AN IMPORTANCE SAMPLING METHOD FOR PORTFOLIOS OF CREDIT RISKY ASSETS

William J. Morokoff

New Product Research
Moody's KMV
99 Church St.
New York, NY 10007, U.S.A.

ABSTRACT

The distribution of possible future losses for a portfolio of credit risky corporate assets, such as bonds or loans, shows strongly asymmetric behavior and a fat tail as the consequence of the limited upside of credit (the promised coupon payment) and substantial downside if the corporation defaults. Because of correlation, it is not possible to fully diversify away this fat tail. Detailed correlation models require Monte Carlo simulation to determine the loss distribution for a credit portfolio. This paper describes an importance sampling method that provides substantial speed up for computing economic capital, the rare event quantile of the loss distribution that must be held in reserve by a lending institution for solvency. The method, based solely on correlation information, provides accuracy in the tail while maintaining suitable performance for statistics related to the center of the distribution. It is also suitable for long/short portfolios.

1 INTRODUCTION

Portfolios of bonds or loans generally show asymmetric returns, with limited upside and substantial but rare downside risk associated with multiple corporate defaults. Accurate assessment of how large potential losses on a credit portfolio may be plays a key role in the management of financial institutions with large credit portfolios. For example, a large bank with tens or hundreds of thousands of loans will often be required to have sufficient capital to withstand 99.9% of potential losses. Questions of to whom to lend and what rates to charge are often linked to the effect of the loan on the capital requirement and the associated return on capital.

For very large, homogenous portfolios, it is possible to derive a limiting distribution for the portfolio losses (Vasicek 1991). For portfolios that are smaller or are inhomogeneous with respect to position size, default probability, default correlation or recovery, it is usually necessary to use Monte Carlo simulation to determine the loss distribution unless substantial simplifying assumptions are made.

For large inhomogeneous portfolios, the standard Monte Carlo simulation of the underlying stochastic processes may require days of computation time to provide sufficiently accurate calculations of tail events and tail statistics. This can substantially limit the feasibility of stress testing portfolios around composition or parameters. Therefore, variance reduction methods that can reduce computation time from days to minutes or hours are of substantial interest.

For this paper, we consider a simulation derived from the structural model framework for credit first proposed by Merton (1974). Details of the model framework and the associated portfolio model can be found in Duffie and Singleton (2003) and Kealhofer and Bohn (2001). In Section 2 we provide an overview of the portfolio model and associated simulation. Section 3 describes an importance sampling method based on manipulation the correlation structure to induce greater default correlation, thereby leading to sampling more extreme loss events from the large loss tail of the distribution. Section 4 describes the application of this approach to a simple default/no default credit portfolio and discusses the convergence results and speed up over the standard Monte Carlo method. Section 5 discusses the choice of a key parameter that determines the performance of the importance sampling method. Finally, Section 6 contains a summary and conclusions about the proposed method.

2 PORTFOLIO MODEL

The key questions in modeling credit portfolio risk are i) how much will each credit instrument change in value over a given horizon, say one year, and ii) how are these value changes correlated across the portfolio. To answer these questions there are various modeling options available depending on what fundamental stochastic variables are chosen and what process is selected to determine their evolution.

For this paper, we consider the Structural Model approach first proposed in Merton (1974). Here the fundamental stochastic variable is the underlying franchise, or asset,

value of a corporation. A firm defaults when its asset value fall below a certain default point derived from its outstanding liabilities. The credit quality of the firm, and the value of firm's bonds and loans, increases as the distance between the firm's asset value and its default point increases. The stochastic process driving the evolution of the asset values of a pool of firms is often taking to be correlated geometric Brownian motion.

A detailed discussion of modeling credit portfolio risk in this framework can be found in Kealhofer and Bohn (2001). The basic idea is that over a single time horizon, the firm asset returns e_i for a portfolio of corporate credits are drawn from a multi-variate Normal distribution with a given correlation matrix. The distribution can be assumed to be normalized to mean zero and marginal variance 1. Those firms whose asset return falls below a certain threshold are considered to have defaulted during the period in question. Their value in default is drawn from a recovery distribution. The bonds or loans of firms with asset returns above the default threshold are valued based on their credit quality implied by their sampled asset value at horizon. The default threshold α_i for each name is directly related to the default probability p_i by the formula

$$\alpha_i = \Phi^{-1}(p_i),\qquad(1)$$

where $\Phi(\cdot)$ is the standard cumulative Normal distribution function. To be specific, we will consider here a fixed horizon of one year so that all default probabilities refer to the probability of a firm defaulting over the next year. The assumption of a Normal copula can be relaxed, although in practice elliptical correlation structures are almost always used.

