

EFFICIENT IMPORTANCE SAMPLING FOR REDUCED FORM MODELS IN CREDIT RISK

Achal Bassamboo

Kellogg School of Management
Northwestern University
Evanston, IL, 60208, U.S.A.

Sachin Jain

Amaranth Group Inc.
One American Lane
Greenwich, CT, 06831, U.S.A.

ABSTRACT

In this paper we study the problem of estimating probability of large losses in the framework of doubly stochastic credit risk models. We derive a logarithmic asymptote for the probability of interest in a specific asymptotic regime and propose an asymptotically optimal importance sampling algorithm for efficiently estimating the same. The numerical results in the last section corroborate our theoretical findings.

1 INTRODUCTION

With the expansion in the corporate credit risk market and the surge in the trading of securitized products, such as Collateralized Debt Obligations (CDOs), there is a need for sophisticated credit risk models. The pricing and risk management of these products requires the computation of portfolio loss distributions. The current industry standard for modeling portfolio losses is to have a non-stochastic (maybe time varying) intensity for each obligor (see Andersen, Sidenius, and Basu 2003 and Li 2001) boot-strapped from the current term structure of CDS spreads for the individual names and to correlate the default times using a *copula* function. Some commonly used copula functions are the normal copula and the student-t copula. See Li (2001) for the use of copula functions in modeling correlation of default times. One shortcoming of the copula framework is that, in its current implementation, the intensity (and hence the credit spread) dynamics are omitted, thus ignoring the risk associated with changes in credit spreads.

In the current paper, we focus on the class of doubly stochastic models (see Duffie and Singleton 2003) where the dynamics of the intensity process are governed by affine-jump diffusion processes. A stochastic intensity model, though captures spread dynamics, tends to be computationally intensive for the purpose of pricing and risk management since it requires one to simulate the path of the default intensity. In order to reduce the computational burden, we apply the idea of *importance sampling* for estimating proba-

bilities of large portfolio losses and for pricing CDOs using simulation. Importance sampling is a variance reduction technique for rare event simulation and is based on the idea of conducting the simulation under a new measure for which the event of interest occurs more frequently.

The most relevant work to ours is the recent paper by Glasserman and Li (2005) on efficient simulation for credit portfolio. They propose an importance sampling algorithm for computing large loss probabilities in a static normal copula model, where the dependency of the default times are modeled through a normal copula function. They work in a homogeneous framework where all firms have the same default probabilities. In this paper we analyze a dynamic reduced form model and also allow for heterogeneity.

The key contributions of the paper can be summarized as:

1. We derive a logarithmic asymptote for large loss probabilities for doubly stochastic reduced form models in a specific asymptotic regime.
2. We present an asymptotically optimal change of measure for estimating probabilities of large number of defaults for doubly stochastic reduced form models.
3. We apply our analysis to the case of doubly stochastic Poisson processes where the intensity process follows an affine jump diffusion and is allowed to be different for different obligors.

We also conducted a numerical study to illustrate the efficacy of the proposed importance sampling technique for estimating the probability of large losses.

The paper is organized as follows. In Section 2 and 2 we explain our problem setup and discuss reduced form models, respectively. In Section 3 we provide some preliminaries necessary for understanding simulation using importance sampling. We also provide a notion of optimality in that section. In Section 4 we present our proposed algorithm for a general setup and prove its optimality. In Section 5, we apply

the results of Section 4 to the setting of doubly stochastic reduced form models where the intensity process follows an affine jump diffusion. Section 6 contains numerical results. Proof of all the theorems has been omitted for the purpose of brevity.

2 PROBLEM DEFINITION

Fixing a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an information filtration $\{\mathcal{F}_t : t \geq 0\}$ satisfying the usual conditions (see Protter 2004 for details), let $X = \{X(t) : t \geq 0\}$ be a time-homogeneous Markov process in \mathbb{R}^d , for some integer $d \geq 1$. Consider a system with N obligors. For the i^{th} obligor define the default intensity $\lambda_i(t) = f_i(X(t))$. The default time τ_i of the i^{th} obligor is modeled as the first jump time of a doubly stochastic process $K_i(\cdot)$ with intensity λ_i . Under the doubly stochastic assumption, conditional on $\{\lambda_i(t) : t > 0\}$, $K_i(\cdot)$ is a Poisson process with intensity λ_i . In particular,

$$P(\tau_i > t) = E[P(K_i(t) = 0 | \lambda_i)] = E[e^{-\int_0^t \lambda_i(s) ds}].$$

Intuitively, $\lambda_i(t)\Delta t$ is the probability of default of obligor i in the interval $[t, t + \Delta t)$ conditional on survival upto time t . Moreover, in the modeling of multi-obligor default times, X could include common economy wide, industry specific or firm specific factors. The default time correlation is generated through dependence of firms' intensities on the common factors. A nice feature of intensity models is that they enable one to model the joint movement of the obligors' default probabilities. Note that under the doubly stochastic setting conditional on the covariates X , the default times of the obligors are independent.

The probability of interest is $P\left(\sum_{i=1}^N Z_i \geq b\right)$, where Z_i is an indicator of default by time T for the i^{th} obligor and b is the default threshold. The simulation of the probability of interest is conducted in two stages. In stage 1, we simulate the dynamics of the underlying common factors and in stage 2, we generate default indicators. Since the single name default probability is small, the event of interest is rare. Thus the effort required to get a good estimate of the probability of interest using naive simulation is tremendous, and one has to resort to variance reduction techniques.

