\ 4, & k = 21, \ldots, 40, \\ 9, & k = 41, \ldots, 60, \\ 16, & k = 61, \ldots, 80, \\ 25, & k = 81, \ldots, 100. \end{cases} \tag{20}$$

This is a deliberately lumpy profile that will lead to significant variation in marginal risk contributions.

A loss level of $x = 80$ corresponds roughly to a 99.9% VaR because $P(L \geq 80)$ is about 0.1%. The probability that the loss exactly equals 80 is about 0.03%, so in calculating VaR contributions (conditional expectations given $L = x$) we are indeed conditioning on a rare event.

Figure 2 compares the performance of the IS estimators and ordinary Monte Carlo for VaR contributions (left panel) and expected shortfall contributions (right panel). In each case, the horizontal axis lists the obligors from 1 to 100 and the vertical axis shows the estimated risk contribution for each obligor. The two solid lines show 95% confidence intervals for each obligor using IS; the two dotted lines show 95% confidence intervals using ordinary Monte Carlo. Both are calculated using 250,000 replications. This rather large number of replications is needed to get meaningful confidence intervals using ordinary Monte Carlo because we are conditioning on rare events; far fewer replications would suffice using IS. Indeed, the figures clearly show a substantial reduction in variability using IS. From the IS estimates, it is evident that the risk contributions are the same for consecutive blocks of obligors, though this is not at all clear from the ordinary Monte Carlo estimates.

Example 2. We modify Example 1 by making it a single-factor model in which each obligor's latent variable ξ_k has a correlation of 0.5 with the common factor. This dependence pushes more mass into the tail of the loss distributions, so we increase the loss threshold to $x = 100$. The mean shift for the common factor is $\mu_* = 2.00$. The results are illustrated in Figure 3. The average variance reduction ratio in this example, averaged over the 100 obligors, is about 20.

Example 3. To test the effect of changing the exposures, we modify Example 2 by letting the c_k increase linearly from 1 to 100. We set the loss threshold at $x = 500$ which makes $P(L \geq x)$ about 1.1%. The granularity of the c_k makes $P(L = x)$ too small to be meaningful, so we replace the event $\{L = x\}$ with $\{|L - x| \leq 1\}$. This event has probability 0.02%. The results are illustrated in Figure 4.

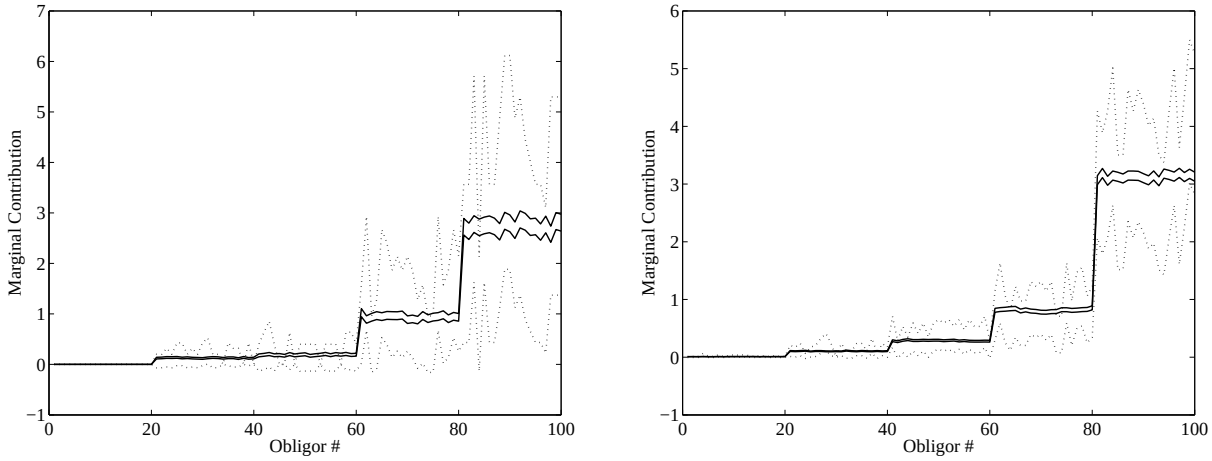

Figure 2: Comparison of IS (solid) and ordinary Monte Carlo (dotted) estimates for VaR contributions (left) and expected shortfall contributions (right) in Example 1.

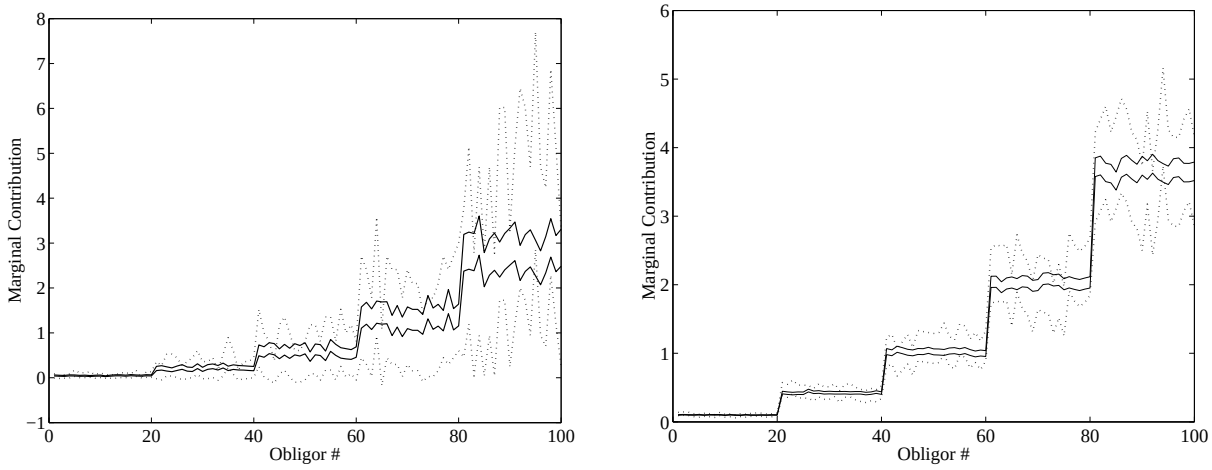

Figure 3: Comparison of IS (solid) and ordinary Monte Carlo (dotted) estimates for VaR contributions (left) and expected shortfall contributions (right) in Example 2.

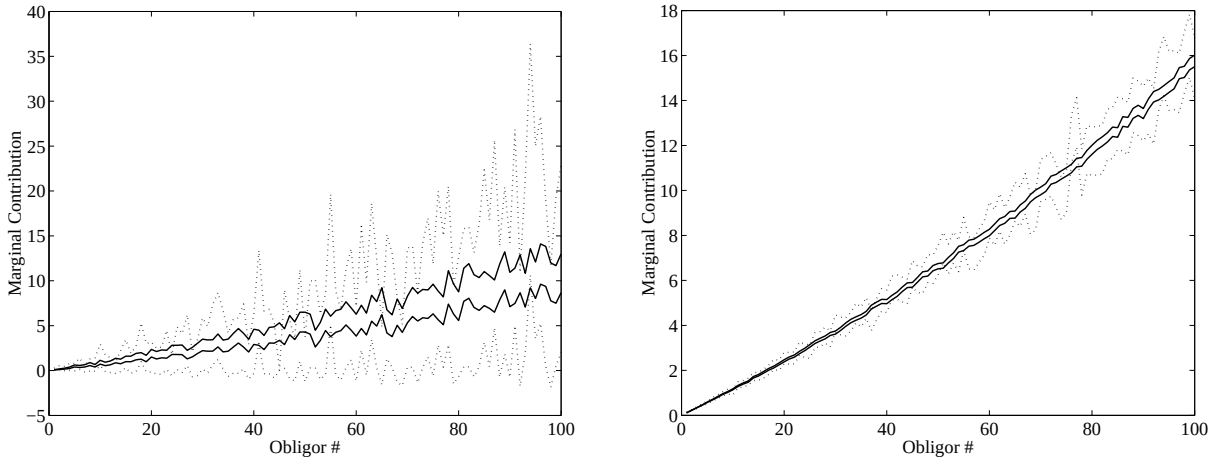

Figure 4: Comparison of IS (solid) and ordinary Monte Carlo (dotted) estimates for VaR contributions (left) and expected shortfall contributions (right) in Example 3.

