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

Achal Bassamboo*
Stanford University

Sandeep Juneja†
Tata Institute of Fundamental Research

Assaf Zeevi‡
Columbia University

Submitted: July 2005. Revised: February 20, 2006

Abstract

We consider the risk of a portfolio comprised of 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 heterogenous mix of obligors, and the fact that default events are rare and mutually dependent makes 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.

Short Title: Portfolio Credit Risk

Keywords: Portfolio, credit, asymptotics, expected shortfall, simulation, importance sampling, rare events, risk management.

*Graduate School of Business, e-mail: achalb@stanford.edu
†School of Technology and Computer Science, e-mail: juneja@tifr.res.in
‡Graduate School of Business, e-mail: assaf@gsb.columbia.edu

1 Introduction

Market conditions over the past few years combined with regulatory arbitrage have lead to significant interest and activity in trading and transferring of credit-related risk. Since 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. While there are numerous copula functions that can serve such a purpose, the *normal copula*, which assumes 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, Finger and Bhatia (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. BCBS (2002)]; see also Li (2000), the survey paper by Crouhy, Galai and Mark (2000) and the recent monograph by McNeil, Frey and Embrechts (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 non-negligible probability; for further discussion see Embrechts, Lindskog and McNeil (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 t-distributed random variable may register large co-movements [see further discussion in Embrechts et al. (2003), Frey and McNeil (2003) and Glasserman, Heidelberger and Shahabuddin (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 heterogenous 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 semi-parametric, 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 pricing of various instrument such as credit baskets, CDO's, 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) that builds on simulation studies.]

The 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. Section 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 section 6 contains some discussion and concluding remarks. Proofs are relegated to two appendices: Appendix A contains the proof of the main results and Appendix B gives proofs of auxiliary results.

Related literature and positioning of the present 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 homogenous portfolios within the normal copula framework. Threshold models with non normal dependence structure have been recently proposed and studied by Frey, McNeil and Nyfeler (2001) and Frey and McNeil (2003). The latter also formulates non-normal threshold models for credit portfolios which 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, Frey and Embrechts (2005).

While 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 which are more accurate. In addition, we develop these sharp asymptotics for expected shortfall, a risk measure which 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 which are much easier to implement and yet enjoy good variance reduction properties. As we indicate in Section 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 multi-factor 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 homogenous 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 type 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 semi-parametric 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 so as 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 Section 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 $\mathbb{E}[L - x | L > x]$, 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 iid 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_i's; 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 section 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 the current 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 $(\eta_i : 1 \leq i \leq n)$ denote iid random variables and let Z denote another random variable independent of $(\eta_i : 1 \leq i \leq n)$. 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 non-negative random variable independent of Z and $(\eta_i : 1 \leq i \leq n)$ and its probability density function (pdf) $f_W(\cdot)$ 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 iid 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_i's 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 been recently 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 pdf is given by

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

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

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

This pdf 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 $(X_i : 1 \leq i \leq n)$ 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

Since it is virtually impossible to exactly compute the probability of large portfolio losses and the associated expected shortfall, we focus on an asymptotic regime which 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 section 3.1 and subsequently in section 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

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

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

In Section 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 $b > 0$. Hence as the size of the portfolio, n, grows large, the individual probability of default decreases in a manner that is governed by the function $f(n)$, and the loss level of interest, nb, scales up with the size of the portfolio. In Section 3.4, we extend our analysis to develop sharp asymptotics for $\mathbb{E}[L_n - x | L_n > x]$.

We assume that $f(n)$ increases at a sub-exponential 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, while 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 non-negative sequence $((e_i, a_i) : i \geq 1)$ 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 $\sum_{j \leq |\mathcal{V}|} q_j = 1$).*

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. While 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 correspond to class j. (In the remainder of the paper we ignore the non-integrality of $q_j n$ for simplicity and clarity of exposition.)

3.2 Sharp asymptotics for the probability of large portfolio losses

Let $\bar{e} = \sum_{j \leq |\mathcal{V}|} e_j q_j$, 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 precisely in Appendix A.1. Essentially, conditioned on $Z = z$, $w(z)$ denotes the threshold value so that for $W \in (0, \frac{w(z)}{f(n)}]$ the mean loss from the portfolio is greater than b; for $W \in (\frac{w(z)}{f(n)}, \infty)$, 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*

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

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

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

Using the assumption given in (3) we have that

$$\begin{aligned} \mathbb{P}(W \leq w(z)/f(n)) \quad &\approx \quad \int_0^{w(z)/f(n)} \alpha x^{\nu-1} dx \qquad (7) \\ &\approx \quad \frac{\alpha}{\nu}\left(\frac{w(z)}{f(n)}\right)^{\nu}, \qquad (8) \end{aligned}$$

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 1 in the limit as $n \to \infty$. Plugging (8) in (6), the sharp asymptotic (5) for the portfolio risk follows.

Implication of extremal dependence. Theorem 1 may be re-expressed 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 non-negative sequences $(a_n : n \geq 1)$ and $(b_n : n \geq 1)$ when $a_n/b_n \to 1$ as $n \to \infty$.) Inspection of the expressions in Appendix A.1 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 (since such values increase the propensity for joint defaults in the portfolio). In contrast, the correlation between obligors only affects the magnitude of the constant pre-multiplier; 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_i's), the bounds on the function $w(z)$ are linearly dependent on ρ [see (27) and (28) in Appendix A.1]. 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 Appendix A.1. 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 Assumption 1 as well as the distributional assumptions on (Z, η, W) hold true. Then*

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

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.,

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

The asymptotic may be briefly understood as follows:

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

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 value of W and Z) and since in this region W is small, its pdf may be approximated using (3).

4 Large Portfolio Loss: Importance Sampling Simulation

As we illustrate later in Section 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.

Since the probability of large portfolio lossses 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 Section 4.2, we show that this is indeed the case. We first focus on efficiently estimating $P(L_n > nb)$ efficiently 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 (though in our analysis we do not require that the distribution of Z have a density function). Let $(p_j : j \le |\mathcal{V}|)$ denote the probabilities associated with the Bernoulli variables $(\mathbb{I}\{X_i > a_i f(n)\} : i \le n)$, 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). 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 $(\mathbb{I}\{X_i > a_i f(n)\} : i \le n)$, and hence $\mathbb{I}\{A_n\}$, using density functions $\tilde{f}_Z(\cdot)$, $\tilde{f}_W(\cdot)$ and probabilities $(\tilde{p}_j : j \le |\mathcal{V}|)$, where the distribution of W may depend upon the generated value of Z, and the distribution of the Bernoulli success probabilities may depend upon 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

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

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 $(\tilde{L}\mathbb{I}\{A_n\} : n \ge 1)$ under probability $\tilde{\mathbb{P}}$ are said to estimate the sequence of probabilities $(\mathbb{P}(A_n) : n \geq 1)$ 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 $(\tilde{L}\mathbb{I}\{A_n\} : n \geq 1)$ under probability $\tilde{\mathbb{P}}$ are said to be *asymptotically optimal* with respect to the sequence of probabilities $(\mathbb{P}(A_n) : n \geq 1)$ if

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

Since, $\tilde{\mathbb{E}}(\tilde{L}^2 \mathbb{I}\{A_n\}) \geq (\tilde{\mathbb{E}}[\tilde{L}\mathbb{I}\{A_n\}])^2 = \mathbb{P}(A_n)^2$, 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 Sections 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 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 Section 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 (though 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 pdf $f_X(\cdot)$, 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 $\theta \geq 0$, 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 $\sum_{j \leq |\mathcal{V}|} e_j q_j p_j < b$, consider the new default probabilities $(p_j^{\theta^*} : j \leq |\mathcal{V}|)$, where $\theta^* > 0$ is the unique solution to the equation

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

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 $i \leq n$, using naive simulation or importance sampling. For a fixed positive constant ξ, put $\tilde{w}(z) = \max(\xi, w(z))$.

