OPERATIONS RESEARCH
Vol. 56, No. 3, May–June 2008, pp. 593–606
ISSN 0030-364X | EISSN 1526-5463 | 08 | 5603 | 0593

DOI 10.1287/opre.1080.0513

Portfolio Credit Risk with Extremal Dependence: Asymptotic Analysis and Efficient Simulation

Achal Bassamboo
Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston, Illinois 60208, a-bassamboo@kellogg.northwestern.edu

Sandeep Juneja
Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai-400005, India, juneja@tifr.res.in

Assaf Zeevi
Graduate School of Business, Columbia University, New York, New York 10027, assaf@gsb.columbia.edu

We consider the risk of a portfolio comprising loans, bonds, and financial instruments that are subject to possible default. In particular, we are interested in performance measures such as the probability that the portfolio incurs large losses over a fixed time horizon, and the expected excess loss given that large losses are incurred during this horizon. Contrary to the normal copula that is commonly used in practice (e.g., in the CreditMetrics system), we assume a portfolio dependence structure that is semiparametric, does not hinge solely on correlation, and supports extremal dependence among obligors. A particular instance within the proposed class of models is the so-called t-copula model that is derived from the multivariate Student t distribution and hence generalizes the normal copula model. The size of the portfolio, the heterogeneous mix of obligors, and the fact that default events are rare and mutually dependent make it quite complicated to calculate portfolio credit risk either by means of exact analysis or naïve Monte Carlo simulation. The main contributions of this paper are twofold. We first derive sharp asymptotics for portfolio credit risk that illustrate the implications of extremal dependence among obligors. Using this as a stepping stone, we develop importance-sampling algorithms that are shown to be asymptotically optimal and can be used to efficiently compute portfolio credit risk via Monte Carlo simulation.

Subject classifications: portfolio; credit; asymptotics; expected shortfall; simulation; importance sampling; rare events; risk management.
Area of review: Financial Engineering.
History: Received July 2005; revision received February 2006; accepted June 2006.

1. Introduction

Market conditions over the past few years combined with regulatory arbitrage have led to significant interest and activity in trading and transferring of credit-related risk. Because most financial institutions are exposed to multiple sources of credit risk, a portfolio approach is needed to adequately measure and manage this risk. One of the most fundamental problems in this context is that of modeling dependence among a large number of obligors (consisting, for example, of companies to which a bank has extended credit), and assessing the impact of this dependence on the likelihood of multiple defaults and large losses.

The event of default for an individual obligor within the portfolio is often captured using so-called threshold models. These models stipulate that an obligor defaults when an appropriate state variable exceeds (or falls below) a suitably chosen threshold. This idea underlies essentially all models that descend from Merton's seminal firm-value work (cf. Merton 1974). The state variables associated with each obligor are typically modeled using latent variables that may arise from factor analysis, and thus summarize common macroeconomic or industry-specific effects. The dependence structure that governs the resulting multivariate default distribution is called a *copula function*. In particular, a copula decouples the risk associated with the portfolio dependence structure from the individual risks of each obligor. Although there are numerous copula functions that can serve such a purpose, the *normal copula*, which assumes that the latent variables follow a multivariate normal distribution, is one of the most widely used models in practice. It has been incorporated into many popular risk-management systems such as J. P. Morgan's CreditMetrics (cf. Gupton et al. 1997), Moody's KMV system (cf. Kealhofer and Bohn 2001), and is also prominently featured in the latest Basel accords that regulate capital allocation in banks (cf. Basel Committee on Banking Supervision 2002). See also Li (2000), the survey paper by Crouhy et al. (2000), and the recent monograph by McNeil et al. (2005).

In recent years, empirical work has argued that financial variables often exhibit stronger dependence than that captured in the correlation-based normal model. The stronger linkage is often manifested in large joint movements. In particular, in the credit risk context it has been argued that the main source of risk in large balanced loan portfolios is the occurrence of many near simultaneous defaults—what might be termed as "extreme credit risk." These observations strongly suggest that in many instances the normal copula may not be an adequate way to model dependencies. An attractive alternative to the normal model is one based on the multivariate Student t distributions, known as the *t-copula model*. While generalizing the normal copula model, the t-based model remains simple, parsimonious, and analytically tractable. Recent work has shown that, at least in certain instances, this model provides a better fit to empirical financial data in comparison with the normal copula (see, e.g., Mashal and Zeevi 2003).

Unlike the normal copula, the t-based model supports *extremal dependence* between the underlying variables. Roughly speaking, this means that variables may simultaneously take on very large (or small) values with nonnegligible probability; for further discussion, see Embrechts et al. (2003). A useful interpretation of extremal dependence follows from the construction of a multivariate t distribution as a ratio of a multivariate normal and the square root of a scaled chi-squared random variable. When the denominator takes values close to zero, coordinates of the associated vector of the t-distributed random variable may register large co-movements (see further discussion in Embrechts et al. 2003, Frey and McNeil 2003, and Glasserman et al. 2002). Hence, the chi-squared random variable plays the role of a "common multiplicative shock."

This paper is concerned with consequences of extremal dependence on the risk of large heterogeneous credit portfolios. The model that we stipulate builds on the latent variable approach and blends in a common multiplicative shock. The distributional assumptions we make are quite general, and the model is, hence, reasonably flexible. One can view the copula structure that arises from our model as being essentially semiparametric, wherein a designated parameter captures the extent of extremal dependence present in the portfolio (roughly speaking, this parameter governs the behavior of common shock distribution near zero). The t-copula model discussed above is one particular instant that is contained within our model. The main objective of this paper is to derive tractable procedures for computing common risk measures such as the probability of large portfolio losses and the *expected shortfall*, i.e., the expected excess loss given that there are large portfolio losses. The latter also plays an important role in the pricing of various instruments such as credit baskets, collateralized debt obligations, and options on credit baskets. The approach we take is first to develop asymptotic approximations, which in turn form the basis for devising provably efficient importance-sampling (IS) algorithms for estimation of the above portfolio performance measures. In doing so, we are also able to articulate in a mathematically precise manner the effects of extremal dependence on the portfolio risk, and contrast this to the more standard normal-based theory.

The main contributions of this paper include the following.

- We derive sharp asymptotics for two common risk measures: the portfolio loss distribution and expected shortfall (see Theorems 1 and 2). These results illustrate in a precise manner how extremal dependence affects the portfolio risk in a manner that is quite different from the normal copula model .
- We construct two IS algorithms to efficiently estimate the risk of a portfolio via simulation. The first is an algorithm that uses an exponential twist, and the second algorithm uses a variant of hazard-rate twisting (see, e.g., Juneja and Shahabuddin 2006 for a discussion on these importance-sampling techniques). Both algorithms are shown to achieve maximal variance reduction in a suitable asymptotic sense: the first in the stronger sense of bounded relative error (see Theorem 3), and the second in the weaker sense of logarithmic efficiency (see Theorem 4). The second algorithm has significant computational advantages over the first.

Numerical results illustrate the asymptotic results and performance of the algorithms, as well as their respective merits.

Based on the results detailed above, we also contrast the t-copula and the normal copula models in a simple single-factor setting. When the inputs to both models are identical, i.e., the obligors have the same marginal default probabilities and latent variables have a correlation of $0 < \rho < 1$, then we conclude the following: if the probability of large losses is of order $\mathcal{O}(p)$ in the t-copula model, then under the normal copula model it is of order $\mathcal{O}(p^{1/\rho^2})$. This dramatic difference clearly illustrates the importance of specifying the correct credit risk model. (See also the discussion in Frey and McNeil 2001, which builds on simulation studies.)

This paper is organized as follows. This section ends with a review of related literature that places the contributions of this paper within the context of existing work. Section 2 describes the model. Sections 3 and 4 contain our main results: the former derives the asymptotics and the latter describes the IS algorithms and investigates their performance. Section 5 presents numerical results, and §6 contains some discussion and concluding remarks. Proofs are relegated to two appendices that can be found in the online companion: Appendix A contains the proof of the main results and Appendix B gives proofs of auxiliary results. The online companion can be found at http://or.journal.informs.org/.

Related Literature and Positioning of This Paper. Threshold-based models for portfolio credit risk are widely used in practice; see, for example, CreditMetrics (cf. Gupton et al. 1997) and Moody's KMV system (cf. Kealhofer and Bohn 2001), both of which use the normal copula as a model for the portfolio dependence structure. The recent work by Glasserman and Li (2003) develops large deviation asymptotics for the probability of large losses, and importance-sampling simulation procedures for homogeneous portfolios within the normal copula framework.

Threshold models with nonnormal dependence structure have recently been proposed and studied by Frey et al. (2001) and Frey and McNeil (2003). The latter also formulates nonnormal threshold models for credit portfolios that are based on a mixing distribution; our common shock model falls into this category. Frey and McNeil (2003) also describes an approach to modeling seniority trenches. For further references on this topic, see the recent monograph by McNeil et al. (2005).