Example 4. We now construct an 11-factor version of Example 2. The exposures are as in (20). The first factor is a market-wide factor, and each obligor has a coefficient of 0.3 on this factor. In addition, the first ten obligors have a coefficient of 0.8 on the second factor, the next ten obligors has a coefficient of 0.8 on the next factor, and so on. Thus, each obligor is sensitive to the market-wide factor and to one additional factor. The block structure of the dependence in this example ensures that the model cannot be well-approximated using a small number of factors. We set the loss threshold at $x = 250$, for which $P(L \geq x)$ is about 0.1% (99.9% VaR) and $P(L = x)$ is about 0.03%.

The mean shift for this example is

$$\mu_* = (1.6214, 0.0002, 0.0002, 0.0009, 0.0009, 0.0018, 0.0018, 0.0028, 0.0028, 2.1563, 2.1563)^\top.$$

Observe that this primarily shifts the mean of the first factor and the last two. To interpret this solution, think of μ_* as the least costly outcome of Z leading to large losses (losses near x) when the cost associated with $Z = z$ is $z^\top z/2$. A large outcome of the first factor is cost effective because the first factor affects all obligors. Large outcomes of the last two factors are cost effective not because those factor affect more obligors (each of the last 10 factors affects exactly 10 obligors), but rather because the last two factors affect the obligors with the highest exposures. Those are the obligors most likely to have defaulted when we condition on a large loss level. Also, observe that μ_* correctly detects the symmetry in the effect of consecutive pairs of factors. The results obtained using μ_* and 250,000 replications are illustrated Figure 5. The ordinary Monte Carlo results show not only wider confidence intervals but also a high degree

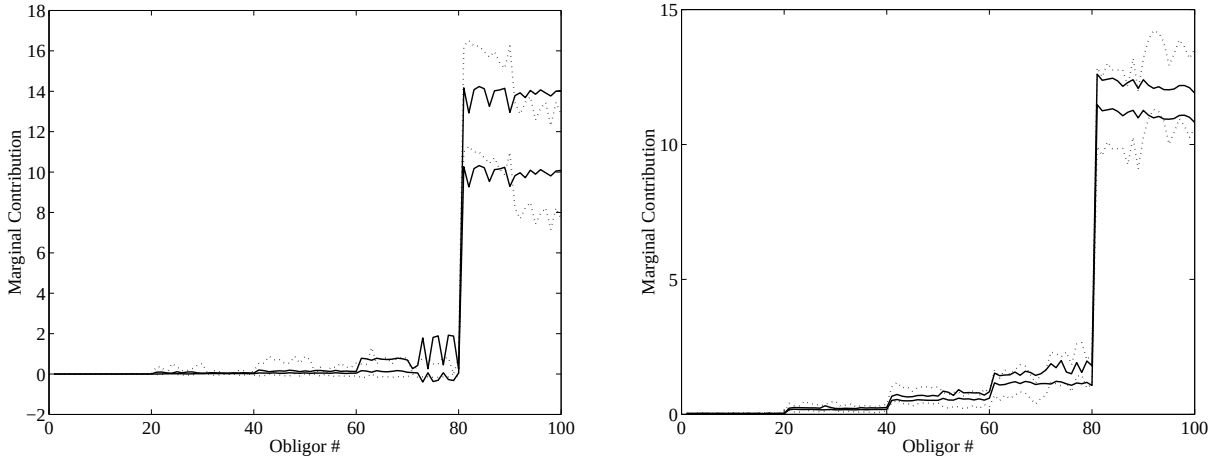

Figure 5: Comparison of IS (solid) and ordinary Monte Carlo (dotted) estimates for VaR contributions (left) and expected shortfall contributions (right) in Example 4.

of sampling variability in the estimated intervals.

7 Shrinking the Factor Variance

In the IS algorithms of the previous section, we shift the mean of the factors Z but otherwise leave their distribution unchanged. The "optimal" density for Z is the one in (12); we chose the normal distribution to have the same mode as the optimal one. It is natural to consider whether changing the covariance matrix of the factors improves the approximation. Changing the covariance matrix is particularly appealing in estimating a conditional expectation given $L = x$: to have more samples fall near a given point, we would like to reduce the variability of Z.

The optimal density (as in (12)) is the conditional density of Z, conditioned on either $L \geq x$ or $L = x$, depending on the context. One way to select an approximating normal distribution is to select one that has the same first and second moments as the conditional density. These conditional moments are generally unknown but could be estimated from a preliminary set of runs. Using ordinary Monte Carlo, one could estimate these conditional moments by storing the values of Z on which $L = x$ or $L \geq x$ and then calculating the sample mean and covariance of the stored Zs. This would require a large number of replications. A more effective way would use importance sampling, starting with the distribution $N(\mu_*, I)$ used above. The conditional moments can be estimated using importance sampling, producing estimates $\tilde{\mu}$ and $\tilde{\Sigma}$, say; these can then be used to update the importance sampling distribution to $N(\tilde{\mu}, \tilde{\Sigma})$.

An alternative is to use the relation between the covariance matrix of a normal distribution

and the Hessian (matrix of second derivatives) of the log density at its mode. The log of the $N(\mu, \Sigma)$ density (for nonsingular Σ) is

$$y \mapsto -\tfrac{1}{2}(y-\mu)^\top \Sigma^{-1}(y-\mu) + \text{constant},$$

from which we see that the Hessian at μ is $-\Sigma^{-1}$.

In Figure 1, we select μ_* as the mode of $F_x^o(z) - z^2/2$. This corresponds to approximating $\exp(F_x^o(z) - z^2/2)$ by (a multiple of) the normal density $N(\mu_*, 1)$. In approximating this function by a normal density $N(\mu_*, \sigma^2)$, we can use the relation between the second derivative and the variance and choose σ^2 to satisfy

$$-\frac{1}{\sigma^2} = \frac{\partial^2}{\partial z^2}\left(F_x^o(z) - \tfrac{1}{2}z^2\right) = \frac{\partial^2}{\partial z^2}F_x^o(z) - 1.$$

More generally, in the multifactor case this corresponds to setting

$$\Sigma = \left(I - \nabla_2 F_x^o(\mu_*)\right)^{-1}, \tag{21}$$

assuming the inverse on the right exists and the resulting matrix is positive definite. The notation $\nabla_2 F_x^o(\mu_*)$ indicates the Hessian of $F_x^o(\cdot)$ at μ_*. Initial numerical experiments reported in Li [14] indicate that this is an effective choice in estimating $P(L = x)$.