Importance Sampling Algorithm 1

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

2. Generate a sample of W using the density f_W^* obtained by exponentially twisting f_W with parameter $-\theta_{Z,n}$, where
$$\theta_{Z,n} = \frac{\nu f(n)}{\tilde{w}(Z)}.$$
Later in the section we justify this choice of the twisting parameter based on asymptotic considerations.

3. For each $i \leq n$, 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

$$L_* = \exp[\theta_{Z,n} W + \Lambda_W(-\theta_{Z,n})] \prod_{j \leq |\mathcal{V}|} \left(\frac{p_j}{p_j^*}\right)^{Y_j q_j n} \left(\frac{1-p_j}{1-p_j^*}\right)^{(1-Y_j)q_j n}. \tag{11}$$

The main result of this section is the following.

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

$$\limsup_{n\to\infty} f(n)^{2\nu} \mathbb{E}^* L_*^2 \mathbb{I}\{A_n\} < \infty. \tag{12}$$

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 $\mathbb{P}(W \leq \frac{w(z)}{f(n)})$ 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 $\{W \leq \frac{w(z)}{f(n)}\}$. It is a standard practice in importance sampling to select a parameter θ that minimizes the uniform bound on the likelihood ratio, since, e.g., this also minimizes the corresponding upper bound on the second moment $\mathbb{E}^*[L^2 \mathbb{I}\{W \leq \frac{w(z)}{f(n)}\}]$. Let $\tilde{\theta} > 0$ denote the parameter minimizing $\theta \frac{w(z)}{f(n)} + \Lambda_W(-\theta)$. 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 $\Lambda'_W(-\theta) = -\frac{\nu}{\theta}$ and $\tilde{\theta} = \frac{\nu f(n)}{w(z)}$. In the more general setting when f_W only satisfies (3), $\tilde{\theta} \sim \frac{\nu f(n)}{w(z)}$ as $n \to \infty$ is easily established, e.g., by the use of Tauberian Theorems [see pp. 442-445 Feller (1970)]. 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 $\frac{w(z)}{f(n)}$.

Recall that obligor i defaults if $X_i \geq a_i f(n)$. 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, that 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 $\mathbb{P}(V \leq x) = \mathbb{P}(W \geq 1/x)$ and hence the pdf of V, i.e., $f_V(\cdot)$ satisfies the relation

$$f_V(x) = \frac{1}{x^2} f_W(1/x) = \frac{\alpha}{x^{\nu+1}}(1+o(1)), \tag{13}$$

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

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

for $x \leq c_1$, and

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

for $x \geq c_1$, 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

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

2. For each $i \leq n$, 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

$$\bar{L} = \frac{f_V(V)}{\bar{f}_V(V)} \prod_{j \leq |\mathcal{V}|} \left(\frac{p_j}{\bar{p}_j}\right)^{Y_j q_j n} \left(\frac{1-p_j}{1-\bar{p}_j}\right)^{(1-Y_j) q_j n}. \tag{14}$$

We then have the following result.

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

$$\frac{\log \bar{\mathbb{E}}[\bar{L}^2 \mathbb{I}\{A_n\}]}{\log f(n)} \to -2\nu \quad \textit{as } n \to \infty. \tag{15}$$

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 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. Re-expressing the pdf $f_V(x)$ as $h(x)\exp(-\mathcal{H}(x))$, where $h(x) = \frac{f_V(x)}{1-F_V(x)}$ 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 pdf

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

(Note that the hazard rate function $\mathcal{H}$ is non-decreasing.) 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 $\mathbb{P}(V > \frac{f(n)}{w(z)})$ efficiently. Using the hazard rate twisted distribution f_V^θ, the associated likelihood ratio equals $\frac{1}{(1-\theta)}\exp(-\theta\Lambda_V(x))$ and this is upper bounded by

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

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

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

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

$$\exp\left[-\frac{\mathcal{H}(x)}{\mathcal{H}(\frac{f(n)}{w(z)})}\right]. \tag{16}$$

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

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

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 as determining $w(z)$ can be potentially computationally expensive. Then the corresponding pdf equals

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

This is quite similar to the pdf proposed in Algorithm 2. The pdf $\bar{f}_V(x)$ is set to $f_V(x)$ for $x \leq c_1$ 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 pdf different from f_V in this region as long as the ratio $f_V(x)/\bar{f}_V(x)$ remains bounded from above for $x \leq c_1$.

4.2 Importance sampling for expected shortfall

Denote the expected shortfall $\mathbb{E}[L_n - nb | L_n > nb]$, 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 Section 4.1.1, for estimating $\beta(n, b)$. The analysis easily extends to importance sampling algorithm 2.

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

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

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

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

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

$$\sigma^2(n, b) = \frac{\sigma_1^2(n, b)}{\mu_2^2(n, b)} + \frac{\mu_1^2(n, b)\sigma_2^2(n, b)}{\mu_2^4(n, b)} + 2\frac{\sigma_{12}(nb)\mu_1(n, b)}{\mu_2^3(n, b)}, \tag{17}$$

with

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

The definition of bounded relative error may be modified to include the estimation of expected shortfall as follows: The sequence estimators $(\widehat{\beta}_m(n,b) : n \geq 1)$, under the probability measure $\mathbb{P}^*$ are said to estimate the sequence of performance measures $(\beta(n,b) : n \geq 1)$ 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 Algorithm 1 and Algorithm 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, while Algorithm 1 takes on average three times more time 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 1) and each η_i is normally distributed with mean 0 and variance 9. (We set the value of variance to 9 instead of 1 simply to ensure that the loss probability is sufficiently large to be practically relevant). The random variables W, Z and $(\eta_i : i \leq n)$ are mutually independent so that $X = (X_1, \ldots, X_n)$ has a multi-dimensional Student t-distribution, with the dependence structure given by a t-copula.

5.1 Implementation issues

Recall that the pdf of W has the form

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

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

$$f_{W,\theta}(x) = \frac{2k^{k/2}x^{k-1}}{2^{k/2}\Gamma(k/2)}e^{-\theta x - kx^2/2 + \Lambda_W(-\theta)}, \quad x \geq 0. \tag{18}$$

Recall that $\Lambda_W(\cdot)$ is the log-moment generating function of W and $\theta = \nu f(n)/\widetilde{w}(Z)$. Since 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 pdf to update the likelihood ratio. This is done using numerical integration. Since the latter causes computation burden we compute it off-line.

Algorithm 2 is implemented by generating V using the IS density

$$\widetilde{f}_V(x) = \begin{cases} 0.025 & x \in [0, 0.5] \\ \frac{K}{x^{1+1/\log f(n)}} & \text{otherwise,} \end{cases} \tag{19}$$

where K is the normalizing constant given by $\frac{\log f(n)}{2^{1/\log f(n)}}$. Its 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.

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

Table 1: Performance of Algorithm 1 and 2 as a function of the degrees of freedom ν. Variance reduction is measured relative to naive simulation.

Table 2 shows the comparison of Algorithm 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 $e_i = 1$.

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

Table 2: Performance of Algorithm 1 and 2 as a function of correlation ρ. Variance reduction is measured relative to naive simulation.

Table 3 shows the comparison of Algorithm 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, 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 the as n increases the accuracy of the sharp asymptotic improves.

