Tail Approximations for Portfolio Credit Risk

PAUL GLASSERMAN

PAUL GLASSERMAN is the Jack R. Anderson professor of business at Columbia Business School in New York City.

This research develops approximations for the distribution of losses from default in a normal copula framework for credit risk, with particular emphasis on approximating low probabilities of large losses. The starting point is an approximation of the rate of decay of the tail of the loss distribution for multifactor heterogeneous portfolios.

This decay rate is used in three approximations: a homogeneous single-factor approximation, a saddlepoint heuristic, and a Laplace approximation. The first of these methods is the fastest, but the last two can be surprisingly accurate at very low loss probabilities. The accuracy of the methods depends primarily on the structure of the correlation matrix in the normal copula. Calculation of the decay rate requires solving an optimization problem with as many variables as there are factors driving the correlations, and this is the main computational step underlying these approximations.

Also derived is a two-moment approximation that fits a homogeneous portfolio by matching the mean and variance of the loss distribution. While the complexity of the other approximations depends primarily on the number of factors, the complexity of the two-moment approximation depends primarily on the size of the portfolio.

Calculation of the distribution of losses from default—especially the tail of the distribution—is a prerequisite to calculation of value at risk, shortfall risk, and related risk measures. Approximations for the distribution of losses are developed here in a normal copula framework for credit risk of the type associated with the CreditMetrics approach now in widespread use.

The starting point is an exponential upper bound on the tail of the loss distribution and an approximation of its rate of decay. I use these to develop several related approximations of the entire distribution. The simplest uses the decay rate to match a single-factor homogeneous portfolio to a general portfolio. The limit of the loss distribution for the homogeneous portfolio as the number of obligors increases is then used to approximate the loss distribution of the original portfolio.

A second approximation is based on a heuristic application of a saddlepoint method for approximating tail probabilities, and a third applies the classic Laplace method for approximating integrals. I also derive a two-moment approximation that fits a homogeneous portfolio by matching the mean and variance of the loss distribution; this approximation is largely unrelated to the others.

In numerical tests, none of these methods is found to be consistently more accurate than the others. It does appear, however, that when the various methods give similar results, they are all consistent with the exact values (as estimated by Monte Carlo). Thus, the approximations serve as informal checks on one another. That is, getting very different values from the various approximations provides a

warning that at least one value is not accurate. The saddlepoint heuristic and the Laplace approximation are particularly well suited to approximating very low probabilities.

The key determinant of the accuracy of the approximations (excluding the two-moment method) is the correlation structure in the normal copula used to model dependence among obligors. The approximations work best with smoothly distributed correlations and have difficulty when the correlations fall into a small number of widely dispersed clumps. The correlation matrix is often specified through a matrix of factor loadings; the approximations work best when these loadings are such that large losses tend to result from movement of the factors in a single, most likely direction. The approximations work less well when different directions of factor movements can result in losses of similar magnitude with similar probability, as can occur with highly structured block-diagonal loading matrices.

Computationally, the most demanding step in evaluating the exponential upper bound and the approximate decay rate is the solution of an unconstrained non-linear optimization problem. If the correlation matrix is specified through a factor model, then the dimension of the optimization problem equals the number of factors. For models with, say, 50 factors, the optimization problem can usually be solved in seconds from an arbitrary starting point, and much faster from a good starting point. Good starting points are determined primarily by the matrix of factor loadings. As this matrix is unlikely to change much from one day to the next, a bank that calculates risk measures daily would have a good starting point available from the previous day's calculation.

The two-moment approximation requires evaluation of bivariate normal probabilities for calculation of the variances of the exact and approximating loss distributions. For the exact distribution, the variance calculation involves $O(m^2)$ bivariate normal probabilities, where m is the number of obligors in the portfolio. The complexity of this method is thus determined primarily by the size of the portfolio rather than the number of factors.

In developing approximations, I have in mind applications in risk measurement and risk management. This influences the modeling framework we consider and also our focus on tail probabilities. It should be noted, however, that for pricing many CDOs, the key step is calculating the distribution of losses at multiple dates (as in Andersen, Sidenius, and Basu [2003]). The methods we propose here are thus potentially relevant to valuing CDOs as well.

I. PORTFOLIO CREDIT RISK: NORMAL COPULA MODEL

I consider the distribution of losses from default over a fixed time horizon—one year, for example. The notation is as follows:

m = number of obligors to which the portfolio is exposed;
Y_k = default indicator for the k-th obligor;
= 1 if the k-th obligor defaults, and 0 otherwise;
p_k = marginal probability that the k-th obligor defaults;
c_k = loss resulting from default of the k-th obligor; and
$L = c_1 Y_1 + \ldots + c_m Y_m$ = total loss from defaults.

The objective is to calculate (or approximate) the tail distribution $P(L > x)$, especially for large values of x. (We take the loss to be a positive quantity and thus look at the upper tail of the distribution.) The individual default probabilities p_k are assumed known, either from credit ratings or from the market prices of corporate bonds or credit default swaps. We take the c_k to be known constants for simplicity, although it would suffice to know the distribution of $c_k Y_k$.

The focus of most credit risk modeling lies in capturing the dependence among the default indicators $Y_1, \ldots, Y_m$. In the normal copula model (as in Gupton, Finger, and Bhatia [1997] and Li [2000]), dependence is introduced through a multivariate normal vector $(X_1, \ldots, X_m)$ of latent variables. Each default indicator is represented as

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

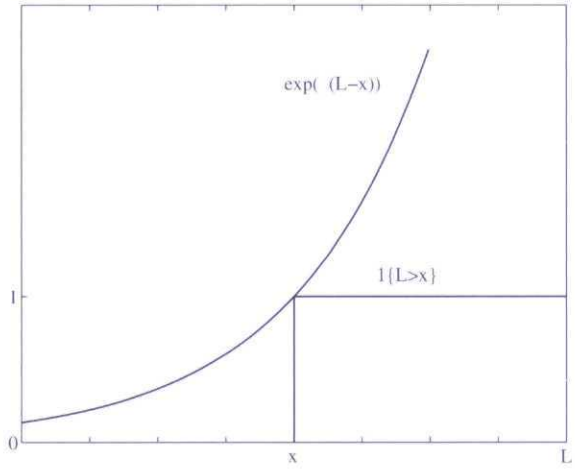
with x_k chosen to match the marginal default probability p_k. The threshold x_k is sometimes interpreted as a default boundary of the type arising in Merton [1974]. Without loss of generality, we take each X_k to have a standard normal distribution and set $x_k = \Phi^{-1}(1 - p_k)$, with Φ the cumulative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

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

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

EXHIBIT 1

Exponential Upper Bound on Indicator of Event that Loss L Exceeds Level x



$$X_k = a_{k1}Z_1 + \cdots + a_{kd}Z_d + b_k\epsilon_k \qquad (1)$$

where

$Z_1, \ldots, Z_d$ are systematic risk factors, each having an $N(0, 1)$ (standard normal) distribution;

$a_k1, \ldots, a_{kd}$ are the factor loadings for the k-th obligor, assumed non-negative;

$b_k = \sqrt{1 - (a_{k1}^2 + \cdots + a_{kd}^2)}$ so that X_k is $N(0, 1)$; and

ϵ_k is an idiosyncratic risk associated with the k-th obligor, also $N(0, 1)$ distributed.

Write A for the $m \times d$ loading matrix with (k, j)-th entry a_{kj}. The correlation between X_k and X_ℓ, $\ell \uparrow k$, is given by the (k, ℓ)-th entry of $AA^\top$. The underlying factors Z_j are sometimes given economic interpretations (as industry or regional risk factors, for example), but not always.

II. APPROXIMATING THE DECAY RATE

I now turn to how to approximate the rate of decay of the tail of the loss distribution $P(L > x)$ as x increases. The random variable L is bounded by the maximum loss

$$\ell_{\max} = \sum_{k=1}^{m} c_k$$

so $P(L > x) = 0$ for all x larger than this maximum. But because default probabilities are generally low, the tail probability $P(L > x)$ becomes negligible at values of x much smaller than ℓ_{max}. It is therefore meaningful to discuss the rate of decrease of the tail, even though L is bounded.

First, we will define a family of upper bounds on $P(L > x|Z)$, the tail of the conditional loss distribution, given the factor outcome Z. These bounds depend on a parameter θ. Next, we find the value of the parameter giving the tightest bound for given x and Z. Then, we remove the dependence on the factor outcome Z by finding the most likely value of Z leading to losses larger than x. This provides our basic approximation of the tail of the loss distribution near a given threshold x, which we then extend in various ways.

Tail Bound

An important tool in the analysis is the *conditional cumulant generating function* of L:

$$\psi(\theta, z) = \log \mathsf{E}[\exp(\theta L)|Z = z], \quad \theta \in \Re, z \in \Re^d$$

where $Z = (Z_1, \ldots, Z_d)^\top$ is the vector of risk factors in (1). This function becomes relevant to bounding the tail of the loss distribution through the observation that for any $\theta \geq 0$ and any loss level x:

$$\mathbf{1}\{L > x\} \leq e^{\theta(L-x)} \qquad (2)$$

This is illustrated in Exhibit 1. Taking the conditional expectation of both sides of (2), we get

$$P(L > x|Z) \leq \mathsf{E}[e^{\theta(L-x)}|Z] = e^{\psi(\theta,Z)-\theta x} \qquad (3)$$

so $\psi(\theta, Z)$ incorporates information about the tail of the conditional loss distribution.

To calculate the conditional cumulant generating function, observe that, conditional on the risk factors, the default indicators $Y_1, \ldots, Y_d$ become independent. The conditional default probabilities are

$$\begin{aligned} p_k(z) &= P(Y_k = 1|Z = z) \\ &= P(X_k > \Phi^{-1}(1 - p_k)|Z = z) \\ &= \Phi\left(\frac{a_k z + \Phi^{-1}(p_k)}{b_k}\right) \end{aligned} \qquad (4)$$

where $a_k = (a_{k1}, \ldots, a_{kd})$ is the (row) vector of factor loadings for the k-th obligor, and $\Phi^{-1}(1 - p_k) = -\Phi^{-1}(p_k)$ by

the symmetry of the normal distribution. The conditional cumulant generating function is then

$$\begin{aligned}\psi(\theta, z) &= \log \mathsf{E}\left[\exp\left(\theta \sum_{k=1}^{m} c_k Y_k\right) | Z = z\right] \\ &= \sum_{k=1}^{m} \log \mathsf{E}[\exp(\theta c_k Y_k) | Z = z] \\ &= \sum_{k=1}^{m} \log\left(1 + p_k(z)(e^{\theta c_k} - 1)\right) \end{aligned} \tag{5}$$

For each z, $\psi(\cdot, z)$ has certain basic properties shared by all cumulant generating functions. In particular, it passes through the origin; it is convex; and its derivative at the origin is the (conditional) mean of L, which in this case is

$$\frac{\partial}{\partial\theta}\psi(0, z) = \mathsf{E}[L|Z = z] = \sum_{k=1}^{m} p_k(z)c_k \tag{6}$$

Because L is bounded, $\psi(\theta, z)$ is finite for all θ and z.