To further simplify the model, we consider here a "Default/No Default" implementation. This means that all credit exposures are priced at par, i.e. $1 per unit notional, unless the firm defaults. In other words, a $10 million notional loan to a non-defaulted company with always have market value $10 million. In reality, loans are often valued at a discount to the par value. This leads to an additional source of uncertainty in the portfolio valuation that we will not consider here.

Further simplifying assumptions for this article will be that the portfolios are homogeneous in exposure size. For example, for a portfolio of N loans, it is assumed that each loan accounts for $1/N$ of the total exposure. Also, it will be assumed that the Loss Given Default (LGD) values of a defaulted loans (equal to 1 − recovery) are independent and identically distribution according to a Beta distribution. Finally it is assumed that the random LGD values are independent of the random asset returns.

Under these assumptions, relative to the total portfolio exposure size, the value of the portfolio is 1 under scenarios with no defaults and between 0 and 1 for scenarios with defaults. The loss over the horizon on the portfolio can

therefore be expressed using the default thresholds of Equation (1) as

$$L = \frac{1}{N}\sum_{i=1}^{N} LGD_i\, \mathrm{X}(\alpha_i - e_i),\qquad(2)$$

where X is the indicator function defined as

$$\mathrm{X}(a) = 0 \quad \text{for } a < 0$$
$$\mathrm{X}(a) = 1 \quad \text{for } a \geq 0.$$

Thus the portfolio loss for a given default scenario is just the losses on the defaulted names normalized by the total portfolio exposure.

It follows for the independence of the iid Loss Given Default samples and the asset returns that the mean of the Loss distribution, known as the Expected Loss EL, is given by

$$EL = \overline{LGD}\ \overline{p}.\qquad(3)$$

Here $\overline{LGD}$ is the expected value of the Loss Given Default random sample. The quantity $\overline{p}$ is the portfolio average default probability given by

$$\overline{p} = \frac{1}{N}\sum_{i=1}^{N} p_i.$$

It is also possible derive a formula for the loss distribution standard deviation, known as Unexpected Loss UL. If JDP_{ij} is the Joint Default Probability of two firms, defined as

$$JDP_{ij} = E\big[\mathrm{X}(\alpha_i - e_i)\cdot\mathrm{X}(\alpha_j - e_j)\big],$$

then because the asset returns have a multi-variate Normal distribution, we have

$$JDP_{ij} = \Phi^{(2)}(\alpha_i,\alpha_j,\rho_{ij}).$$

Here $\Phi^{(2)}(\cdot,\cdot,\rho)$ is the standard cumulative bivariate Normal distribution function with correlation ρ. The UL for the portfolio can then be computed as

$$\begin{aligned}UL^2 &= E\big[(L-EL)^2\big]\\ &= \overline{LGD}^2\left(\overline{JDP} - \overline{p}^2\right) + \\ &\quad \frac{1}{N}\left(\overline{LGD^2}\ \overline{p} - \overline{LGD}^2\ \overline{JDP}\right).\end{aligned}\qquad(4)$$

The quantity $\overline{JDP}$ is the portfolio average Joint Default Probability obtained by averaging the JDP_{ij} over all $N(N-1)/2$ distinct pairs.

Historical default data show that defaults are correlated by industry and region as well as to general economic conditions. It is therefore common to use a factor model to describe correlated asset returns. A description of the model developed by Moody's KMV can be found in Kealhofer and Bohn (2001), while a comparison with other correlation models is discussed in Zeng and Zhang (2001). For this paper, we assume that the factor model has the following structure. There exist M independent $N(0,1)$ standard Normal factors $z^T = (z_1, \cdots, z_M)$ describing the systematic risk to which all firms are exposed with varying weights $\beta_i^T = (\beta^i{}_1, \cdots, \beta^i{}_M)$. The weights are assumed to be normalized so that $\beta_i^T \beta_i = 1$. Each firm is also exposed to its own idiosyncratic risk, modeled by an additional independent $N(0,1)$ random variable ε_i, one for each firm. The percentage of the asset return variance explained by the systematic factors is the firm's R-square value r_i^2. The standard Normal asset return e_i can therefore be expressed as

$$e_i = \sqrt{r_i^2}\, \beta_i^T z + \sqrt{1 - r_i^2}\, \varepsilon_i.$$

Under this model it is easy to show that the pair wise asset return correlation is given by

$$\rho_{ij} = \sqrt{r_i^2 r_j^2}\, \beta_i^T \beta_j.$$

If we define Γ to be the diagonal matrix with $\Gamma_{ii} = r_i^2$, and B to be the matrix $B = [\beta_1 \cdots \beta_N]$, then the correlation matrix P can be expressed as

$$P = \Gamma^{1/2} B^T B\ \Gamma^{1/2} + I - \Gamma. \qquad (5)$$

The quantity I in Equation (5) is the identity matrix. Note that this assumes that all the credit exposures are issued by distinct firms and no firms are directly related to other firms (e.g. subsidiaries). These assumptions simplify the accounting but do not impact the effectiveness of the method.