Appendix B of Glasserman [7] shows that the Hessian of $F_x^o(\cdot)$ is given by

$$\nabla_2 F_x^o(z) = \nabla_2 \psi(\theta_x, z) + \frac{\partial}{\partial \theta}\nabla\psi(\theta_x, z)^\top \nabla\theta_x(z).$$

with

$$\nabla\theta_x(z) = -\nabla\frac{\partial}{\partial\theta}\psi(\theta_x, z)/\frac{\partial^2}{\partial\theta^2}\psi(\theta_x, z).$$

Thus, the Hessian can be evaluated purely in terms of derivatives of ψ. This is important because ψ is given explicitly by (9), so its derivatives are easily evaluated. Evaluation of this part of (21) is thus practical. There is no guarantee that the inverse in (21) exists; we have generally found that it does exist at sufficiently large values of x. Also, whether Σ actually "shrinks" variance relative to the identity matrix depends on $\nabla_2 F_x^o(\mu_*)$ being negative semidefinite, which requires that F_x^o be concave near μ_*. This property is also important in the approximations of Glasserman [7].

The use of the modified variance in importance sampling is illustrated in Figure 6 for Examples 2–4, where we compare IS based solely on shifting the factor mean with IS based on shifting the mean and changing the covariance matrix. The figure indicates some modest further variance reduction from the change in covariance.

For Figure 6, we have moved the loss threshold to $x = 300$ for Example 2, where the benefit of shrinking the factor standard deviation (from 1 to 0.294) should be greater. For Example 3, (21) gives a standard deviation of 0.3463. The covariance matrix we used for IS with Example 4 is

$$\begin{pmatrix}
0.791 & 0.000 & 0.000 & 0.003 & 0.003 & 0.006 & 0.006 & 0.009 & 0.009 & -0.296 & -0.296 \\
0.000 & 1.002 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 1.002 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.003 & 0.000 & 0.000 & 1.009 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.001 & -0.001 \\
0.003 & 0.000 & 0.000 & 0.000 & 1.009 & 0.000 & 0.000 & 0.000 & 0.000 & -0.001 & -0.001 \\
0.006 & 0.000 & 0.000 & 0.000 & 0.000 & 1.020 & 0.000 & 0.000 & 0.000 & -0.003 & -0.003 \\
0.006 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.020 & 0.000 & 0.000 & -0.003 & -0.003 \\
0.009 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.031 & 0.000 & -0.004 & -0.004 \\
0.009 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.031 & -0.004 & -0.004 \\
-0.296 & 0.000 & 0.000 & -0.001 & -0.001 & -0.003 & -0.003 & -0.004 & -0.004 & 1.048 & -0.823 \\
-0.296 & 0.000 & 0.000 & -0.001 & -0.001 & -0.003 & -0.003 & -0.004 & -0.004 & -0.823 & 1.048
\end{pmatrix}.$$

This covariance matrix differs from the identity matrix primarily in two ways: it assigns smaller variance to the first factor, and it introduces negative correlation between the factors whose means have been substantially increased — the first and the last two. In particular, the rather large negative covariance of -0.823 between the last two factors suggests that the event $\{L = x\}$ (in this case with $x = 250$) occurs because of a large value of one of the last two factors, but not both. This phenomenon is also reflected in the pattern of estimated confidence intervals using ordinary Monte Carlo (the dotted lines) in the left panel of Figure 5.

8 Asymptotic Approximation

In this section, we develop a hybrid method that combines an asymptotic approximation with Monte Carlo. The asymptotic method approximates conditional risk contributions, given the factors; we use Monte Carlo solely to integrate the approximation over the distribution of the factors.

8.1 A General Approximation

We formulate a rigorous result assuming homogeneous obligors and then apply the same techniques to develop approximations for general portfolios. In order to consider asymptotics as the portfolio size grows, we assume an infinite sequence $\{Z, X, X_1, X_2, \ldots\}$ with the X_i interpreted (as before) as loss random variables, though not necessarily of the form $c_i Y_i$. We impose the following conditions:

(i) $X, X_1, X_2, \ldots$ are conditionally i.i.d. given Z;

(ii) $0 \le X \le b$ for some $b > 0$, and $P(X < \epsilon|Z)$ and $P(X > b - \epsilon|Z)$ are positive with probability 1, for all $\epsilon > 0$.

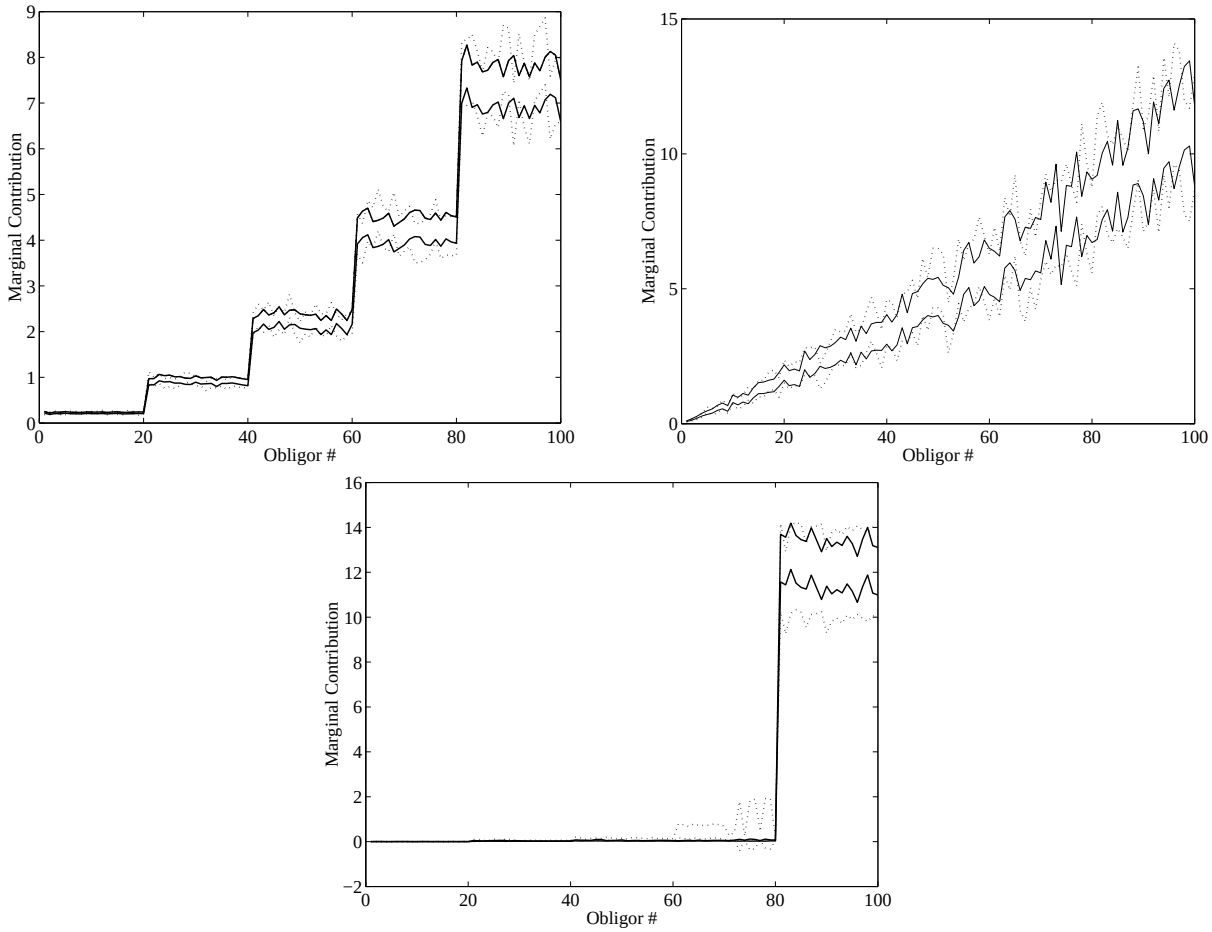

Figure 6: IS confidence intervals for VaR contributions using shifted factor mean (dotted) and both shifted mean and modified factor variance (solid). Left panel shows Example 2 with $x = 300$; right panel shows Example 4.