	Algorithm 1		Algorithm 2		Asymptote
n	Prob. est. [95% C.I.]	Var. reduction	Prob. est. [95% C.I.]	Var. reduction	
100	$2.49 \times 10^{-3}[\pm 3.2\%]$	29	$2.57 \times 10^{-3}[\pm 3.6\%]$	11	2.15×10^{-3}
250	$1.06 \times 10^{-5}[\pm 3.5\%]$	$7,331$	$1.04 \times 10^{-5}[\pm 5.3\%]$	$1,291$	8.80×10^{-6}
500	$1.66 \times 10^{-7}[\pm 3.1\%]$	4.5×10^{5}	$1.62 \times 10^{-7}[\pm 6.9\%]$	$49,740$	1.37×10^{-7}
1000	$2.38 \times 10^{-9}[\pm 3.3\%]$	2.9×10^{7}	$2.30 \times 10^{-9}[\pm 7.2\%]$	3.2×10^{6}	2.15×10^{-9}

Table 3: Performance of Algorithm 1 and 2 together with the sharp probability asymptotic derived in Theorem 1 as a function of n. Variance reduction is measured relative to naive simulation.

Table 4 shows the comparison of Algorithm 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 $e_i = 1$.

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

Table 4: Performance of Algorithm 1 and 2 as a function of b. Variance reduction is measured relative to naive simulation.

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 which 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 $n = 250$, $f(n) = \sqrt{n}$, $\rho = 0.25$, $b = 0.25$,

n	$\widehat{\beta}(n, b)$ [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
1000	48.8[±1.6%]	48.8
2000	95.3[±1.7%]	97

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

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.

df	$\hat{\beta}(n, b)$[95% C.I.]	Var. reduction	$\mathbb{P}(L_n > nb)$
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}

Table 6: Performance of IS Algorithm 1 for estimating expected shortfall as a function of the degrees of freedom ν. Variance reduction is measured relative to naive simulation. (∗) denotes that the event of interest was not observed in any sample path using naive simulation.

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 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 $X_i \geq g(n)$, 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 in Appendix A), 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 1 in the limit as $n \to \infty$. Otherwise, the large loss probability is decaying at an exponential rate in n. The sub-exponential rate of decay of $\mathbb{P}(Z > g(n)/\rho + z_b)$ clearly dominates, and consequently we have that

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

so that

$$\lim_{n\to\infty} \frac{\log \mathbb{P}(L_n > nb)}{g(n)} = -\frac{1}{2\rho^2}. \tag{20}$$

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 sub-exponential 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 $f(n) \sim \frac{c}{\epsilon(n)^{1/\nu}}$ for some constant c. Hence, from Theorem 1

$$\lim_{n\to\infty} \frac{\log \mathbb{P}(L_n > nb)}{\log \epsilon(n)} = 1. \tag{21}$$

Consider now the normal copula model. Since $\rho Z + \sqrt{1-\rho^2}\eta_i$ 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 (20) we observe that,

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

When contrasting this with the t-copula model asymptotic in (21) one observes that since $\rho < 1$, 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.

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 (21) and (22), for small values of ρ the normal copula model significantly underestimates the loss probability compared to the t-copula model.

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

Table 7: Large loss probability under t-copula and normal copula based models. The number in $[\cdot]$ represents the 95% confidence interval.

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.

Multi-factor model. In our analysis for notational simplicity we restricted ourselves to a single factor model. The results generalize to the multi-factor 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 iid 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_i's.

For instance, the sharp asymptotic in Theorem 1 generalizes to:

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

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 $w \in (0, \frac{w(z)}{f(n)})$, the mean loss from a portfolio conditional on $Z = z$ and $W = w$ is greater than b. (When this is not true for any $w \geq 0$ for a given z, $w(z)$ is set to zero, as in the one dimensional analysis.)

Exponential growth of $f(n)$. In our analysis we assume that $f(n)$ increases at a sub-exponential rate and Z is a light-tailed random variables. This ensures that the rare event happens primarily when W takes small values, while 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 extension that merit further investigation.

A Proofs of the Main Results

A.1 Preliminaries

We first introduce some preliminary notation and observations that are useful in proving the main theorems. Let

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

Note that this probability is non-decreasing in z and is non-increasing in w. Let

$$\begin{aligned} r(w,z) &:= \sum_{j\leq|\mathcal{V}|} e_j q_j p_{w,z,j} \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} e_i p_{w,z,i}, \end{aligned} \tag{23}$$

where the limit follows from Assumption 1. For $w > 0$, $r(w,z)$ denotes the limiting average portfolio loss (as $n \to \infty$) when $W = \frac{w}{f(n)}$ and $Z = z$. Note that $r(w,z)$ is non-decreasing in z and non-increasing in w.

Conditional on $Z = z$ and $W = w/f(n)$, Hoeffding's inequality [cf. Dembo and Zeitouni (1993)] can be used to bound the probability that the random variable $\frac{1}{n}\sum_{i=1}^{n} e_i \mathbb{I}\{X_i > a_i f(n)\}$ deviates significantly from its mean $r(w, z)$. In particular, for $\epsilon > 0$, there exists a constant $\beta > 0$ such that

$$P_{w,z}\left(\left|\frac{1}{n}\sum_{i=1}^{n} e_i \mathbb{I}\{X_i > a_i f(n)\} - r(w, z)\right| \geq \epsilon\right) \leq \exp(-n\beta), \tag{24}$$

for all n sufficiently large, where $P_{w,z}$ denotes the the original probability measure conditioned on $Z = z$ and $W = w/f(n)$. Furthermore, this inequality holds with the same constant β, uniformly for all (w, z) for which $r(w, z)$ is unchanged.

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

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

(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 since 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

$$r(w, z) = b. \tag{25}$$

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

To perform asymptotic analysis, we need additional notation obtained by perturbing certain parameters. For each δ, let z_{b_δ} denote the unique solution to

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

Note that $z_{b_0} \equiv z_b$, and z_{b_δ} is a decreasing function of δ. Further, we have $z_{b_\delta} \to z_b$ as $\delta \to 0$. Let $w_\delta(z) \geq 0$ denote the unique solution to the equation $r(w, z) = b - \delta$ for $z \geq z_{b_\delta}$. Note that $w(z) = w_0(z)$, $w_\delta(z)$ is a strictly increasing function of z for $z \geq z_{b_\delta}$, and using continuity and monotonicity of $r(w, z)$ in w, we have

$$w_\delta(z) \to w(z) \tag{26}$$

as $\delta \to 0$. The following upper bound on $w_\delta(z)$ is useful in the analysis that follows,

$$w_\delta(z) \leq \frac{\rho}{\min_i a_i}(z - z_{b_\delta}) \quad \text{for all } z > z_b. \tag{27}$$

To see why this is true, note that for each i,

$$a_i \frac{\rho(z - z_{b_\delta})}{\min_i a_i} - \rho z \geq -\rho z_{b_\delta}.$$

It then follows from the definition of $r(\cdot, \cdot)$ that $r(\frac{\rho}{\min_i a_i}(z - z_{b_\delta}), z) \leq b - \delta$, from which (27) follows. In a similar manner it is easy to establish that

$$w_\delta(z) \geq \frac{\rho}{\max_i a_i}(z - z_b) \quad \text{for all } z > z_b. \tag{28}$$

A.2 Proof of Theorem 1

Fix $\delta > 0$. Let $A_n = \{L_n > nb\}$. We decompose the probability of the event $\{A_n\}$ as follows

$$\mathbb{P}(A_n) = \mathbb{P}(A_n,\, Z \leq z_b) + \mathbb{P}\left(A_n,\, W > \frac{w_\delta(Z)}{f(n)},\, Z \geq z_b\right) + \mathbb{P}\left(A_n,\, W \leq \frac{w_\delta(Z)}{f(n)},\, Z > z_b\right). \tag{29}$$

We divide the remaining part of the proof into four steps. The first and the second step show that the first and second term on the right hand side of (29), respectively, are asymptotically negligible. The third and the fourth step develop upper and lower bounds on the third term on the right-hand-side of (29).