In later sections, we apply our analysis to the case where each obligor belongs to one of the N_c classes and the default intensity of the i^{th} obligor belonging to class j is given by $\lambda_i^{(j)}(t) = w_j \cdot X(t) + \epsilon_i^{(j)}(t)$ where $X = (X(t) : t \geq 0)$ is a stochastic process of "common factors," where $X(t) \in \mathbb{R}_+^d$, w_j is the weight on common factors for all firms belonging to class j and $\epsilon_i^{(j)}$ is the idiosyncratic component. Here, we allow for each class of obligors to weigh differently on each of the common factors. Thus, in the case where each of the common factor represents the riskiness of different

industries we would expect IBM to weigh more heavily on the technology factor as compared to the automotive factor, on which firms like GM and Ford would be heavily weighed. The idiosyncratic factors are assumed to follow affine-jump diffusions and represent the firm's own riskiness independent of all other factors.

3 IMPORTANCE SAMPLING PRELIMINARIES

In this section, we state some definitions from the simulation literature. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\{X_n : n \geq 1\}$ be a sequence of random variables adapted to a filtration $\{\mathcal{F}_n : n \geq 1\}$ defined on this space. Let M be a stopping time with respect to the filtration $\{\mathcal{F}_n : n \geq 1\}$. We are interested in the expectation of the random variable $\mathbb{I}(X_1, \dots, X_M)$, where $\mathbb{I}(\cdot)$ is an indicator function of a rare event, i.e. $\gamma = \mathbb{E}[\mathbb{I}(X_1, \dots, X_M)]$. To estimate this expectation by naive simulation, we generate m i.i.d. samples of the function $\mathbb{I}(X_1, \dots, X_M)$ and average them to get an unbiased estimate $\hat{\gamma}_m$. The relative error of the estimator (defined as the ratio of standard deviation and mean) is given by $\sqrt{\frac{(1-\gamma)}{m\gamma}}$. As $\gamma \rightarrow 0$, the number of simulation runs must increase without bound in order to have a small relative error.

Importance Sampling (IS) is a variance reduction technique which is particularly efficient in the computation of small probabilities. Consider another probability measure $\tilde{\mathbb{P}}$ on the same sample space (Ω, \mathcal{F}) such that $\mathbb{P} \ll \tilde{\mathbb{P}}$, i.e., \mathbb{P} is absolutely continuous w.r.t. $\tilde{\mathbb{P}}$, on the set $\{\omega : M < \infty, \mathbb{I}(X_1, \dots, X_M) = 1\}$. It is then a straight forward application of Wald's identity that

$$\mathbb{E}[\mathbb{I}(X_1, \dots, X_M)] = \tilde{\mathbb{E}}[\mathbb{I}(X_1, \dots, X_M)L_M], \quad (1)$$

where $L_M = \left. \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} \right|_{\mathcal{F}_M}$ is the Radon-Nikodym derivative restricted to \mathcal{F}_M , and $\tilde{\mathbb{E}}$ is the expectation operator under the measure $\tilde{\mathbb{P}}$. In light of (1), we can generate i.i.d samples of $\mathbb{I}(X_1, \dots, X_M)L_M$ under the measure $\tilde{\mathbb{P}}$ and the average of these is an unbiased estimate of $\mathbb{E}[\mathbb{I}(X_1, \dots, X_M)]$. The variance of this estimator is $\text{Var}_{\tilde{\mathbb{P}}}[\mathbb{I}(X_1, \dots, X_M)L_M]$. We refer to $\tilde{\mathbb{P}}$ as the Importance Sampling (IS) change-of-measure and L_M as the corresponding likelihood ratio. By choosing the IS change-of-measure appropriately we can substantially reduce the variance of the estimator.

3.1 Asymptotic Optimality

Consider a sequence of rare-events indexed by a parameter u . Let \mathbb{I}_u be the indicator of this rare event, and suppose $\mathbb{E}[\mathbb{I}_u] \rightarrow 0$ as $u \rightarrow \infty$. Let $\tilde{\mathbb{P}}$ be an IS distribution and L be the corresponding likelihood ratio. Let $Z_u = L\mathbb{I}_u$.

Definition 1 (asymptotic optimality) A sequence of IS estimators is said to be asymptotically optimal if

$$\frac{\log \widetilde{\mathbb{E}}[Z_u^2]}{\log \widetilde{\mathbb{E}}[Z_u]} \rightarrow 2 \text{ as } u \rightarrow \infty. \tag{2}$$

3.2 Exponential Twisting

In this subsection, we describe the concept of exponential twisting which will be used in the chapter repeatedly. Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let X be a random variable defined on this space. Let \mathbb{Q} be another probability measure defined on the same space.

Definition 2 (exponential twisting) The measure \mathbb{Q} is said to be an exponentially twisted measure of \mathbb{P} with respect to the random variable X , by a parameter θ , if

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{\exp(\theta X)}{\mathbb{E}^{\mathbb{P}}[\exp(\theta X)]}.$$

Let $M(\theta)$ represent the log-moment generating function, i.e., $M(\theta) = \log \mathbb{E}^{\mathbb{P}}[\exp(\theta X)]$, then the mean of the random variable X under the measure \mathbb{Q} is given by $M'(\theta)$.