Define the conditional cumulant generating function of the X_i by setting

$$\psi_X(\theta, Z) = \log \mathsf{E}[\exp(\theta X)|Z], \quad -\infty < \theta < \infty.$$

The boundedness assumption in (ii) more than suffices to ensure finiteness of $\psi_X(\theta, Z)$. Let $G_0(\cdot, z)$ denote the conditional distribution of X given $Z = z$,

$$G_0(x, z) = P(X \le x|Z = z), \text{ for all } x.$$

We embed $G_0(\cdot, z)$ in an exponential family of distributions $\{G_\theta(\cdot, z), -\infty < \theta < \infty\}$ by setting

$$G_\theta(x, z) = \int_{-\infty}^{x} e^{\theta u - \psi_X(\theta, z)}\, dG_0(u, z).$$

These have the property (standard for exponential families) that

$$\int_{\infty}^{\infty} x\, dG_\theta(x, z) = \frac{\partial}{\partial\theta}\psi_X(\theta, z).$$

For any $0 < q < b$, let $\theta_q(z)$ be the unique solution to

$$\frac{\partial}{\partial\theta}\psi_X(\theta_q(z), z) = q; \tag{22}$$

assumption (ii) ensures that exactly one such solution indeed exists. The parameter $\theta_q(z)$ picks out the member of the exponential family $\{G_\theta(\cdot, z), -\infty < \theta < \infty\}$ whose mean is q.

Let $L_m = X_1 + \cdots + X_m$. We let m increase and consider the conditional distribution of any fixed set of losses $X_1, \ldots, X_k$, conditional on Z and on either $L_m = qm$ or $L_m \ge qm$. A general class of results may be paraphrased as stating that the conditional distribution of, e.g., X_1 converges to the exponentially twisted distribution with mean q. In other words, conditioning on the sum (in our case, the total loss) changes the distribution of the summands in a predictable way. We record precise versions of these assertions in the following:

Proposition 3 *Suppose (i) and (ii) hold and $0 < q < b$. (a) If X has a lattice distribution given $Z = z$ with $P(X = q|Z = z) > 0$, then for any fixed k, as $m \to \infty$,*

$$P(X_1 \le x_1, \ldots, X_k \le x_k|L_m = qm, Z = z) \to \prod_{i=1}^{k} G_{\theta_q(z)}(x_i, z). \tag{23}$$

Also, for $k \le m$,

$$\mathsf{E}[X_k|L_m = qm, Z = z] = \int_{-\infty}^{\infty} x\, dG_{\theta_q(z)}(x, z) = q. \tag{24}$$

(b) If, given $Z = z$, X has a lattice distribution or a density, then

$$P(X_1 \le x_1, \ldots, X_k \le x_k|L_m \ge qm, Z = z) \to \prod_{i=1}^{k} G_{\theta_q^+(z)}(x_i, z). \tag{25}$$

Also, as $m \to \infty$,

$$\mathsf{E}[X_k | L_m \geq qm, Z = z] \to \int_{-\infty}^{\infty} x \, dG_{\theta_q^+(z)}(x, z) = q. \tag{26}$$

Proof. Once we condition on $Z = z$, the X_i are conditionally i.i.d. and we can apply results for i.i.d. sequences. The limit in (23) then follows from Zabell [20]. The limit in (25) in the density case follows from Van Campenhout and Cover [19] who also cite earlier work of O.E. Lanford. The lattice case is treated in Cover and Thomas [3]. The equality in (24) follows from the fact that $X_1, \ldots, X_m$ are exchangeable, given $L_m = qm$ and $Z = z$. The limit in (26) follows from (25) and the boundedness of the X_i: convergence of distributions implies convergence of expectations for bounded random variables. □.

If we interpret the limits in the proposition as approximations, then in each case the result states that in conditioning on a large loss for the whole portfolio (and also the factor outcome), we may approximate the conditional loss distribution for an individual obligor by exponentially twisting the original distribution loss distribution for that obligor.

8.2 The Hetereogeneous Case

The limits in Proposition 3 provide approximations for conditional expectations for fixed m. Although the proposition assumes homogeneous obligors, the approximations extend in a natural way to the heterogeneous case, as we now explain.

The limits in (25) and (26) suggest the approximation

$$\mathsf{E}[X_k | L_m \geq qm, Z = z] \approx \mathsf{E}\left[X_k \exp\left(\theta_q^+(Z) L_m - m\psi_X(\theta_q^+(Z), Z)\right) | Z = z\right]. \tag{27}$$

In more detail, the right side becomes

$$\begin{aligned} &\mathsf{E}\left[X_k \exp\left(\theta_q^+(Z) L_m - m\psi_X(\theta_q^+(Z), Z)\right) \middle| \, Z = z\right] \\ &= \mathsf{E}\left[X_k \exp\left(\theta_q^+(Z) X_k - \psi_X(\theta_q^+(Z), Z)\right) \middle| \, Z = z\right] \\ &= \int_{-\infty}^{\infty} x \, dG_{\theta_q^+(z)}(x, z), \end{aligned} \tag{28}$$

the first equality using the conditional independence of the X_i and the second using the definition of G_θ. From (28), we see that the approximation in (27) is the limit in (26).

For loss levels X_i that are conditionally identically distributed, $m\psi_X$ is the conditional cumulant generating function of the total loss L; in generalizing (27) to the heterogeneous case, we therefore replace $m\psi_X$ with $\psi = \psi_{X_1} + \cdots + \psi_{X_m}$, the conditional cumulant generating

function of the total loss L. If we set the loss threshold x equal to qm, then equation (22) defining θ_q coincides with equation (10) defining θ_x. With these substitutions, we generalize (27) to

$$\mathsf{E}[X_k|L \geq x, Z = z] \approx \mathsf{E}\left[X_k \exp\left(\theta_x^+(Z)L - \psi(\theta_x^+(Z), Z)\right)|Z = z\right], \tag{29}$$

with θ_x as in (10). Written this way, the approximation does not rely on the X_i being conditionally identically distributed.

The interpretation of (29) and of Proposition 3 is that conditioning on a large loss for the whole portfolio changes the distribution of the individual losses in a predictable way. In particular, conditional on a large loss $L \geq x$ (and on the factor outcome $Z = z$), the individual losses X_k look like they came from the distribution determined by $\theta_x^+(z)$, rather than the original distribution.

Taking the expectation of both sides of (29), we get the approximation

$$\mathsf{E}[X_k|L \geq x] \approx \mathsf{E}\left[X_k \exp\left(\theta_x^+(\tilde{Z})L - \psi(\theta_x^+(\tilde{Z}), \tilde{Z})\right)\right]. \tag{30}$$

Here, $\tilde{Z}$ has the distribution of Z conditional on $L \geq x$. Comparing (23) and (25), we similarly get

$$\mathsf{E}[X_k|L = x] \approx \mathsf{E}\left[X_k \exp\left(\theta_x(\tilde{Z})L - \psi(\theta_x(\tilde{Z}), \tilde{Z})\right)\right], \tag{31}$$

but with $\tilde{Z}$ having the distribution of Z conditional on $L = x$. The two approximations nearly coincide when x is large (i.e., when $\theta_x(z) > 0$) because conditional on $L \geq x$, the total loss L is likely to be very near x if x is large. In applying these ideas, we will introduce an adjustment that differentiates the two cases.