The bound (3) holds for all $\theta \geq 0$. To get the tightest bound, we optimize the choice of θ and define, for fixed z and any $x \in (0, \ell_{max})$:

$$F_x(z) = \min_{\theta \geq 0} (\psi(\theta, z) - \theta x) = -\max_{\theta \geq 0} (\theta x - \psi(\theta, z)) \tag{7}$$

The function $x \mapsto -F_x(z)$, called the Legendre-Fenchel dual of the function $\theta \mapsto \psi(\theta, z)$, is a standard tool in approximations of low probabilities. (See, e.g., Dembo and Zeitouni [1998]. The usual definition takes the maximum over all real θ; the two definitions coincide for large x. For our purposes, it is more convenient to restrict θ to be positive from the outset.)

Because $\psi(\cdot, z)$ is strictly convex, the maximum in (7) is attained at just one point, which we denote by $\theta_x(z)$ and call the *(conditional) twisting parameter*. (As I explain in Glasserman [2004, p. 531], this parameter determines the amount by which the (conditional) default probabilities must be adjusted or twisted to make the (conditional) expected loss greater than or equal to x.) If the slope of $\psi(\cdot, z)$ at the origin is higher than x, then the maximum is attained at $\theta_x(z) = 0$; otherwise, it is attained at the unique solution to the first-order condition:

$$\frac{\partial}{\partial\theta}\psi(\theta, z) - x = 0$$

Using (6), we may therefore write

$$\theta_x(z) = \begin{cases} \text{unique } \theta \text{ such that } \partial\psi(\theta, z)/\partial\theta = x, & x > \mathsf{E}[L|Z = z] \\ 0, & x \leq \mathsf{E}[L|Z = z] \end{cases} \tag{8}$$

The twisting parameter $\theta_x(z)$ may be loosely interpreted as a measure of how much higher the loss level x is than the conditionally expected loss $\mathsf{E}[L|Z = z]$. With (8) we may write F_x as:

$$F_x(z) = \psi(\theta_x(z), z) - \theta_x(z)x \tag{9}$$

Our interest in this function stems from the observations in the Proposition.

Proposition

1. For any $0 < x < \ell_{max}$
$$P(L > x) \leq \mathsf{E}\left[e^{F_x(Z)}\right]$$
2. *The function* $z \mapsto F_x(z)$ *is (coordinatewise) increasing with* z, *decreasing with* x, and $F_x(z) \leq 0$ for all z.
3. $F_x(z) = 0$ *if and only if* $\mathsf{E}[L|Z = z] \geq x$.
4. If $F_x(\cdot)$ is concave, then
$$P(L > x) \leq e^{-J(x)} \tag{10}$$

where

$$J(x) = -\max_z\{F_x(z) - \tfrac{1}{2}z^\top z\} \tag{11}$$

These assertions are substantiated in Appendix A; here we offer some interpretation. For each loss level x, $\exp[F_x(z)]$ should be viewed as a measure of the rarity or likelihood of the event $\{L > x\}$, conditional on the factors Z taking the value z (recall Equation (3) and Exhibit 1). For this reason, we call $F_x(z)$ (or $|F_x(z)|$) the *conditional rate* at x. When $\mathsf{E}[L|Z = z] \geq x$, the third statement above says that the event is not rare because $\exp[F_x(z)] = 1$. Also, the second statement says that the event $\{L > x\}$ becomes less likely as x increases and more likely as the factor level z increases. (Recall that all the factor loadings a_{kj} are non-negative.) Similarly, $\exp(-J(x))$ should be viewed as a measure of the *unconditional* rarity of $\{L > x\}$. The point z_x at which the maximum in (11) occurs:

$$z_x = \operatorname*{argmax}_z\{F_x(z) - \tfrac{1}{2}z^\top z\} \tag{12}$$

may be interpreted as the most likely value of the factors leading to losses greater than x.

When (10) holds, it seems reasonable to take

$$\dot{J}(x) = -\frac{d}{dx}\left(\max_z\{F_x(z) - \tfrac{1}{2}z^\top z\}\right) \tag{13}$$

as an approximation of the decay rate of $P(L > x)$, in the sense that (an upper bound on) $\log P(L > x)$ has a slope of $-\dot{J}(x)$ at the point x—the derivative $\dot{J}(x)$ tells us something about how quickly the tail of the loss distribution $P(L > x)$ is decreasing at x. This is the premise of our first approximation. In fact, we view the rate $\dot{J}(x)$ as potentially more useful than the bound itself in (10). Even when the bound is too loose to be informative, the decay of the bound will often parallel the decay of the probability itself.

The validity of the concavity assumption underlying (10) is determined primarily by the loading matrix A, and fully satisfying this condition turns out to be restrictive. Concavity holds in a single-factor model of identical obligors, but it virtually never holds in a multifactor model. It would hold if all rows of the loading matrix A were multiples of each other, but such a case could be reduced to a single-factor model. We will see some evidence of the ways the structure of the matrix A determines deviations from concavity and the accuracy of our approximations.

Inspection of the proof of the proposition shows, however, that concavity is a bit more than what we need to justify (10). It would be enough for Equation (A-1) to hold at $z = z_x$, the maximizer of $F_x(z) - z^\top z/2$. What we need, loosely speaking, is for $F_x(z) - z^\top z/2$ to drop off quickly as z moves away from z_x, and this (admittedly vague) property can hold even if concavity fails. There are also ways to justify expressions like the right side of (10) as asymptotic approximations (as in Glasserman, Heidelberger, and Shahabuddin [1999] and Glasserman and Li [2003]), but I do not pursue that route here.

Properties of the Bound and Rate

From the expression in (13), it is not evident how one would calculate the derivative defining $\dot{J}(x)$. It turns out that this derivative is calculated automatically as a by-product of the maximization over z required to calculate $J(x)$.

Before explaining this point, I discuss some further properties of the rate F_x and its calculation. Evaluation of $F_x(z)$ requires finding the twisting parameter $\theta_x(z)$ by solving the equation in (8) if $x > \Sigma_k p_k(z)c_k$. Although $\theta_x(z)$ does not have a simple closed-form expression, finding it numerically is just a scalar root-finding problem. The derivatives of $\psi(\cdot, z)$ can be written out explicitly and used to implement a second-order Newton method, for example (see Appendix B). If some of the c_k are very large, then terms of the form $\exp(\theta c_k)$ can create numerical difficulties, so it is a good idea to choose units for the losses in which the c_k are not too large.

Evaluation of $J(x)$ requires optimization of $F_x(z) - z^\top z/2$ over z, which is the most computationally demanding step in the approximations. The optimization is faciliated by the availability of explicit expressions for the first- and second-order derivatives of $F_x(z)$ with respect to the components of z. These are detailed in Appendix B. With these expressions, I optimize using a line search Newton method with Armijo backtracking as in Nocedal and Wright [1999, pp. 35-42]. When there are many factors, it may be preferable to use a quasi-Newton method to avoid evaluation of the full Hessian.

Let z_x maximize $F_x(z) - z - z^\top z/2$, so that

$$J(x) = -\left(F_x(z_x) - \tfrac{1}{2}z_x^\top z_x\right) \tag{14}$$

In Appendix B, we show that

$$\dot{J}(x) = \theta_x(z_x) \tag{15}$$

To evaluate $F_x(z)$ we need to find the twisting parameter $\theta_x(z)$, so in the course of computing the optimizer z_x we evaluate $\theta_x(z_x)$ anyway. Thus, the derivative $\dot{J}(x)$ is available as a by-product of the optimization. This is an important simplification, particularly for the method below.

III. FITTING A HOMOGENEOUS PORTFOLIO

In this section and the next two, I present strategies for converting the properties in the Proposition to approximations for $P(L > x)$. First, I use the infinite-obligor limit of a homogeneous portfolio to approximate $P(L > x)$. The decay rate $\dot{J}(x)$ is used to select the approximating portfolio.

Consider, then, a model with identical default probabilities ($p_k \equiv p$), identical exposures ($c_k \equiv 1$), and identical loadings ($a_{k1} \equiv \rho$) on a single factor Z. Each default indicator takes the form

$$Y_k = \mathbf{1}\{\rho Z + \sqrt{1-\rho^2}\epsilon_k > \Phi^{-1}(1-p)\}, \quad k = 1, \ldots, m \tag{16}$$

with conditional default probabilities

$$p_k(z) = p(z) = P(Y_k = 1|Z = z)$$
$$= \Phi\left(\frac{\rho z + \Phi^{-1}(p)}{\sqrt{1-\rho^2}}\right)$$

Let $L_m = Y_1 + \ldots + Y_m$ denote the loss for this portfolio and extend the Y_k to an infinite sequence Y, Y_1, Y_2, ..., of the form in (16) using independent and identically distributed $N(0, 1)$ random variables ϵ, ϵ_1, ϵ_2.... Because the Y_k are conditionally iid (given Z) and the c_k are all equal to 1, we have

$$P(L_m/m > q) \to P(\mathsf{E}[Y|Z] > q) =$$
$$P(p(Z) > q) = 1 - G_{p,\rho}(q)$$

where

$$G_{p,\rho}(q) = \Phi\left(\frac{\Phi^{-1}(q)\sqrt{1-\rho^2} - \Phi^{-1}(p)}{\rho}\right) \tag{17}$$

for $0 < q < 1$ and $G_{p,\rho}(0) = 0$. This is the loss distribution for the infinite-obligor limit of a homogeneous portfolio. This distribution has two parameters, p and ρ.

This distribution has been derived and applied by authors including Vasicek [1991], Lucas et al. [2001], Schönbucher [2001], and Kalkbrenner, Lotter, and Overbeck [2004]. It has also been incorporated into the Basel Committee's internal ratings-based approach to credit risk.

Approximation

In using this distribution as an approximation for a general portfolio, I interpret q to be the loss level expressed as a fraction of the maximum possible loss $\ell_{\max}$. Thus, I seek an approximation of the form:

$$P(L > x) \approx 1 - G_{p,\rho}(x/\ell_{\max}), \quad 0 < x < \ell_{\max}$$

for which I need to select the parameters p and ρ.

I choose p to match the expected loss $\mathsf{E}[L]$. On the one hand, we have

$$\int_0^{\ell_{\max}} P(L > x)\, dx = \mathsf{E}[L] = \sum_{k=1}^{m} p_k c_k$$

and on the other hand we have

$$\int_0^{\ell_{\max}} [1 - G_{p,\rho}(x/\ell_{\max})]\, dx = \ell_{\max} \int_0^1 [1 - G_{p,\rho}(q)]\, dq$$
$$= p\ell_{\max} = p\sum_{k=1}^{m} c_k$$

To equate these two means, I choose $p = \bar{p}$ with

$$\bar{p} = \sum_{k=1}^{m} p_k c_k \Big/ \sum_{k=1}^{m} c_k \tag{18}$$

This is the exposure-weighted average of the default probabilities, and is thus appealing on intuitive grounds as well.