4 A TWO-STAGE SIMULATION ALGORITHM AND ITS OPTIMALITY

In this section we derive results for simulating the probability of interest in a general setup and prove the asymptotic optimality of the proposed algorithm. The results of this section are subsequently applied to the specific setup where the dynamics of the intensity process is governed by an affine jump diffusion. Consider a “sequence of portfolios” indexed by κ . In the κ^{th} system, the number of obligors are κ , and the random variable Z_κ denotes the number of defaults observed upto some finite time T . Our goal is to efficiently estimate the probability of a large fraction b of defaults, i.e., $\mathbb{P}(Z_\kappa \geq \kappa b)$.

In the setting of reduced form doubly stochastic models our problem can be modeled in two stages. In the first stage, a d -dimensional random variable A_κ from distribution $G_\kappa(\cdot)$ is generated. In the second stage, conditional on A_κ the number of defaults Z_κ is generated from the conditional distribution $F_\kappa(\cdot; A_\kappa)$, where $F_\kappa(y; A_\kappa) = \mathbb{P}(Z_\kappa \leq y | A_\kappa)$. We elaborate on the specifics of A_κ , and the distributions $G_\kappa(\cdot)$ and $F_\kappa(\cdot; A_\kappa)$ in the context of doubly stochastic poisson processes in later sections.

This section is organized as follows. We first derive a logarithmic asymptote for the probability of interest and then describe an asymptotically optimal importance sampling algorithm for estimating the same using simulation.

Define $M_\kappa^{(1)}(\theta_1) = \log \mathbb{E}[e^{\theta_1 \cdot A_\kappa}]$, $M_\kappa^{(2)}(\theta_2; p) = \log \mathbb{E}[e^{\theta_2 \cdot Z_\kappa} | A_\kappa = p]$, where θ_1 and p are vectors in \mathbb{R}_+^d and θ_2 is a scalar taking value in \mathbb{R}_+ .

We make the following technical assumptions:

- (A1) $M^{(1)}(\theta_1) := \lim_{\kappa \rightarrow \infty} \kappa^{-1} M_\kappa^{(1)}(\kappa \theta_1)$ exists on the extended real line for all $\theta_1 \in \mathbb{R}_+^d$. Further, there exists $\delta > 0$ such that $M^{(1)}(\delta) < \infty$ and $M^{(1)}$ is essentially smooth.
- (A2) $M^{(2)}(\theta_2; p) := \lim_{\kappa \rightarrow \infty} \kappa^{-1} M_\kappa^{(2)}(\theta_2; p)$ exists on the extended real line for all $\theta_2 \in \mathbb{R}_+$ and $p \in \mathbb{R}_+^d$. Further, there exists $\delta_p > 0$ such that $M^{(2)}(\delta_p; p) < \infty$ and $M^{(2)}(\cdot; p)$ is essentially smooth for all p .
- (A3) We assume that $\mathbb{P}(Z_\kappa > x | A_\kappa = y_1) \leq \mathbb{P}(Z_\kappa > x | A_\kappa = y_2)$, for all $x \in \mathbb{R}_+$ and $y_1, y_2 \in \mathbb{R}_+^d$ satisfying $y_1 < y_2$. The inequality $y_1 < y_2$ holds iff the inequality holds co-ordinate wise.

Assumptions (A1) and (A2) ensure that a large deviation principle holds for random variables A_κ and Z_κ conditional on A_κ , respectively. The third assumption (A3) implies that $Z_\kappa | A_\kappa$ is stochastically increasing in A_κ .

The following theorem provides an asymptotic lower bound for the probability of interest.

Theorem 1 Under assumptions (A1), (A2) and (A3) we have

$$\liminf_{\kappa \rightarrow \infty} \frac{\log \mathbb{P}(Z_\kappa \geq \kappa b)}{\kappa} \geq \sup_{y \in \mathbb{R}_+^d} \left[\inf_{\theta_1 \in \mathbb{R}_+^d} (M^{(1)}(\theta_1) - \theta_1 \cdot y) + \inf_{\theta_2 \in \mathbb{R}_+} (M^{(2)}(\theta_2; y) - \theta_2 b) \right]. \tag{3}$$

We shall later show that under technical conditions, the lower bound of Theorem 1 is achieved. Let the function on the right-hand-side as a function of b be represented by $\psi(b)$. An intuitive explanation of the above result is as follows: the first term on the right hand side is the asymptote for the probability that the vector A_κ exceeds y , and the second term on the right hand side is the asymptote for the probability that the losses exceed κb conditional on the event $\{A_\kappa = y\}$. The sum of the two terms thus is a logarithmic estimate of the probability of the joint event $\{A_\kappa \geq y\} \cap \{Z_\kappa > \kappa b\}$. Taking the supremum over y 's gives us the most likely value of y for which the event of interest $\{Z_\kappa > \kappa b\}$ occurs.

We now outline the importance sampling algorithm. Let y^*, θ_1^* and θ_2^* be the arguments where the supremum is achieved in (3). For technical reasons, we assume in this section and henceforth that all extremums are attainable and unique. Define the function $\theta_2^*(y) = \arg \min_{\theta_2 \in \mathbb{R}_+} (M^{(2)}(\theta_2; y) - \theta_2 b)$.

The proposed IS algorithm involves two stage exponential twisting.