8.3 A Hybrid Method

We now specialize to the case $X_k = Y_k c_k$ in which Y_k is a default indicator and c_k is a fixed loss given default. In this setting, (31) becomes

$$\mathsf{E}[X_k|L = x] = c_k \mathsf{E}[Y_k|L = x] \approx c_k \mathsf{E}\left[p_k(\theta_x(Z), Z)\right] \tag{32}$$

with $p_k(\theta, z)$ as in (2). This suggests a hybrid method in which we use Monte Carlo to "integrate out" the factors Z, but conditional on Z we use the approximation $p_k(\theta_x(Z), Z)$, rather than a ratio estimator like the one in (5). This has the enormous advantage that it allows us to estimate a conditional expectation given $L = x$ from *all* replications, not just the few on which L equals x.

To implement this, we need to generate Z from the correct distribution. In passing from (29) to (30), we are integrating over the distribution of Z given $L \geq x$ (or, in the case of (31)

the distribution of Z given $L = x$). So, in calculating the expectation on the right side of (32), we need to generate Z from the appropriate conditional distribution.

Let $\phi(z)$ denote the unconditional density of Z and $\phi(z|L \geq x)$ the conditional density given $L \geq x$. The conditional density is given by Bayes' rule:

$$\phi(z|L \geq x) \propto P(L \geq x|Z = z)\phi(z). \tag{33}$$

In the case of multivariate normal Z, this becomes

$$\phi(z|L \geq x) \propto P(L \geq x|Z = z) \exp(-z^\top z).$$

Much as in our discussion of importance sampling, we approximate this distribution by a multivariate normal $N(\mu, I)$ with μ chosen to match the mode of the conditional distribution — more precisely, we choose μ as in (16). This leads to the following algorithm to estimate $\mathsf{E}[Y_k c_k|L = x]$, combining the approximation $p_k(\theta_x(Z), Z)$ with Monte Carlo using the approximate conditional distribution of Z:

Algorithm 8.1: Approximate VaR Contributions

1. Find μ_* by solving (16)
2. Repeat for replications $i = 1, \ldots, n$
   (a) Generate Z from $N(\mu_*, I)$
   (b) Calculate $p_k(Z)$ as in (2), $k = 1, \ldots, m$
   (c) Solve for $\theta_x(Z)$ in (11) and calculate $p_k(\theta_x(Z), Z)$, $k = 1, \ldots, m$
   (d) Set $\hat{v}_k = \hat{v}_k + p_k(\theta_x(Z), Z)c_k$
3. Return $\hat{v}_k = \hat{v}_k/n$

To approximate $\mathsf{E}[Y_k c_k|L \geq x]$, we make two modifications. The first is to replace $\theta_x(Z)$ with $\theta_x^+(Z)$ in Step 2c. Recall that $\theta_x(z)$ is negative when the conditional expected loss $\mathsf{E}[L|Z = z]$ exceeds x. Using a negative value of the parameter makes sense in estimating a conditional expectation given $L = x$, because doing so pulls large values of L down towards x. But for a conditional expectation given $L \geq x$, large values of L are as useful as samples that land on x, so we replace $\theta_x(z)$ with zero when it is negative.

The second modification is an optional adjustment at the last step. The asymptotics in Proposition 3 do not distinguish between conditioning on $L = x$ and $L \geq x$. In fact, the

conditional default probabilities should be larger, on average, in the second case, so we correct for this.

If the approximation

$$p_k(\theta_x^+(z), z) \approx P(Y_k = 1 | Z = z, L \geq x)$$

held exactly, then we would have equality between

$$\mathsf{E}\left[\sum_{k=1}^{m} p_k(\theta_x^+(\tilde{Z}), \tilde{Z}) c_k\right] \quad \text{and} \quad \mathsf{E}[L | L \geq x].$$

Let $\bar{M}$ denote the average value of $\sum_{k=1}^{m} p_k(\theta_x^+(Z), Z)c_k$ over n replications, with $Z \sim N(\mu_*, I)$ or $Z \sim N(\mu_*, \Sigma)$ to approximate $\tilde{Z}$, which has the distribution of Z conditional on $L \geq x$. We would like to make the adjustment

$$p_k(\theta_x^+(Z), Z) \to p_k(\theta_x^+(Z), Z) \frac{\mathsf{E}[L | L \geq x]}{\bar{M}}$$

to the approximate conditional default probabilities. We do not know $\mathsf{E}[L|L \geq x]$, but we can estimate it from the simulation itself. This is what the optional steps do in the following algorithm:

Algorithm 8.2: Approximate Shortfall Contributions

1. Find μ_* by solving (16)
2. Repeat for replications $i = 1, \ldots, n$
   (a) Generate Z from $N(\mu_*, I)$
   (b) Calculate $p_k(Z)$ as in (2), $k = 1, \ldots, m$
   (c) Solve for $\theta_x(Z)$ in (11) and calculate $p_k(\theta_x^+(Z), Z)$, $k = 1, \ldots, m$
   (d) [Optional]
      i. Set $\bar{M} = \bar{M} + \sum_{k=1}^{m} p_k(\theta_x^+(Z), Z)c_k$
      ii. Generate default indicators $Y_1, \ldots, Y_m$ using probabilities $p_k(\theta_x^+(Z), Z)$, $k = 1, \ldots, m$
      iii. Calculate loss L and likelihood ratio ℓ and set $Num = Num + L\mathbf{1}\{L \geq x\}\ell$ and $Den = Den + \mathbf{1}\{L \geq x\}\ell$
   (e) Set $\hat{s}_k = \hat{s}_k + p_k(\theta_x(Z), Z)c_k$
3. Return $\hat{s}_k = \hat{s}_k / n$

4. [Optional] Return $\hat{s}_k = \hat{s}_k Num/(Den \times \bar{M})$

Remarks

1. The optional adjustment (which we use in all our examples) scales the estimated risk contributions of all obligors by the same amount.

2. Rather than solve for $\theta_x(z)$ and then take the positive part, one can first check if

$$\sum_{k=1}^{m} p_k(z)c_k > x$$

and solve for $\theta_x(z)$ only if this holds. If this inequality fails, $\theta_x(z)$ is in $(-\infty, 0]$ and $\theta_x^+(z) = 0$.

3. In the algorithms above, the $\hat{v}_k$ are sample means of i.i.d. replications, so supplementing them with confidence intervals is straightforward using their sample standard deviations. The estimator of Algorithm 8.2 is slightly more complicated because of the multiplicative adjustment, but a central limit theorem for it can be established along the lines of Propositions 1 and 2.

8.4 Numerical Examples

We illustrate the performance of the asymptotic approximation using the example portfolios of Section 6.3. Figure 7 shows results for Example 1 (which has independent obligors); Figure 8 applies to the single-factor model in Example 2; Figure 9 shows results for Example 3; and results for the 11-factor model of Example 4 appear in Figure 10. Each figure compares the approximation with a 95% confidence interval estimated using importance sample. In each case, the approximation is sufficiently accurate to be useful.

The figures suggest that the asymptotic approximation is more accurate for VaR contributions (conditioned on $L = x$) than for expected shortfall contributions (conditioned on $L \geq x$). This makes sense, because the asymptotics in Proposition 3 implicitly approximate conditioning on $L \geq x$ by conditioning on $L = x$. The approximations in the figure distinguish between the two primarily by a scale factor: the final (optional) step in Algorithm 8.2 scales the risk contributions so they sum (approximately) to $\mathsf{E}[L|L \geq x]$.

An alternative way to approximate shortfall contributions $\mathsf{E}[X_k|L \geq x]$ would be to take a weighted average of approximate VaR contributions $\mathsf{E}[X_k|L = x_i]$ at several thresholds x_i. The weight for $\mathsf{E}[X_k|L = x_i]$ should approximate the probability that L is closest to that x_i given that $L \geq x$. These probabilities can be estimated using importance sampling.