Although our work focuses on a general model for extremal dependence, and in that sense is quite distinct from the normal copula model studied in Glasserman and Li (2003), it also shares several common threads with their paper. As in Glasserman and Li (2003), our work also develops an asymptotic regime, but in contrast to their work, which derives logarithmic-scale large deviations asymptotics, we establish sharp asymptotic approximations that are more accurate. In addition, we develop these sharp asymptotics for expected shortfall, a risk measure that is of wide interest in practice. The IS techniques we develop in this paper emphasize the common shock structure of our model, and hence are significantly different from those in Glasserman and Li (2003). Our exponential twisting IS procedure exhibits bounded relative error, a stronger notion of asymptotic optimality than that established by Glasserman and Li (2003), and we also explore further IS techniques based on hazard-rate twisting that are much easier to implement, and yet enjoy good variance reduction properties. As we indicate in §6.2, due to the common shock structure of our model, the asymptotic analysis, as well as the proposed importance-sampling techniques, generalize easily to the multifactor model. In contrast, the work of Glasserman and Li (2003) is restricted to the single-factor case, and does not extend easily to a setting with multiple factors. Finally, we derive IS algorithms for the expected shortfall of credit portfolios, which can be used for both risk management and pricing purposes and provide efficient computational tools for large problems.

In the specific case of a t-copula, the recent work of Schloegel and O'Kane (2005) uses an asymptotic approximation for a homogeneous portfolio loss distribution, and for this approximation derives explicit formulas for the density of the loss distribution. The nature of the approximation is based on the strong law of large numbers, conditioned on the common shock variable (see also the general asymptotic detailed in Frey and McNeil 2003, Proposition 4.5). It is worth noting that these types of approximations end up relying on the entire distribution of the common shock in a potentially complicated manner, and must typically be evaluated numerically. In particular, they do not explicitly articulate the effects of extremal dependence that are present in t-copula models. In contrast, our common shock model hinges on a more general semiparametric assumption for the mixing distribution, which encompasses several cases of practical interest, including the t-copula. Unlike the work of Schloegel and O'Kane (2005), our asymptotic approximation for the tail of the loss distribution is simple enough to elucidate the effects of extremal dependence in a precise and intuitive manner, and is also quite accurate and easy to compute. Hence, by focusing on the tail distribution, one can both generalize the scope of the model and obtain simpler approximations.

2. Problem Formulation

2.1. The Portfolio Structure and Loss Distribution

Consider a portfolio of loans consisting of n obligors. Our interest centers on the distribution of losses from defaults over a fixed time horizon. The probability of default for the ith obligor over the time horizon of interest is $p_i \in (0, 1)$, and is used as an input to our model. This value is often set based on the average historical default frequency of companies with similar credit profiles. The associated exposure to default of counterparty i is assumed to be given by $e_i > 0$, that is, the default event results in a fixed and given loss of e_i monetary units. (We note that it is easy to generalize the main results of the paper to the case where the loss size is random under mild regularity conditions.) To keep the analysis simple, we ignore degradation in the quality of the loan, e.g., due to rating downgrades, but such generalizations are straightforward.

For the determination of the portfolio loss distribution, the specification of dependence between defaults is of paramount importance. The dependence model that we consider is closely related to the widely used CreditMetrics model; see Gupton et al. (1997), Crouhy et al. (2000), and Li (2000). In particular, we assume that there exists a vector of underlying latent variables $\{X_1, \ldots, X_n\}$ so that the ith default occurs if X_i exceeds some given threshold x_i (the distributional assumptions related to the latent variables will be discussed in §2.2). The loss incurred from defaults is then given by

$$L = e_1 \mathbb{I}\{X_1 > x_1\} + \cdots + e_n \mathbb{I}\{X_n > x_n\}, \tag{1}$$

where $\mathbb{I}\{\cdot\}$ is the indicator function. The threshold x_i is chosen according to the marginal default probabilities so that $\mathbb{P}(X_i > x_i) = p_i$. In this paper, our interest is in developing sharp asymptotics and efficient simulation techniques to estimate the probability of large losses, $\mathbb{P}(L > x)$, and the expected excess loss conditioned on large portfolio losses (commonly referred to as the expected shortfall of the portfolio) given by $\mathbb{E}[L - x \mid L > x]$, for a large threshold x.

The normal copula model that is widely used in the financial industry and that forms the basis of the CreditMetrics and KMV models assumes that the vector of latent variable follows a multivariate normal distribution. Hence, the dependence between the default events is determined by the correlation structure of the latent variables, in particular, the dependence structure of the vector $(\mathbb{I}\{X_i > x_i\}, \ldots, \mathbb{I}\{X_n > x_n\})$ can be represented with a normal copula (cf. Embrechts et al. 2003). The underlying correlations are often specified through a linear factor model

$$X_i = c_{i1}Z_1 + \cdots + c_{id}Z_d + c_i\eta_i,$$

where: (i) $Z_1, \ldots, Z_d$ are i.i.d. standard normal random variables that measure, for example, global, country, and industry effects impacting all companies; (ii) $c_{i1}, \ldots, c_{id}$ are the loading factors; (iii) η_i is a normal random variable that captures idiosyncratic risk, and is independent of the Z_is; and (iv) c_i and the loading factors are chosen so that the variance of X_i is equal to one (without loss of generality). To keep the notation simple, we restrict attention to single-factor models ($d = 1$); as we discuss in §6, the extension of our analysis and results to multiple-factor models is not difficult.

The multivariate normal that underlies CreditMetrics/KMV provides a limited form of dependence between obligors, which, in particular, may not assign sufficient probability to the occurrence of many simultaneous defaults in the portfolio. As indicated in the introduction, one of the primary objectives of this paper is to extend the normal copula model to incorporate "stronger" dependence among obligors, so that the corresponding dependence structure is more in line with recently proposed models of extremal dependence (see, e.g., Frey and McNeil 2001 and Embrechts et al. 2003) and empirical findings (see, e.g., Mashal and Zeevi 2003), both of which suggest consideration of t-copula models and the like over the normal copula.

2.2. Extremal Dependence

Let $(\eta_i\colon 1 \leqslant i \leqslant n)$ denote i.i.d. random variables and let Z denote another random variable independent of $(\eta_i\colon 1 \leqslant i \leqslant n)$. Fix $0 < \rho < 1$ and put

$$X_i = \frac{\rho Z + \sqrt{1-\rho^2}\,\eta_i}{W}, \quad i = 1, \ldots, n, \tag{2}$$

where W is a nonnegative random variable independent of Z and $(\eta_i\colon 1 \leqslant i \leqslant n)$, and its probability density function (p.d.f.) $f_W(\cdot)$ satisfies

$$f_W(w) = \alpha w^{\nu-1} + o(w^{\nu-1}) \quad \text{as } w \downarrow 0, \tag{3}$$

for some constants $\alpha > 0$ and $\nu > 0$. Here and in what follows, we write $h(x) = o(g(x))$ if $h(x)/g(x) \to 0$ as $x \to 0$ or as $x \to \infty$, where the limit considered is obvious from the context. If Z and $\{\eta_i\}$ are i.i.d. having a normal distribution and W is removed from (2), then this model reduces to a single-factor latent variable instance of CreditMetrics/KMV. As alluded to earlier, our aim is to model economies where the dependence amongst obligor defaults is primarily due to *common shocks*, and this is captured in (3) through the random variable W. When W takes values close to zero, all the X_is are likely to be large, leading to many simultaneous defaults. The parameter ν measures the likelihood of common shocks: smaller values imply a higher probability that W takes values close to zero. This class of models has recently been proposed in the context of credit risk modeling (cf. Frey and McNeil 2001 and references therein); in the particular instance where (Z, η_i) follow a bivariate normal distribution, this is often referred to as a mean-variance normal mixture, with $1/W$ providing the mixing distribution.

Example 1. Let W follow a Gamma(β, γ) distribution, with $\gamma, \beta > 0$, whose p.d.f. is given by

$$f_W(x) = \frac{\beta^\gamma x^{\gamma-1}}{\Gamma(\gamma)} e^{-\beta x}, \quad x \geqslant 0.$$

Then, this distribution satisfies (3) with $\nu = \gamma$, $\alpha = \beta^\gamma/\Gamma(\gamma)$.

Example 2. For a positive integer k, let $W = \sqrt{k^{-1}\text{Gamma}(1/2, k/2)}$ so that

$$f_W(x) = \frac{2k^{k/2}x^{k-1}}{2^{k/2}\Gamma(k/2)} e^{-kx^2/2}, \quad x \geqslant 0.$$

This p.d.f. satisfies (3) with $\nu = k$, $\alpha = 2(k/2)^{k/2}/\Gamma(k/2)$.

Note that for $\gamma = k/2$ and $\beta = 1/2$, the distribution discussed in Example 1 is chi-squared with k degrees of freedom (df). Note that when a linear combination of Z and η_i follows a normal distribution and W has the distribution given in Example 2, then the vector $(X_i\colon 1 \leqslant i \leqslant n)$ follows a multivariate t-distribution, whose dependence structure is given by a t-copula with k degrees of freedom.

