

Interacting Particle Systems for the Computation of Rare Credit Portfolio Losses

René Carmona · Jean-Pierre Fouque ·
Douglas Vestal

the date of receipt and acceptance should be inserted later

Abstract In this paper, we introduce the use of interacting particle systems in the computation of probabilities of simultaneous defaults in large credit portfolios. The method can be applied to compute small historical as well as risk neutral probabilities. It only requires that the model be based on a background Markov chain for which a simulation algorithm is available. We use the strategy developed by Del Moral and Garnier in [5] for the estimation of random walk rare events probabilities. For the purpose of illustration, we consider a discrete time version of a first passage model for default. We use a structural model with stochastic volatility, and we demonstrate the efficiency of our method in situations where importance sampling is not possible or numerically unstable.

Keywords Interacting Particle Systems · Rare Defaults · Monte Carlo Methods · Credit Derivatives · Variance Reduction

Mathematics Subject Classification (2000) 60H35 · 91B70

Work supported by NSF-FRG-DMS-0455982

R. Carmona
Bendheim Center for Finance, Department ORFE, Princeton University, Princeton, NJ
08544
E-mail: rcarmona@princeton.edu

J.-P. Fouque
Department of Statistics and Applied Probability, University of California, Santa Barbara,
CA 93106-3110
E-mail: fouque@pstat.ucsb.edu

D. Vestal
Julius Finance, 5 Penn Plaza, Level 23 NY, NY 10001
E-mail: doug.vestal@juliusfinance.com

1 Introduction

The past few years have seen an explosion in the credit markets. Subsequently, the field of credit risk and credit derivatives research has substantially increased. As the credit derivative products have grown in complexity, so has the need for fast and accurate numerical methods to reliably value derivatives and quantify their risks.

In this paper, we follow the interacting particle system approach of [5], and we adapt it to the computation of loss probabilities for large credit portfolios. Our interest in this problem is motivated by the uncontrolled growth of the market of Collateralized Debt Obligations (CDOs) and the lack of understanding of the risk profiles of some of these structures. Typically, a CDO is a complex derivative that pools together different loans ($N = 125$ is the size of a typical pool), and sells exposure to different default levels of the portfolio, the so-called *tranches*. The tranching of the default risk makes it possible to create derivatives with high ratings, concentrating the default risk on the riskier lower tranches first hit by the defaults in the portfolio. Moreover, the segmentation of the risk enables the buyer to tailor the purchase to her risk appetite. Both features were important factors in the sudden growth of the CDO market.

The main difficulty in pricing CDOs is the high-dimensional nature of the problem. To accurately price CDO tranches, the joint distribution of the defaults is needed. Moreover, even if this joint distribution were to be found explicitly, there would be no guarantee that the values of the expectations needed to compute tranche spreads could be found analytically. One has to rely on numerical schemes. Due to the high-dimensional nature of the problem, partial differential equations based methods are ruled out and Monte Carlo methods are heavily relied upon.

While Monte Carlo methods are easy to implement, even in high-dimensions, they do suffer from slow convergence. In addition, due to the rare nature of multiple defaults, the computational problem is exacerbated since many Monte Carlo simulations are needed to observe occurrences of the joint default of many names. Therefore, variance reduction and efficiency become very important for these Monte Carlo computations.

The main variance reduction technique used in Monte Carlo computations is importance sampling. There have been many successful applications of importance sampling in credit risk [1, 10]. However, most authors have concentrated on multi-factor Gaussian copula models or reduced-form models, and the main difficulty with implementing importance sampling remains computing the change of measure under which random variables need to be simulated. More information about importance sampling and its inherent shortcomings can be found in [9] and the references therein.

The use of Feynman-Kac path measures, and their subsequent interacting particle systems implementations may have their origin in the development of particle methods in stochastic filtering. Despite the difficulties in turning computational algorithms into rigorous mathematical theorems, the remarkable successes of the method in situations where linear techniques failed mis-

erably granted credibility and popularity to interacting particle systems (IPS for short). In [5], Del Moral and Garnier used a general IPS algorithm for the computation of probabilities of rare events, and illustrated the potential of their algorithm with the detailed analysis of random walk tail events.