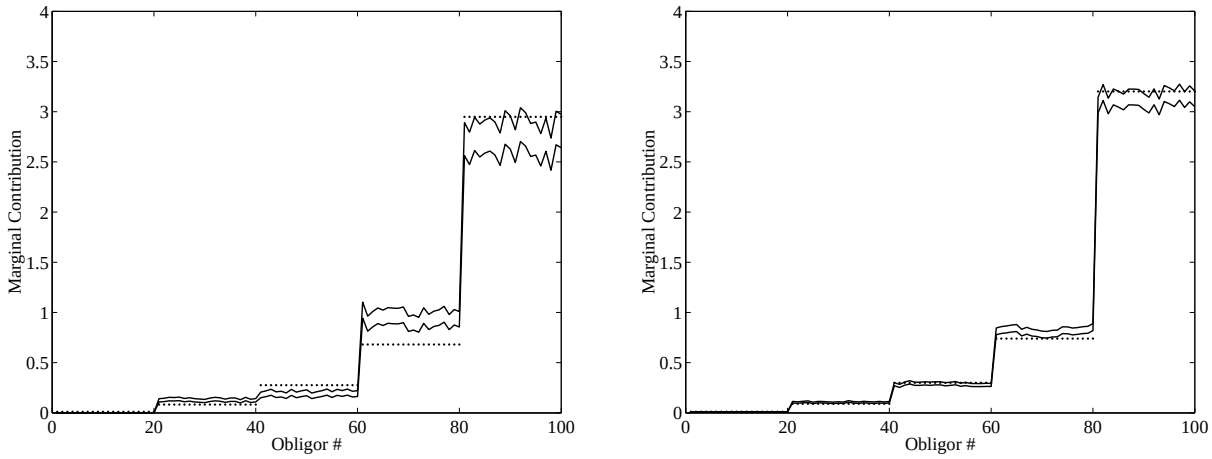

Figure 7: The dotted line shows the hybrid approximation. IS confidence intervals (solid lines) are included for reference. The graphs shows VaR contributions (left) and expected shortfall contributions (right) in Example 1.

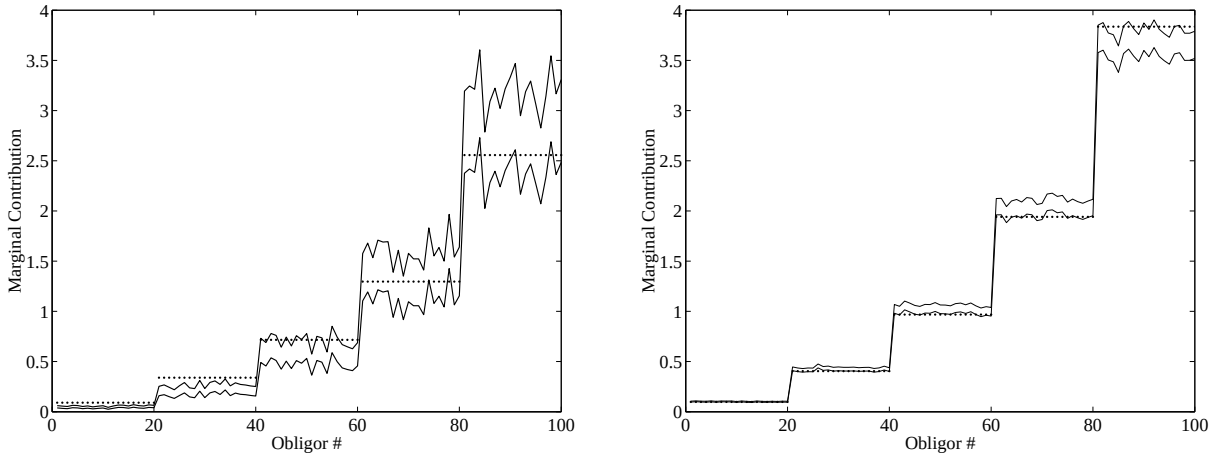

Figure 8: The dotted line shows the hybrid approximation. IS confidence intervals (solid lines) are included for reference. The graphs shows VaR contributions (left) and expected shortfall contributions (right) in Example 2.

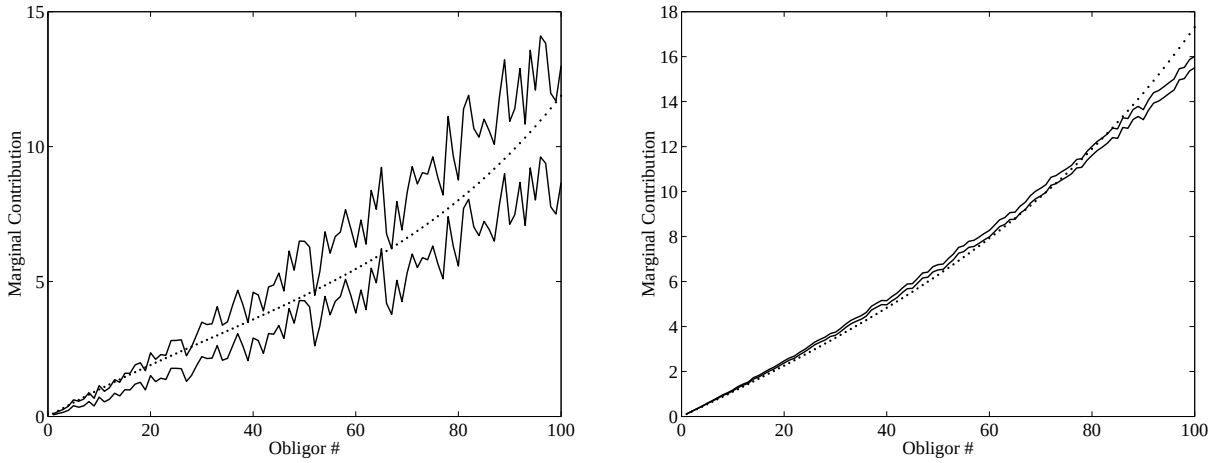

Figure 9: The dotted line shows the hybrid approximation. IS confidence intervals (solid lines) are included for reference. The graphs shows VaR contributions (left) and expected shortfall contributions (right) in Example 2.

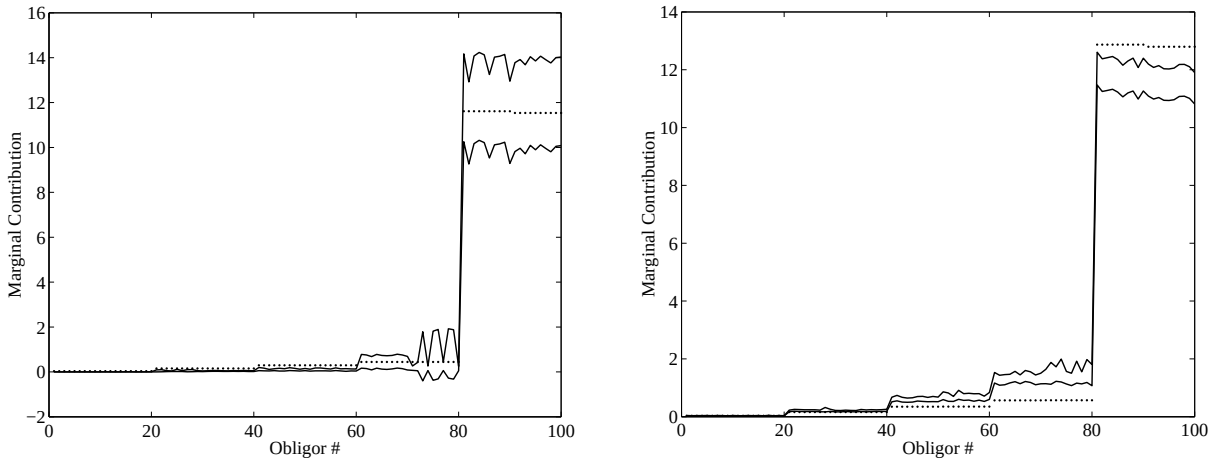

Figure 10: The dotted line shows the hybrid approximation. IS confidence intervals (solid lines) are included for reference. The graphs shows VaR contributions (left) and expected shortfall contributions (right) in Example 4.