Step 1. We show that

$$\lim_{n\to\infty} f(n)^\nu \mathbb{P}(A_n,\, Z \leq z_b) = 0. \tag{30}$$

Fix $\epsilon > 0$. The probability term in (30) may be re-expressed as the sum of

$$\mathbb{P}(A_n,\, Z \leq z_b, W > \epsilon/f(n)), \tag{31}$$

and

$$\mathbb{P}(A_n,\, Z \leq z_b, W \leq \epsilon/f(n)). \tag{32}$$

First consider the probability (31). Note that for $w > \epsilon$, $z \leq z_b$, $r(w, z) \leq r(\epsilon, z_b) < r(0, z_b) = b$, since $r(w, z_b)$ is strictly decreasing in w. From (24), there exists a $\beta > 0$ such that $P_{w,z}(A_n) \leq \exp(-\beta n)$ uniformly for all $w \geq \epsilon$ and $z \leq z_b$. Hence, the same bound holds for the probability (31).

The expression in (32) is upper bounded by $\mathbb{P}(W \leq \epsilon/f(n))$ which in light of (3) is upper bounded by $(\alpha(1+\epsilon)\epsilon^\nu)/(\nu f(n)^\nu)$ for n sufficiently large. Thus, $\lim_{n\to\infty} f(n)^\nu \mathbb{P}(A_n,\, Z \leq z_b)$ is upper bounded by $\alpha(1+\epsilon)\epsilon^\nu/\nu$. Since ϵ is arbitrary we get (30).

Step 2. We show that

$$\lim_{n\to\infty} f(n)^\nu \mathbb{P}(A_n,\, W \geq \frac{w_\delta(Z)}{f(n)},\, Z > z_b) = 0. \tag{33}$$

Note that for $w \geq w_\delta(z)$ and $z \geq z_b$, $r(w, z) \leq b - \delta$. Thus, as discussed in (24), there exists a constant $\beta > 0$ so that $P_{w,z}(A_n) \leq \exp(-n\beta)$ for all $w \geq w_\delta(z)$ and $z \geq z_b$. Now, $f(n)^\nu \exp(-n\beta)$

is a bounded sequence that converges to 0, so (33) follows by the use of the bounded convergence theorem.

Step 3. We now develop an asymptotic upper bound on the third term on the right hand side of (29), which in turn gives an asymptotic upper bound on the probability of A_n. To this end, we show that for $\delta > 0$,

$$\limsup_{n\to\infty} f(n)^\nu \mathbb{P}\left(A_n,\, W \le \frac{w_\delta(Z)}{f(n)},\, Z > z_{b_0}\right) \le \frac{\alpha}{\nu}\int_{z\in(z_b,\infty)} w_\delta(z)^\nu dF_Z(z). \tag{34}$$

Note that

$$\mathbb{P}\left(A_n,\, W \le \frac{w_\delta(Z)}{f(n)},\, Z > z_b\right) \le \mathbb{P}\left(W \le \frac{w_\delta(Z)}{f(n)},\, Z \ge z_b\right). \tag{35}$$

For any $0 < \kappa < 1$, this is upper bounded by

$$\int_{z\in(z_b,f(n)^\kappa)}\int_{w\le\frac{w_\delta(z)}{f(n)}} f_W(w)dwdF_Z(z) + \mathbb{P}(Z \ge f(n)^\kappa). \tag{36}$$

Note from (3) that for any $\epsilon > 0$ and n sufficiently large, $f_W(w) \le \alpha(1+\epsilon)w^{\nu-1}$ for $0 \le w \le \frac{w_\delta(z)}{f(n)}$ and $z \in (z_b, f(n)^\kappa)$. (This follows since $w_\delta(z)$ increases at most at a linear rate as a function of z). Thus, for sufficiently large n, (36) is upper bounded by

$$\frac{\alpha(1+\epsilon)}{\nu f(n)^\nu}\int_{z\in(z_b,\infty)} w_\delta(z)^\nu dF_Z(z) + \mathbb{P}(Z \ge f(n)^\kappa).$$

The upper bound in (34) follows by multiplying above by $f(n)^\nu$, taking limits as $n \to \infty$, noting that ϵ is arbitrary and Z is light tailed.

Using the above three steps together with (29) establishes that

$$\limsup_{n\to\infty} f(n)^\nu \mathbb{P}(A_n) \le \frac{\alpha}{\nu}\int_{z\in(z_b,\infty)} w_\delta(z)^\nu dF_Z(z).$$

Note that the left hand side is independent of δ; $w_\delta(z)$ is bounded from above by a linear function in z; $w_\delta(z) \to w(z)$ as $\delta \to 0$; and Z is light tailed. Using the dominated convergence theorem when letting $\delta \to 0$, we deduce that

$$\limsup_{n\to\infty} f(n)^\nu \mathbb{P}(A_n) \le \frac{\alpha}{\nu}\int_{z\in(z_b,\infty)} w(z)^\nu dF_Z(z).$$

Step 4. We now prove the following lower bound

$$\liminf_{n\to\infty} f(n)^\nu \mathbb{P}\left(A_n,\, W \le \frac{w_\delta(Z)}{f(n)},\, Z > z_b\right) \ge \frac{\alpha}{\nu}\int_{z\in(z_b,\infty)} w(z)^\nu dF_Z(z). \tag{37}$$

Let

$$I_{n,-\hat{\delta}} := \mathbb{P}\left(A_n,\, W \le \frac{w_{-\hat{\delta}}(Z)}{f(n)},\, Z \ge z_{b_{-\hat{\delta}}}\right),$$

$\hat{\delta} > 0$. Recall that $z_{b_{-\hat{\delta}}} \geq z_b$. Thus, $\mathbb{P}\left(A_n,\, W \leq \frac{w_{\hat{\delta}}(Z)}{f(n)},\, Z > z_b\right)$ is bounded below by $I_{n,0}$, which in turn is bounded below by $I_{n,-\hat{\delta}}$. Next we will find a lower bound for $\liminf_{n\to\infty} f(n)^\nu I_{n,-\hat{\delta}}$.

Note that for $0 < \kappa < 1$, $I_{n,-\hat{\delta}}$ is lower bounded by

$$\int_{z\in(z_{b_{-\hat{\delta}}},f(n)^\kappa)} \int_{w\leq \frac{w_{-\hat{\delta}}(z)}{f(n)}} P_{wf(n),z}(A_n) f_W(w)dwdF_Z(z). \tag{38}$$

Further, as the conditional probability $P_{w,z}(A_n)$ is non-increasing in w, then for $w \leq \frac{w_{-\hat{\delta}}(z)}{f(n)}$ we have $P_{wf(n),z}(A_n) \geq P_{w_{-\hat{\delta}}(z),z}(A_n)$. Also note that for any $\epsilon > 0$, and for n sufficiently large $f_W(w) \geq \alpha(1-\epsilon)w^{\nu-1}$ for $0 \leq w \leq \frac{w_{-\hat{\delta}}(z)}{f(n)}$ for $z \in (z_{b_{-\hat{\delta}}}, f(n)^\kappa)$. Thus, for sufficiently large n, $I_{n,-\hat{\delta}}$ is lower bounded by

$$\frac{\alpha(1-\epsilon)}{\nu f(n)^\nu} \int_{z\in(z_{b_{-\hat{\delta}}},f(n)^\kappa)} w_{-\hat{\delta}}(z)^\nu P_{w_{-\hat{\delta}}(z),z}(A_n)dF_Z(z). \tag{39}$$