I choose ρ to match the approximate decay rate of the tail of L. For this, I pick a relatively high loss level x_1 and calculate $\dot{J}(x_1)$ for the original portfolio. Recall that I interpret this as an approximate decay rate in the sense that

$$\frac{d}{dx}\log P(L > x_1) \approx -\dot{J}(x_1) \tag{19}$$

I therefore choose ρ so that

$$\frac{d}{dq}\log(1 - G_{\bar{p},\rho}(x_1/\ell_{\max})) \approx -\dot{J}(x_1)\ell_{\max} \tag{20}$$

(Because L is discrete, its distribution is not differentiable and we are ignoring this in (19). The function $J(x)$ is in a sense a smoothed approximation of $\log P(L > x)$.)

Given the value of $\dot{J}(x_1)$, one can solve numerically for a value of ρ that makes the two sides of (20) equal. A further approximation makes it possible to write the required value of ρ explicitly, with no evident impact on the quality of the overall approximation. At high values of x, $1 - \Phi(x)$ is well approximated by $\phi(x)/x$, with ϕ the standard normal density. Accordingly, from (17) we have

$$\log(1 - G_{\bar{p},\rho}(q)) \approx -\tfrac{1}{2}\left(\frac{\Phi^{-1}(q)\sqrt{1-\rho^2} - \Phi^{-1}(p)}{\rho}\right)^2$$

Differentiating with respect to q and using the resulting derivative on the left side of (20) yields a quadratic equation for $\delta = \sqrt{1-\rho^2}$. The required root is given by

$$\delta = \frac{-b + \sqrt{b^2 + 4a\dot{J}(x_1)}}{2a} \tag{21}$$

with

$$a = \dot{J}(x_1) + \frac{\Phi^{-1}(q_1)}{\phi(\Phi^{-1}(q_1))\ell_{\max}}$$

$$b = -\Phi^{-1}(\bar{p})/\phi(\Phi^{-1}(q_1))\ell_{\max}$$

$q_1 = x_1/\ell_{\max}$ and then $\rho = \sqrt{1-\delta^2}$.

I can summarize the approximation in steps as follows:

1. Choose a high value of x_1 (where $P(L > x_1) \approx 0.001 - 0.01$) and find
   $z_{x_1} = \operatorname{argmax}_z\{F_{x_1}(z) - z^\top z/2\}$
   using, e.g., a Newton method with derivatives as in Appendix B.
2. Set $\dot{J}(x_1) = \theta_{x_1}(z_{x_1})$ with $\theta_x(z)$ as in (8). (This is available as a by-product of Step 1.)
3. Set $\rho_1 = \sqrt{1-\delta^2}$ with δ the root of the quadratic equation given in (21).
4. Set $\bar{p} = \Sigma_k p_k c_k/\ell_{\max}$.
5. Return the approximation

$$P(L > x) \approx 1 - \Phi\left(\frac{\Phi^{-1}(x/\ell_{\max})\sqrt{1-\rho_1^2} - \Phi^{-1}(\bar{p})}{\rho_1}\right) \quad (22)$$

for all $0 < x < \ell_{\max}$.

It should be noted that this procedure requires just one optimization over z to compute a single decay rate $\dot{J}(x_1)$. Once ρ_1 is selected using this decay rate, (22) provides an approximation of the entire curve of $P(L > x)$.

In choosing x_1, we use a rule of the form

$$x_1 = \sum_{k=1}^{m} p_k c_k + \nu \sum_{k=1}^{m} c_k \sqrt{p_k(1-p_k)} \quad (23)$$

If $Y_1, \ldots, Y_k$ were perfectly correlated, this would set x_1 at ν standard deviations above the mean. We generally use $\nu = 1/2$ or $\nu = 2$ because this usually results in a low value for $P(L > x_1)$. Yet (23) and these values of ν are fairly arbitrary, and we will see that one can often do better with better information about where $P(L > x)$ becomes low. There are usually a wide range of values of x_1 that produce very similar approximations. If $\dot{J}(x_1) < 0$ then x_1 is too small and should be increased.

As x approaches 0, the approximation on the right side of (22) approaches 1; the distribution for the infinite-obligor limit assigns probability 0 to a loss of exactly 0. The actual loss L is 0, with positive probability, however. We therefore cannot expect the approximation to be accurate for low values of x. But even though the approximation in (22) invariably starts at a higher level than the true tail probability, it drops very quickly and can be surprisingly accurate for even moderately high values of x.

The optimization problem that defines $J(x)$ may have multiple local maxima, but these are ignored by (22). A natural way to try to incorporate them in (22) uses a mixture of M homogeneous distributions of the form:

$$P(L > x) \approx \sum_{i=1}^{M} w_i(1 - G_{\bar{p},\rho_i}(x/\ell_{\max})) \quad (24)$$

where the weights w_i sum to 1. Each ρ_i is the value fit using a local maximum z_i. The w_i should give less weight to z_i far from the origin, weighting them in inverse proportion to their norms, for example. I have found this extension helpful in some examples with two local maxima, but to limit the number of alternatives I do not pursue this extension here.

Numerical Examples

I illustrate the approximations discussed so far with some numerical results for test portfolios (used in subsequent sections as well).

Portfolios A1 and A2. As a first illustration, consider a portfolio (which I call A1) of $m = 1{,}000$ obligors in a ten-factor model. The marginal default probabilties and exposures have the form:

$$p_k = 0.01 \cdot (1 + \sin(16\pi k/m)), \quad k = 1, \ldots, m \quad (25)$$

$$c_{k\psi} = (\lceil 5k/m \rceil)^2, \quad k = 1, \ldots, m \quad (26)$$

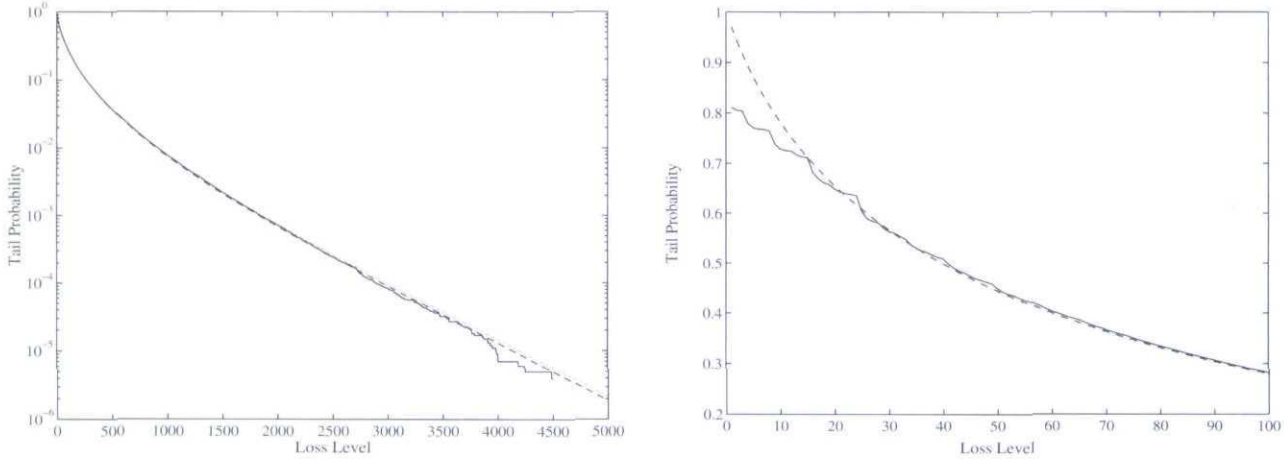
Thus, the marginal default probabilities fluctuate between 0 and 2% with a mean of 1%, and the possible exposures are 1, 4, 9, 16, and 25, with 200 obligors at each level. This is a lumpy exposure profile compared with the fixed value of c_k implicit in the homogeneous approximation.

Finally, for the factor loading matrix, we generate the a_{kj} independently and uniformly from the interval $(0, 1/\sqrt{d})$, $d = 10$. The upper limit of this interval ensures that the sum of squared entries in each row of A does not exceed 1.

In fitting a homogeneous portfolio, we need a rule for selecting the point x_1 at which to match the decay rate. To provide a systematic comparison, we use (23)

EXHIBIT 2

Homogeneous Approximations for Portfolio A1



Solid lines are Monte Carlo estimates.
The left panel displays probability on a log scale. The right panel gives a magnified view at low loss levels.

with $\nu = 1/2$ and $\nu = 2$. As already noted, these values are fairly arbitrary, and they do not necessarily provide the best possible approximation.

The left panel in Exhibit 2 compares the approximations with loss probabilities estimated using Monte Carlo simulation. The horizontal axis shows loss levels, and the vertical axis shows the probability that the loss exceeds each level, on a logarithmic scale. The solid line shows the Monte Carlo estimates; the dashed line is the approximation at $\nu = 1/2$ ($x_1 = 579$, $\rho_1 = 0.4889$); and the dotted line is the approximation at $\nu = 2$($x_1 = 2005$, $\rho_1 = 0.4859$). The three curves are nearly identical over a wide range of loss levels and tail probabilities.

The right panel of Exhibit 2 shows a magnified view of the distributions at low loss levels. The homogeneous approximations have no chance of matching the correct value at 0, but they catch up with the correct values remarkably quickly. (The wobbly pattern in the Monte Carlo curve results from the lumpy exposure profile.)

The excellent performance of the approximation in Exhibit 2 is typical of many examples I have tested in which the elements of A are randomly generated. To illustrate this point, I consider a modification referred to as Portfolio A2. The modified portfolio has just 100 obligors. The marginal default probabilities are as in (25) but multiplied by 2.5 so that they lie between 0 and 5%. These changes should make the portfolio less amenable to the approximation. The exposures have the lumpy form in (26). I consider a five-factor version of the model, with each entry of the loading matrix A uniformly distributed between 0 and $1/\sqrt{5}$.

The results are displayed in Exhibit 3. The two approximating curves (calculated at $\nu = 1/2$ and $\nu = 2$) are virtually indistinguishable and are indicated by the dashed line. The solid line shows Monte Carlo estimates.

Portfolio B. The next example has five factors and 496 obligors, grouped into sets of sizes 256, 128, 64, 32, and 16. Within each group, the obligors have identical factor loadings, given by the row vectors:

$$(1/\sqrt{6},\ 1/\sqrt{6},\ 1/\sqrt{6},\ 1/\sqrt{6},\ 1/\sqrt{6})$$
$$(1/\sqrt{5},\ 1/\sqrt{5},\ 1/\sqrt{5},\ 1/\sqrt{5},\ 0)$$
$$(1/\sqrt{4},\ 1/\sqrt{4},\ 1/\sqrt{4},\ 0,\ 0)$$
$$(1/\sqrt{3},\ 1/\sqrt{3},\ 0,\ 0,\ 0)$$
$$(1/\sqrt{2},\ 0,\ 0,\ 0,\ 0)$$

The exposures c_k are constant within groups, taking the values 25, 16, 9, 4, and 1. The default probabilities are inversely proportional to the exposures, with $p_k = 0.05/c_k$.