9 Approximation Without Monte Carlo

The methods of Section 8 approximate the conditional default probability of each obligor (given a large loss level and the factor outcome), but use Monte Carlo to integrate out the distribution of the factors. In this section, we take the approximations one step further and dispense entirely with Monte Carlo. Rather than integrate over the distribution of factors, we evaluate the approximations of Section 8 at a single outcome of the factors — namely, the "most likely" outcome μ_* leading to large losses. Related ideas are investigated in Glasserman [7] as a means of approximating the overall risk in a portfolio (as opposed to the individual risk contributions). As explained there, this may be viewed as a simplified application of the classical Laplace approximation for integrals. The full Laplace approximation could be applied using the candidate covariance matrix in (21), provided this matrix is indeed positive definite.

By evaluating the steps in Algorithm 8.1 at just the single outcome $Z = \mu_*$, we arrive at the following approximation:

Algorithm 9.1: Approximate VaR Contributions

1. Find μ_* by solving (16); set $z = \mu_*$
2. Calculate $p_k(z)$ as in (2), $k = 1, \ldots, m$
3. Solve for $\theta_x(z)$ in (11) and calculate $p_k(\theta_x(z), z)$, $k = 1, \ldots, m$
4. Return $p_k(\theta_x(Z), Z)c_k$, $k = 1, \ldots, m$

In Algorithm 8.2, one of the ways we distinguished between VaR contributions and expected shortfall contributions involved scaling the approximate contributions to set their sum approximately equal to $\mathsf{E}[L|L \geq x]$, or rather an estimate of this ratio. This is the optional adjustment in Algorithm 8.2, calculated from the simulated losses themselves. Once we dispense entirely with Monte Carlo, we have no obvious mechanism for applying a similar adjustment. In the algorithm that follows, we adopt as a crude approximation

$$\mathsf{E}[L|L \geq x] \approx x + \mathsf{E}[L] = x + \sum_{k=1}^{m} p_k c_k.$$

We then approximate expected shortfall contributions as VaR contributions at this shifted threshold x.

Algorithm 9.2: Approximate Shortfall Contributions

1. Find μ_* by solving (16) at $x + \sum_k p_k c_k$; set $z = \mu_*$

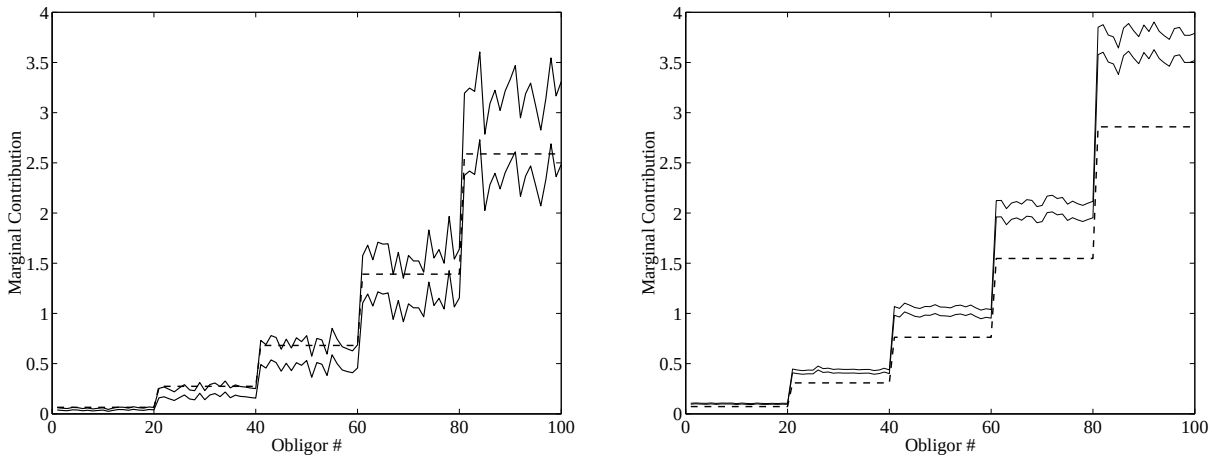

Figure 11: The dashed line shows the approximation. IS confidence intervals (solid lines) are included for reference. The graphs shows VaR contributions (left) and expected shortfall contributions (right) in Example 2.

2. Calculate $p_k(z)$ as in (2), $k = 1, \ldots, m$

3. Solve for $\theta_x(z)$ in (11) and calculate $p_k(\theta_x^+(z), z)$, $k = 1, \ldots, m$

4. Return $p_k(\theta_x^+(Z), Z)c_k$, $k = 1, \ldots, m$

In this algorithm, we evaluate the twisted default probabilities at $\theta_x^+(z)$ to be consistent with Algorithm 8.2. But it is not hard to see that $\theta_x(z)$ is positive at $z = \mu_*$ (as reflected in the fact that $F_x(\mu_*) < 0$ in Figure 1), so in this case $\theta_x = \theta_x^+$.

We illustrate these approximations through examples. For Example 1, the approximations of this section coincide with those of Section 8 because there is no dependence on Z; thus, Figure 2 applies to Algorithms 8.1–8.2. Results for Examples 2–4 are shown in Figures 11–13. The approximations are very good for the VaR contributions, particularly keeping in mind that they are extremely fast.

The approximations are less accurate for expected shortfall contributions (the right panel in each figure) than for VaR contributions. This is at least in part due to the fact that we do not know the conditional mean $\mathsf{E}[L|L \geq x]$, to which the expected shortfall contributions should sum.

The improvement in the approximation that could result from a better estimate of the conditional mean is illustrated for Example 3 in Figure 14. In this example, we have $x = 500$ and $\mathsf{E}[L] = 50.5$, so Algorithm 9.2 evaluates approximate shortfall contributions at 550.5. Using simulation, we estimate that $\mathsf{E}[L|L \geq 500]$ is about 713. Evaluating the approximation at this

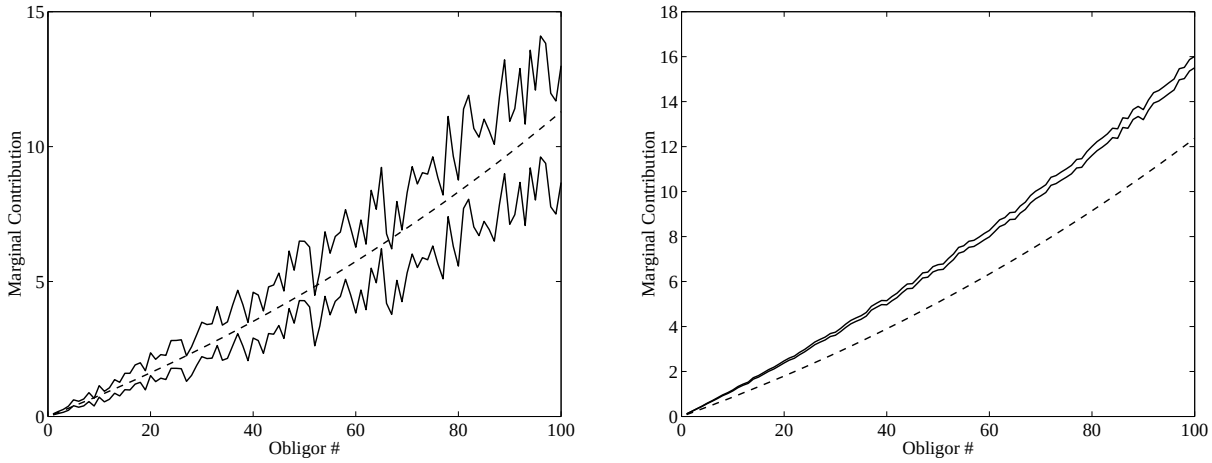

Figure 12: The dashed line shows the approximation. IS confidence intervals (solid lines) are included for reference. The graphs shows VaR contributions (left) and expected shortfall contributions (right) in Example 3.