Technical Assumptions. Let $F_Z(\cdot)$ and $F_\eta(\cdot)$ denote the distribution functions of Z and η_i, respectively. For notational ease, let η denote a random variable independent of W and Z with an identical distribution to η_i. In what follows, we restrict Z and η to be light tailed, i.e., $1 - F_Z(x)$ and $1 - F_\eta(x)$ are both upper bounded by an exponentially decaying term as $x \to \infty$. Further, with regard to the "noise" variable η, we make the following technical assumption: the distribution of η possesses a probability density function $f_\eta(\cdot)$ such that $f_\eta(x) > 0$ for all $x \in \mathbb{R}$. (The latter assumption is made to facilitate analysis and can be generalized at the expense of further technical details.) In what follows, we refer to (3) together with the above conditions collectively as *the distributional assumptions* associated with our model.

3. Large Portfolio Loss: Asymptotic Analysis

Because it is virtually impossible to exactly compute the probability of large portfolio losses and the associated expected shortfall, we focus on an asymptotic regime that is of practical interest and supports a tractable analysis.

This regime is one where the portfolio of interest is comprised of a "large number" of obligors, each individual obligor defaults with "small" probability, and the focus is on "large" portfolio losses. The mathematical meaning of these terms is spelled out in §3.1. Subsequently, in §3.2 we describe the main results.

3.1. Preliminaries

Let $f(x)$ denote an increasing function so that $f(x) \to \infty$ as $x \to \infty$. Fix n (the number of obligors in the portfolio), and let $\{a_1, \ldots, a_n\}$ be strictly positive constants. Set the default thresholds for the individual obligors to be $x_i^n = a_i f(n)$, so that obligor i defaults if $X_i > a_i f(n)$ and obligors may have different marginal default probabilities. The overall portfolio loss is given by

$$L_n = e_1 \mathbb{1}\{X_1 > a_1 f(n)\} + \cdots + e_n \mathbb{1}\{X_n > a_n f(n)\}, \tag{4}$$

where e_i, $i = 1, \ldots, n$, is the exposure associated with the ith obligor.

In §3.2, we analyze the probability that L_n takes on large values when n is large; in particular, we focus on the probability of the event $\{L_n > nb\}$ for $b > 0$. In §3.4, we extend our analysis to develop sharp asymptotics for $\mathbb{E}[L_n - x \mid L_n > x]$. We note that the regime we are focusing on here is suitable for analyzing the risk of large portfolios in which each obligor defaults with very small likelihood.

We assume that $f(n)$ increases at a subexponential rate so that $f(n)\exp(-\beta n)$ is a bounded sequence that converges to zero as $n \to \infty$ for all $\beta > 0$. By suitably selecting the function $f(n)$, we can model portfolios of varying credit ratings classes. For example, letting $f(n)$ increase polynomially in n, we can model a portfolio with high-quality obligors, whereas if $f(n)$ increases, say, at a logarithmic rate, then the loans are considered more risky.

To deal with the heterogeneity among obligors, captured by the sequences $\{e_i, a_i\}_{i=1}^n$, we impose the following assumption.

Assumption 1. *The nonnegative sequence $((e_i, a_i)\colon i \geqslant 1)$ takes values in a finite set $\mathcal{V}$, with cardinality $|\mathcal{V}|$. In addition, the proportion of each element $(e_j, a_j) \in \mathcal{V}$ in the portfolio converges to $q_j > 0$ as $n \to \infty$ (so that $\sum_{j \leqslant |\mathcal{V}|} q_j = 1$).*

In practice, the loan portfolio may be partitioned into a finite number of homogeneous loans based on factors such as industry, quality of risk, and exposure sizes. Assumption 1 allows this flexibility. Although our analysis easily generalizes to the case where each obligor corresponds to the pair (e_j, a_j) with probability q_j, and e_j is a light-tailed random variable, we avoid overburdening the notation by simply assuming a constant exposure level e_j, and that for a given portfolio a fraction q_j of the obligors corresponds to class j. (In the remainder of the paper, we ignore the nonintegrality of $q_j n$ for simplicity and clarity of exposition.)

3.2. Sharp Asymptotics for the Probability of Large Portfolio Losses

Let $\bar{e} = \sum_{j \leqslant |\mathcal{V}|} e_j q_j$, i.e., the limiting average loss when all the obligors default. Recall that the portfolio loss, L_n, is given in (4). The following theorem derives a sharp asymptotic for the probability of large portfolio losses. The function $w(z)$ used in the statement of the theorem is defined in §3.5. Essentially, conditioned on $Z = z$, $w(z)$ denotes the threshold value so that for $W \in (0, w(z)/f(n)]$, the mean loss from the portfolio is greater than b; for $W \in (w(z)/f(n), \infty)$, the mean portfolio loss is less than b.

Theorem 1. *Fix $0 < b < \bar{e}$, and let Assumption 1 as well as the distributional assumptions on (Z, η, W) hold true. Then,*

$$f(n)^\nu \mathbb{P}(L_n > nb) \to \frac{\alpha}{\nu} \int_{-\infty}^{\infty} w(z)^\nu \, dF_Z(z) \quad \text{as } n \to \infty. \tag{5}$$

We note that a similar result can be derived when $f(n)$ is not growing with n, in which case the probability of large portfolio losses no longer vanishes asymptotically; see further discussion in §6.2.

3.3. Discussion

Intuition and an Informal Proof Sketch. The proof follows from behavior of W relative to the threshold $w(z)$. On the event $\{W > w(Z)/f(n)\}$, the mean portfolio loss is less than b, and hence due to Chernoff's bound the probability of large loss is exponentially decaying in n. The event $\{0 < W < w(Z)/f(n)\}$ is significant from the point of view of large losses, and it occurs with probability

$$\int_{-\infty}^{\infty} \mathbb{P}\left(0 < W \leqslant \frac{w(z)}{f(n)}\right) dF(z). \tag{6}$$

Using the assumption given in (3), we have that

$$\mathbb{P}(W \leqslant w(z)/f(n)) \approx \int_0^{w(z)/f(n)} \alpha x^{\nu - 1} \, dx \tag{7}$$

$$\approx \frac{\alpha}{\nu} \left(\frac{w(z)}{f(n)}\right)^\nu, \tag{8}$$

neglecting lower-order terms. Conditioned on the event $\{0 < W < w(Z)/f(n)\}$, the mean loss from the portfolio is greater than b. Hence, due to the law of large numbers, the event of large loss $\{L_n > nb\}$ happens with probability one in the limit as $n \to \infty$. Plugging (8) into (6), the sharp asymptotic (5) for the portfolio risk follows.

Implication of Extremal Dependence. Theorem 1 may be reexpressed as

$$\mathbb{P}(L_n > nb) \sim \frac{1}{f(n)^\nu} \frac{\alpha}{\nu} \int_{-\infty}^{\infty} w(z)^\nu \, dF_Z(z). \tag{9}$$

(We say that $a_n \sim b_n$ for nonnegative sequences $(a_n\colon n \geqslant 1)$ and $(b_n\colon n \geqslant 1)$ when $a_n/b_n \to 1$ as $n \to \infty$.) Inspection of the expressions in Appendix A in the online companion reveals that when $a_i \equiv a$ for all i, then $w(z) = \rho(z - z_b)/a$ for some constant z_b that depends on b. Hence, it is evident that the asymptotic behavior of the portfolio risk is governed mostly by ν, i.e., the likelihood that the common shock W takes values near the origin and obligors tend to default simultaneously. In particular, as is evident in the above asymptotic, smaller values of ν lead to a higher portfolio risk (because such values increase the propensity for joint defaults in the portfolio). In contrast, the correlation between obligors only affects the magnitude of the constant premultiplier; as expected, higher values of correlation increase the magnitude of this constant. We note that even when obligor default probabilities are not identical (and characterized by different a_is), the bounds on the function $w(z)$ are linearly dependent on ρ (see the first subsection in the online companion). Thus, even in this case, it is clear that the probability of large losses is far more sensitive to ν than to ρ. One consequence of this observation is that greater accuracy is needed in estimating ν in comparison to ρ to get a reasonable approximation for the probability of large portfolio losses. (For examples of such estimation results in the context of the t-copula model, see Mashal and Zeevi 2003.)

3.4. Sharp Asymptotics for the Expected Shortfall

Theorem 1 is the key to establishing an asymptotic for the expected shortfall in Theorem 2. The function $r(w, z)$ used in the statement of Theorem 2 is defined precisely in §3.5. Essentially, $r(w, z)$ denotes the mean loss from the portfolio conditioned on $Z = z$ and $W = w/f(n)$. Let $(Y)^+ := \max(0, Y)$.

Theorem 2. *Fix $0 < b < \bar{e}$, and suppose that Assumption 1 as well as the distributional assumptions on (Z, η, W) hold true. Then,*

$$\frac{\mathbb{E}[L_n - nb \mid L_n > nb]}{n} \to \psi(b, \nu) \quad a.s.$$

as $n \to \infty$, where

$$\psi(b, \nu) := \frac{\nu \int_{-\infty}^{\infty} \int_0^{w(z)} (r(w, z) - b)^+ w^{\nu-1}\, dw\, dF_Z(z)}{\int_{-\infty}^{\infty} w(z)^\nu\, dF_Z(z)}.$$

The theorem asserts that the expected shortfall grows roughly linearly in the size of the portfolio n, i.e.,

$$\mathbb{E}[L_n - nb \mid L_n > nb] \sim n\psi(b, \nu).$$

The asymptotic may be briefly understood as follows:

$$\mathbb{E}[L_n - nb \mid L_n > nb] = \frac{\mathbb{E}[(L_n - nb) I(L_n > nb)]}{P(L_n > nb)}. \tag{10}$$

The asymptotic for the denominator is derived in Theorem 1. The numerator may be asymptotically approximated by noting that the set of values of W and Z, for which the mean portfolio loss is less than b, contributes negligibly to it (because, in that region, the probability of $\{L_n > nb\}$ decays exponentially with n). On the remaining set, the portfolio loss amount may be replaced by its conditional expectation (conditioned on the value of W and Z), and because in this region W is small, its p.d.f. may be approximated using (3).