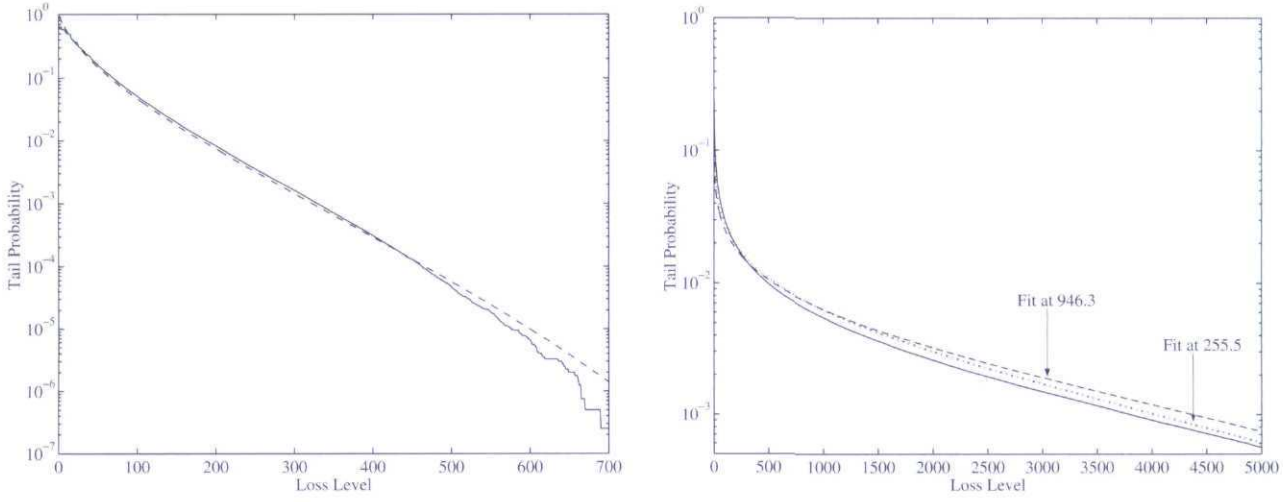
This structure is designed to be quite far from a single-factor model, but it also differs from Portfolios A1 and A2 in that the randomization in those models made all factors equally important. Here, the first three factors are of roughly equal importance, and the last two factors are significant but have somewhat less impact.

The results are displayed in the right panel of Exhibit 3, based on $\nu = 1/2$ ($x_1 = 255.5$) and $\nu = 2$ ($x_1 = 946.3$). Although not quite as effective as in A1 and A2, the curves provide decent approximations, with the one fit at $x_1 = 255.5$ a bit more accurate at higher loss levels.

Portfolio C. While it is useful to see examples where an approximation works well, it is also instructive to see examples where it does not. Our next portfolio is there-

EXHIBIT 3

Homogeneous Approximations for Portfolios A2 (left) and B (right)



Solid lines are Monte Carlo estimates.

fore specifically designed to defeat any attempt to fit a homogeneous approximation to the actual loss distribution.

We consider a two-factor model with $m = 1{,}000$ obligors and all $c_k = 1$. The first 150 obligors have a marginal default probability of 5%, a loading of 0.8 on the first factor, and a loading of zero on the second factor. The last 850 obligors have a marginal default probability of 0.1%, and they are sensitive only to the second factor, with a common loading of 0.7. This example is intended to be difficult rather than realistic.

For loss levels below 150, this portfolio behaves nearly like a single-factor model with just the first 150 obligors. At higher loss levels, the second factor and the remaining obligors become relevant. As a result, the loss distribution (as estimated through simulation) has an inflection near 150; see the solid line in Exhibit 4. The inflection begins a bit below 150, around 146.

Exhibit 4 compares several homogeneous approximations against Monte Carlo estimates. As before, I include the approximations fit using $\nu = 1/2$ and $\nu = 2$ in (23), which give $x_1 = 38.1$ and $x_1 = 127.5$, respectively. (The figure has two curves fit at 127.5. At this point, the relevant one is the lower curve.) Both of these values of x_1 are smaller than 150, so for comparison we include the curve fit at 200, above the inflection.

In all three cases, the approximation does what it is designed to do, in the sense that in each case the slope of the approximating curve runs parallel to that of the true curve at the corresponding value of x_1. But, because of the unusual shape of the distribution, this strategy does not result in accurate approximations.

The curve fit at 127.5 (the lowest one) is particularly vulnerable to the shape of the distribution. Because 127.5 is near the inflection, this approximation sees only the rapid decline in the tail probability as the loss level approaches 150. To try to match this rate of decline, it selects a much lower value of ρ (0.3170, compared with 0.6263 at 38.1 and 0.5870 at 200), and thus ends up underestimating the tail of the loss distribution. The other curve fit at 127.5 uses only the first 150 obligors in computing ρ (which results in $\rho = 0.8040$). It therefore fits the initial portion of the tail distribution quite well, but of course completely misses losses above 150.

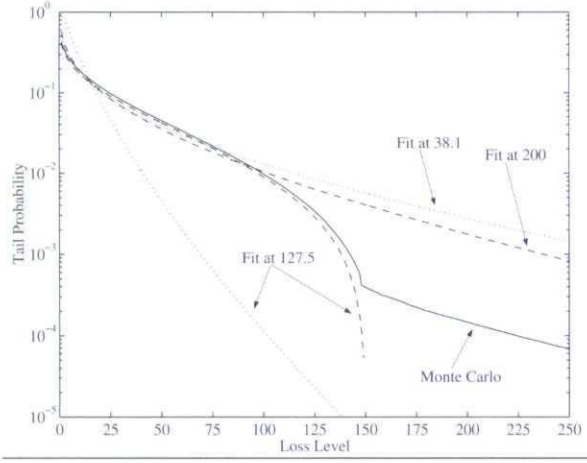
Portfolios D1 and D2. The next example, like Portfolio C, is designed to be difficult and to highlight some properties of the homogeneous approximation. I consider a portfolio of 1,000 obligors in a 21-factor model. The marginal default probabilities fluctuate as in (25), and the exposures c_k increase linearly from 1 to 100 as k increases from 1 to 1,000. The $1{,}000 \times 21$ factor loading matrix has the block structure:

$$A = \left(\begin{array}{c|ccc|c} R & F & & & G \\ & & \ddots & & \vdots \\ & & & F & G \end{array}\right), \quad G = \begin{pmatrix} g & & \\ & \ddots & \\ & & g \end{pmatrix} \tag{27}$$

with R a column vector of 1,000 entries, all equal to 0.8; F a column vector of 100 entries, all equal to 0.4; G a 100×10 matrix; and g a column vector of 10 entries, all equal

EXHIBIT 4

Homogeneous Approximations for Portfolio C



to 0.4. Think of the first factor as a marketwide factor, the next ten factors (the F block) as industry factors, and the last ten (the G block) as geographic factors. Each obligor then has a loading of 0.4 on one industry factor and on one geographic factor, as well as a loading of 0.8 on the marketwide factor. There are 100 combinations of industries and regions, and exactly ten obligors fall in each combination.

The highly structured form of the matrix A and the extreme variation among the exposures c_k make this a challenging example for the homogeneous portfolio approximation. This is illustrated in the left panel of Exhibit 5. The approximations fit using $\nu = 1/2$ ($x_1 = 2684$) and $\nu = 2$ ($x_1 = 9282$) do a poor job. We obtain a better fit at $x_1 = 30{,}000$, but it too degrades beyond probabilities of about 0.5%.

The right panel of Exhibit 5 shows results for a slightly perturbed version of the same portfolio. We replace each entry a_{kj} of the original loading matrix with

$$0.9a_{kj} + 0.1(U_{kj} - 0.5) \tag{28}$$

where the U_{kj} are independently and uniformly distributed over the unit interval. The average absolute change in the entries of A is 0.03, and the maximum absolute change is 0.13. The change in A is therefore small. The results in the right panel of Exhibit 5 indicate that adding a small amount of noise in this way makes the model more amenable to approximation.

I am not suggesting one should deliberately add noise to factor loadings. The purpose is instead to illustrate the circumstances that are favorable or unfavorable to the approximation.

The approximation works better with greater variability in the factor loadings. A loading matrix estimated from data would not have the highly structured form of the matrix A in (27) unless this structure is imposed on the estimation procedure. Companies are often sensitive to more than one industry factor and more than one geographic factor, so even with this interpretation a real loading matrix is unlikely to have the sharply delineated block structure in (27). We expect estimated loading matrices to look more like the perturbed version, and thus to be more amenable to approximation.

IV. SADDLEPOINT HEURISTIC

In its simplest form, the approximation I next propose is

$$P(L > x) \approx 1 - \Phi(\sqrt{2J(x)}) \tag{29}$$

Explaining how I arrive at this expression and how it might be refined requires a digression on saddlepoint approximations. I refer to (29) (and (31), below) as a saddlepoint heuristic to distinguish it from rigorous approximations of the type in Jensen [1995].

Background

Consider an arbitrary random variable V with cumulant generating function

$$\kappa(\theta) = \log \mathrm{E}[\exp(\theta V)]$$

The argument used in Item 1 of the Proposition similarly shows that $P(V > v) \leq \exp[-I(v)]$ where:

$$I(v) = \max_{\alpha}(\alpha v - \kappa(\alpha)) = \alpha_v v - \kappa(\alpha_v) \tag{30}$$

and α_v is the point at which the maximum is attained. Saddlepoint methods convert this bound on tail probabilities into approximations using higher-order derivatives of κ.