1. In the first stage, generate A_κ from the exponential twisted distribution $G_\kappa^*(\cdot)$ where

$$\frac{dG_\kappa^*(x)}{dG_\kappa(x)} = \exp(\kappa\theta_1^* \cdot x - M_\kappa^{(1)}(\kappa\theta_1^*)).$$

2. Conditional on A_κ generate Z_κ from the exponential twisted distribution $F_\kappa^*(\cdot; A_\kappa)$ where

$$\frac{dF_\kappa^*(x; A_\kappa)}{dF_\kappa(x; A_\kappa)} = \exp \left[\theta_2^*(A_\kappa)x - M_\kappa^{(2)}(\theta_2^*(A_\kappa); A_\kappa) \right].$$

3. The likelihood ratio L is given by

$$L = \exp \left(-\kappa\theta_1^* \cdot A_\kappa + M_\kappa^{(1)}(\kappa\theta_1^*) - \theta_2^*(A_\kappa)Z_\kappa + M_\kappa^{(2)}(\theta_2^*(A_\kappa); A_\kappa) \right).$$

If $Z_\kappa \geq \kappa b$ count L else count 0. Average over a number of runs to obtain an estimate and confidence interval.

In order to prove the asymptotic optimality of the proposed importance sampling algorithm, we make the following technical assumptions.

- (A4) The set $\mathcal{B} := \text{cl}(\{y \in \mathbb{R}_+^d : \theta_2^*(y) > 0\})$ is a compact set, where $\text{cl}(\mathcal{A})$ represents the closure of set \mathcal{A} .
- (A5) The function $\theta_2^*(y)$ is a continuous function of y . The sequence of functions $\{\kappa^{-1}M_\kappa^{(2)}(\theta_2^*(y); y) : \kappa = 1, 2, \dots\}$ is uniformly equicontinuous for $y \in \mathcal{B}$.
- (A6) We assume that

$$\psi(b) = \inf_{\theta_1 \in \mathbb{R}_+^d} \sup_{y \in \mathbb{R}_+^d} \left[(M^{(1)}(\theta_1) - \theta_1 \cdot y) + \inf_{\theta_2 \in \mathbb{R}_+} (M^{(2)}(\theta_2; y) - \theta_2 b) \right]. \quad (4)$$

To interpret Assumption (A4), consider the conditional probability $\mathbb{P}(Z_\kappa > \kappa b | A_\kappa = y)$ as a function of y . Assumption (A4) implies that the aforementioned conditional probability decays exponentially only on a compact set \mathcal{B} . (On the complement set \mathcal{B}^c , the conditional probability either does not decay with y or has a decay rate slower than exponential, implying that the event is “not rare.”) Assumption (A6) implies that the supremum and the infimum in (4) can be interchanged.

Theorem 2 Under assumptions (A1)-(A6), we have

$$\frac{\log \mathbb{P}(Z_\kappa \geq \kappa b)}{\kappa} \rightarrow \psi(b),$$

$$\frac{\log \mathbb{E}^{\mathbb{Q}} [L^2 \mathbb{I}_{\{Z_\kappa \geq \kappa b\}}]}{\kappa} \rightarrow 2\psi(b),$$

as $\kappa \rightarrow \infty$.

The above theorem, thus, establishes that the proposed change of measure is asymptotically optimal for estimating $\mathbb{P}(Z_\kappa > \kappa b)$.

5 AFFINE JUMP DIFFUSIONS

In this section we apply our analysis to the case where defaults are generated by a doubly stochastic poisson process and the intensities follow affine jump diffusions. We also derive the exact dynamics of the intensities under the importance sampling measure. In this section the “original measure” would refer to the measure under which we wish to compute the expectations (and hence do the simulation) *i.e.* it would be the physical measure when we want to estimate risk measures for a portfolio and it would be the risk-neutral measure when we would be pricing derivatives.

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an information filtration \mathcal{F}_t . Let $X = \{X(t) : t \geq 0\}$ be a time-homogeneous Markov process in some state space $D \subset \mathbb{R}^n$, satisfying the following stochastic differential equation

$$dX(t) = \mu(X(t))dt + \sigma(X(t))dW(t) + dJ(t),$$

where W is an \mathcal{F}_t -standard Brownian motion in \mathbb{R}^n , $\mu : D \rightarrow \mathbb{R}^n$, $\sigma : D \rightarrow \mathbb{R}^{n \times n}$ and J is a pure jump process whose jumps have a fixed probability distribution ν on \mathbb{R}^n and arrive with intensity $\{\gamma(X(t)) : t \geq 0\}$ for some $\gamma : D \rightarrow [0, \infty)$. For $c \in \mathbb{R}^n$, the set of n -tuples, let $\Gamma(c) = \int_{\mathbb{R}^n} \exp(c \cdot z) d\nu(z)$. Let \mathcal{G}_t be the minimal filtration generated by the process X .

X is said to follow an “affine” jump-diffusion if

- $\mu(x) = K_0 + K_1x$ for $K = (K_0, K_1) \in \mathbb{R}^n \times \mathbb{R}^{n \times n}$,
- $(\sigma(x)\sigma(x)^T)_{ij} = (H_0)_{ij} + (H_1)_{ij}x$ for $H = (H_0, H_1) \in \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n \times n}$,
- $\gamma(x) = l_0 + l_1x$ for $l = (l_0, l_1) \in \mathbb{R} \times \mathbb{R}^n$.