3.5. Definitions of the Auxiliary Functions Used in the Statements of the Main Theorems

Let

$$p_{w,z,i} := \mathbb{P}\left(\eta_i > \frac{a_i W f(n) - \rho Z}{\sqrt{1-\rho^2}} \,\middle|\, W = \frac{w}{f(n)}, Z = z\right) = \mathbb{P}\left(\eta_i > \frac{a_i w - \rho z}{\sqrt{1-\rho^2}}\right).$$

Note that this probability is nondecreasing in z and is nonincreasing in w. Let

$$r(w, z) := \sum_{j \leqslant |\mathcal{V}|} e_j q_j p_{w,z,j} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n e_i p_{w,z,i}, \tag{11}$$

where the limit follows from Assumption 1. For $w > 0$, $r(w, z)$ denotes the limiting average portfolio loss (as $n \to \infty$) when $W = w/f(n)$ and $Z = z$. Note that $r(w, z)$ is nondecreasing in z and nonincreasing in w.

Recall that $\bar{e} = \sum_{j \leqslant |\mathcal{V}|} e_j q_j$. Let z_b denote the unique value of z that solves

$$\bar{e}\mathbb{P}\left(\eta \geqslant \frac{-\rho z}{\sqrt{1-\rho^2}}\right) = b.$$

(Note that our assumption that η has a positive density function on the real line ensures that there exists a unique z_b that solves the above equation.) The term z_b assumes significance in our analysis because for $Z < z_b$, the event of average loss exceeding b remains a rare event for all values $W > 0$. Let $w(z)$ be defined as the unique solution to

$$r(w, z) = b. \tag{12}$$

Note that $w(z)$ is strictly positive for each $z > z_b$. Note also that for $w \leqslant w(z)$, under $P_{w,z}$ the average loss amount $(1/n) \sum_{i=1}^n e_i \mathbb{I}\{X_i > a_i f(n)\}$ in the limit as $n \to \infty$ has mean that is greater than or equal to b, and hence the probability of large loss is no longer a rare event. We set $w(z) = 0$ for $z \leqslant z_b$.

4. Large Portfolio Loss: Importance-Sampling Simulation

As we illustrate later in §5 through numerical examples, the asymptotics presented in Theorems 1 and 2 can lead to significant inaccuracies in assessing the probability of large portfolio losses and the expected shortfall. Hence, Monte Carlo methods become an attractive alternative to accurately estimate these performance measures.

Because the probability of large portfolio losses is typically small, naive simulation would require a very large number of runs to achieve a satisfactory variance for the estimate. As in other rare event estimation problems, importance sampling often provides an efficient means of generating low-variance estimates, essentially by placing further probability mass on the rare event of interest and then suitably unbiasing the resultant simulation output. Our approach to estimating the expected shortfall via Monte Carlo simulation exploits its ratio representation (10). Note that the samples generated to estimate the numerator $E(L_n - nb)I(L_n > nb)$ take positive value only when large losses occur. Hence, the importance-sampling technique that works well in estimating the probability of large losses $P(L_n > nb)$ may be expected to work well in estimating $E(L_n - nb)I(L_n > nb)$ as well. In §4.2, we show that this is indeed the case. We first focus on efficiently estimating $P(L_n > nb)$ as $n \to \infty$.

4.1. Importance Sampling for Loss Probability

For notational convenience, assume that Z and W have probability density functions $f_Z(\cdot)$ and $f_W(\cdot)$, respectively (although in our analysis we do not require that the distribution of Z have a density function). Let $(p_j\colon j \leqslant |\mathcal{V}|)$ denote the probabilities associated with the Bernoulli variables $(\mathbb{I}\{X_i > a_i f(n)\}\colon i \leqslant n)$, as a function of the generated Z and W. We suppress this dependence from the notation for ease of presentation (this dependence is explicitly displayed in the proofs given in Appendix A in the online companion). For notational purposes, let $A_n = \{L_n > nb\}$ denote the event in which portfolio losses exceed a level nb in a portfolio with n obligors. Suppose that under an importance-sampling distribution we generate samples of Z, W and the Bernoulli variables $(\mathbb{I}\{X_i > a_i f(n)\}\colon i \leqslant n)$, and hence $\mathbb{I}\{A_n\}$, using density functions $\tilde{f}_Z(\cdot)$, $\tilde{f}_W(\cdot)$ and probabilities $(\tilde{p}_j\colon j \leqslant |\mathcal{V}|)$, where the distribution of W may depend on the generated value of Z, and the distribution of the Bernoulli success probabilities may depend on the generated values of Z and W (this dependence is also suppressed in the notation here). Let $\tilde{\mathbb{P}}$ denote the corresponding probability measure. The sample output then equals $\tilde{L}\mathbb{I}\{A_n\}$, where $\tilde{L}$ denotes the *unbiasing* likelihood ratio (Radon-Nikodym derivative of $\mathbb{P}$, the original probability measure, w.r.t. $\tilde{\mathbb{P}}$) and equals

$$\frac{f_Z(Z) f_W(W)}{\tilde{f}_Z(Z)\tilde{f}_W(W)} \prod_{j \leqslant |\mathcal{V}|} \left(\frac{p_j}{\tilde{p}_j}\right)^{Y_j q_j n} \left(\frac{1-p_j}{1-\tilde{p}_j}\right)^{(1-Y_j) q_j n},$$

where $Y_j q_j n$ denotes the number of defaults in class j obligors.

We now discuss two standard characterizations of performance for importance-sampling estimators. The sequence of estimators $(\tilde{L}\mathbb{I}\{A_n\}\colon n \geqslant 1)$ under probability $\tilde{\mathbb{P}}$ are said to estimate the sequence of probabilities $(\mathbb{P}(A_n)\colon n \geqslant 1)$ with *bounded relative error* if

$$\limsup_{n\to\infty} \frac{\sqrt{\tilde{\mathbb{E}}[\tilde{L}^2 \mathbb{I}\{A_n\}]}}{\mathbb{P}(A_n)} < \infty,$$

where $\tilde{\mathbb{E}}$ denotes expectation with respect to the probability distribution $\tilde{\mathbb{P}}$. Note that $\tilde{\mathbb{E}}[\tilde{L}\mathbb{I}\{A_n\}] = \mathbb{P}(A_n)$. This, together with the condition above, implies that the computational effort needed to estimate the probability to a fixed degree of relative accuracy remains bounded no matter how rare the event is (i.e., independent of the value of n; see, e.g., Heidelberger 1995).

The sequence of estimators $(\tilde{L}\mathbb{I}\{A_n\}\colon n \geqslant 1)$ under probability $\tilde{\mathbb{P}}$ is said to be *asymptotically optimal* with respect to the sequence of probabilities $(\mathbb{P}(A_n)\colon n \geqslant 1)$ if

$$\lim_{n\to\infty} \frac{\log \tilde{\mathbb{E}}(\tilde{L}^2 \mathbb{I}\{A_n\})}{\log \mathbb{P}(A_n)} = 2.$$

Because $\tilde{\mathbb{E}}(\tilde{L}^2\mathbb{I}\{A_n\}) \geqslant (\tilde{\mathbb{E}}[\tilde{L}\mathbb{I}\{A_n\}])^2 = \mathbb{P}(A_n)^2$, asymptotic optimality implies *asymptotic zero variance* on a logarithmic scale. Note that if $\tilde{p}$ has bounded relative error, then it is also asymptotically optimal.

As discussed in the previous section, the key to the occurrence of the large loss events in the portfolio corresponds to W taking small values so that the mean loss, conditioned on W and Z, exceeds a level b. In §§4.1.1 and 4.1.2, we describe two different IS algorithms for estimating $\mathbb{P}(A_n)$ that judiciously assign large probability to this event to reduce simulation variance. The first algorithm generates a new distribution of W by *exponentially twisting* the original one (see, e.g., Heidelberger 1995 for an introduction to exponential twisting). We prove that this results in an estimator that has bounded relative error. The second algorithm derives a new distribution for W by approximately *hazard-rate twisting* the original distribution of $1/W$ (see Juneja and Shahabuddin 2002 for an introduction to hazard-rate twisting), and we show that it results in an estimator that is asymptotically optimal. This suggests that the first algorithm may perform better than the second, and we indeed observe this to be the case in our empirical experiments reported in §5. We note that the second algorithm may have significant implementation advantages that will be discussed briefly in what follows.