3 IMPORTANCE SAMPLING METHOD

Importance sampling is a technique used in Monte Carlo simulation to reduce the number of simulation runs required to achieve a given accuracy. The basic idea is to change the distribution from which the random samples are taken so that a greater concentration of samples are chosen from the region in sample space that has the greatest impact on the calculation. For computing economic capital in credit portfolios, this means sampling scenarios that lead to large losses. See Glasserman and Li (2003) for alternate approach to applying importance sampling to credit portfolios, Kalos and Whitlock (1986) and Glasserman (2003) for an overview of importance sampling methodology, and Owen and Zhou (1999) for a discussion of various importance sampling issues.

As an example, consider computing capital at the 99.9% confidence level. Assume we know the capital value C, and are computing the probability of exceeding a loss of C. For a standard Monte Carlo simulation, 100,000 simulation runs would produce on average 100 sample points with losses greater than C. If the actual number of points exceeding C is N_S, then the estimate of the probability for exceeding C is $\hat{p} = N_S/100,000$. Using importance sampling, however, we might sample 5000 points with losses greater than C. Associated with each sample point we define a weight w_i such that the estimate for the probability of exceeding C becomes $\hat{p} = \left(\sum w_i \right)/100,000$, where the sum is taken over all samples that exceed C. For standard Monte Carlo, all the weights are 1, while for a good implementation of importance sampling, there will be many more points than N_S, and the associated weights will all be approximately equal and less than 1.

3.1.1 Implementation Strategy - Theory

Choosing an appropriate importance sampling distribution from which to sample (and which determines the corresponding weight function) is somewhat of an art. It is very problem dependent. Concentration sample points in one region necessarily means that other regions will be less represented; problems that depend strongly on these depleted regions may then show a decrease in accuracy of the simulation. Also, if a large number of samples appear in the region of interest, but with very small weights, while a few samples have very large weights, the accuracy will suffer. This latter problem is a common occurrence in high dimensional problems.

To address these problems, we consider here applying importance sampling to a single dimension, leaving the other dimensions unchanged. If we disregard the randomness of LGD draws, then a portfolio with N (unrelated) facilities has in effect N dimensions, each represented in the simulation by a standard Normal draw for its asset return. The goal is to find a single dimension associated with these correlated asset returns which has the largest impact on the portfolio value.

Also, when dealing with Normal distributions, a common importance sampling technique used when more samples in the tails are required is to scale up the variance. In low dimensions (in our case a single dimension) with reasonable scale up factors (say increasing the standard deviation by a factor of two), the effect is to widen the distribution and put more points further out, but still maintain the general shape so that no region is extremely depleted.

An initial thought on how to apply low dimensional importance sampling to this problem led to the idea of using the first global factor in the factor model, or possibly the first most dominant factors (as measured by the size of the average weights on the factors) from the factor model. The problem here is that these factors do not dominate enough. Increasing their variance does tend to push the samples toward more defaults, but the independent firm specific risk, for firms with R-square values in the usual range of 10 to 50%, tends to cloud this effect; and making the variance scale up large leads to the problem of many samples with very small weights. Ideally we would like to have the firm specific draws correlated so that we achieve scenarios in which many default events occur together.

The best way to achieve this effect is to orthogonalize the covariance matrix of the asset returns (i.e., the correlation matrix), and work with the eigenvector decomposition. The eigenvector direction corresponding to the largest eigenvalue is exactly the single dimension that has the largest impact on the portfolio. In this orthogonal framework we can scale up the variance in one coordinate direction (corresponding to the largest eigenvalue) independently of the other dimensions. This leads to sampling scenarios with substantially higher numbers of defaults. In addition, the weight function, derived below, takes a particularly simple form.

In order to derive the formulas, let us assume that we have the true correlation matrix for the asset returns of the facilities in the portfolio and its eigenvalue - eigenvector decomposition. For large portfolios these are unwieldy to work with; methods for addressing these problems are discussed in the next section on practical implementation. Let P be the correlation matrix, Q be the orthogonal matrix whose columns are the orthonormal eigenvectors of P, and Λ be the diagonal matrix of eigenvalues sorted such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0$. Then

$$P = Q\Lambda Q^T.$$

Let q_1 be the first column of Q, corresponding to the largest eigenvalue so that $Pq_1 = \lambda_1 q_1$.