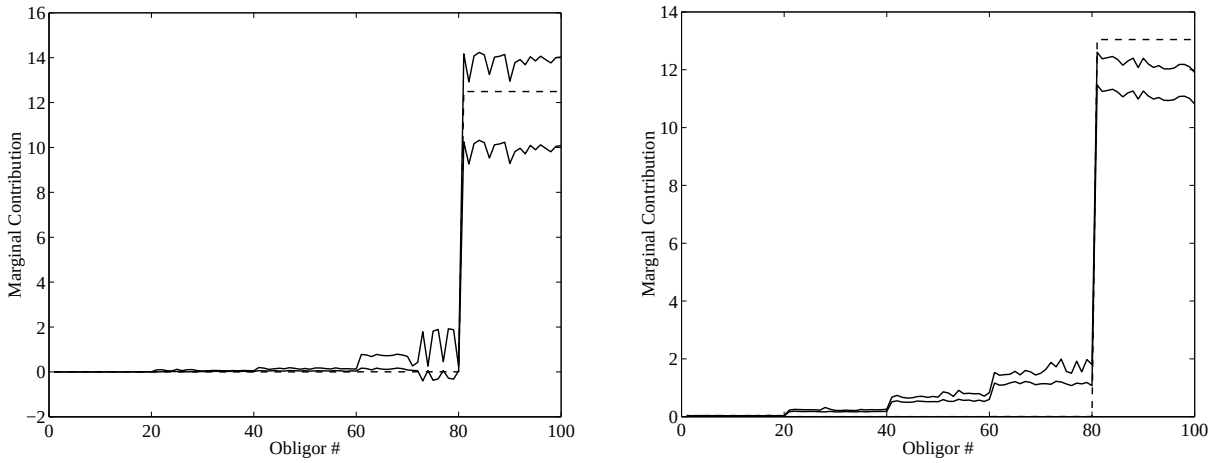

Figure 13: The dashed line shows the approximation. IS confidence intervals (solid lines) are included for reference. The graphs shows VaR contributions (left) and expected shortfall contributions (right) in Example 4.

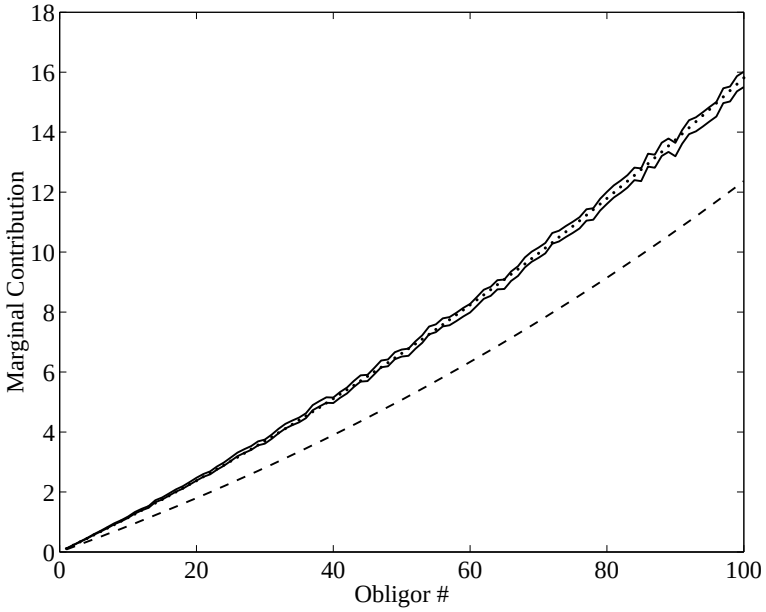

Figure 14: The dotted line shows approximate shortfall contributions for Example 3 calculated using $x = 713$, the estimated value of $\mathsf{E}[L|L \geq 500]$. The dashed line shows the approximation at $x = 550.5$. The solid lines are IS confidence intervals.

threshold produces the dotted line in the figure, which is nearly indistinguishable from the Monte Carlo results.

10 Summary

We have developed importance sampling methods and approximations for calculating the marginal risk contributions used to allocated portfolio credit risk to individual obligors. We have developed these methods for VaR and expected shortfall in the setting of the Gaussian copula model of portfolio credit risk. We summarize the application of these methods as follows:

- The importance sampling methods are very effective in improving precision in estimating marginal risk contributions. They are particularly effective at high loss quantiles.

- Combining importance sampling with an asymptotic approximation to conditional default probabilities leads to smoother estimates of the marginal risk contributions. These are quite effective both for VaR and shortfall contributions.

- Similar ideas can be used to approximate marginal risk contributions without Monte Carlo. These approximations are very fast and quite accurate for marginal VaR contributions. They are somewhat less accurate for marginal shortfall contributions.

References

[1] Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999) Coherent measures of risk, *Mathematical Finance* 9:203–228.

[2] Basel Committee on Bank Supervision (2003), The New Basel Capital Accord, third consultative document, www.bis.org.

[3] Cover, T.M., and Thomas, J. (1991) *Elements of Information Theory*, Wiley, New York.

[4] Dembo, A., and Zeitouni, O. (1998) *Large Deviations Techniques and Applications,* Second Edition, Springer-Verlag, New York.

[5] Denault, M. (2001) Coherent allocation of risk capital, *Journal of Risk* 4:1–34.

[6] Garman, M. (1999) Taking VAR to pieces, in *Hedging with Trees*, Risk Publications, London.

[7] Glasserman, P. (2004) Tail approximations for portfolio credit risk, *Journal of Derivatives* 12(2):24–42.

[8] Glasserman, P., and Li, J. (2003) Importance sampling for portfolio credit risk, *Management Science*, to appear.

[9] Gupton, G., Finger, C., and Bhatia, M. (1997) CreditMetrics Technical Document, J.P. Morgan & Co., New York.

[10] Kalkbrener, M. (2003) An axiomatic approach to capital allocation, *Mathematical Finance*, to appear.

[11] Kalkbrener, M., Lotter, H., and Overbeck, L. (2004) Sensible and efficient capital allocation for credit portfolios, *Risk* 17:S19–S24.

[12] Kurth, A., and Tasche, D. (2003) Contributions to credit risk, *Risk* (March) 84–88.

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

[14] Li, J. (2004) *Importance sampling for portfolio credit risk*, Doctoral dissertation, Columbia Business School, New York.

[15] Litterman, R. (1999) Hot spots and hedges, *Hedging with Trees*, Risk Publications, London.

[16] Merton, R.C. (1974) On the pricing of corporate debt: the risk structure of interest rates, *Journal of Finance* 29:449–470.

[17] Morokoff, W.J. (2004) An importance sampling method for portfolios of credit risky assets, pp.1668–1676 in *Proceedings of the Winter Simulation Conference*, IEEE Press, Piscataway, New Jersey.

[18] Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

[19] Van Campenhout, J.M., and Cover, T.M. (1981) Maximum entropy and conditional probability, *IEEE Transactions on Information Theory* 27:483–489.

[20] Zabell, S.L. (1980) Rates of convergence for conditional expectations, *Annals of Probability* 8:928–941.