When, conditional on (W, Z), the mean loss is less than b, it may be a good practice (although not essential for the asymptotic optimality of the algorithms) to generate the corresponding Bernoulli random variables under an exponentially twisted distribution so that the event A_n is no longer rare, and the mean loss under the new distribution equals b. For any random variable X with p.d.f. $f_X(\cdot)$, the associated distribution that is exponentially twisted by parameter θ has the form

$$\exp(\theta x - \Lambda_X(\theta)) f_X(x),$$

where $\Lambda_X(\cdot)$ denotes the log-moment generating function of X. For $\theta \geqslant 0$, let $\Lambda_j(\theta)$ denote $\log(\exp(\theta e_j)p_j + (1-p_j))$. It is well known, and easily checked through differentiation, that $\Lambda_j(\cdot)$ is strictly convex when $0 < p_j < 1$ (see, e.g., Dembo and Zeitouni 1993). Let

$$p_j^\theta = \frac{\Lambda_j'(\theta)}{e_j} = \frac{\exp(\theta e_j)p_j}{\exp(\theta e_j)p_j + (1-p_j)} = \exp(\theta e_j - \Lambda_j(\theta))p_j,$$

where e_j is the exposure to the jth obligor, and p_j the probability that the jth obligor defaults. Put $1 - p_j^\theta = \exp(-\Lambda_j(\theta))(1-p_j)$, and note that p_j^θ is strictly increasing in θ. For the case where the mean loss $\sum_{j \leqslant |\mathcal{V}|} e_j q_j p_j < b$, consider the new default probabilities $(p_j^{\theta^*}\colon j \leqslant |\mathcal{V}|)$, where $\theta^* > 0$ is the unique solution to the equation

$$\sum_{j \leqslant |\mathcal{V}|} e_j q_j p_j^\theta = b.$$

This choice of twisting parameter induces a probability distribution under which the mean loss is b, hence, the event of incurring such loss in a sample is no longer rare. In what follows, we suppress the dependence of θ^* on w and z in the notation, although it is noteworthy that θ^* increases with w and decreases with z.

4.1.1. An Algorithm Based on Exponential Twisting. This algorithm consists of three stages. First, a sample of Z is generated using the original distribution. Depending on the value of Z, a sample of W is generated using appropriate importance sampling. Depending on the value of samples of Z and W, samples of the Bernoulli variable $\mathbb{I}\{X_i > a_i f(n)\}$ are generated for $i \leqslant n$, using naive simulation or importance sampling. For a fixed positive constant ξ, put $\widetilde{w}(z) = \max(\xi, w(z))$.

Importance-Sampling Algorithm 1

Step 1. Generate a sample of Z according to the original distribution $F_Z(\cdot)$.

Step 2. Generate a sample of W using the density f_W^* obtained by exponentially twisting f_W with parameter $-\theta_{Z,n}$, where

$$\theta_{Z,n} = \frac{\nu f(n)}{\widetilde{w}(Z)}.$$

Later in the section, we justify this choice of the twisting parameter based on asymptotic considerations.

Step 3. For each $i \leqslant n$, generate samples of $\mathbb{I}\{X_i > a_i f(n)\}$ independent of each other using the distribution: $p_i^* = p_i$ if the mean loss under the generated W and Z is greater than b; and using $p_i^* = p_i^{\theta^*}$ otherwise.

Let $\mathbb{P}^*$ denote the probability measure corresponding to this algorithm and $\mathbb{E}^*$ the expectation operator under this measure. Again, let $Y_j q_j n$ denote the number of class j defaults in a single simulation run. The likelihood ratio is then given by

$$L_* = \exp[\theta_{Z,n} W + \Lambda_W(-\theta_{Z,n})] \cdot \prod_{j \leqslant |\mathcal{V}|} \left(\frac{p_j}{p_j^*}\right)^{Y_j q_j n} \left(\frac{1-p_j}{1-p_j^*}\right)^{(1-Y_j) q_j n}. \tag{13}$$

The main result of this section is the following.

THEOREM 3. *Under Assumption* 1 *and the distributional assumptions on* (Z, η, W),

$$\limsup_{n\to\infty} f(n)^{2\nu} \mathbb{E}^* L_*^2 \mathbb{I}\{A_n\} < \infty. \tag{14}$$

In view of Theorem 1, which provides the tail asymptotic for the probability of the event $A_n = \{L_n > nb\}$, we conclude that the proposed importance-sampling algorithm has bounded relative error.

On the Choice of the Exponential Twisting Parameter in Algorithm 1. Conditional on $Z = z$, our importance-sampling problem essentially reduces to that of estimating $\mathbb{P}(W \leqslant w(z)/f(n))$ efficiently. If W is generated using a distribution obtained by exponential twisting by an amount $-\theta$ ($\theta > 0$), then the associated likelihood ratio $L = \exp[\theta W + \Lambda_W(-\theta)]$ is upper bounded by

$$\exp\left[\theta \frac{w(z)}{f(n)} + \Lambda_W(-\theta)\right]$$

on the set $\{W \leqslant w(z)/f(n)\}$. It is a standard practice in importance sampling to select a parameter θ that minimizes the uniform bound on the likelihood ratio because, e.g., this also minimizes the corresponding upper bound on the second moment $\mathbb{E}^*[L^2 \mathbb{I}\{W \leqslant w(z)/f(n)\}]$. Let $\tilde{\theta} > 0$ denote the parameter minimizing $\theta(w(z)/f(n)) + \Lambda_W(-\theta)$. Then,

$$\Lambda_W'(-\tilde{\theta}) = -\frac{w(z)}{f(n)}.$$

Note that

$$\Lambda_W'(-\theta) = -\frac{\int_0^\infty w e^{-\theta w} f_W(w)\,dw}{\int_0^\infty e^{-\theta w} f_W(w)\,dw}.$$

Suppose that $f_W(w) = \alpha w^{\nu-1}$. Then, it is easily seen that

$$\int_0^\infty w e^{-\theta w} f_W(w)\,dw = \frac{\alpha \Gamma(\nu+1)}{\theta^\nu}$$

and

$$\int_0^\infty e^{-\theta w} f_W(w)\,dw = \frac{\alpha \Gamma(\nu)}{\theta^{\nu-1}}.$$

It then follows that $\Lambda_W'(-\theta) = -\nu/\theta$ and $\tilde{\theta} = \nu f(n)/w(z)$. In the more general setting when f_W only satisfies (3), $\tilde{\theta} \sim \nu f(n)/w(z)$ as $n \to \infty$ is easily established, e.g., by the use of Tauberian Theorems (see Feller 1970, pp. 442–445).

Also note that $\Lambda'(\theta)$ denotes the mean of W under the distribution obtained by exponentially twisting f_W by an amount θ. Hence, twisting by an amount $-\theta_{Z,n}$ roughly sets the mean of W to equal $w(z)/f(n)$.

Recall that obligor i defaults if $X_i \geqslant a_i f(n)$. Equivalently, this probability equals $\mathbb{P}(\rho Z + \sqrt{1-\rho^2}\eta_i - W a_i f(n) > 0)$. Glasserman et al. (2002) devised exponential twisting-based importance-sampling techniques that consider analogous probabilities. Our framework is different from theirs; and our approach, which focuses on "making" W take small values, provides greater insight into how the large losses occur.

4.1.2. An Algorithm Based on Hazard-Rate Twisting. Let $V = 1/W$. Note that $\mathbb{P}(V \leqslant x) = \mathbb{P}(W \geqslant 1/x)$, and hence the p.d.f. of V, i.e., $f_V(\cdot)$ satisfies the relation

$$f_V(x) = \frac{1}{x^2} f_W(1/x) = \frac{\alpha}{x^{\nu+1}}(1+o(1)), \tag{15}$$

where $o(1) \to 0$ as $x \to \infty$. Define

$$\bar{f}_V(x) = f_V(x)$$

for $x \leqslant c_1$, and

$$\bar{f}_V(x) = (1 - F_V(c_1)) c_1^{1/\log f(n)} \frac{1}{\log f(n)} \frac{1}{x^{1+1/\log f(n)}}$$

for $x \geqslant c_1$, where c_1 is chosen so that $f_V(x)/\bar{f}_V(x)$ remains upper bounded by a constant for all x. The importance-sampling algorithm builds on this new distribution for V; later in the section we justify our choice of $\bar{f}_V(x)$.

Importance-Sampling Algorithm 2

Step 1. Generate a sample of Z from the original $F_Z(\cdot)$ and generate a sample of V using $\bar{f}_V(\cdot)$.

Step 2. For each $i \leqslant n$, generate the samples of $\mathbb{I}\{X_i > a_i f(n)\}$ independently with $\bar{p}_i = p_i$, if the mean loss under the generated V and Z is greater than b, and with $\bar{p}_i = p_i^{\theta^*}$ otherwise.

Let $\bar{\mathbb{P}}$ denote the probability distribution corresponding to this algorithm. Recall that $Y_j q_j n$ denotes the number of class j defaults. The likelihood ratio of $\mathbb{P}$ w.r.t. $\bar{\mathbb{P}}$ is given by

$$\bar{L} = \frac{f_V(V)}{\bar{f}_V(V)} \prod_{j \leqslant |\mathcal{V}|} \left(\frac{p_j}{\bar{p}_j}\right)^{Y_j q_j n} \left(\frac{1-p_j}{1-\bar{p}_j}\right)^{(1-Y_j) q_j n}. \tag{16}$$

We then have the following result.