We also have that for $z \geq z_{b_{-\hat{\delta}}}$ and $r(w_{-\hat{\delta}}(z), z) = b + \hat{\delta}$, the probability $\mathbb{P}_{w_{-\hat{\delta}}(z),z}(A_n) \to 1$ as $n \to \infty$ by the law of large numbers and $\mathbb{I}\{z < f(n)^\kappa\} \to 1$ as $n \to \infty$. Taking limits in (39) and appealing to the bounded convergence theorem we have

$$\liminf_{n\to\infty} f(n)^\nu I_{n,-\hat{\delta}} \geq \frac{\alpha(1-\epsilon)}{\nu f(n)^\nu} \int_{z\in(z_{b_{-\hat{\delta}}},\infty)} w_{-\hat{\delta}}(z)^\nu dF_Z(z).$$

Thus, for all $\hat{\delta}$ and ϵ sufficiently small we have

$$\liminf_{n\to\infty} f(n)^\nu \mathbb{P}\left(A_n,\, W \leq \frac{w_\delta(Z)}{f(n)},\, Z > z_b\right) \geq \frac{\alpha(1-\epsilon)}{\nu} \int_{z\in(z_{b_{-\hat{\delta}}},\infty)} w_{-\hat{\delta}}(z)^\nu dF_Z(z).$$

Taking first $\epsilon \to 0$ followed by $\hat{\delta} \to 0$, we get (37). (The fact that $w_{\hat{\delta}}(z)$ is bounded by a linear function of z allows the determination of the second limit using the dominated convergence theorem.) Combining Step 4 with the upper bound completes the proof of Theorem 1. ■

A.3 Proof of Theorem 2.

Using Theorem 1, it suffices to show that

$$\frac{f(n)^\nu}{n}\mathbb{E}[(L_n - nb)^+] \to \alpha \int_{z_b}^{\infty} \int_0^{w(z)} \left(\sum_{j=1}^{|\mathcal{V}|} e_j q_j p_{w,z,j} - b\right) w^{\nu-1}dwdF_Z(z), \tag{40}$$

as $n \to \infty$. Here, $(Y)^+ := \max(0, Y)$.

Fix $\delta > 0$. We decompose the left hand side of (40) into the following three terms

$$\begin{aligned} \mathbb{E}[(L_n - nb)^+] \;=\;& \mathbb{E}[(L_n - nb)^+\mathbb{I}\{z \leq z_b\}] + \mathbb{E}\left[(L_n - nb)\mathbb{I}\left\{W > \frac{w_\delta(Z)}{f(n)},\, Z \geq z_b\right\}\right] \qquad (41)\\ &+\mathbb{E}\left[(L_n - nb)\mathbb{I}\left\{W \leq \frac{w_\delta(Z)}{f(n)},\, Z > z_b\right\}\right]. \end{aligned}$$

We divide the remaining part of the proof into four steps. The first and the second step show that the first and second term on the right hand side of (41), respectively, are asymptotically negligible. The third and the fourth step develop upper and lower bounds on the third term on the right-hand-side of (41).

Step 1. We show that

$$\lim_{n\to\infty} \frac{f(n)^\nu}{n}\mathbb{E}[(L_n - nb)^+\mathbb{I}\{Z \le z_b\}] = 0. \tag{42}$$

Note that $n^{-1}(L_n - nb) < (\max_i e_i - b)$. Thus, we have

$$\begin{aligned} \frac{f(n)^\nu}{n}\mathbb{E}[(L_n - nb)^+\mathbb{I}\{Z \le z_b\}] &= \mathbb{E}[(L_n - nb)\mathbb{I}\{L_n > nb,\, Z \le z_b\}] \\ &\le f(n)^\nu(\max_i e_i - b)\mathbb{P}(L_n > nb,\, Z \le z_b). \end{aligned}$$

The assertion in (42) now follows from Step 1 of proof of Theorem 1.

Step 2. We show that

$$\lim_{n\to\infty} \frac{f(n)^\nu}{n}\mathbb{E}\left[(L_n - nb)^+\mathbb{I}\left\{W > \frac{w_\delta(Z)}{f(n)},\, Z \ge z_b\right\}\right] = 0. \tag{43}$$

As in Step 1, the left hand side is bounded above by

$$\lim_{n\to\infty} (\max_i e_i - b) f(n)^\nu \mathbb{P}\left(L_n > nb,\, W > \frac{w_\delta(Z)}{f(n)},\, Z \ge z_b\right),$$

which by Step 2 of the proof of Theorem 1, gives (43).

Step 3. We now develop an asymptotic upper bound on the third term on the right hand side of (41), which in turn gives an asymptotic upper bound on $\mathbb{E}[L_n - nb]$. To this end, we show that for $\delta > 0$,

$$\lim_{n\to\infty} \frac{f(n)^\nu}{n}\mathbb{E}\left[(L_n - nb)^+\mathbb{I}\left\{W \le \frac{w_\delta(Z)}{f(n)},\, Z > z_{b_0}\right\}\right] \le \alpha\int_{z_b}^{\infty}\int_0^{w(z)}\left(\sum_{j=1}^{|\mathcal{V}|} e_j q_j p_{w,z,j} - b\right) w^{\nu-1} dw dF_Z(z).$$

To see this, note that

$$\mathbb{E}\left[(L_n - nb)^+\mathbb{I}\left\{W \le \frac{w_\delta(Z)}{f(n)},\, Z > z_{b_0}\right\}\right] = \int_{z_b}^{\infty}\int_0^{\frac{w_\delta(Z)}{f(n)}} \mathbb{E}_{f(n)w,z}[(L_n - nb)^+] f_W(w) dF_Z(z). \tag{44}$$

For any $0 < \kappa < 1$, this is upper bounded by

$$\int_{z_b}^{f(n)^\kappa}\int_0^{\frac{w_\delta(z)}{f(n)}} \mathbb{E}_{f(n)w,z}[(L_n - nb)^+] f_W(w) dw dF_Z(z) + \mathbb{E}[(L_n - nb)^+\mathbb{I}\{Z \ge f(n)^\kappa\}]. \tag{45}$$

Note from (3) that for any $\epsilon > 0$ there exists n sufficiently large such that $f_W(w) \le \alpha(1+\epsilon)w^{\nu-1}$ for $0 \le w \le \frac{w_\delta(z)}{f(n)}$ and $z \in (z_b, f(n)^\kappa)$. (This follows since $w_\delta(z)$ increases at most at a linear rate as a function of z). Thus, for sufficiently large n, (45) is upper bounded by

$$\alpha(1+\epsilon)\int_{z_b}^{\infty}\int_0^{\frac{w_\delta(z)}{f(n)}} \mathbb{E}_{w,z}[(L_n - nb)^+] w^{\nu-1} dw dF_Z(z) + n(\max_i e_i - b)\mathbb{P}(Z \ge f(n)^\kappa).$$

The last term multiplied by $\frac{f(n)^\nu}{n}$ vanishes in the limit as $f(n)^\nu \mathbb{P}(Z \geq f(n)^\nu) \to 0$ as $n \to \infty$. Next consider the first term, changing the variable and letting $y = wf(n)$ we get

$$\frac{\alpha(1+\epsilon)}{f(n)^\nu} \int_{z_b}^{\infty} \int_{0}^{w_\delta(z)} \mathbb{E}_{y,z}[(L_n - nb)] y^{\nu-1} dy dF_Z(z).$$

The desired upper bound follows by multiplying the above by $f(n)^\nu/n$, taking limits as $n \to \infty$, noting that ϵ is arbitrary, L_n/n is bounded, and the fact that

$$\lim_{n\to\infty} \mathbb{E}_{y,z}\left[\left(\frac{L_n}{n} - nb\right)^+\right] = \left(\sum_{j=1}^{|\mathcal{V}|} e_j q_j p_{y,z,j} - b\right)^+.$$

Using the above three steps together with (41) establishes that

$$\lim_{n\to\infty} \frac{f(n)^\nu}{n} \mathbb{E}[(L_n - nb)^+] \leq \alpha \int_{z_b}^{\infty} \int_{0}^{w_\delta(z)} \left(\sum_{j=1}^{|\mathcal{V}|} e_j q_j p_{w,z,j} - b\right) w^{\nu-1} dw dF_Z(z).$$