The problem we are interested in solving is to find the probability $p(J)$ that the (normalized) loss value will lie in a given interval J. For example, we would typically take $J = [C, \infty)$ to find the probability that losses exceed the capital level C. This probability can be expressed as the integral

$$p(J) = \int_{R^N} \chi_J\left(L(e)\right)\ N(e\,|\,0, P)\ de. \qquad (6)$$

Here $N(e\,|\,0, P)$ is the Normal density function for vector e with mean zero and covariance matrix P:

$$N(e\,|\,0, P) = \frac{1}{\sqrt{(2\pi)^N\,|P|}} \exp\left(-.5\ e^T P^{-1} e\right).$$

The notation $|P|$ indicates the determinant of P. The function $\chi_J(\bullet)$ is the characteristic function of the interval J, and $L(e)$ is the (normalized) value of the portfolio loss given the asset returns e.

To apply importance sampling to this problem, we represent this integral as

$$p(J) = \int_{R^N} \chi_J\left(L(e)\right)\ w(e)\ N(e\,|\,0, \tilde{P})\ de,$$

where $\tilde{P}$ is another positive definite symmetric $N \times N$ matrix and the weight function is defined as

$$
\begin{aligned}
w(e) &= \frac{N(e\,|\,0, P)}{N(e\,|\,0, \tilde{P})} \\
&= \sqrt{\frac{|\tilde{P}|}{|P|}}\ \exp\left(-.5\ e^T \left[P^{-1} - \tilde{P}^{-1}\right] e\right)
\end{aligned}.
$$

The key to the method described here is in how $\tilde{P}$ is chosen. Let us define

$$\tilde{P} = Q\tilde{\Lambda}Q^T,$$

where $\tilde{\lambda}_1 = \sigma^2 \lambda_1$ and $\tilde{\lambda}_j = \lambda_j$, $j \geq 2$. In other words, $\tilde{P}$ is the covariance matrix which results from scaling up the largest eigenvalue by a factor σ^2. From the properties of orthogonal matrices, it follows that

$$w(e) = \sigma\ \exp\left(-.5\ \left(1 - \frac{1}{\sigma^2}\right)\frac{\left(q_1^T e\right)^2}{\lambda_1}\right).$$

Note that when $e \sim N\left(e\,|\,0, \tilde{P}\right)$, it is easy to show that $q_1^T e / \sqrt{\lambda_1} \sim N\left(\varepsilon\,|\,0, \sigma^2\right)$. This makes it possible to com-

pute what percentage of samples will have weight less than one as a function of σ. A necessary condition for improved accuracy is that over the interval J, the conditional expectation $E_J\left(w^2\right) < 1$ when the expectation is with respect to the $\tilde{P}$ measure.

In order to implement this importance sampling approach, we need to be able to sample $e \sim N\left(e \,|\, 0, \tilde{P}\right)$. Now suppose we have a random sample $e^* \sim N\left(e^* \,|\, 0, P\right)$. This is just the standard asset return sample that can be obtained from the factor model as

$$e^* = \Gamma^{1/2} B^T z + \left(I - \Gamma\right)^{1/2} \varepsilon \,, \qquad (7)$$

where z is an $M \times 1$ standard iid Normal vector and ε is an $N \times 1$ standard iid Normal vector. If we define the diagonal matrix Σ such that $\Sigma_{11} = \sigma$, and $\Sigma_{jj} = 1$, $j \geq 2$, and we define

$$e = \left(Q \Sigma Q^T\right) e^* \,,$$

then it is easy to show that $e \sim N\left(e \,|\, 0, \tilde{P}\right)$. With a little algebra we have that

$$e = \left(\sigma - 1\right)\left(q_1^T e^*\right) q_1 + e^* \,. \qquad (8)$$

Moreover, we also have that

$$w(e) = \sigma \ \exp\left(-.5 \ \left(\sigma^2 - 1\right)\frac{\left(q_1^T e^*\right)^2}{\lambda_1}\right). \qquad (9)$$

Thus we can generate the scaled up sample e and the appropriate weight function from the original unscaled sample e^*, the largest eigenvalue, the corresponding orthonormal eigenvector, and the scale up factor.

3.1.2 Practical Implementation

In order to implement the above equations it is necessary to first determine the largest eigenvalue and corresponding eigenvector. There is a simple iterative procedure known as the Power Method which does just this. See Golub and Van Loan (1996) for a detailed account of this method. Moreover, it only requires matrix-vector multiplication, i.e. computing $P*v$ for some vector v. This is important because this multiplication can be carried out without explicitly computing the potentially very large matrix P, but rather can be based on the factor model which requires storing only the factor loadings and R-squares.