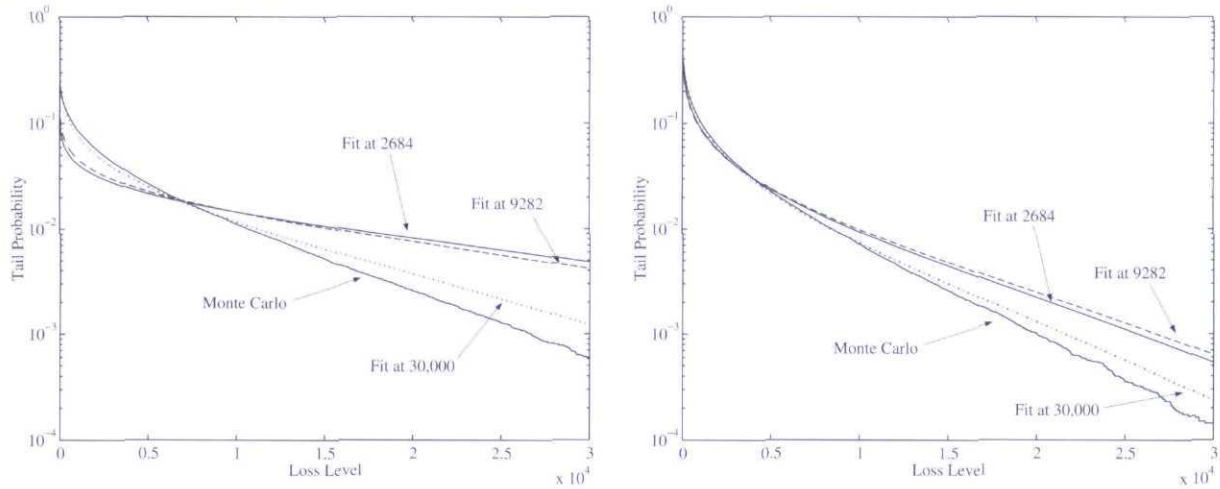
Of the many possible saddlepoint approximations in the literature, we consider the Lugannani-Rice formula, which is

$$P(V > v) \approx 1 - \Phi(r) + \phi(r)\left(\frac{1}{\lambda} - \frac{1}{r}\right)$$

with

EXHIBIT 5

Homogeneous Portfolio Approximations for Portfolios D1 (left) and D2 (right)



$$r = \sqrt{2(\alpha_v v - \kappa(\alpha_v))} = \sqrt{2I(v)}$$

and

$$\lambda = \alpha_v \sqrt{\kappa''(\alpha_v)} = I'(v)/\sqrt{I''(v)}$$

The last equality follows from the Legendre-Fenchel transformation in (30), along the same lines used to derive (B-3) in Appendix B.

Saddlepoint approximations have been used to measure portfolio credit risk by Martin, Thompson, and Browne [2001] in a model with independent obligors (or in a finite mixture of such models) and by Gordy [2002] in the CreditRisk+ model for which the cumulant generating function is known. A direct application is not possible because of the intractability of the cumulant generating function of L. Rather than try to evaluate or approximate this cumulant generating function, we approximate its Legendre-Fenchel transform.

Approximation

Set $V = \Phi^{-1}(L/\ell_{\max})$. The Proposition implies that

$$P(V < v) \leq \exp\left(-J(\Phi(v)\ell_{\max})\right)$$

so let

$$I(v) = J(\Phi(v)\ell_{\max})$$

In a purely heuristic approximation, I proceed as though $I(v)$ were in fact the Legendre-Fenchel transform of the cumulant generating function of V. The resulting approximation is

$$P(L > x) = 1 - \Phi(\sqrt{2J(x)}) + \phi(\sqrt{2J(x)})\left(\frac{\sqrt{I''(v)}}{I'(v)} - \frac{1}{\sqrt{2J(x)}}\right) \tag{31}$$

with $v = \Phi^{-1}(x/\ell_{\max})$

$$I'(v) = \dot{J}(x)\phi(v)\ell_{\max}$$

and

$$I''(v) = \ddot{J}(x)\phi^2(v)\ell_{\max}^2 - \dot{J}(x)\phi(v)v\ell_{\max}$$

I noted previously that $\dot{J}(x) = \theta_x(z_x)$. Although an explicit expression is also available for $\ddot{J}$ in (46), it is somewhat simpler to use a central difference approximation of the form

$$\ddot{J}(x) \approx \frac{J(x+h) - J(x-h)}{2h}$$

If $I(v)$ were in fact the Legendre-Fenchel dual of a cumulative generating function, it would be convex; $I''(v)$ would always be positive; and the $\sqrt{I''(v)}$ appearing in the approximation would always be meaningful. But no such guarantee applies to our heuristic use of J. Also, the last term in (31) is often quite small; omitting it yields the simplified approximation in (29).

Adjusting the Mean

Neither (29) nor (31) is consistent with the known value of the expected loss, $\mathsf{E}[L] = \sum_k p_k c_k$. To make them consistent with $\mathsf{E}[L]$, I propose scaling these approximations. This sometimes improves accuracy.

Consider any approximation of the form $P(L > x) \approx 1 - G(x)$. If we could calculate

$$m_G = \int_0^\infty (1 - G(x))\, dx$$

we could make the approximation consistent with the mean loss through the adjustment

$$P(L > x) \approx \frac{1}{m_G} \sum_{k=1}^{m} p_k c_k \,(1 - G(x)))$$

where integrating both sides over x from 0 to ∞ now yields $\mathsf{E}[L]$. In practice, accurate calculation of m_G may be difficult, especially if each evaluation of G is time-consuming. Also, the approximations in (29) and (31) are intended for large values of x, but the integral of the approximations would be heavily influenced by their values at low x where the probabilities $P(L > x)$ are greatest and the approximations least accurate.

To avoid undertaking a numerical integration of the approximations, we apply a simple adjustment based on the homogeneous portfolio distribution $G_{\bar{p},\rho}$. We fix $\bar{p}$ as in (18) and choose ρ to fit a multiple of $1 - G_{\bar{p},\rho}$ to the approximation $1 - G$. We then use this multiple to scale $1 - G$. In more detail, the approach includes steps as follows:

1. Choose points x_i and evaluate $G(x_i)$, $i = 1, \ldots, n$ (e.g., $n = 20$).
2. Find ρ to minimize the sum of squared differences
$$\sum_{i=1}^{n-1} \left(\log \frac{1 - G_{\bar{p},\rho}(x_{i+1}/\ell_{\max})}{1 - G_{\bar{p},\rho}(x_i/\ell_{\max})} - \log \frac{1 - G(x_{i+1})}{1 - G(x_i)} \right)^2$$
3. Calculate the average ratio
$$\bar{r} = \frac{1}{n} \sum_{i=1}^{n} \log \frac{1 - G_{\bar{p},\rho}(x_i/\ell_{\max})}{1 - G(x_i)}$$
4. Return the scaled approximation $e^{\bar{r}}(1 - G(x_i))$, $i = 1, \ldots, n$.

Step 2 finds the value of ρ for which $\log(1 - G_{\bar{p},\rho}(x/\ell_{\max})$ is most nearly parallel to $\log(1 - G(x))$. Step 3 calculates the average displacement between the two curves, and Step 4 shifts $\log(1 - G)$ by this amount. If $1 - G(x)$ were exactly equal to a multiple of some $\log(1 - G_{\bar{p},\rho}(x/\ell_{\max})$, the adjusted approximation would integrate to $\mathsf{E}[L]$.

A drawback of this adjustment is that it is very sensitive to the choice of points x_i at which the approximation is evaluated. In numerical examples we find that the adjustment can improve the accuracy of the approximation, but it can also make it worse. Its application requires care.

V. LAPLACE APPROXIMATION

The method here uses Item 1 of the Proposition. Calculating the upper bound $\mathsf{E}[\exp(F_x(Z))]$ is in principle a problem of integrating over the distribution of the vector of risk factors Z. This is impractical unless there are very few risk factors, especially since a separate integration is required at each loss level x.

The Laplace approximation (as in, e.g., Wong [1989, Section IX.5]) provides an alternative to numerical integration. It is based on a quadratic approximation of the logarithm of the integrand in

$$\mathsf{E}[e^{F_x(Z)}] = (2\pi)^{-d/2} \int_{\Re^d} e^{F_x(z) - z^\top z/2}\, dz$$

The quadratic approximation results from a Taylor expansion of the exponent $F_x(z) - z^\top z/2$ near the point at which this expression is maximized, i.e., near:

$$z_x = \operatorname*{argmax}_z \{F_x(z) - \tfrac{1}{2} z^\top z\}$$

This is precisely the optimization problem that appears in the definition of $J(x)$. The Laplace approximation is then

$$\mathsf{E}[e^{F_x(Z)}] \approx \frac{e^{-J(x)}}{\sqrt{\det(I - \nabla_2 F_x(z_x))}} \tag{32}$$

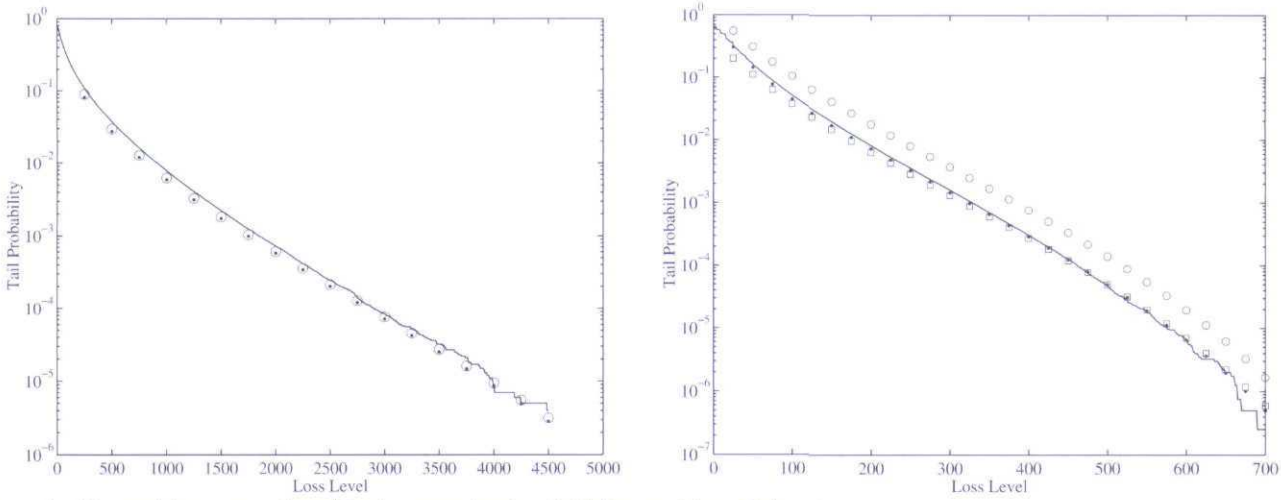
where $\nabla_2 F_x$ denotes the Hessian of F_x with respect to z. The Hessian can be calculated using formulas in Appendix B.

I apply the right side of (32) as an approximation of $P(L > x)$, although I have derived it as an approximation to a bound on this tail probability. The method in the last subsection can be used to adjust the approximation to try to match the known value of $\mathsf{E}[L]$.

Because the Laplace approximation is based on a quadratic approximation of $F_x(z) - z^\top z/2$, it works best if this function is at least concave. To put this another way,

EXHIBIT 6

Saddlepoint Heuristic (filled circles) and Laplace Approximation (empty circles) **for Portfolios A1 (left) and A2 (right)**



Squares in right panel show mean-adjusted Laplace approximation. Solid lines are Monte Carlo estimates.

the approximation uses information about the integrand only near its maximum, ignoring the possibility of local maxima. The impact of local maxima is driven primarily by the structure of the factor loading matrix A. We can reasonably expect that $F_x(z) - z^\top z/2$ will be strictly concave in the neighborhood of the global maximizer z_x, so that $\nabla_2 F_x(z_x) - I$ will be negative-definite and the denominator in (32) strictly positive.