Let $f(x) = \rho \cdot x$ for $\rho \in \mathbb{R}^n$. From Duffie, Pan, and Singleton (2000) we know that, under technical regularity conditions

$$\phi(X_t, t, T, \rho) = E \left[\exp \left(\int_t^T f(X(s)) ds \right) \middle| \mathcal{F}_t \right],$$

$$= \exp(\alpha(t, \rho) + \beta(t, \rho) \cdot X_t),$$

where α and β satisfy the following ODEs

$$\begin{aligned} \dot{\beta}(t, \rho) &= -\rho - K_1^T \beta(t, \rho) - \frac{1}{2} \beta(t, \rho)^T H_1 \beta(t, \rho) - l_1(\Gamma(\beta(t, \rho)) - 1), \\ \dot{\alpha}(t, \rho) &= -K_0 \cdot \beta(t, \rho) - \frac{1}{2} \beta(t, \rho)^T H_0 \beta(t, \rho) - l_0(\Gamma(\beta(t, \rho)) - 1), \\ \beta(T, \rho) &= 0, \quad \alpha(T, \rho) = 0. \end{aligned} \tag{6}$$

The system of ODEs (5)-(6) can be efficiently solved numerically using Runge Kutta.

5.1 Homogeneous Model

We first consider a homogeneous model with N obligors where the default intensity of each obligor is identical, i.e., $\lambda_i(s) = w \cdot X(s)$ for all i . Under the doubly stochastic assumption and the homogeneity assumed here, conditional on the path of the intensities or equivalently the ‘‘common factors’’ X , the defaults are independent and the default probability of all the obligors is identical and is given by

$$\begin{aligned} P(\tau_i < T | \mathcal{G}_T) &= 1 - \exp\left(-\int_0^T \lambda(s) ds\right) \\ &= 1 - \exp\left(-w \cdot \int_0^T X(s) ds\right). \end{aligned} \tag{7}$$

Since conditional on \mathcal{G}_T the defaults are independent and the default probabilities are identical, the number of defaults Z conditional on \mathcal{G}_T are distributed binomial with number of trials equal to the number of firms and probability given by (7).

In the notation of the general setup introduced in Section 4, $A = \int_0^T X(s) ds$ and, Z conditioned on A is Binomially distributed with parameters $(N, p(A))$, where $p(A) = 1 - \exp(-w \cdot A)$.

5.1.1 Asymptotic Regime

In this subsection, we will consider a sequence of portfolios indexed by κ . In the κ^{th} portfolio, there are κ obligors. We assume that all obligors have identical default intensity, given by $\lambda^{(\kappa)}(t) = w \cdot X^{(\kappa)}(t)$, where $X^{(\kappa)}(t)$ is an affine stochastic process satisfying

$$dX^{(\kappa)}(t) = \mu^{(\kappa)}(X^{(\kappa)}(t))dt + \sigma^{(\kappa)}(X^{(\kappa)}(t))dW(t) + dJ^{(\kappa)}(t), \tag{8}$$

where

- $\mu^{(\kappa)}(x) = K_0^{(\kappa)} + K_1^{(\kappa)}x$, where $K_0^{(\kappa)} = K_0$ and $K_1^{(\kappa)} = K_1$ for $K = (K_0, K_1) \in \mathbb{R}^n \times \mathbb{R}^{n \times n}$,
- $(\sigma^{(\kappa)}(x)\sigma^{(\kappa)}(x)^T)_{ij} = (H_0^{(\kappa)})_{ij} + (H_1^{(\kappa)})_{ij}x$ where $H_0^{(\kappa)} = H_0/\kappa$ and $H_1^{(\kappa)} = H_1/\kappa$ for $H = (H_0, H_1) \in \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times n \times n}$,
- $J^{(\kappa)}$ is a pure jump process whose jumps have a fixed probability distribution $\nu^{(\kappa)}$ and jump intensity $\gamma^{(\kappa)}(X^{(\kappa)}(t))$, where $\nu^{(\kappa)}$ satisfy $\nu^{(\kappa)}(\mathcal{A}) = \nu(\kappa\mathcal{A})$ which implies $\Gamma^{(\kappa)}(c) = \Gamma(\frac{c}{\kappa})$. Moreover, $\gamma^{(\kappa)}(x) = (l_0^{(\kappa)} + l_1^{(\kappa)}x)$ where $l_0^{(\kappa)} = \kappa l_0$ and $l_1^{(\kappa)} = \kappa l_1$ for $l = (l_0, l_1) \in \mathbb{R} \times \mathbb{R}^n$.

The number of defaults Z_κ is Binomial($\kappa, p(A_\kappa)$), where $A_\kappa = \int_0^T X_\kappa(s) ds$.

Assumptions (A1)-(A6) are easily verified for this setup. Moreover,

$$\begin{aligned} M^{(1)}(\theta_1) &= \log \phi(X(0), 0, T, \theta_1) \\ &= \alpha(0, \theta_1) + \beta(0, \theta_1) \cdot X(0), \\ M^{(2)}(\theta_2; y) &= \log[p(y)(e^{\theta_2} - 1) + 1]. \end{aligned}$$

5.1.2 Importance Sampling Algorithm

In the spirit of Section 4, let y^* , θ_1^* and θ_2^* be the extremums of the following optimization problem

$$\begin{aligned} \sup_{y \in \mathbb{R}_+^d} \left[\inf_{\theta_1 \in \mathbb{R}_+} (\alpha(0, \theta_1) + \beta(0, \theta_1) \cdot X(0) - \theta_1 y) \right. \\ \left. + \inf_{\theta_2 \in \mathbb{R}_+} (\log[p(y)(e^{\theta_2} - 1) + 1] - \theta_2 b) \right]. \end{aligned}$$

The first step of the proposed algorithm is to generate A_κ from the exponential twisted distribution (with the twisting parameter $\kappa\theta_1^*$). Since the distribution $G_\kappa(\cdot)$, and hence the distribution $G_\kappa^*(\cdot)$, is not known analytically it would not be feasible to generate A_κ directly. We thus characterize the dynamics of the intensity under the exponentially twisted measure. As explained in Section 4, y^* is the ‘‘most likely’’ value of A_κ for which the event of interest happens.