THEOREM 4. *Under Assumption* 1 *and the distributional assumptions on* (Z, η, W),

$$\frac{\log \bar{\mathbb{E}}[\bar{L}^2 \mathbb{I}\{A_n\}]}{\log f(n)} \to -2\nu \quad \text{as } n \to \infty. \tag{17}$$

In particular, in view of Theorem 1 it follows that the proposed importance-sampling algorithm is asymptotically optimal in the sense that it achieves zero variance on the logarithmic scale.

On the Choice of the Importance-Sampling Density. The broad motivation for the density function defined above is given in Juneja and Shahabuddin (2002), which discusses hazard-rate twisting. Reexpressing the p.d.f. $f_V(x)$ as $h(x)\exp(-\mathcal{H}(x))$, where $h(x) = f_V(x)/(1 - F_V(x))$ denotes the hazard rate and $\mathcal{H}(x) = -\log(1 - F_V(x))$ denotes the hazard function, the distribution corresponding to hazard-rate twisting by an amount θ has the p.d.f.

$$f_V^{\theta}(x) = h(x)(1-\theta)\exp(-(1-\theta)\mathcal{H}(x)).$$

(Note that the hazard-rate function $\mathcal{H}$ is nondecreasing.) The tail distribution function is given by $\exp(-(1-\theta)\mathcal{H}(x))$. Recall that conditioned on $Z = z$, our interest is essentially in estimating the probability $\mathbb{P}(V > f(n)/w(z))$ efficiently. Using the hazard-rate twisted distribution f_V^{θ}, the associated likelihood ratio equals $(1/(1-\theta)) \cdot \exp(-\theta\Lambda_V(x))$, and this is upper bounded by

$$\frac{1}{(1-\theta)} \exp(-\theta\mathcal{H}(f(n)/w(z)))$$

on the set $\{V > f(n)/w(z)\}$. As in Algorithm 1, here we also search for $\tilde{\theta}$ that minimizes this bound. This value can be seen to equal

$$\tilde{\theta} = 1 - \left(\mathcal{H}\left(\frac{f(n)}{w(z)}\right)\right)^{-1}.$$

Then, the IS tail distribution corresponding to hazard-rate twisting by $\tilde{\theta}$ equals

$$\exp\left[-\frac{\mathcal{H}(x)}{\mathcal{H}(f(n)/w(z))}\right]. \tag{18}$$

Note that

$$1 - F_V(x) \sim \frac{\alpha}{\nu x^{\nu}},$$

and hence $\mathcal{H}(x) \sim \nu\log(x)$ as $x \to \infty$.

Equation (18) suggests that our IS tail distribution function should be close to

$$\exp\left[-\frac{\nu\log x}{\nu(\log f(n) - \log w(z))}\right] = x^{-1/(\log f(n) - \log w(z))}.$$

We achieve considerable simplification by ignoring $\log w(z)$ in this expression (on the basis that this is typically dominated by $\log f(n)$). This is important because determining $w(z)$ can be potentially computationally expensive. Then, the corresponding p.d.f. equals

$$\frac{1}{\log f(n)} \frac{1}{x^{1+1/\log f(n)}}.$$

This is quite similar to the p.d.f. proposed in Algorithm 2. The p.d.f. $\bar{f}_V(x)$ is set to $f_V(x)$ for $x \leqslant c_1$ simply to prevent the ratio $f_V(x)/\bar{f}_V(x)$ from "blowing up" for small values of x. The potential for this type of behavior exists when $f_V(x)$ is large or unbounded in this region. For ease of implementation, one may select a p.d.f. different from f_V in this region as long as the ratio $f_V(x)/\bar{f}_V(x)$ remains bounded from above for $x \leqslant c_1$.

4.2. Importance Sampling for Expected Shortfall

Denote the expected shortfall $\mathbb{E}[L_n - nb \mid L_n > nb]$ by $\beta(n,b)$. We discuss how importance sampling may be used to estimate this efficiently. In the interest of space, we only analyze the exponential twisting-based importance-sampling Algorithm 1, described in §4.1.1, for estimating $\beta(n,b)$. The analysis easily extends to importance-sampling Algorithm 2.

Using $\mathbb{P}^*$, generate m i.i.d. samples $((L_n^i, L_*^i): i \leqslant m)$ of (L_n, L_*) and compute the following estimate:

$$\hat{\beta}_m(n,b) = \frac{\sum_{i=1}^m L_*^i (L_n^i - nb)\mathbb{I}\{L_n^i > nb\}}{\sum_{i=1}^m L_*^i \mathbb{I}\{L_n^i > nb\}}.$$

Using the delta method (see, e.g., Serfling 1981), we note that the following central limit theorem holds:

$$\sqrt{m}[\hat{\beta}_m(n,b) - \beta(n,b)] \Rightarrow \sigma(n,b)\mathcal{N}(0,1)$$

as $m \to \infty$, where $\Rightarrow$ denotes convergence in distribution, and

$$\sigma^2(n,b) = \frac{\sigma_1^2(n,b)}{\mu_2^2(n,b)} + \frac{\mu_1^2(n,b)\sigma_2^2(n,b)}{\mu_2^4(n,b)} + 2\frac{\sigma_{12}(nb)\mu_1(n,b)}{\mu_2^3(n,b)}, \tag{19}$$

with

$$\mu_1(n,b) = \tilde{\mathbb{E}}[L_*(L_n - nb)\mathbb{I}\{L_n > nb\}],$$
$$\mu_2(n,b) = \tilde{\mathbb{E}}[L_*\mathbb{I}\{L_n > nb\}],$$
$$\sigma_1^2(n,b) = \tilde{\mathbb{E}}[(L_*(L_n - nb))^2\mathbb{I}\{L_n > nb\}] - \mu_1^2(n,b),$$
$$\sigma_2^2(n,b) = \tilde{\mathbb{E}}[L_*^2\mathbb{I}\{L_n > nb\}] - \mu_2^2(n,b),$$
$$\sigma_{12}(n,b) = \tilde{\mathbb{E}}[L_*(L_n - nb)\mathbb{I}\{L_n > nb\}] - \mu_1(n,b)\mu_2(n,b).$$

The definition of bounded relative error may be modified to include the estimation of expected shortfall as follows: The sequence estimators $(\hat{\beta}_m(n,b): n \geqslant 1)$ under the probability measure $\mathbb{P}^*$ are said to estimate the sequence of performance measures $(\beta(n,b): n \geqslant 1)$ with bounded relative error if

$$\limsup_{n\to\infty} \frac{\sigma(n,b)}{\beta(n,b)} < \infty.$$

Again, if this property holds, then the computational effort (as measured by m) needed to construct a confidence interval of $\beta(n,b)$ with a fixed degree of relative accuracy remains bounded in m.

Theorem 5. *Under Assumption* 1 *and the distributional assumptions on* (Z, η, W), *the proposed IS algorithm based on Algorithm* 1 *has bounded relative error for estimating the expected shortfall* $\beta(n,b)$.

5. Numerical Results

In this section, we compare the performance of Algorithms 1 and 2 with each other and with naive simulation, and investigate sensitivity to ν, ρ, n, and b. The broad conclusions are that both algorithms provide significant improvement over the performance of naive simulation. This improvement increases as the event becomes more rare (e.g., as ν increases or as ρ decreases). This supports our theoretical conclusions that the relative performance, as measured by the ratio of the standard deviation of the estimate to the mean of the estimate, remains well behaved in the two algorithms even as the probability of large losses becomes increasingly rare. We observe that Algorithm 1 provides about 6 to 10 times higher variance reduction compared to Algorithm 2. As mentioned earlier, Algorithm 2 is easier to implement; its per-sample computational effort was found to be on par with naive simulation, whereas Algorithm 1 takes on average three times longer in generating a sample compared to naive simulation.

Motivated by the t-copula model, we set the distribution of W in our numerical experiments as in Example 2, the random variable Z is chosen to follow a standard Normal distribution (mean zero, variance one), and each η_i is normally distributed with mean zero and variance nine. (We set the value of variance to nine instead of one simply to ensure that the loss probability is sufficiently large to be practically relevant.) The random variables W, Z, and $(\eta_i: i \leqslant n)$ are mutually independent so that $X = (X_1, \ldots, X_n)$ has a multidimensional Student t-distribution, with the dependence structure given by a t-copula.

5.1. Implementation Issues

Recall that the p.d.f. of W has the form

$$f_W(x) = \frac{2k^{k/2}x^{k-1}}{2^{k/2}\Gamma(k/2)} e^{-kx^2/2}, \quad x \geqslant 0.$$

For implementation of Algorithm 1, conditional on Z, we need to generate samples from the distribution obtained by exponentially twisting this p.d.f., i.e., from the p.d.f.

$$f_{W,\theta}(x) = \frac{2k^{k/2}x^{k-1}}{2^{k/2}\Gamma(k/2)} e^{-\theta x - kx^2/2 + \Lambda_W(-\theta)}, \quad x \geqslant 0. \tag{20}$$

Recall that $\Lambda_W(\cdot)$ is the log-moment generating function of W and $\theta = \nu f(n)/\tilde{w}(Z)$. Because the cumulative distribution associated with this density function does not have a closed form, it is not straightforward to use the inverse transform methods to generate samples from this distribution. Instead, we use an acceptance-rejection algorithm to generate these random variables, which increases the overall per-sample computational effort for Algorithm 1. Further, we need to evaluate the moment-generating function associated with this p.d.f. to update the likelihood ratio. This is done using numerical integration. Because the latter causes a computation burden, we compute it offline.