In this paper we are concerned with first passage models where it is impractical, if not impossible, to compute explicitly a reasonable importance sampling change of measure explicitly. Examples of such situation abound, e.g. regime switching models with many factors. For the sake of illustration, we choose to work with a first passage model with stochastic volatility. In [6], it is shown that introducing stochastic volatility at different time scales produces realistic yields at short maturities for single-name Credit Default Swaps. A generalization to the multi-name case is presented in [7]. Explicit importance sampling is not an option in this case. Indeed, desirable changes of measure favoring sample paths realizing rare events are highly unlikely to lead to explicit formula. Moreover, standard changes of measure *à la Girsanov* require the volatility to appear in the denominator of the integrands of stochastic integrals that need to be evaluated by numerical approximation methods, creating instabilities when the volatility approaches zero. This happens quite often for the square-root diffusion which we choose as a model for the stochastic volatility.

Already several papers [3, 8] appeared after the first version of this paper was first circulated. They show the strengths of IPS based Monte Carlo computations of small default probabilities, especially when other methods fail. The interested reader is referred to [3] for a systematic comparison with importance sampling.

The rest of the paper is organized as follows. Section 2 gives an overview of Feynman-Kac measures on genealogical path spaces and the associated IPS interpretation. Section 3 discusses the loss process model for large credit portfolios that we chose for the purpose of illustration. It is a discretized version of the standard first passage model of the structural approach. We provide the algorithm and describe in detail its implementation in the case of our stochastic volatility first passage structural model. Finally, Section 4 discusses the numerical results obtained using IPS. We gather theoretical and numerical results, including a precise variance analysis, for the single-name case in the Appendix at the end of the paper.

Acknowledgements. We would like to thank two anonymous referees for insightful and constructive reviews of the first version of this paper. Also, R.C. would like to thank S. Crépey for enlightening discussions on the scales of loss probabilities.

2 Markov Chains, Monte Carlo and IPS

For the sake of completeness, we provide a quick overview of IPS inspired by the presentation in [5], and we sprinkle our exposé with intuitive comments intended to shed light on the characteristics of the IPS approach specific to the computation of small probabilities.

Particle methods depend upon the existence of a *background* Markov chain which we denote $X = \{X_n\}_{n \geq 0}$. This chain is not assumed to be time homogeneous. In fact, the random element X_n takes values in some measurable state space (E_n, \mathcal{E}_n) that can change with n . We denote by $K_n(x_{n-1}, dx_n)$ the Markov transition kernels. Throughout the paper, for any measurable space (E, \mathcal{E}) we denote by $B_b(E)$ the space of bounded, measurable functions.

2.1 Feynman-Kac Path Expectations

We denote by Y_n the history of X_n as defined by

$$Y_n \stackrel{\text{def.}}{=} (X_0, \dots, X_n) \in F_n \stackrel{\text{def.}}{=} (E_0 \times \dots \times E_n), \quad n \geq 0.$$

$\{Y_n\}_{n \geq 0}$ is itself a Markov chain and we denote by $M_n(y_{n-1}, dy_n)$ its transition kernel. For each $n \geq 0$, we choose a multiplicative potential function G_n defined on F_n and we define the Feynman-Kac expectations by

$$\gamma_n(f_n) = \mathbb{E} \left[f(Y_n) \prod_{1 \leq p < n} G_p(Y_p) \right]. \quad (2.1)$$

We denote by $\eta_n(\cdot)$ the corresponding normalized measure defined as

$$\eta_n(f_n) = \frac{\mathbb{E} \left[f_n(Y_n) \prod_{1 \leq k < n} G_k(Y_k) \right]}{\mathbb{E} \left[\prod_{1 \leq k < n} G_k(Y_k) \right]} = \gamma_n(f_n) / \gamma_n(1). \quad (2.2)$$

A very important observation is that

$$\gamma_{n+1}(1) = \gamma_n(G_n) = \eta_n(G_n) \gamma_n(1) = \prod_{p=1}^n \eta_p(G_p).$$