The $N \times N$ correlation matrix of the facility asset returns can be expressed as in Equation (5). In a typical factor model implementation, either there will only be a few factors, or when the number of factors is large (on the order of 100), each firm only has a relatively small number of non-zero factor loadings. Therefore the $N \times M$ matrix B is generally sparse. It is clear then that $P*v$ can be computed with at most three vector-vector multiplies and two sparse matrix-vector multiplies, without any additional storage required (in particular, P need not be computed or stored).

The Power Method works as follows. Define the initial vector v_0 to be the vector of all ones: $v_0 = \left(1, 1, \cdots, 1\right)^T$. The iteration is then

$$\begin{aligned} z &= P*v_k \\ \lambda &= z(i_{\max}), \quad |z(i_{\max})| \geq |z(j)| \ \text{for all} \ j \,. \\ v_{k+1} &= z / \lambda. \end{aligned}$$

This is repeated until λ converges to the largest eigenvalue λ_1 and v_k converges to the corresponding eigenvector v_1. The orthonormal eigenvector is then computed as

$$q_1 = \frac{v_1}{\sqrt{v_1^T v_1}} \,.$$

This method usually converges after a few iterations.

With λ_1 and q_1 available, we can sample $e \sim N\left(e \,|\, 0, \tilde{P}\right)$ and compute $w(e)$. The Monte Carlo simulation runs exactly as in the standard simulation except that the loss value of Equation (2) is evaluated at the sample points given by Equation (8) as opposed to Equation (7), and that the expected values which were previously estimated as

$$E(f) \approx \frac{1}{N} \sum_{i=1}^{N} f(e_i^*)$$

are now estimated as

$$E(f) \approx \frac{1}{N} \sum_{i=1}^{N} f(e_i) w(e_i).$$

In particular, for computing the loss distribution, instead of counting the number of sample points that fall in a given bin, the estimate is now the sum of the weights of the samples points that fall in that bin.

4 APPLICATION TO SAMPLE PORTFOLIO

To illustrate the effectiveness of the method proposed in Section 3, we consider two test cases. In each case the portfolio is constructed by randomly selecting the default probabilities, the factor loadings and the R-square values based on the factor model described in Section 2 with 50 factors. The factoring loadings for each exposure in the portfolio are chosen randomly as follows. The loading on the first factor is sampled uniformly on the interval [0.21, 0.31]. The loadings on the second through fifth factors are chosen uniformly and independently on the interval [0.11, 0.21]. Finally, two factors are chosen uniformly and independently from the remaining 45 factors. If the factors chosen are identical, the factor and the next succeeding factor are selected, unless the two identical factors selected are the last, in which case, factors 6 and 50 are selected. The loadings on these two factors are sampled uniformly and independently on the interval [0, 0.1]. All other factors have a weight of zero. Finally, the factor loadings are normalized to a square sum equal to one.

The R-square value for each exposure is also sampled uniformly and independent on the interval [0.1, 0.4]. The default probability for each exposure is defined in terms of the R-square value according to the formula

$$p_i = 0.01 \cdot \left(\frac{1}{\sqrt{r_i^2}} - 1 \right). \qquad (10)$$

This formula is essentially arbitrary. It gives default probabilities in the range of 0.58% to 2.16%; these are reasonable one-year default probabilities for a portfolio of high yield, non-distressed firms. Equation (10) establishes the general property that higher R-square firms, i.e. firms that are more tied to the overall market, tend to be larger and have lower default probabilities. As stated in Section 2, it is assumed that the *LGD* distribution for each name in the portfolio is Beta distribution with mean 0.5 and standard deviation 0.25.

For the first test case we consider a portfolio of size 1000 exposures. We compare the simulation error for the standard Monte Carlo simulation to error of the importance sampling method described in Section 3 with a scale factor of $\sigma = 2$. For this portfolio, we consider the simulation error in computing the expected and unexpected loss, *EL* and *UL*, for which exact formulas are given in Equations (3) and (4). Table 1 shows the results for the estimated standard simulation error for these quantities based on a 100,000 sample points. Here σ_S refers to the standard deviation of the standard Monte Carlo method integrand so that the actual simulation error would be of size $\sigma_S / \sqrt{K}$ where K is the number of simulation runs. Similarly, σ_{IS} is the stan-

dard deviation of the importance sampling integrand (including the weight function) with respect to the importance sampling distribution. Because the importance sampling method requires only a minor amount of additional computation to evaluate Equations (8) and (9), the computational speed up is given approximately by the ratio of the squared standard errors. For this case, we see that the expected loss could run 2.4 times faster, while for the unexpected loss calculation, the speed up is about a factor of 20.