Algorithm 2 is implemented by generating V using the IS density

$$\tilde{f}_V(x) = \begin{cases} 0.025, & x \in [0, 0.5], \\ \dfrac{K}{x^{1+1/\log f(n)}}, & \text{otherwise}, \end{cases} \tag{21}$$

where K is the normalizing constant given by $\log f(n)/2^{1/\log f(n)}$. It is easy to generate from this density using the inverse transform method. (The range $[0, 0.5]$ and the choice of uniform density in this range is driven by ease of implementation; results were not sensitive to these choices.)

5.2. Performance of the Two Algorithms

In all the experiments in this subsection, for each set of specified parameters, we generate 50,000 samples for Algorithm 1 and 100,000 samples for Algorithm 2. Variance under naive simulation is estimated indirectly by exploiting the observation that for a Bernoulli random variable with success probability p, the variance equals $p(1-p)$. Thus, we use the probability estimated via Algorithm 1 to estimate the variance of each sample under naive simulation. We then estimate the variance reduction obtained by the two algorithms, which is defined as the ratio of the variance of the estimator under the importance-sampling measure to the variance of the estimator under the original measure.

Table 1 shows the comparison of Algorithms 1 and 2 with naive simulation as ν changes. The model parameters are chosen to be $n = 250$, $f(n) = \sqrt{n}$, $\rho = 0.25$, $b = 0.25$, each $a_i = 0.5$, and $e_i = 1$. As mentioned earlier, Algorithm 1 performs much better than Algorithm 2, and both perform significantly better than naive simulation, especially when ν increases and the probability becomes smaller.

Table 2 shows the comparison of Algorithms 1 and 2 with naive simulation as ρ changes. Again, we set $n = 250$, $b = 0.25$, and $f(n) = \sqrt{n}$. The df is kept fixed at 12, each $a_i = 0.5$, and $e_i = 1$.

Table 3 shows the comparison of Algorithms 1 and 2 with naive simulation as n changes. Again, we set $df = 12$, $b = 0.25$, and $f(n) = \sqrt{n}$. The correlation factor ρ is kept fixed at 0.25, each $a_i = 0.5$, and $e_i = 1$. In the last column,

Table 1. Performance of Algorithms 1 and 2 as a function of the degrees of freedom ν.

	Algorithm 1		Algorithm 2	
df	Prob. est. [95% C.I.]	Var. reduction	Prob. est. [95% C.I.]	Var. reduction
4	8.08×10^{-3} [±1.2%]	65	8.16×10^{-3} [±2.2%]	10
8	2.39×10^{-4} [±1.9%]	878	2.40×10^{-4} [±3.6%]	124
12	1.06×10^{-5} [±3.5%]	7,331	1.04×10^{-5} [±5.3%]	1,291
16	6.08×10^{-7} [±4.9%]	52,185	5.71×10^{-7} [±7.2%]	12,935
20	4.51×10^{-8} [±7.5%]	3.01×10^5	4.27×10^{-8} [±10.6%]	7.9×10^4

Note. Variance reduction is measured relative to naive simulation.

Table 2. Performance of Algorithms 1 and 2 as a function of correlation ρ.

	Algorithm 1		Algorithm 2	
ρ	Prob. est. [95% C.I.]	Var. reduction	Prob. est. [95% C.I.]	Var. reduction
0.1	8.58×10^{-6} [±1.9%]	26,013	8.77×10^{-6} [±3.6%]	3,390
0.2	9.74×10^{-6} [±2.5%]	13,134	1.01×10^{-5} [±4.7%]	1,696
0.3	1.18×10^{-5} [±3.5%]	5,158	1.18×10^{-5} [±6.4%]	808
0.4	1.39×10^{-5} [±6.2%]	1,332	1.50×10^{-5} [±8.0%]	454

Note. Variance reduction is measured relative to naive simulation.

we show the value of the sharp asymptotic for the probability of large losses derived in Theorem 1. Note that for $n = 100$, the discrepancy between the true probability as estimated via importance sampling and the sharp asymptotic equals 16%. Further, we observe that as n increases, the accuracy of the sharp asymptotic improves.

Table 4 shows the comparison of Algorithms 1 and 2 with naive simulation as b changes. Again, we set $\rho = 0.25$, $df = 12$, $b = 0.25$, and $f(n) = \sqrt{n}$. The correlation factor n is kept fixed at 250, each $a_i = 0.5$ and $e_i = 1$.

5.3. Expected Shortfall

In this section, we illustrate the accuracy of the expected shortfall asymptote as the number of obligors becomes large and study the efficacy of IS Algorithm 1 for estimating expected shortfall. Table 5 compares the accuracy of the sharp asymptotic of expected shortfall derived in Theorem 2 as a function of n. Model parameters are taken to be $\nu = 4$, $f(n) = \sqrt{n}$, $\rho = 0.25$, each $a_i = 0.5$, and $b = 0.25$. The accuracy improves significantly for large values of n. Note that for $n = 100$ and 250, the expected shortfall is in the range that is of practical significance. However, in this case, the asymptotic of the expected shortfall is not very accurate.

Table 6 compares the performance of the IS Algorithm 1 with naive simulation for estimating expected shortfall as ν varies. The model parameters are $n = 250$, $f(n) = \sqrt{n}$, $\rho = 0.25$, $b = 0.25$, each $a_i = 0.5$, and $e_i = 1$. For each ν, we generate 50,000 samples under the original measure and the IS measure. We then compute the variance reduction obtained by the two algorithms, which is defined as the ratio of the variance of the estimator under the importance-sampling measure to the variance of the estimator under the original measure. We also report the probability of large loss, i.e., $\mathbb{P}(L_n > nb)$. For $df = 12$ and $df = 16$, we observed $L_n < nb$ under naive simulation for all the 50,000 sample paths generated.

6. Discussion and Concluding Remarks

In this section, we first informally contrast the normal copula model with the t-copula model in a simple setting to illustrate the strikingly different conclusions that the two models may reach for certain parameters. This motivates

Table 3. Performance of Algorithms 1 and 2 together with the sharp probability asymptotic derived in Theorem 1 as a function of n.

	Algorithm 1		Algorithm 2		
n	Prob. est. [95% C.I.]	Var. reduction	Prob. est. [95% C.I.]	Var. reduction	Asymptote
100	2.49×10^{-3} [±3.2%]	29	2.57×10^{-3} [±3.6%]	11	2.15×10^{-3}
250	1.06×10^{-5} [±3.5%]	7,331	1.04×10^{-5} [±5.3%]	1,291	8.80×10^{-6}
500	1.66×10^{-7} [±3.1%]	4.5×10^5	1.62×10^{-7} [±6.9%]	49,740	1.37×10^{-7}
1,000	2.38×10^{-9} [±3.3%]	2.9×10^7	2.30×10^{-9} [±7.2%]	3.2×10^6	2.15×10^{-9}

Note. Variance reduction is measured relative to naive simulation.

the importance of selecting the correct credit risk model. We then conclude with some possible extensions to our analysis.

6.1. Contrasting t-Copula with Normal Copula

We first heuristically derive a sharp asymptotic for the probability of large losses in the normal copula model. (For brevity, we only provide a sketch of the argument, noting that the conclusions can easily be made rigorous along the lines of the proof of Theorem 1.) Recall that under the standard normal copula model,

$$X_i = \rho Z + \sqrt{1-\rho^2}\eta_i,$$

where Z and η_i have a standard normal distribution. Suppose that obligor i defaults if $X_i \geqslant g(n)$, where now $g(n)$ is an increasing function such that $g(n)/(\log n)^\beta \to 0$ for some $\beta > 0$. Then, it is easily argued that on the event $\{Z > g(n)/\rho + z_b\}$ (where z_b is a constant defined earlier in §3.5), the mean loss from the portfolio will exceed b. Hence, due to the law of large numbers, the large loss event $\{L_n > nb\}$ happens with probability one in the limit as $n \to \infty$. Otherwise, the large loss probability is decaying at an exponential rate in n. The subexponential rate of decay of $\mathbb{P}(Z > g(n)/\rho + z_b)$ clearly dominates, and consequently we have that

$$\mathbb{P}(L_n > nb) \sim \mathbb{P}(Z \geqslant g(n)/\rho + z_b),$$

so that

$$\lim_{n\to\infty} \frac{\log \mathbb{P}(L_n > nb)}{g(n)} = -\frac{1}{2\rho^2}. \tag{22}$$

Table 4. Performance of Algorithms 1 and 2 as a function of b.

	Algorithm 1		Algorithm 2	
b	Prob. estimate [95% C.I.]	Var. reduction	Prob. estimate [95% C.I.]	Var. reduction
0.1	3.95×10^{-3} [±1.8%]	57	4.01×10^{-3} [±3.2%]	9
0.2	8.77×10^{-5} [±2.4%]	1,493	8.83×10^{-5} [±5.01%]	173
0.3	1.33×10^{-7} [±4.0%]	36,594	1.29×10^{-6} [±6.8%]	6,414

Note. Variance reduction is measured relative to naive simulation.