The approximation may be summarized as follows. At each x:

1. Solve the optimization problem in (12) to find z_x and evaluate $J(x) = F_x(z) - z^\top z/2$.
2. Compute the Hessian $\nabla_2 F_x(z_x)$ using (B-6).
3. Evaluate the determinant $\det(\mathrm{I} - \nabla_2 F_x(z_x))$ and return the approximation

$$P(L > x) \approx \frac{e^{-J(x)}}{\sqrt{\det(I - \nabla_2 F_x(z_x))}}$$

After the approximation has been evaluated at multiple values of x, one may optionally apply the mean adjustment in the last subsection.

This method, like the saddlepoint heuristic, requires resolving the optimization over z (and over θ) at each x at which the approximation is evaluated. Fortunately, the optimal points $[z_x, \theta_x(z_x)]$ usually change slowly with x, so the optimal value for each point provides an excellent starting value for the optimization at the next point. Starting each optimization at the previous optimum can thus reduce the total time required.

Exhibit 6 shows both the saddlepoint heuristic (29) and the Laplace approximation for Portfolios A1 and A2. For Portfolio A1, the two approximations give nearly identical results, and both are very close to the true probabilities as estimated by Monte Carlo (the solid line in the figure). For Portfolio A2, the saddlepoint heuristic is nearly exact, but the Laplace approximation appears to be off by a constant factor (a constant displacement on a log scale). The mean adjustment in the last subsection produces the points marked by squares in the figure. The mean adjustment has a negligible effect on the saddlepoint heuristic in this example, so that case is omitted for clarity.

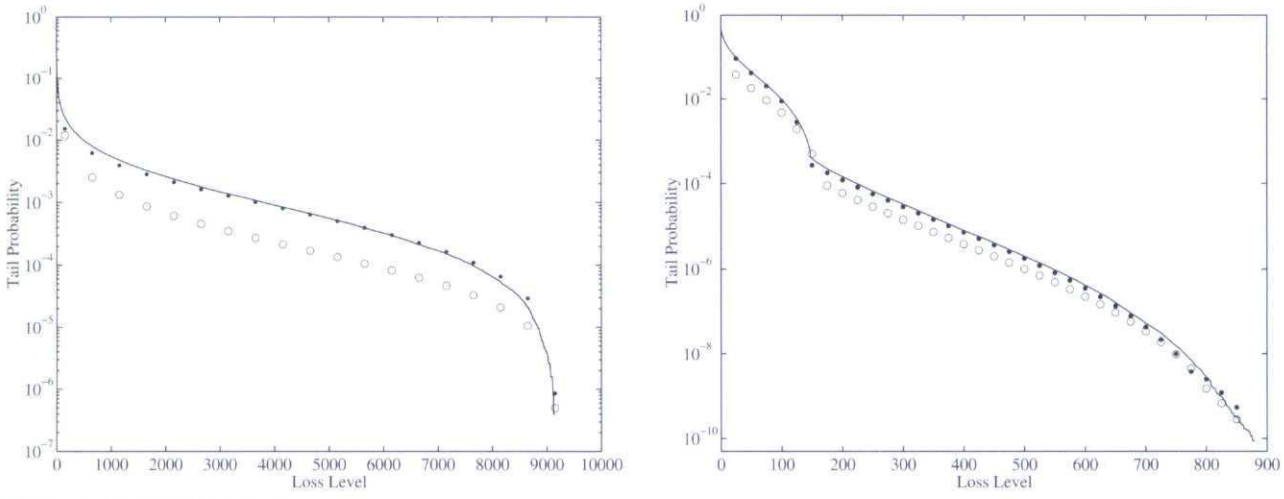
The left panel in Exhibit 7 similarly illustrates the saddlepoint heuristic and Laplace approximation for Portfolio B. The saddlepoint heuristic is again quite accurate. The Laplace approximation sags near the middle of the distribution. Although it is not shown in the graph, the values of $\mathsf{E}[\exp(F_x(Z))]$ are quite close to the true loss probabilities $P(L > x)$. This indicates that the main source of error in the Laplace approximation is the method used to approximate the integral over the normal distribution, rather than the error in using Item 1 in the Proposition as an approximation.

Portfolio C provides a more interesting example. The right panel in Exhibit 7 compares the saddlepoint heuristic (the filled circles) and the Laplace approximation (the empty circles) with Monte Carlo estimates (the solid line). The figure shows the distribution over a very wide range of loss levels.

Despite the unusual shape of the distribution, the saddlepoint heuristic is nearly exact over the full range.

EXHIBIT 7

Saddlepoint Heuristic (filled circles) and Laplace Approximation (empty circles) for Portfolios B (left) and C (right)



Solid lines are Monte Carlo estimates.

The Laplace approximation also manages to trace the shape of the distribution, although again it appears to be displaced over most of the range. Both approximations are remarkably accurate even at probabilities as low as 10^{-10}. (I obtain precise Monte Carlo estimates of such low probabilities using the importance sampling method in Glasserman and Li [2003], which draws on some of the same tools used here to derive approximations.)

Because of the special structure of this portfolio, the optimal solution z_x in (12) undergoes an abrupt change as x approaches 150. As x increases from 0 toward the inflection, the first component of z_x increases while the second stays close to 0—losses in this range are most likely to occur because of large moves in the first factor. At $x = 146$, we get an optimal solution of $z_x = (3.4230, 0.0086)$ but at $x = 147$ the optimum flips to $(0.0345, 3.4412)$ because at higher loss levels the second factor becomes more important than the first. As x continues to increase, both components of z_x increase. These features are automatically detected by both the saddlepoint heuristic and the Laplace approximation, permitting these methods to capture the shape of the loss distribution.

Exhibit 8 shows results for Portfolios D1 (left panel) and D2 (right panel). The filled circles show the saddlepoint heuristic, and the empty circles the Laplace approximation. The dotted line shows the approximation fit at $x_1 = 30{,}000$ (included for comparison).

As in the other approximations, we see greater accuracy in the perturbed case. The saddlepoint heuristic and the Laplace approximation do particularly well at low probabilities.

VI. MATCHING TWO MOMENTS

The last approximation proposed differs from the others in that it does not use the decay rate from the Proposition. Rather, it uses the limiting distribution for a homogeneous portfolio, but it selects ρ by matching variances rather than decay rates.

I noted previously that the distribution $G_{\bar{p},\rho}$ has mean $\bar{p}$. That leaves the parameter ρ free for the choice of variance. I choose ρ so that the variance of the approximating distribution matches that of the actual distribution. In order to do this, we first need to evaluate the two variances.

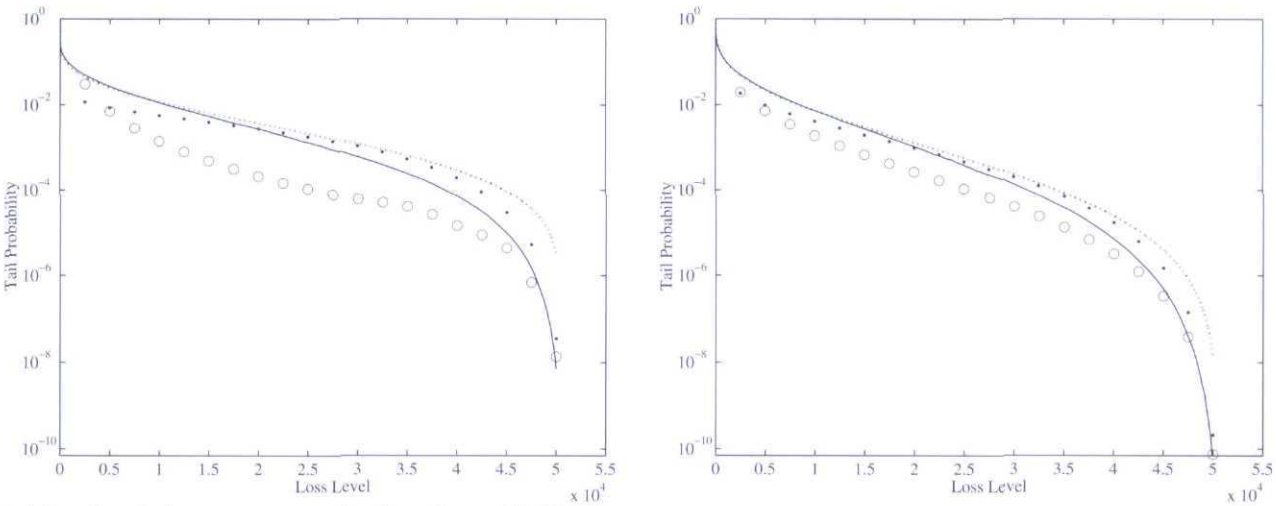
The second moment of the limiting distribution is given by

$$\begin{aligned}
&\int_0^1 q^2\, dG_{\bar{p},\rho}(q) \\
&= 2\int_0^1 q(1 - G_{\bar{p},\rho}(q))\, dq \\
&= 2\int_{-\infty}^{\infty} \Phi(u)\Phi\left(\frac{\Phi^{-1}(\bar{p}) - \sqrt{1-\rho^2}u}{\rho}\right)\phi(u)\, du \\
&= 2P\left(\xi_1 \le \xi,\ \xi_2 \le (\Phi^{-1}(\bar{p}) - \sqrt{1-\rho^2}\xi)/\rho\right) \\
&= 2P\left(\xi_1 - \xi \le 0,\ \rho\xi_2 + \sqrt{1-\rho^2}\xi \le \Phi^{-1}(\bar{p})\right) \\
&= 2\Phi_2\left(0, \Phi^{-1}(\bar{p}); -\sqrt{1-\rho^2}/\sqrt{2}\right)
\end{aligned}$$

where ξ, ξ_1, and ξ_2 are independent $N(0, 1)$ random variables, and $\Phi_2(\cdot, \cdot; r)$ denotes the bivariate normal distribution with correlation r.

EXHIBIT 8

Saddlepoint Heuristic (filled circles) and Laplace Approximation (empty circles) for Portfolios D1 and D2



Dotted lines show the homogeneous approximation with $x_1 = 30{,}000$.

The variance of the limiting distribution is thus given by:

$$\sigma_\rho^2 = 2\Phi_2\left(0, \Phi^{-1}(\bar{p}); -\sqrt{1-\rho^2}/\sqrt{2}\right) - \bar{p}^2 \tag{33}$$

This is the variance of the limiting loss as a fraction of the maximum loss $\ell_{\max}$. The variance of the limiting loss itself is therefore $\ell_{\max}^2 \sigma_\rho^2$.