Therefore, given any bounded measurable function f_n on F_n , we have

$$\gamma_n(f_n) = \eta_n(f_n) \prod_{1 \leq p < n} \eta_p(G_p).$$

Using the notation $G_p^- = 1/G_p$ for the reciprocal of the multiplicative potential function and the above definitions of γ_n and η_n we see that

$$\begin{aligned} \mathbb{E} [f_n(Y_n)] &= \mathbb{E} \left[f_n(Y_n) \prod_{1 \leq p < n} G_p^-(Y_p) \times \prod_{1 \leq p < n} G_p(Y_p) \right] \\ &= \gamma_n \left(f_n \prod_{1 \leq p < n} G_p^- \right) \\ &= \eta_n \left(f_n \prod_{1 \leq p < n} G_p^- \right) \prod_{1 \leq p < n} \eta_p(G_p). \end{aligned} \quad (2.3)$$

Finally, one can check by inspection that the measures $(\eta_n)_{n \geq 1}$ satisfy the nonlinear recursive equation

$$\eta_n = \Phi_n(\eta_{n-1}) \stackrel{\text{def.}}{=} \int_{F_{n-1}} \eta_{n-1}(dy_{n-1}) \frac{G_{n-1}(y_{n-1})}{\eta_{n-1}(G_{n-1})} M_n(y_{n-1}, \cdot),$$

starting from $\eta_1 = M_1(x_0, \cdot)$. This dynamical equation on the space of measures is known as Stettner's equation in filtering theory. We state it to justify the selection/mutation decomposition of each step of the particle algorithm introduced below.

2.2 IPS Interpretation and Monte Carlo Algorithm

Motivated by the above definitions and results, we introduce a very natural interacting path-particle system. For a given integer M , using the transformations Φ_n , we construct a Markov chain $\{\xi_n\}_{n \geq 0}$ whose state $\xi_n = (\xi_n^j)_{1 \leq j \leq M}$ at time n , can be interpreted as a set of M Monte Carlo samples of path-particles

$$\xi_n^j = (\xi_{0,n}^j, \xi_{1,n}^j, \dots, \xi_{n,n}^j) \in F_n = (E_0 \times \dots \times E_n).$$

The transition mechanism of this Markov chain can be described as follows. We start with an initial configuration $\xi_1 = (\xi_1^j)_{1 \leq j \leq M}$ that consists of M independent and identically distributed random variables with distribution,

$$\eta_1(d(y_0, y_1)) = M_1(x_0, d(y_0, y_1)) = \delta_{x_0}(dy_0)K_1(y_0, dy_1),$$

i.e., $\xi_1^j \stackrel{\text{def.}}{=} (\xi_{0,1}^j, \xi_{1,1}^j) = (x_0, \xi_{1,1}^j) \in F_1 = (E_0 \times E_1)$ where the $\xi_{1,1}^j$ are drawn independently of each other from the distribution $K_1(x_0, \cdot)$. Then, the one-step transition taking $\xi_{n-1} \in F_{n-1}^M$ into $\xi_n \in F_n^M$ is given by a random draw from the distribution

$$\mathbb{P}\{\xi_n \in d(y_n^1, \dots, y_n^M) | \xi_{n-1}\} = \prod_{j=1}^M \Phi_n(m(\xi_{n-1}))(dy_n^j), \quad (2.4)$$

where the notation $m(\xi_{n-1})$ is used for the empirical distribution of the ξ_{n-1}^j , i.e.

$$m(\xi_{n-1}) \stackrel{\text{def.}}{=} \frac{1}{M} \sum_{j=1}^M \delta_{\xi_{n-1}^j}.$$