Table 1: Portfolio of 1000 Exposures

	Expected Loss	Unexpected Loss
True Value	0.5496%	0.8845%
σ_S	0.8845%	0.0552%
σ_{IS}	0.5646%	0.0122%
$\sigma_S^2 / \sigma_{IS}^2$	2.45	20.4

Substantial improvement in accuracy in the first two moments in somewhat unexpected in a method designed to increase accuracy of the rare event tail of the loss distribution. This is a consequence of the simplified modeling of the portfolio as Default/No Default. The expected and unexpected loss are both driven by default events that are enhanced under this method. For more realistic portfolio models that capture exposure value change due to credit migration (not just default), as well as changes in the exposure cash flow properties (e.g. time to maturity), we have found that the accuracy of the expected and unexpected loss calculations may deteriorate slightly relative to the standard Monte Carlo approach. This is balanced by the very substantial gains in accuracy of the tail statistics.

For the second test case, we consider a portfolio of 10000 exposures and examine the accuracy of the simulation in determining the tail confidence levels C of the loss distribution at probability levels $q = 1\%$, 0.1% and 0.01%. Again the importance sampling calculation was run based on a scale factor of $\sigma = 2$. Figure 1 shows a graph of the loss distribution on a log scale that was calculated with the importance sampling method. The expected loss for this distribution is 0.552%, while the unexpected loss is 0.87%. A truncated Gaussian distribution with similar mean and variance is also plotted for reference. From the graph it is clear that this is a very fat-tailed distribution given that the 0.1% confidence level is around 8.6% (over 9 standard deviations out, versus about 3 standard deviations for a Normal distribution), and the 0.01% confidence level is around 14% (over 15 standard deviations out compared with around 3.7 for the Normal case).

The error analysis is more straightforward for the integration problem of estimating q given the loss level C than for the order statistics problem of estimating C given q, although the same simulation is used for each case. For
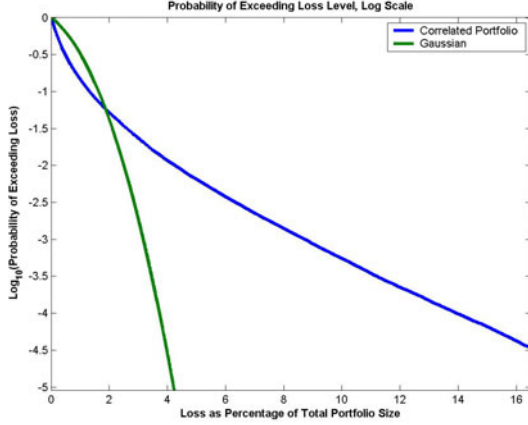
Figure 1: Portfolio Loss Distribution

the standard Monte Carlo method, the relative error of estimating the probability q of exceeding the loss C is determined by standard deviation of the random variable $\chi_J(L(e))$ with respect to the standard distribution (see Equation (6)), where $J = [C, \infty)$. Because this is just an indicator function that has value 1 with probability q, its standard deviation is $\sigma_S = \sqrt{q\,(1-q)}$. The relative error in the estimate of p is then given by

$$E\left(\left(\hat{q}-q\right)^2 / q^2\right)^{1/2} = \frac{\sigma_s}{q\,\sqrt{K}},$$

where K is the number of simulation runs. The relative error for the importance sampling simulation follows the same formula, except that the variance of the estimator is determined from

$$\sigma_{IS}^2 = \int_{R^N} \left(\chi_J\left(L(e)\right)w(e)-q\right)^2 N(e\,|\,0,\tilde{P})de \quad (11)$$

which is approximated by

$$\widehat{\sigma_{IS}^2} = \frac{1}{K-1}\sum_{j=1}^{K}\left(\chi_J\left(L_j\right)\,w_j - \hat{q}\right)^2.$$

Table 2 shows the results for the three quantile levels based on a simulation with 100,000 runs. This includes the loss estimate at each quantile based on the importance sampling calculation along with the two standard deviation confidence interval based on the standard error $m \cdot \sigma_{IS} / \sqrt{K}$ where m is the slope of the loss distribution function. The accuracy of the importance sampling and standard Monte Carlo methods is described by σ_{IS} and

σ_S. Finally, the computation speed up of using importance sampling is estimated by $\left(\sigma_S^2 / \sigma_{IS}^2\right)$.

Table 2: Tail Quantile Results for Portfolio of 10,000 Exposures Based on 100,000 Simulation Runs

Quantile q	1%	0.1%	0.01%
Loss C	4.24%	8.63%	14.0%
2σ Confidence Interval	± 0.05%	± 0.07%	± 0.14%
σ_S / q	9.95	31.6	100.
σ_{IS} / q	3.85	5.63	7.91
$\sigma_S^2 / \sigma_{IS}^2$	6.68	31.5	160.