For the variance of the actual loss distribution, we have

$$\sigma_L^2 = \mathsf{Var}[L] = \sum_{k=1}^{m} c_k^2 \mathsf{Var}[Y_k] + 2\sum_{k=1}^{m-1}\sum_{j=k+1}^{m} c_j c_k \mathsf{Cov}[Y_k, Y_j]$$

with

$$\mathsf{Var}[Y_k] = p_k(1-p_k)$$

and

$$\begin{aligned}\mathsf{Cov}[Y_k, Y_j] &= P(X_k \le \Phi^{-1}(p_k), X_j \le \Phi^{-1}(p_j)) - p_k p \\ &= \Phi_2(\Phi^{-1}(p_k), \Phi^{-1}(p_j); a_k a_j^\top) - p_k p_j\end{aligned}$$

in light of (1). I now solve numerically for ρ satisfying $\ell_{\max}^2 \sigma_\rho^2 = \sigma_L^2$.

The most time-consuming step in this approximation is the calculation of the loss variance σ_L^2, which requires $O(m^2)$ evaluations of the bivariate normal distribution. Agca and Chance [2003] compare several methods for evaluating the bivariate normal, and report that 125,000 evaluations take 7 to 20 seconds. At these speeds, calculation of σ_L^2 takes less than two seconds for $m = 100$ and about 30 to 90 seconds for $m = 1{,}000$—not instantaneous, but not prohibitively slow either. The calculation could be accelerated by bucketing the marginal probabilities p_k and correlations $a_k a_j^\top$ and computing just one value of the bivariate normal for each bucket.

Kalkbrenner, Lotter, and Overbeck [2004] use similar information to fit a homogeneous distribution $G_{\bar{p},\rho}$. They select $p = \bar{p}$ as I do, but they select ρ so that the value of

$$\mathsf{Var}\left[\sum_{k=1}^{m} p_k c_k X_k\right] \tag{34}$$

is the same in the approximating single-factor portfolio as in the original portfolio. This results in the explicit formula

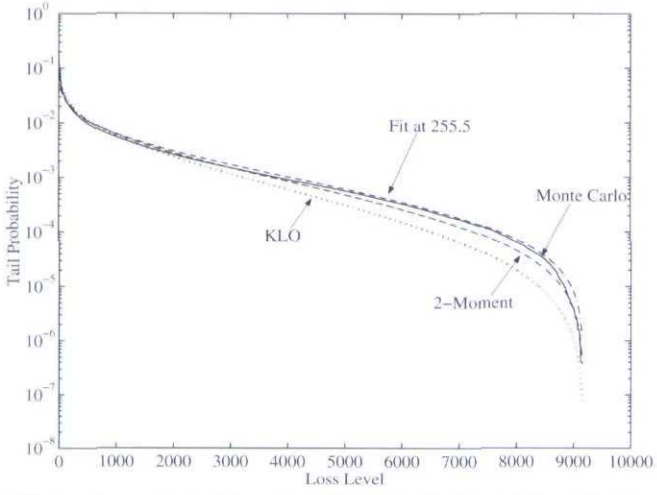
$$\rho^2 = \frac{\sum_{j,k=1}^{m} p_k c_k (a_k a_j^\top) p_j c_j - \sum_{k=1}^{m} p_k^2 c_k^2 b_k^2}{\left(\sum_{k=1}^{m} p_k c_k\right)^2 - \sum_{k=1}^{m} p_k^2 c_k^2} \tag{35}$$

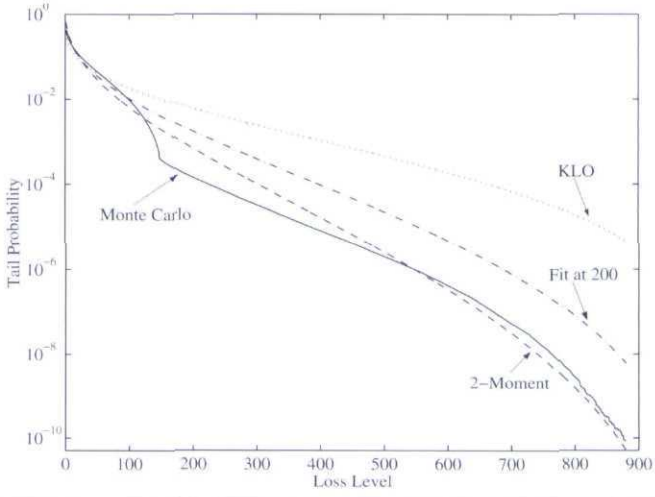
and taking the square root yields ρ. The motivation for matching the variance (34) is unclear, although this may serve as a proxy for the variance of the loss L. Because the variance in (34) requires the covariances between pairs (X_j, X_k) but not pairs (Y_j, Y_k), this method does not require calculation of bivariate normal probabilities and is thus extremely fast.

Matching two moments gives results for Portfolios

EXHIBIT 9

Comparison of Two-Moment Approximation and Other Homogeneous Approximations for Portfolios B (left) and C (right)





A1 and A2 similar to those already presented, so I proceed to the other cases. Exhibit 9 shows two-moment approximations (dashed lines) for Portfolios B and C. For comparison, the figure also includes the best approximations from Section III (fit at 255.5 and 200.0).

The two-moment approximations do quite well. Although no homogeneous approximation can capture the inflection in the actual distribution for Portfolio C, the one selected through the two-moment approximation does a rather good job overall. For this case, the method in Section III is faster, because its complexity is more dependent on the number of factors while the complexity of σ_L^2 depends mainly on m.

Exhibit 9 also shows the approximations resulting from (35), labeled KLO. This method often works very well; its comparison with the method in Section III depends on the value of x_1 at which $\dot{J}(x_1)$ is computed. Like any homogeneous approximation, it has difficulty with Portfolio C.

Exhibit 10 displays similar comparisons for Portfolios D1 (left) and D2 (right). The two-moment approximation works well in both cases. The approximation from Section III fit at 30,000 and the KLO approximation based on (35) are accurate at least out to probabilities as low as 1%. Perturbing the loading matrix improves the accuracy of the approximation in Section III, but it seems to have no effect on the two-moment and KLO approximations.

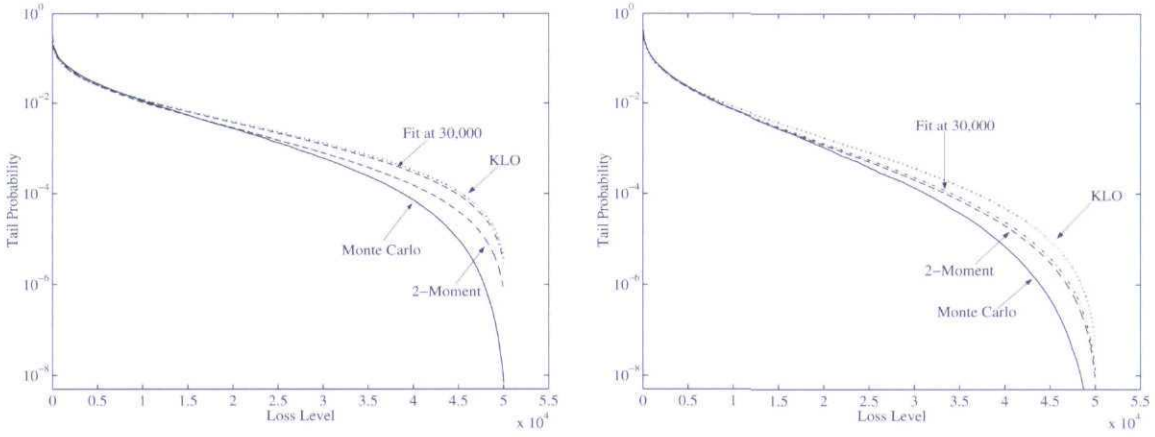
VII. SUMMARY

I have presented several methods for approximating the distribution of losses from default in a normal copula model of credit risk. These approximations provide alternatives to Monte Carlo simulation. Numerical examples indicate that they can be quite accurate. None of the methods proposed here consistently outperforms all the others, but a few patterns emerge from the examples presented here and other numerical tests:

- The method for fitting a homogeneous portfolio based on matching the expected loss and a decay rate is fast. It requires solution of an unconstrained optimization problem in as many variables as there are systematic risk factors. It gives very accurate results with randomly generated factor loading matrices. It works less well with block-diagonal loading matrices, for which different combinations of movements in the underlying factors can have similar effects.
- The saddlepoint heuristic and Laplace approximation are particularly well suited for very low loss probabilities. They are computationally more demanding because they require a separate optimization for each point at which they are evaluated. The saddlepoint heuristic is often more accurate than the Laplace approximation, although not always. The simplified version of the saddlepoint heuristic in (29) appears to be at least as accurate as the full version.
- Fitting a homogeneous portfolio by matching two moments often works well. This method requires $O(m^2)$ evaluations of the bivariate normal distribution, where m is the number of obligors. Unlike the other methods, this one is relatively insensitive to the number of factors.

EXHIBIT 10

Comparison of Two-Moment Approximation and Other Homogeneous Approximations for Portfolios D1 (left) and D2 (right)



A shortcoming of all these methods is the absence of error bounds. This makes it impossible to know whether a particular approximation works well with a particular portfolio without comparing the approximation to results from Monte Carlo simulation. But the effectiveness of the approximations appears to be determined primarily by general features of the factor loading matrix that are unlikely to change often. This makes it possible to test an approximation against Monte Carlo and, if it is found to work well, to continue to use the approximation in the future with only occasional checks against Monte Carlo.

Finally, I should note that the analysis is limited to the normal copula. The normal copula has zero tail-dependence, which in our context means that the default indicators become nearly independent when the marginal default probabilities are very low. It seems reasonable to expect that some of the tools discussed here extend to the t copula, which has strictly positive tail-dependence, but that remains a topic for investigation.

APPENDIX A

Proof of the Proposition

Proof of Proposition.

1. Because (3) holds for all $\theta \geq 0$, it also holds for the minimum over $\theta \geq 0$, so:
   $P(L > x \mid Z) \leq e^{F_x(Z)}$
   Taking expectations of both sides yields the first assertion in the Proposition.
2. Because we have assumed that all factor loadings a_{kj} are non-negative, every $p_k(z)$ in (4) is an increasing function of z, so $\psi(\theta, z)$ is an increasing function of z for all $\theta \geq 0$. Because the maximum in (7) is achieved at $\theta \geq 0$, $F_x(\cdot)$ is also an increasing function. That F_x is decreasing with x is most easily seen from Equation (B-7) in Appendix B. Taking $\theta = 0$ in (7) yields 0, so in taking the maximum over θ we get $F_x(z) \leq 0$.
3. If $x'' \mathsf{E}[L \mid Z = z]$, then the twisting parameter $\theta_x(z)$ is 0, and thus $F_x(z) = 0$. If $x > \mathsf{E}[L \mid Z = z]$, then $\theta_x(z) > 0$. The strict convexity of $\psi(\cdot, z)$ implies that for any $\theta \neq \theta_x(z)$:

$$\psi(\theta, z) > \psi(\theta_x(z), z) + \frac{\partial}{\partial\theta}\psi(\theta_x(z), z)(\theta - \theta_x(z)) = \psi(\theta_x(z), z) + x(\theta - \theta_x(z))$$

the second equality following from (8).