Let

$$\begin{aligned} \xi_T &= \frac{\exp\left(\kappa\theta_1^* \int_0^T X^{(\kappa)}(s) ds\right)}{\exp\left(\alpha^{(\kappa)}(0, \kappa\theta_1^*) + \beta^{(\kappa)}(0, \kappa\theta_1^*) \cdot X^{(\kappa)}(0)\right)} \\ \xi_t &= E[\xi_T | \mathcal{G}_t] \\ &= \frac{\exp\left(\kappa\theta_1^* \int_0^t X^{(\kappa)}(s) ds\right)}{\exp\left(\alpha^{(\kappa)}(0, \kappa\theta_1^*) + \beta^{(\kappa)}(0, \kappa\theta_1^*) \cdot X^{(\kappa)}(0)\right)} \times \\ &\quad \exp\left(\alpha^{(\kappa)}(t, \kappa\theta_1^*) + \beta^{(\kappa)}(t, \kappa\theta_1^*) \cdot X^{(\kappa)}(t)\right). \end{aligned}$$

Under some technical conditions, ξ is a positive martingale. We define a probability measure \mathbb{Q} by $\frac{d\mathbb{Q}}{d\mathbb{P}} = \xi_T$. It is easily verified that under measure \mathbb{Q} the random variable A_κ has the desired distribution given by G_κ^* . As shown in Duffie, Pan, and Singleton (2000) the dynamics of the process $X^{(\kappa)}$ after a change of measure with density process ξ still remain affine. For sake of brevity we do not write down the dynamics of $X^{(\kappa)}$ under measure \mathbb{Q} . The dynamics are exactly the same as in (8) with an addition superscript \mathbb{Q} for all parameters. The parameters are such that:

$$\begin{aligned} K_0^{(\kappa)\mathbb{Q}}(t) &= K_0^{(\kappa)}(t) + H_0^{(\kappa)}(t)\beta^{(\kappa)}(t, T, \kappa\theta_1^*) \\ K_1^{(\kappa)\mathbb{Q}}(t) &= K_1^{(\kappa)}(t) + H_1^{(\kappa)}(t)\beta^{(\kappa)}(t, T, \kappa\theta_1^*) \\ l_0^{(\kappa)\mathbb{Q}}(t) &= l_0^{(\kappa)}(t)\Gamma(\beta^{(\kappa)}(t, T, \kappa\theta_1^*), t) \\ l_1^{(\kappa)\mathbb{Q}}(t) &= l_1^{(\kappa)}(t)\Gamma(\beta^{(\kappa)}(t, T, \kappa\theta_1^*), t) \\ \Gamma^{(\kappa)\mathbb{Q}}(c, t) &= \frac{\Gamma^{(\kappa)}(c + \beta^{(\kappa)}(t, T, \kappa\theta_1^*), t)}{\Gamma^{(\kappa)}(\beta^{(\kappa)}(t, T, \kappa\theta_1^*), t)} \\ H^{(\kappa)\mathbb{Q}}(t) &= H^{(\kappa)}(t). \end{aligned}$$

Since we know that $\beta^{(\kappa)}(t, T, \kappa\theta_1^*) = \kappa\beta(t, T, \theta_1^*)$ and using the dynamics of the κ^{th} system we get

$$\begin{aligned} K_0^{(\kappa)\mathbb{Q}}(t) &= K_0(t) + H_0(t)\beta(t, T, \theta_1^*) \\ K_1^{(\kappa)\mathbb{Q}}(t) &= K_1(t) + H_1(t)\beta(t, T, \theta_1^*) \\ l^{(\kappa)\mathbb{Q}}(t) &= \kappa l(t)\Gamma(\beta(t, T, \theta_1^*), t) \\ \Gamma^{(\kappa)\mathbb{Q}}(\kappa c, t) &= \frac{\Gamma(c + \beta(t, T, \theta_1^*), t)}{\Gamma(\beta(t, T, \theta_1^*), t)} \\ H^{(\kappa)\mathbb{Q}}(t) &= \frac{H(t)}{\kappa}. \end{aligned}$$

We above explicitly write down the dependence of β on the twisting parameter θ_1^* .

Thus under the new measure we find that the drift of the process increases, the jump size frequency increases and the mean jump size increases also; all of which contribute to higher default probability under the new measure.