The results in Table 2 show that the importance sampling method yields a significant improvement in accuracy that increases further into the tail. For example, at the 0.1% level and 100,000 runs, the standard Monte Carlo method gives σ_S =0.001, so that with 96% probability, the estimate of q lies in the interval [0.0008, 0.0012]. Taking the value of C to be 8.63% and the slope of the loss distribution at this point to be about 2000, this corresponds to a confidence interval in the estimate of C to be [8.2%, 9.0%]. In contrast, the importance sampling method gives σ_{IS} =0.0056, so that with 96% probability we have the estimate for q in the interval [0.00096, 0.00104]. The corresponding confidence interval for C is [8.56%, 8.70%]. In terms of computational speed, the same accuracy obtained with standard Monte Carlo at 100,000 runs could be achieved with the importance sampling method using around 3200 runs for a speed up of about a factor of 31.

Further out in the tail the accuracy improvement becomes even more pronounced. At the 0.01% level, in order to achieve a standard error of 10% relative accuracy in estimating this probability, it is necessary to use about 1 million simulation runs. This gives on average 100 points in the tail exceeding this 0.01% tail. For the importance sampling simulation, this accuracy can be achieved with under 10,000 runs. For this simulation, there were over 3200 sample points in this tail.

5   SCALE PARAMETER SELECTION

The question of the optimal choice for the scale parameter used in the importance sampling method is clearly problem dependent. However, by considering Equation (9) for the importance sampling weight function and properties of the quantile confidence levels we would like to estimate, it is possible to derive some general guidelines.

We begin by observing that the quantity $\left(q_1^T e^*\right)/\sqrt{\lambda_1}$ in Equation (9) is a $N(0,1)$ standard Normal random variable. It follows that the weight can also be considered a random variable with probability distribution function

$$P(\omega < w) = 2\bullet\left(1 - \Phi\left(\sqrt{\frac{-2\log(w/\sigma)}{\sigma^2 - 1}}\right)\right) \qquad (12)$$

for $0 \le w \le \sigma$. We consider now how this fact can be used to select σ to minimize the error of the quantile estimation while balancing the need to maintain accuracy for properties that depend on the center of the distribution.

Basis for this approach is the observation that the weight associated with a random sample from the loss distribution is generally decreasing with increasing loss. This is not strictly true because of the independence of the *LGD* draw from the weight, as well as because of the inhomogeneous nature of the portfolio. However, there is a dominant trend that the weights decrease as the losses increase. Therefore it makes sense to look for a threshold level a for the weights corresponding to the confidence level C for the loss at quantile q such that

$$q = E\left(\chi_{[0,a]}(w)\,w\right),$$

where the expectation is with respect to the distribution in Equation (12). Thus given σ and a quantile q, we can solve for a. Once a has been determined, it is then possible to compute

$$\sigma_{IS} = \left[E\left(\left(\chi_{[0,a]}(w)\,w - q\right)^2\right)\right]^{1/2}.$$

The goal is to minimize this quantity with respect to the scale factor for a given quantile.

Unless the only requirement of the calculation is a specific quantile estimation, the desire to improve the accuracy of the quantile estimator must be balanced by the need to maintain accuracy for quantities that depend on samples from the center of the asset return distribution. It is often that case that the same simulation is used to estimate statistics that depend on all parts of the distribution; in particular, the unexpected loss and derivatives of *UL* with respect to exposure size play an important role in credit risk management. Therefore we will also consider the accuracy of estimating the basic identity

$$1 = E(w)$$

in the importance sampling simulation. The standard error estimate for this calculation,

$$\sigma_w = \left[E\left((w-1)^2\right)\right]^{1/2},$$

can be shown to be an increasing function of the scale parameter. A criterion for determining a suitable scale factor is to minimize the quantity

$$\sigma_w + \sigma_{IS}/q, \qquad (13)$$

which is the sum of the relative errors, as a function of the scale parameter σ. Table 3 shows that for the quantile $q = 0.1\%$, this quantity is minimized for a scale parameter around 3, although the relative error is fairly flat for values of 2 or greater. A calculation of the variance estimator from Equation (11) for a simulation with scale factor 3 showed that the estimate of the standard error for the 0.1% quantile was indeed about 15% lower than the calculation with scale factor 2, while the error in estimating the expected weight was about 50% higher. However, because the relative error of the quantile estimation problem (the σ_{IS}/q term of Equation (13)) is about 5 times larger than the relative error of the weight calculation, the net gain in accuracy for the quantile calculation was 0.88 while the net loss in accuracy for the expected weight calculation was 0.37, so that by the criterion of Equation (13), the choice of scale factor 3 is better.