We are now in a position to compare the asymptotic derived on the basis of the normal copula model with the t-copula model. We fix common input data, i.e., ρ and the marginal probabilities of default p_i for each obligor. For simplicity, we assume that the marginal probability of default for obligor i equals $\epsilon(n)$, where $\epsilon(n)$ decays to zero at a subexponential rate. Then, if

$$\mathbb{P}\left(\frac{\rho Z + \sqrt{1-\rho^2}\eta_i}{W} > f(n)\right) = \epsilon(n),$$

it can be seen that $f(n) \sim c/\epsilon(n)^{1/\nu}$ for some constant c. Hence, from Theorem 1,

$$\lim_{n\to\infty} \frac{\log \mathbb{P}(L_n > nb)}{\log \epsilon(n)} = 1. \tag{23}$$

Now consider the normal copula model. Because $\rho Z + \sqrt{1-\rho^2}\eta_i$ has a standard normal distribution, it follows that if

$$\mathbb{P}(\rho Z + \sqrt{1-\rho^2}\eta_i > g(n)) = \epsilon(n),$$

then $g(n) \sim -2\log \epsilon(n)$. Thus, from (22) we observe that

$$\lim_{n\to\infty} \frac{\log \mathbb{P}(L_n > nb)}{\log \epsilon(n)} = \frac{1}{\rho^2}. \tag{24}$$

When contrasting this with the t-copula model asymptotic in (23), one observes that because $\rho < 1$, the normal copula model *underestimates* the probability of large losses compared to the t-copula model for large n. In particular, in the t-copula asymptotic, the correlation ρ does not affect the rate (and appears only as a multiplicative constant), whereas in the normal copula case, the rate itself is affected.

Table 5. The expected shortfall and its asymptotic as a function of the number of obligors (n).

n	$\hat{\beta}(n, b)$ [95% C.I.]	Asymptotic
100	5.4 [±1.3%]	4.8
250	13.0 [±1.3%]	12.3
500	24.9 [±1.5%]	24.4
1,000	48.8 [±1.6%]	48.8
2,000	95.3 [±1.7%]	97

Table 6. Performance of IS Algorithm 1 for estimating expected shortfall as a function of the degrees of freedom ν.

df	$\hat{\beta}(n,b)$ [95% C.I.]	Var. reduction	$\mathbb{P}(Ł_n > nb)$
4	13.20 [±1.5%]	62	8.06×10^{-3}
8	7.84 [±2.6%]	743	2.41×10^{-4}
12	5.81 [±4.1%]	(∗)	1.07×10^{-5}
16	4.67 [±6.9%]	(∗)	6.18×10^{-7}

Note. (∗) denotes that the event of interest was not observed in any sample path using naive simulation. Variance reduction is measured relative to naive simulation.

We now verify this observation through a numerical experiment. Set $n = 100$ and $b = 0.1$. For the standard t-copula model, set $f(n) = \sqrt{n}$, $a_i = 0.5$, and $e_i = 1$ for all i. For each ρ, $g(n)$ for the standard normal copula model is chosen so that the single name default probability is equal to that of the t-copula model. The probability of large losses for both models, as ρ varies, is estimated via simulation. The results are presented in Table 7. (Importance-sampling techniques were used to efficiently estimate these probabilities.) As indicated by (23) and (24), for small values of ρ, the normal copula model significantly underestimates the loss probability compared to the t-copula model.

6.2. Possible Extensions

In this paper, we considered a common shock-based model for measuring portfolio credit risk. This model generalizes the t-copula model that is increasingly used for modelling extremal dependence amongst obligors. We developed sharp asymptotics and importance-sampling techniques to estimate the probability of large losses and the expected shortfall in this setting. We now list some of the possible extensions of our analysis.

Multifactor Model. In our analysis for notational simplicity, we restricted ourselves to a single-factor model. The results generalize to the multifactor setting with

$$X_i = \frac{c_{i1}Z_1 + \cdots + c_{id}Z_d + c_i\eta_i}{W},$$

where: $(Z_1, \ldots, Z_d)$ are i.i.d. standard normal random variables, $c_{i1}, \ldots, c_{id}$ are the loading factors, and η_i is a normal random variable that captures idiosyncratic risk and is independent of the Z_is. For example, the sharp asymptotic in Theorem 1 generalizes to

$$\lim_{n\to\infty} f(n)^\nu \mathbb{P}(L_n > b) = \frac{\alpha}{\nu}\int_{z\in\mathbb{R}^d} w(z)^\nu \, dF_Z(z),$$

Table 7. Large loss probability under t-copula- and normal copula-based models.

ρ	t-copula	Normal copula
0.25	1.84×10^{-3} [±3.1%]	5.16×10^{-9} [±0.66%]
0.5	2.67×10^{-3} [±4.7%]	1.41×10^{-4} [±0.99%]
0.75	3.33×10^{-3} [±6.3%]	2.15×10^{-3} [±0.88%]

Note. The number in [·] represents the 95% confidence interval.

where F_Z denotes the d-dimensional multivariate distribution of $(Z_1, \ldots, Z_d)$, and for $z \in \mathbb{R}^d$, $w(z)$ denotes the threshold so that if $w \in (0, w(z)/f(n))$, the mean loss from a portfolio conditional on $Z = z$ and $W = w$ is greater than b. (When this is not true for any $w \geqslant 0$ for a given z, $w(z)$ is set to zero, as in the one-dimensional analysis.)

On the Role of the Function $f(n)$. The assumption that $f(n)$ increases without bound ensures that the probability of large portfolio losses diminishes as n increases. The importance-sampling schemes developed in §5 are geared toward efficient simulation of this probability, and hinge on the asymptotes derived in §4. If $f(n)$ was taken to be constant, say, then a similar expression to that on the right-hand side of (5) can be derived, but the probability of large losses is no longer "small." Our analysis further assumed that $f(n)$ increases at a subexponential rate and Z is a light-tailed random variables. This ensures that the rare event happens primarily when W takes small values, whereas Z and the η_i essentially do not play any role in its occurrence. This implies that correlations and idiosyncratic effects play less of a role in the occurrence of large losses vis-a-vis the common shock. However, there can be models where correlations and/or idiosyncratic effects play an important role in the occurrence of the rare event. In certain scenarios, one may expect these other models to be more realistic, and hence are important extensions that merit further investigation.

7. Electronic Companion

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

References

Basel Committee on Banking Supervision. 2002. Report. The New Basel Capital Accord. Basel, Switzerland.

Crouhy, M., D. Galai, R. Mark. 2000. A comparative analysis of current credit risk models. *J. Banking Finance* **24** 59–117.

Dembo, A., O. Zeitouni. 1993. *Large Deviations Techniques and Applications*. Jones and Bartlett Publishers, Boston.

Embrechts, P., F. Lindskog, A. J. McNeil. 2003. Modelling dependence with copulas and applications to risk management. S. Rachev, ed. *Handbook of Heavy Tailed Distributions in Finance*, Chapter 8. Elsevier, New York, 329–384.

Feller, W. 1970. *An Introduction to Probability Theory and Its Applications*, Vol. 1. John Wiley & Sons, New York.

Frey, R., A. McNeil. 2001. Modelling dependent defaults. ETH E-Collection, http://e-collection.ethbib.ethz.ch/show?type=bericht&nr=273.

Frey, R., A. McNeil. 2003. Dependent defaults in models of portfolio credit risk. *J. Risk* **6** 59–92.

Frey, R., A. J. McNeil, M. Nyfeler. 2001. Copulas and credit risk models. *Risk* **14** 111–114.

Glasserman, P., J. Li. 2003. Importance sampling for portfolio credit risk. *Management Sci.* **51** 1643–1656.

Glasserman, P., P. Heidelberger, P. Shahabuddin. 2002. Portfolio value-at-risk with heavy-tailed risk factors. *Math. Finance* **12** 239–269.

Gupton, G., C. Finger, M. Bhatia. 1997. Credit metrics technical document. Technical report, J. P. Morgan & Co., New York.

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

Juneja, S., P. Shahabuddin. 2002. Simulating heavy-tailed processes using delayed hazard rate twisting. *ACM TOMACS* **12** 94–118.

Juneja, S., P. Shahabuddin. 2006. Rare event simulation techniques: An introduction and recent advances. S. Henderson, B. Nelson, eds. *Handbook on Simulation*, Chapter 11. Elsevier, 291–350.

Kealhofer, S., J. Bohn. 2001. Portfolio management of credit risk. Technical report, Working paper, Moody's KMV, New York.

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

Mashal, R., A. Zeevi. 2003. Beyond correlation: Extreme co-movements between financial assets. Working paper, Columbia University, New York.

McNeil, A. J., R. Frey, P. Embrechts. 2005. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton, NJ.

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

Schloegel, L., D. O'Kane. 2005. A note on the large homogenous portfolio approximation with the student-t copula. *Finance Stochastics* **9** 577–584.

Serfling, R. J. 1981. *Approximation Theorems of Mathematical Statistics*. Wiley-Interscience, New York.