In the second stage of the algorithm, conditional on A_κ , we generate Z_κ from the exponentially twisted (by parameter $\theta_2^*(A_\kappa)$) binomial distribution. Under the new distribution, Z_κ is still binomial with parameters κ and $p^*(A_\kappa)$ where

$$p^*(A_\kappa) = p(A_\kappa) \frac{e^{\theta_2^*(A_\kappa)}}{p(A_\kappa)(e^{\theta_2^*(A_\kappa)} - 1) + 1}.$$

Substituting the optimal choice of $\theta_2^*(A_\kappa)$ in the above relation one obtains

$$p^*(A_\kappa) = \max(p(A_\kappa), b).$$

The likelihood ratio L is given by

$$\begin{aligned} \exp\left(-\kappa\left(\theta_1^* \cdot \int_0^T X(s)ds - \alpha(0, \theta_1^*) - \beta(0, \theta_1^*) \cdot X(0)\right)\right) \\ \times \left(\frac{p(A_\kappa)}{p^*(A_\kappa)}\right)^{Z_\kappa} \left(\frac{1-p(A_\kappa)}{1-p^*(A_\kappa)}\right)^{\kappa-Z_\kappa}. \end{aligned}$$

5.2 Heterogeneous Model

We now return to our original problem described in section 2.

Proposition 1 *Let $\tau_i^{(j)}$ be the default time of the i^{th} obligor belonging to the j^{th} class. Conditional on the path of the common factor the default time distribution is given by*

$$\begin{aligned} P(\tau_i^{(j)} < T | \mathcal{G}_T) &= 1 - \exp\left(-w_j \int_0^T X(s)ds\right) \times \\ &\exp(\alpha_j(T) + \beta_j(T)\epsilon_i^{(j)}(0)). \end{aligned}$$

As in the previous subsection, $A = \int_0^T X(s)ds$. Because of heterogeneity the number of defaults are no longer binomial distributed. In fact $Z = \sum_{i=1}^N Z_i^{(j)}$, where $Z_i^{(j)}$ is an indicator of default by time T for the i^{th} obligor belonging to the j^{th} class. Conditional on A , $Z_i^{(j)}$ is a Bernoulli random variable with probability $p_j(A)$ where

$$p_j(A) = 1 - e^{-w_j \cdot A} \exp(\alpha_j(T) + \beta_j(T)\epsilon_i^{(j)}(0)).$$

5.2.1 Asymptotic Regime

As in the the case of the homogeneous model we will consider a sequence of portfolios indexed by κ . In the κ^{th} portfolio, there are κ obligors and N_c classes. Each obligor belongs to one of the N_c classes. Let the proportion of each obligors of each class be denoted by $\varphi_j : j = 1, \dots, N_c$ such that $\sum_j \varphi_j = 1$. We assume that the i^{th} obligor belonging to class j has default intensity given by $\lambda_i^{(j)(\kappa)}(t) = w_j \cdot X^{(\kappa)}(t) + \epsilon_i^{(j)}(t)$, where $\epsilon_i^{(j)}$ is the idiosyncratic component and $X^{(\kappa)}(t)$ is an affine stochastic process having dynamics as in the homogeneous case.

The number of defaults Z_κ is the sum of κ bernoulli random variables *i.e.*

$$Z_\kappa = \sum_{i=1}^{\kappa} Z_{\kappa,i}^{(j)},$$

where $Z_{\kappa,i}^{(j)}$ is an indicator of default by time T for the i^{th} obligor belonging to class j . Conditional on A_κ , $Z_{\kappa,i}^{(j)}$ is a

Bernoulli random variable with probability $p_j(A_\kappa)$ and

$$p_j(A_\kappa) = 1 - e^{-w_j \cdot A_\kappa} \exp(\alpha_j(T) + \beta_j(T)e^{(j)}(0)).$$

Again, Assumptions (A1) - (A6) are easily verified. And we have

$$\begin{aligned} M^{(1)}(\theta_1) &= \log \phi(X(0), 0, T, \theta_1) \\ &= \alpha(0, \theta_1) + \beta(0, \theta_1) \cdot X(0). \end{aligned}$$

$M^{(2)}(\theta_2; y)$ is given by

$$M^{(2)}(\theta_2; y) = \sum_{j=1}^{N_c} \varphi_j \log[p_j(y)(e^{\theta_2} - 1) + 1]$$

5.2.2 Importance Sampling Algorithm

As in the previous subsection, let y^*, θ_1^* and θ_2^* be the extremum of the following optimization problem

$$\sup_{y \in \mathbb{R}_+^d} \left[\inf_{\theta_1 \in \mathbb{R}_+} (\alpha(0, (0, \theta_1)) + \beta(0, (0, \theta_1)) \cdot X(0) - \theta_1 y) + \inf_{\theta_2 \in \mathbb{R}_+} \left(\sum_{j=1}^{N_c} \varphi_j \log[p_j(y)(e^{\theta_2} - 1) + 1] - \theta_2 b \right) \right].$$

Once θ_1^* is computed the dynamics of the state variables under the twisted distribution can be derived in exactly the same manner as in the homogeneous case. For the second stage, conditional on A_κ , $Z_{\kappa,i}^{(j)}$ is a Bernoulli random variable with probability $p_j^*(A_\kappa)$ where

$$p_j^*(A_\kappa) = p_j(A_\kappa) \frac{e^{\theta_2^*(A_\kappa)}}{p_j(A_\kappa)(e^{\theta_2^*(A_\kappa)} - 1) + 1}.$$

Thus, under the proposed change of measure the single name default probability is atleast as large as level b .