Table 3: Effect of Scale Parameter σ on Relative Error in Expected Weight and 0.1% Quantile Estimation

σ	a	σ_w	σ_{IS}/q	$\sigma_w + \sigma_{IS}/q$
1.1	0.430	0.12	19.2	19.4
1.5	0.074	0.45	6.92	7.37
2	0.035	0.72	4.48	5.20
3	0.025	1.09	3.53	4.62
4	0.025	1.37	3.51	4.88
5	0.028	1.60	3.69	5.29
6	0.031	1.81	3.91	5.72

6 CONCLUSIONS

This paper describes a method for improving the accuracy of a Monte Carlo simulation used to determine the probability distribution of losses on a portfolio of credit risky assets. The importance sampling method developed here relies on introducing a scalar parameter into the asset correlation model that may be adjusted to increase correlations, thereby inducing a greater number of correlated defaults and thus producing samples further out in the loss tail. When applied to the simplified Default/No Default portfolio model described here, the method reduces the

number of simulation runs by a factor of 30 compared with a standard Monte Carlo simulation for the typical problem of estimating the 0.1% quantile.

The method described here has the distinct advantage of relying only on the correlation structure of the model, not on any valuation properties (exposure size, cash flows, etc.) of the individual credits in the portfolio nor on any direct properties of the portfolio loss distribution itself. In particular, the method does not require shifting the mean of the samples from the distribution to heavily favor one tail; therefore the method is applicable to portfolios with both long and short positions. Underlying the method is the implicit assumption that the credit portfolio is not dominated by a small number of large exposures and that there is a significant amount of systematic risk (i.e. correlation) in the portfolio. Thus the portfolio loss is the sum of a large number of roughly similar random losses on each exposure. Without correlation, this distribution would be approximately Normal; the fat loss tail is then driven by the correlation. Credit portfolios of interest in practice always have these properties.

An additional advantage of this method is that it does not favor tail losses to the exclusion of samples in the center of the distribution. This means that one simulation can be run to determine a range of interesting portfolio statistics in addition to the far tail quantiles. The method essentially pushes probability mass into the tail by pressing down on the most probable center part of the distribution; however, the basic shape of the original probability distribution in the underlying asset returns remains the same, so that no region covered by the original distribution is neglected.

Finally, this method is applicable more broadly than just to the model described here. The assumption of a multi-variate Normal distribution for the asset returns may be relaxed. In principal this approach could apply to a multi-variate distribution for which the correlation matrix is used in the sampling process, although the calculation of the weight function may be substantially more difficult. However, the multi-variate t distribution, which is the most common alternative to the Normal distribution, shares the same elliptical correlation structure as the Normal distribution, so the implementation of the method is essentially the same as described here.

ACKNOWLEDGMENTS

REFERENCES

Duffie, D., and K. J. Singleton. 2003. *Credit Risk: Pricing, Measurement and Management.* Princeton: Princeton University Press.

Glasserman, P. 2003. *Monte Carlo Methods in Financial Engineering.* New York: Springer Verlag.

Glasserman, P., and J. Li. 2003. Importance sampling for portfolio credit risk. Available online via <http://www-1.gsb.columbia.edu/faculty/pglasserman/Other/> (accessed August 1, 2004).

Golub, G. H., and C. F. Van loan. 1996. *Matrix Computations*. 3rd ed. Baltimore: The Johns Hopkins University Press.

Kalos, M. H., and P. A. Whitlock. 1986. *Monte Carlo Methods, Volume I: Basics.* New York: John Wiley & Sons.

Kealhofer, S. , and J. Bohn. 2001. Portfolio management of default risk. Available online via <http://www.moodyskmv.com/research/portfoliotheory.html> (accessed August 1, 2004).

Merton, R. C. 1974. On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance* 29.

Owen, A. B., and Y. Zhou. 1999. Safe and effective importance sampling. Available online via <http://www-stat.stanford.edu/~owen//reports/safe.pdf> (accessed August 1, 2004).

Vasicek, O. 1991. Limiting loan loss probability distributions. Available online via <http://www. Moodyskmv.com/research/portfoliotheory. html> (accessed August 1, 2004).

Zeng, B., and J. Zhang. 2001. An empirical assessment of asset correlation models. Available online via <http://www.moodyskmv.com/research/portfoliotheory.html> (accessed August 1, 2004).

AUTHOR BIOGRAPHY

WILLIAM MOROKOFF is a director at Moody's KMV in the research department where he heads the New Product Research Group. His research focuses on Monte Carlo simulation applied to risk management and derivatives pricing in finance. He holds a Ph.D. in Mathematics from the Courant Institute at New York University. He can be contacted at <William.Morokoff@mkmv.com>.