At $\theta = 0$ (where $(\psi(0, z) = 0)$ this reads

$$0 > \psi(\theta_x(z), z) - x\theta_x(z)$$

i.e., $F_x(z) < 0$.

For assertion (4), we use the fact that if $F_x(.)$ is concave, then for any z:

$$F_x(Z) \leq F_x(z) + \nabla F_x(z)(Z - z) \qquad \text{(A-1)}$$

where the row vector ∇F_x is the gradient of F_x with respect to z. (More generally, the argument holds using subgradients if differentiability were to fail.) Exponentiating both sides and taking expectations, we get

$$\mathsf{E}[e^{F_x(Z)}] \leq \exp\left(F_x(z) - \nabla F_x(z)z + \tfrac{1}{2}\nabla F_x(z)\nabla F_x(z)^\top\right) \qquad \text{(A-2)}$$

Now suppose z_x maximizes the concave function $F_x(z) - z^\top z/2$. Then $z_x^\top$ satisfies the first-order conditions $z_x^\top = \nabla F_x(z_x)$. Because (A-2) holds, in particular, at $z = z_x$, we then have

$$\mathsf{E}[e^{F_x(Z)}] \le \exp\left(F_x(z_x) - \tfrac{1}{2} z_x z_x^\top\right)$$

which is $\exp(-J(x))$. □

APPENDIX B

Calculation of Derivatives

In this appendix, I record expressions for derivatives that are useful in optimizing over θ (to evaluate F_x), optimizing over z (to evaluate J), and computing approximations. I use the symbol ∇ to denote gradients with respect to z and a dot (as in $\dot{J}$) to denote derivatives with respect to x. I follow the convention that gradients are row vectors. As in (4), a_k denotes the k-th row of the loading matrix A and is therefore a row vector. All other vectors are assumed to be column vectors.

Derivatives of ψ and θ

Directly from (5) we get

$$\frac{\partial}{\partial\theta}\psi(\theta, z) = \sum_{k=1}^{m} \frac{p_k(z) c_k e^{\theta c_k}}{1 + p_k(z)(e^{\theta c_k} - 1)} \tag{B-1}$$

and

$$\frac{\partial^2}{\partial\theta^2}\psi(\theta, z) = \sum_{k=1}^{m} \frac{p_k(z)(1 - p_k(z)) c_k^2 e^{\theta c_k}}{[1 + p_k(z)(e^{\theta c_k} - 1)]^2}$$

Also:

$$\nabla\psi(\theta, z) = \sum_{k=1}^{m} \frac{e^{\theta c_k} - 1}{1 + p_k(z)(e^{\theta c_k} - 1)} \nabla p_k(z) \tag{B-2}$$

where (4) gives

$$\nabla p_k(z) = \phi\left(\frac{a_k z + \Phi^{-1}(p_k)}{\sqrt{1 - a_k a_k^\top}}\right) \frac{a_k}{\sqrt{1 - a_k a_k^\top}}$$

For all z at which $\mathsf{E}[L \mid z] < x$, the relation (8) defines $\theta_x(z)$ as a differentiable function of x and z. The expressions that follow apply on this set. Where $\mathsf{E}[L \mid z] > x$, $\theta_x(z)$, is identically zero and so too are its derivatives. By differentiating both sides of (8) with respect to x, we get

$$\frac{\partial^2}{\partial\theta^2}\psi(\theta_x(z), z)\dot{\theta}_x(z) = 1$$

so

$$\dot{\theta}_x(z) = 1 \Big/ \frac{\partial^2}{\partial\theta^2}\psi(\theta_x(z), z) \tag{B-3}$$

For the gradient with respect to z, we instead differentiate (8) with respect to z to get

$$\nabla\frac{\partial}{\partial\theta}\psi(\theta_x, z) + \frac{\partial^2}{\partial\theta^2}\psi(\theta_x, z)\nabla\theta_x = 0$$

In the first term, ∇ takes the gradient with respect to the second argument of ψ. By rearranging this equation we get:

$$\nabla\theta_x(z) = -\nabla\frac{\partial}{\partial\theta}\psi(\theta_x, z) \Big/ \frac{\partial^2}{\partial\theta^2}\psi(\theta_x, z) \tag{B-4}$$

Derivatives of F_x

We continue to work on the set $\{(x, z)\colon \mathsf{E}[L \mid z] < x\}$ because $F_x(z)$ is identically 0 on the complement of this set. By differentiating (9), we get:

$$\begin{aligned}\nabla F_x(z) &= \frac{\partial}{\partial\theta}\psi(\theta_x(z), z)\nabla\theta_x(z) + \nabla\psi(\theta_x(z), z) - \nabla\theta_x(z)x \\ &= \nabla\psi(\theta_x(z), z) \end{aligned} \tag{B-5}$$

in view of (8).

Write $\nabla_2 F_x$ for the Hessian F_x. From (B-5) we get

$$\nabla_2 F_x(z) = \nabla_2\psi(\theta_x, z) + \frac{\partial}{\partial\theta}\nabla\psi(\theta_x, z)^\top \nabla\theta_x(z) \tag{B-6}$$

The second term is the matrix formed by the outer product of (B-1) and (B-4). The first term is

$$\nabla_2\psi(\theta, z) =$$

$$\sum_{k=1}^{m} \frac{e^{\theta c_k}(1 + p_k(e^{\theta c_k} - 1))\nabla_2 p_k(z) - e^{2\theta c_k}\nabla p_k(z)^\top \nabla p_k(z)}{[1 + p_k(z)(e^{\theta c_k} - 1)]^2}$$

with

$$\nabla_2 p_k(z) = -\left(\frac{a_k z + \Phi^{-1}(p_k)}{1 - a_k a_k^\top}\right)\phi\left(\frac{a_k z + \Phi^{-1}(p_k)}{\sqrt{1 - a_k a_k^\top}}\right) a_k^\top a_k$$

and

$$\nabla p_k^\top(z)\nabla p_k(z) = \phi^2\left(\frac{a_k z + \Phi^{-1}(p_k)}{\sqrt{1 - a_k a_k^\top}}\right) a_k^\top a_k$$

Also:

$$\frac{\partial}{\partial x}F_x(z) = \frac{\partial}{\partial\theta}\psi(\theta_x, z)\dot{\theta}_x - \dot{\theta}_x x - \theta_x(z) = -\theta_x(z) \tag{B-7}$$

Derivatives of J

With z_x as in (14), we have

$$J(x) = \theta_x(z_x)x - \psi(\theta_x(z_x), z_x) + \tfrac{1}{2} z_x^\top z_x$$

Differentiation yields

$$\begin{aligned}\dot{J}(x) &= \dot{\theta}_x x + \nabla\theta_x \dot{z}_x x + \theta_x - \frac{\partial}{\partial\theta}\psi(\theta_x(z), z)[\dot{\theta}_x + \nabla\theta_x \dot{z}_x] - \\ &\quad \nabla\psi(\theta_x, z_x)\dot{z}_x + z_x^\top \dot{z}_x \\ &= \theta_x - [\nabla\psi(\theta_x, z_x) - z_x^\top]\dot{z}_x \\ &= \theta_x\end{aligned}$$

with θ_x evaluated at z_x in each instance. Here, the second equality follows from (8) and the third from (B-5) and the first-order condition $\nabla F_x(z_x) = z_x^\top$. We have thus verified the important simplification in (15).

For the second derivative of J, we have

$$\ddot{J}(x) = \dot{\theta}_x(z_x) + \nabla\theta_x(z_x)\dot{z}_x \tag{B-8}$$

To evaluate $z_x^\top$ (the derivative of z_x with respect to x), we differentiate both sides of the first-order conditions $\nabla F_x(z_x) = z_x^\top$ to get

$$\frac{\partial}{\partial x}\nabla F_x(z_x)^\top + \nabla_2 F_x(z_x)\dot{z}_x = \dot{z}_x$$

We can use (B-7) to replace the first term with $-\nabla\theta_x(z_x)^\top$ and thus get

$$\dot{z}_x = -(I - \nabla_2 F_x(z_x))^{-1}\nabla\theta_x(z_x)^\top$$

provided the indicated matrix inverse exists. Using this in (B-8), we get

$$\ddot{J}(x) = \dot{\theta}_x - \nabla\theta_x(I - \nabla_2 F_x(z_x))^{-1}\nabla\theta_x(z_x)^\top \tag{B-9}$$

The derivatives of θ_x can be evaluated as in (B-3) and (B-4).

ENDNOTE

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

REFERENCES

Agca, S., and D.M. Chance. "Speed and Accuracy Comparison of Bivariate Normal Distribution Approximations for Option Pricing." *Journal of Computational Finance*, 6 (2003), pp. 61-96.

Andersen, L., J. Sidenius, and S. Basu. "All Hedges in One Basket." *Risk*, 16 (2003), pp. 67-72.

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

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

Glasserman, P., P. Heidelberger, and P. Shahabuddin. "Asymptotically Optimal Importance Sampling and Stratification for Pricing Path-Dependent Options." *Mathematical Finance*, 9 (1999), pp. 117-152.

Glasserman, P., and J. Li. "Importance Sampling for Portfolio Credit Risk." *Management Science*, forthcoming 2004.

Gordy, M.B. "Saddlepoint Approximation of CreditRisk+." *Journal of Banking and Finance*, 26 (2002), pp. 1335-1353.

Gupton, G., C. Finger, and M. Bhatia. "CreditMetrics Technical Document." J.P. Morgan & Co., 1997.

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

Kalkbrenner, M., H. Lotter, and L. Overbeck. "Sensible and Efficient Capital Allocation for Credit Portfolios." *Risk*, 17 (2004), pp. S19-S24.

Li, D. "On Default Correlation: A Copula Function Approach." *The Journal of Fixed Income*, 9 (2000), pp. 43-54.

Lucas, A., P. Klaasen, P. Spreij, and S. Straetmans. "An Analytical Approach to Credit Risk in Large Corporate Bond and Loan Portfolios." *Journal of Banking and Finance*, 25 (2001), pp. 1635-1664.

Martin, R., K. Thompson, and C. Browne. "Taking to the Saddle." *Risk*, 14 (2001), pp. 91-94.

Merton, R.C. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance*, 29 (1974), pp. 449-470.

Nocedal, J., and M. Wright. *Numerical Optimization*. New York: Springer-Verlag, 1999.

Schönbucher, P. "Factor Models: Portfolio Credit Risks When Defaults Are Correlated." *The Journal of Risk Finance*, 3 (2001), pp. 45-56.

Vasicek, O. "Limiting Loan Loss Probability Distribution." KMV Corporation, 1991.

Wong, R. *Asymptotic Approximations of Integrals*. San Diego, CA: Academic Press, 1989.