6 NUMERICAL RESULTS

We conducted simulations to estimate the loss probability in a doubly stochastic setting where the common factor follows a CIR process. We present here the case for a 1-factor homogeneous model. Specifically,

$$dX(t) = \kappa(\theta - X(t))dt + \sigma\sqrt{X(t)}dB(t),$$

The parameters we chose are $\kappa = 0.2$, $\theta = 100bp$, $\sigma = 0.06$, $\lambda(0) = \theta$ and $T = 5$ years. The number of firms m is set to 100. The weight on the common factor for each firm is the same and is equal to 1. There is no idiosyncratic component. At these parameter choice the

5 year single obligor default probability is 4.86%. We conduct 10000 simulation runs to estimate the probability for different threshold levels. For each threshold level, besides performing naive simulation and *two stage optimal importance sampling*, we also performed simulation when only the first or the second stage twisting is performed. For each threshold level, we note the variance reduction obtained for the different simulation methodologies viz-a-viz naive simulation.

Table 1: Affine Diffusion

b	$p(y^*)$	$\hat{\gamma}_b$	Both Stage	1 st Stage	2 nd Stage
10%	0.079	6.55E-2±3.17%	5	2	2
15%	0.109	8.43E-3±3.79%	31	4	2
20%	0.139	8.24E-4±4.28%	254	26	3
25%	0.169	6.21E-5±4.73%	2760	135	74
30%	0.198	3.78E-6±5.10%	39039	761	86

In Table 1, the first column is the threshold level b . The second column $p(y^*)$ is the expected default probability conditional on the path of intensities under the new measure. Mathematically,

$$p(y^*) = \mathbb{E}^{\mathbb{Q}} \left[1 - \exp \left(- \int_0^T X(s) ds \right) \right]$$

where \mathbb{Q} is the new measure obtained after twisting the distributions. The third column $\hat{\gamma}_b$ is the estimate of probability at the desired threshold level along with the corresponding 95% confidence interval for the estimator. We report here the 95% confidence interval obtained from the simulation results for two stage optimal importance sampling. The standard error of the estimator for naive simulation is

$$\sqrt{\frac{\hat{\gamma}_b(1 - \hat{\gamma}_b)}{n}}$$

where n is the number of simulation runs, equal to 10000 here.

The fourth column reports the variance reduction obtained for the two-stage importance sampling algorithm versus naive simulation. The fifth and sixth columns report the variance reduction obtained when using importance sampling algorithms for only the 1st and 2nd stages respectively, over the naive simulation.

When conducting 1st stage importance sampling only, we switch off the second stage twisting i.e. $\theta_2^* = 0$. For the first stage we change the measure such that $1 - e^{-y^*} = b$ and θ_1^* solves the following optimization problem

$$\inf_{\theta_1 \in \mathbb{R}_+} (\alpha(0, (0, \theta_1)) + \beta(0, (0, \theta_1)) \cdot X(0) + \theta_1 \log(1 - b)).$$

When conducting 2^{nd} stage importance sampling only, we switch off the first stage change of measure i.e. $\theta_1^* = 0$. We simulate the common factor with parameters as in the original measure. Conditional on the path of the intensities, we perform the 2^{nd} stage twisting exactly as in the case of the two-stage importance sampling.

As the threshold level b increases the probability of interest decreases. From the results in Table 1, we observe that the variance reduction achieved from all the three importance sampling schemes over naive simulation increases many folds as the event becomes rarer. In the case when the probability of interest is around 0.6 basis point (for $b = 0.25$), variance reduction of the order of 2760 is achieved by our proposed algorithm. In contrast 1^{st} and 2^{nd} stage importance sampling achieve only a small fraction of that. To put the above numbers in perspective the variance reduction could be viewed equivalently as reduction in number of runs, to achieve the same level of accuracy, by the same factor.

We also observe that in our parameter choice the 1^{st} stage importance sampling scheme outperforms the 2^{nd} stage importance sampling scheme for all threshold levels. Whereas, our proposed two stage importance sampling scheme outperforms both 1^{st} and 2^{nd} stage importance sampling by significant factors. The relative gain obtained by using two stage importance sampling increases as the event of interest becomes rarer.

A key observation is that under the two stage importance sampling measure $p(y^*)$ varies between $[0.7b, 0.8b]$. The ratio of $p(y^*)$ to b decreases as b increases. Note that in the 1^{st} stage importance sampling algorithm applied to the homogenous case $p(y^*)$ is such that the expected default probability is simply the threshold level b . Whereas for Algorithm 2, $p(y^*)$ is simply the single name default probability under the original measure.

REFERENCES

- Andersen, L., J. Sidenius, and S. Basu. 2003. All your hedges in one basket. *Risk* 67–72.
- Duffie, D., J. Pan, and K Singleton. 2000. Transform analysis and asset pricing for affine jump diffusions. *Econometrica* **58**, 1343–1376.
- Duffie, D., and K. Singleton. 2003. *Credit Risk: Pricing, Measurement, and Management*, Princeton University Press, Princeton, NJ.
- Glasserman, P., and J. Li. 2005. Importance sampling for portfolio credit risk. *Management Science*. To appear.
- Li, D. 2001. On default correlation: a copula function approach. *Journal of Fixed Income* **9**, 43–54.
- Protter, P. 2004. *Stochastic Integration and Differential Equations*, 2nd ed., Springer-Verlag, New York, NY.

AUTHOR BIOGRAPHIES

ACHAL BASSAMBOO is Assistant Professor in managerial economics and decision sciences at Kellogg School of Management. His research interests are stochastic systems, revenue management and rare event simulation. His e-mail address is <a-bassamboo@northwestern.edu>.

SACHIN JAIN is a Risk Strategist with Amaranth LLC, a Greenwich, CT based multi strategy hedge fund. His research interests include credit derivatives, simulation and stochastic systems. His e-mail address is <sjain@amaranthllc.com>.