Note that the left hand side is independent of δ; $w_\delta(z)$ is bounded from above by a linear function in z; $w_\delta(z) \to w(z)$ as $\delta \to 0$; and Z is light tailed. Using the dominated convergence theorem when letting $\delta \to 0$, we deduce that $\lim_{n\to\infty} \frac{f(n)^\nu}{n} \mathbb{E}[(L_n - nb)^+]$

$$\leq \alpha \int_{z_b}^{\infty} \int_{0}^{w(z)} \left(\sum_{j=1}^{|\mathcal{V}|} e_j q_j p_{w,z,j} - b\right) w^{\nu-1} dw dF_Z(z).$$

Step 4. We now prove the following lower bound

$$\liminf_{n\to\infty} \frac{f(n)^\nu}{n} \mathbb{E}\left[(L_n - nb)^+ \mathbb{I}\left\{W \leq \frac{w_\delta(Z)}{f(n)},\, Z > z_{b_0}\right\}\right] \tag{46}$$

$$\geq \alpha \int_{z_b}^{\infty} \int_{0}^{w(z)} \left(\sum_{j=1}^{|\mathcal{V}|} e_j q_j p_{w,z,j} - b\right) w^{\nu-1} dw dF_Z(z).$$

To see this, note that for a given $\widetilde{\delta} > 0$, there exists N such that $\mathbb{E}_{w,z}[L_n] \geq n\left(\sum_{j=1}^{|\mathcal{V}|} e_j p_{w,z,j} - \widetilde{\delta}\right)$ for all $n > N$. Thus, we have that the left-hand-side of (46) is lower bounded by

$$\begin{aligned} &\liminf_{n\to\infty} f(n)^\nu \int_{z_b}^{\infty} \int_{0}^{\frac{w(z)}{f(n)}} \left(\sum_{j=1}^{|\mathcal{V}|} p_{f(n)w,z,j} e_j - b - \widetilde{\delta}\right) f_W(w) dw dF_Z(z) \\ \geq \quad & \alpha(1-\epsilon) \liminf_{n\to\infty} f(n)^\nu \int_{z_b}^{\infty} \int_{0}^{\frac{w(z)}{f(n)}} \left(\sum_{j=1}^{|\mathcal{V}|} p_{f(n)w,z,j} e_j - b - \widetilde{\delta}\right) w^{\nu-1} dw dF_Z(z), \end{aligned}$$

for any $\epsilon > 0$. The last inequality follows from (3). Let $y = f(n)w$. Thus the above expression equals

$$\alpha(1-\epsilon) \int_{z_b}^{\infty} \int_{0}^{w(z)} \left(\sum_{j=1}^{|\mathcal{V}|} e_j q_j p_{y,z,j} - b - \widetilde{\delta}\right) y^{\nu-1} dy dF_Z(z).$$

Taking limits as $\epsilon \to 0$ and $\tilde{\delta} \to 0$, we get the desired result. This completes the proof. ■

A.4 Proof of Theorem 3

Lemma 1 and Lemma 2 are useful in proving Theorem 3. We need some preliminaries before we state these lemmas (the proofs of the lemmas are relegated to Appendix B). On the set $\{W > \frac{w(Z)}{f(n)}\}$, let

$$\begin{aligned}\hat{L} &= \prod_{j\le|\mathcal{V}|}\left(\frac{p_{Wf(n),Z,j}}{p^{\theta^*}_{Wf(n),Z,j}}\right)^{Y_jq_jn}\left(\frac{1-p_{Wf(n),Z,j}}{1-p^{\theta^*}_{Wf(n),Z,j}}\right)^{(1-Y_j)q_jn}\\ &= \exp\left(-n(\theta^*\sum_{j\le|\mathcal{V}|}Y_jq_je_j-\sum_{j\le|\mathcal{V}|}q_j\Lambda_j(\theta^*))\right). \qquad (47)\end{aligned}$$

Note that $A_n = \{\sum_{j\le|\mathcal{V}|}Y_jq_je_j \ge b\}$. It follows that

$$\hat{L}\mathbb{I}\{A_n\} \le \exp\left[-n(\theta^*b-\sum_{j\le|\mathcal{V}|}q_j\Lambda_j(\theta^*))\right]\mathbb{I}\{A_n\} \quad \text{a.s.} \qquad (48)$$

Observe that $\theta b-\sum_{j\le|\mathcal{V}|}\Lambda_j(\theta)$ is a strictly concave function that equals 0 at $\theta=0$ and is maximized at θ^* so that

$$\theta^*b-\sum_{j\le|\mathcal{V}|}\Lambda_j(\theta^*) > 0. \qquad (49)$$

Lemma 1 *Suppose that there exist positive constants K_1 and β_1 and a non-negative function $g(n,w,z)$ such that,*

$$p_{Wf(n),Z,j} \le K_1\exp(-\beta_1 g(n,W,Z)) \quad a.s.$$

Then, there exist positive constants K_2, β_2 such that

$$\hat{L}\mathbb{I}\{A_n\} \le K_2^n\exp[-\beta_2 ng(n,W,Z)]\mathbb{I}\{A_n\} \quad a.s.$$

Lemma 2

$$\limsup_{n\to\infty} f(n)^{2\nu}\int_z \exp(2\Lambda_W(-\theta_{z,n}))dF_Z(z) < \infty. \qquad (50)$$

Proof of Theorem 3. Recall that for a positive constant ξ, $\tilde{w}(z)=\max(\xi,w(z))$. Fix constants $K_3, K_4>0$. To prove the theorem we re-express

$$\mathbb{E}^*{L_*}^2\mathbb{I}\{A_n\} = \mathbb{E}^*\left[{L_*}^2\mathbb{I}\left\{A_n, W\le\frac{K_3\tilde{w}(Z)+K_4}{f(n)}\right\}\right]+\mathbb{E}^*\left[{L_*}^2\mathbb{I}\left\{A_n, W>\frac{K_3\tilde{w}(Z)+K_4}{f(n)}\right\}\right].$$

The proof is divided into two steps.

Step 1. For constants $K_3, K_4>0$, we establish that

$$\limsup_{n\to\infty} f(n)^{2\nu}\widetilde{\mathbb{E}}\left[L_*^2\mathbb{I}\left\{A_n, W\le\frac{K_3\tilde{w}(Z)+K_4}{f(n)}\right\}\right] < \infty. \qquad (51)$$

From (48) and (49), it follows that

$$\prod_{j\le|\mathcal{V}|}\left(\frac{p_{Wf(n),Z,j}}{p^*{}_{Wf(n),Z,j}}\right)^{Y_jq_jn}\left(\frac{1-p_{Wf(n),Z,j}}{1-p^*{}_{Wf(n),Z,j}}\right)^{(1-Y_j)q_jn}\mathbb{I}\{A_n\}\le\mathbb{I}\{A_n\}.$$

From this and (11) we have that on the set $\{A_n, W\le\frac{K_3\tilde{w}(Z)+K_4}{f(n)}\}$,

$$L_*\le\exp(\Lambda_W(-\theta_{Z,n}))\exp[\nu(K_3+K_4\xi^{-1})].$$

Integrating ${L_*}^2$ over this set under $\mathbb{P}^*$, (51) follows from (50).

Step 2. We show that for $K_3, K_4>0$ with K_3 and K_4 sufficiently large,

$$\lim_{n\to\infty}f(n)^{2\nu}\widetilde{\mathbb{E}}\left[{L_*}^2\mathbb{I}\left\{A_n, W\ge\frac{K_3\tilde{w}(Z)+K_4}{f(n)}\right\}\right]<\infty. \tag{52}$$

Recall that

$$p_{Wf(n),Z,i}=\mathbb{P}\left(\eta>\frac{a_iWf(n)-\rho Z}{\sqrt{1-\rho^2}}\middle|W,Z\right)\le\mathbb{P}\left(\eta>\frac{\min_i a_iWf(n)-\rho Z}{\sqrt{1-\rho^2}}\middle|W,Z\right)\quad\text{a.s.}$$

Since, η is light tailed, there exist constants K_5 and β_3 such that

$$\mathbb{P}\left(\eta>\frac{\min_i a_iWf(n)-\rho Z}{\sqrt{1-\rho^2}}\middle|W,Z\right)\le K_5\exp(-\beta_3(\min_i a_iWf(n)-\rho Z))\quad\text{a.s.}$$

This and Lemma 1 above imply that there exist constants $K_6>0$ and $\beta_4>0$ such that

$$\hat{L}\mathbb{I}\{A_n\}\le K_6^n\exp(-n\beta_4(\min_i a_iWf(n)-\rho Z))\mathbb{I}\{A_n\}\quad\text{a.s.,}$$

where $\hat{L}$ is as defined in (47). It follows that

$$L_*\mathbb{I}\{A_n\}\le\exp\left[\frac{\nu Wf(n)}{\tilde{w}(Z)}+\Lambda_W(-\theta_{Z,n})\right]K_6^n\exp(-n\beta_4(\min_i a_iWf(n)-\rho Z))\mathbb{I}\{A_n\}\quad\text{a.s..}$$

We restrict our discussion to the set $\{A_n, W\ge\frac{K_3\tilde{w}(Z)+K_4}{f(n)}\}$. Note that,

$$\begin{aligned}L_*\quad\le\quad&\exp\Big[\Lambda_W(-\theta_{Z,n})+\frac{\nu(K_3\tilde{w}(Z)+K_4)}{\tilde{w}(Z)}+\frac{\nu f(n)(W-\frac{K_3\tilde{w}(Z)+K_4}{f(n)})}{\xi}\Big]\\ \times\quad&K_6^n\exp\Big[-n\beta_4\{\min_i a_if(n)\big(W-\frac{K_3\tilde{w}(Z)+K_4}{f(n)}\big)\\ +\quad&(K_3\min_i a_i\tilde{w}(Z)-\rho Z)+\min_i a_iK_4\}\Big]\quad\text{a.s.}\end{aligned}$$

Note that $\frac{\nu(K_3\tilde{w}(Z)+K_4)}{\tilde{w}(Z)}$ is upper bounded by a constant for all Z. Select K_3 large enough so that $K_3\min_i a_i\tilde{w}(Z)-\rho Z$ is bounded from below by a non-negative number for all Z. Again note that for n large enough,

$$\frac{\nu f(n)(W-\frac{K_3\tilde{w}(Z)+K_4}{f(n)})}{\xi}\le n\beta_4\min_i a_if(n)\left(W-\frac{K_3\tilde{w}(Z)+K_4}{f(n)}\right)\quad\text{a.s.,}$$

and select K_4 large so that we have $K_6^n \leq \exp(K_4\beta_4 \min_i a_i n)$. Then, it follows that on the set $\{A_n, W \geq \frac{K_3\tilde{w}(Z)+K_4}{f(n)}\}$, $L_* \leq K_7 \exp[\Lambda_W(-\theta_{z,n})]$ for n large enough, for an appropriate constant $K_7 > 0$. Integrating ${L_*}^2$ over this set under $\mathbb{P}^*$ (52) follows from (50). The result asserted in the theorem follows from (51) and (52) by selecting K_3 and K_4 sufficiently large. This completes the proof. ■

A.5 Proof of Theorem 4

Fix $\delta > 0$ and a positive constant ξ. Let $\tilde{w}_\delta(z) = \max(\xi, w_\delta(z))$ for $z \geq z_{b_\delta}$ and let $\tilde{w}_\delta(z) = \xi$ for $z < z_{b_\delta}$. The proof is divided into two steps.

Step 1. We first establish

$$\lim_{n\to\infty} \frac{1}{\log f(n)} \log \bar{\mathbb{E}}[\bar{L}^2 \mathbb{I}\Big\{A_n, V > \frac{f(n)}{\tilde{w}_\delta(Z)}\Big\}] \leq -2\nu. \tag{53}$$

Fix $0 < \kappa < 1$. Note that the likelihood ratio is uniformly upper bounded by K, and thus on the set $\{Z \geq f(n)^\kappa\}$,

$$\bar{\mathbb{E}}\bar{L}^2 \mathbb{I}\{Z \geq f(n)^\kappa\} \leq K^2 \mathbb{P}(Z \geq f(n)^\kappa).$$

Since Z is light tailed, it follows that

$$\lim_{n\to\infty} \frac{1}{\log f(n)} \log \bar{\mathbb{E}}\bar{L}^2 \mathbb{I}\{Z \geq f(n)^\kappa\} = -\infty.$$

On the set $\{A_n, V > f(n)/2\tilde{w}(Z), Z \leq f(n)^\kappa\}$ the likelihood ratio is upper bounded by $f_V(v)/\bar{f}_V(v)$ and hence for sufficiently large n it is in turn bounded by

$$\frac{\alpha \log f(n)}{(1 - F_V(c_1))c_1^\nu} \frac{1}{v^{\nu - 1/\log f(n)}}(1 + o(1)),$$

where $o(1) \to 0$ as $v \to \infty$ and hence as $n \to \infty$ on this set. This is then upper bounded by

$$\log f(n) \left(\frac{\tilde{w}_\delta(Z)}{f(n)}\right)^\nu, \tag{54}$$

for all n sufficiently large. Squaring this upper bound on the likelihood ratio, multiplying it with the indicator $\mathbb{I}\{A_n, V > \frac{f(n)}{\tilde{w}_\delta(Z)}, Z \leq f(n)^\kappa\}$, taking expectation with respect to Z and V, we get that

$$\bar{\mathbb{E}}\left[\bar{L}^2 \mathbb{I}\left\{A_n, V > \frac{f(n)}{\tilde{w}_\delta(Z)}, Z \leq f(n)^\kappa\right\}\right] \leq \Big(\frac{\log f(n)}{f(n)^\nu}\Big)^2 \bar{\mathbb{E}}[\tilde{w}(Z)^{2\nu}].$$

Finally, taking logarithms of the resultant bound, dividing by $\log f(n)$ and taking the limit as $n \to \infty$ we get (53).

Step 2. To complete the proof we next establish

$$\lim_{n\to\infty} \frac{1}{\log f(n)} \log \bar{\mathbb{E}}\bar{L}^2 \mathbb{I}\Big\{A_n, V \leq \frac{f(n)}{\tilde{w}_\delta(Z)}\Big\} = -\infty. \tag{55}$$

Recall that there exists a finite positive constant K such that

$$\frac{f_V(v)}{\bar{f}_V(v)} \leq K,$$

for all v. Also note that on the set $\{V \leq \frac{f(n)}{\tilde{w}_\delta(Z)}\}$, or equivalently, $\{W \geq \frac{\tilde{w}_\delta(Z)}{f(n)}\}$, we have $r(Wf(n), z) \leq b - \bar{\xi}$ for some $\bar{\xi}$. Thus, from Hoeffding's inequality it is easily seen that there exists a $\beta > 0$ such that

$$\mathbb{P}\Big(A_n, V \leq \frac{f(n)}{\tilde{w}_\delta(Z)}\Big) \leq \exp(-\beta n).$$

This establishes (55) and concludes the proof. ■

A.6 Proof of Theorem 5.

Using Theorem 2, it suffices to prove that

$$\limsup_{n\to\infty} \frac{\sigma^2(n,b)}{n^2} < \infty.$$

To this end, we will prove that each term on the right-hand-side of (17) scaled by n^2 is finite. Consider the first term on the right-hand-side of (17). We first observe that

$$\limsup_{n\to\infty} \frac{f(n)^{2\nu}}{n^2}\widetilde{\mathbb{E}}[L_*^2(L_n - nb)^2\mathbb{I}\{L_n > nb\}] < \infty. \tag{56}$$

To see this, note that $(L_n - nb)^2 \leq (\max_j e_j n)^2$. Also,

$$\limsup_{n\to\infty} f(n)^{2\nu}\widetilde{\mathbb{E}}[L_*^2\mathbb{I}\{L_n > nb\}] < \infty$$

follows from Theorem 3 which states that the proposed algorithm has bounded relative error for estimating $\mathbb{P}(L_n > nb)$. Thus, we have

$$\limsup_{n\to\infty} \frac{\sigma_1^2(n,b)}{n^2\mu_2^2(n,b)} < \infty,$$

since

$$\begin{aligned} \limsup_{n\to\infty} \frac{\sigma_1^2(n,b)f(n)^{2\nu}}{n^2} &< \infty, & (57)\\ \text{and} \quad \limsup_{n\to\infty} \mu_2^2(n,b)f(n)^{2\nu} &< \infty. & (58)\end{aligned}$$

Here (57) follows from (56) and (58) follows from Theorem 1. Similarly,

$$\limsup_{n\to\infty} \frac{\mu_1^2(n,b)\sigma_2^2(n,b)}{n^2\mu_2^4(n,b)} < \infty.$$

For the last term, note that

$$\limsup_{n\to\infty} \frac{f(n)^{2\nu}\widetilde{\mathbb{E}}[L_*^2(L_n - nb)\mathbb{I}\{L_n > nb\}]}{n} < \infty.$$

Therefore, $\limsup_{n\to\infty} \frac{\sigma_{12}(nb)\mu_1(n,b)}{n^2\mu_2^3(n,b)} < \infty$, and the proof is complete. ■

B Proofs of Side Lemmas

Proof of Lemma 1. From (48) and (49), it follows that for $\theta > 0$,

$$\hat{L}\mathbb{I}\{A_n\} \le \exp\Big[-n(\theta b - \sum_{j \le |\mathcal{V}|} q_j \Lambda_j(\theta))\Big]\mathbb{I}\{A_n\} \quad \text{a.s.}$$

For $\theta = \frac{\beta_1 g(n,W,Z)}{\max_i e_i}$, it follows that $\exp(\theta e_j) p_{W,Z,j}$ is bounded by a constant and hence

$$\exp[\sum_{j \le |\mathcal{V}|} \Lambda_j(\theta))]$$

is bounded by a constant. The result follows. ■

Proof of Lemma 2. For any $\epsilon > 0$, $\frac{1}{2}\int_z \exp(2\Lambda_W(-\theta_{z,n}))dF_Z(z)$ is upper bounded by

$$\int_z \Big(\int_{w<\epsilon} \exp(-\theta_{z,n} w) f_W(w) dw\Big)^2 dF_Z(z) + \int_z \Big(\int_{w \ge \epsilon} \exp(-\theta_{z,n} w) f_W(w) dw\Big)^2 dF_Z(z). \qquad (59)$$

Consider the second term in right hand side. Using Jensen's inequality, it is upper bounded by

$$\int_z \int_{w \ge \epsilon} \exp(-2\theta_{z,n} w) f_W(w) dw dF_Z(z).$$

For $0 < \kappa < 1$ this can in turn be bounded by

$$\int_{z \le f(n)^\kappa} \int_{w \ge \epsilon} \exp(-2\theta_{z,n}\epsilon) f_W(w) dw dF_Z(z) + \mathbb{P}(Z \ge f(n)^k).$$

Since Z is light tailed, it follows that $\lim_{n\to\infty} f(n)^{2\nu}\mathbb{P}(Z \ge f(n)^k) = 0$. Also note that since $\tilde{w}(f(n)^\kappa)$ is upper bounded by a constant times $f(n)^\kappa$, it follows that when $z \le f(n)^\kappa$, $\theta_{z,n}$ is lower bounded by a constant times $f(n)^{1-\kappa}$, and hence, $f(n)^{2\nu}\exp(-2\theta_{z,n}\epsilon)\mathbb{I}\{z \le f(n)^\kappa\}$ is uniformly bounded for all n and converges to zero as $n \to \infty$. By the bounded convergence theorem

$$\lim_{n\to\infty} f(n)^{2\nu} \int_{z \le f(n)^\kappa} \exp(-2\theta_{z,n}\epsilon) dF_Z(z) = 0.$$

Now consider the first term in right hand side of (59). Note that due to (3), for $\epsilon > 0$ sufficiently small, there exists a $\tau > 0$ such that this term is upper bounded by

$$\alpha^2(1+\tau)^2 \int_z \Big(\int_{w \le \epsilon} \exp(-\theta_{z,n} w) w^{\nu-1} dw\Big)^2 dF_Z(z). \qquad (60)$$

Setting $y = \theta_{z,n} w$, (60) equals

$$\alpha^2(1+\tau)^2 \int_z (\frac{\tilde{w}(z)}{df(n)})^{2\nu} \Big(\int_{w \le \theta_{z,n}\epsilon} \exp(-y) y^{\nu-1} dy\Big)^2 dF_Z(z).$$

This is upper bounded by

$$\alpha^2(1+\tau)^2 \int_z \Big(\frac{\tilde{w}(z)}{df(n)}\Big)^{2\nu} dF_Z(z) \Big(\int_w \exp(-y) y^{\nu-1} dy\Big)^2,$$

and the result follows. ■

References

Basel Committee on Banking Supervision. (2002), 'The New Basel Capital Accord,' Basel.

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

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

Embrechts, P., Lindskog, F. and McNeil, A. J. (2003), Modelling dependence with copulas and applications to risk management, *in* S. Rachev, ed., 'Handbook of Heavy Tailed Distributions in Finance', Elsevier, NY.

Feller, W.(1970), *An Introduction to Probability Theory and Its Applications, Volume 1*, John Wiley & Sons.

Frey, R. and McNeil, A. (2001), 'Modelling Dependent Defaults'. Working Paper, Universität and ETH Zürich.

Frey, R. and McNeil, A.(2003), 'Dependent deafults in models of portfolio credit risk', *Journal of Risk*, **6**, 59–92.

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

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

Glasserman, P. and Li, J. (2003), 'Importance sampling for portfolio credit risk', *Manag. Sci.*, **51**, 1643–1656.

Gupton, G., Finger, C. and Bhatia, M. (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. and Shahabuddin, P. (2002), 'Simulating heavy-tailed processes using delayed hazard rate twisting', *ACM TOMACS* **12**, 94–118.

Juneja, S. and Shahabuddin, P. (2006), 'Rare Event Simulation Techniques: An Introduction and Recent Advances', To appear in *Handbook on Simulation*. Elsevier. Editors: Shane Henderson and Barry Nelson.

Kealhofer, S. and Bohn, J. (2001), Portfolio management of credit risk, Technical report, KMV working paper, New York.

Li, D. (2000), ‘On default correlation: a copula function approach’, *J. of Fixed Income* **9**, 43–54.

Mashal, R. and Zeevi, A. (2003), ‘Beyond correlation: Extreme co-movements between financial assets’. Working paper, Columbia University.

McNeil, A. J., Frey, R. and Embrechts, P. (2005), *Quantitative Risk Management: concepts, techniques and tools* , Princeton University Press, Princeton.

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

Schloegel, L. and O’Kane, D. (2005), ‘A note on the large homogenous portfolio approximation with the Student-t copula’, *Finance and Stochastics* **9**, 577–584.

Serfling, R. J. (1981), *Approximation Theorems of Mathematical Statistics*, Wiley-Interscience