From the definition of Φ_n , one can see that (2.4) is the superposition of a selection procedure followed by a mutation given by the transition of the original Markov chain. More precisely:

$$\xi_{n-1} \in F_{n-1}^M \xrightarrow{\text{selection}} \hat{\xi}_{n-1} \in F_{n-1}^M \xrightarrow{\text{mutation}} \xi_n \in F_n^M.$$

where the selection stage is performed by choosing randomly and independently M path-particles

$$\hat{\xi}_{n-1}^j = (\hat{\xi}_{0,n-1}^j, \hat{\xi}_{1,n-1}^j, \dots, \hat{\xi}_{n-1,n-1}^j) \in F_{n-1},$$

according to the Boltzmann-Gibbs measure

$$\sum_{j=1}^M \frac{G_{n-1}(\xi_{0,n-1}^j, \dots, \xi_{n-1,n-1}^j)}{\sum_{k=1}^M G_{n-1}(\xi_{0,n-1}^k, \dots, \xi_{n-1,n-1}^k)} \delta_{(\xi_{0,n-1}^j, \dots, \xi_{n-1,n-1}^j)}, \quad (2.5)$$

and for the mutation stage, each selected path-particle $\hat{\xi}_{n-1}^j$ is extended as follows

$$\begin{aligned} \xi_n^j &= (\hat{\xi}_{n-1}^j, \xi_{n,n}^j) \\ &= ((\hat{\xi}_{0,n-1}^j, \dots, \hat{\xi}_{n-1,n-1}^j), \xi_{n,n}^j) \in F_n = F_{n-1} \times E_n, \end{aligned}$$

where $\xi_{n,n}^j$ is a random variable with distribution $K_n(\hat{\xi}_{n-1,n-1}^j, \cdot)$. In other words, the transition step is a mere extension of the path particle with an element drawn at random using the transition kernel K_n of the original Markov chain. All of the mutations are performed independently. But most importantly, all these mutations are happening with the original transition distribution of the chain. This is in sharp contrast with importance sampling where the Monte Carlo transitions are from twisted transition distributions obtained from a Girsanov-like change of measure. So from a practical point of view, a black-box providing random samples from the original chain transition distribution is enough for the implementation of the IPS algorithm: no need to know the details of such a generation.

A result of [4] reproduced in [5] states that for each fixed time n , the empirical historical path measure

$$\eta_n^M \stackrel{\text{def.}}{=} m(\xi_n) = \frac{1}{M} \sum_{j=1}^M \delta_{(\xi_{0,n}^j, \xi_{1,n}^j, \dots, \xi_{n,n}^j)}$$

converges in distribution, as $M \rightarrow \infty$, toward the normalized Feynman-Kac measure η_n . Moreover, there are several propagation of chaos estimates that ensure that $(\xi_{0,n}^j, \xi_{1,n}^j, \dots, \xi_{n,n}^j)$ are asymptotically independent and identically distributed with distribution η_n [4]. This justifies for each measurable function \tilde{f}_n on F_n , the choice of

$$\gamma_n^M(\tilde{f}_n) = \eta_n^M(\tilde{f}_n) \prod_{1 \leq p < n} \eta_p^M(G_p). \quad (2.6)$$

for a particle approximation of the expectation $\gamma_n(\tilde{f}_n)$. The main properties of the particle approximation γ_n^M are stated in the following lemma whose statement is borrowed from [5].

Lemma 1 ([5]). *Under the assumption*

$$\sup_{(y_n, \bar{y}_n) \in F_n^2} G_n(y_n)/G_n(\bar{y}_n) < \infty,$$

γ_n^M is an unbiased estimator for γ_n , in the sense that for any $p \geq 1$ and $f_n \in B_b(F_n)$ with $\|f_n\| \leq 1$, we have

$$\mathbb{E}\{\gamma_n^M(\tilde{f}_n)\} = \gamma_n(\tilde{f}_n),$$

and in addition

$$\sup_{M \geq 1} \sqrt{M} \mathbb{E}[|\gamma_n^M(\tilde{f}_n) - \gamma_n(\tilde{f}_n)|^p]^{1/p} \leq c_p(n),$$

for some constant $c_p(n) < \infty$ whose value does not depend on the function \tilde{f}_n .

We refer to [5] for a complete proof of this result. From formula (2.3), it is clear that starting from a function f_n on F_n , we will apply the above result to

$$\tilde{f}_n = f_n \prod_{p < n} G_p^-. \quad (2.7)$$

3 Loss Distributions of Credit Portfolios

In this section, we explain how we use the interacting particle system approach described in the previous section to the computation of the probabilities of rare credit losses in a large portfolio of credit sensitive instruments modeled in the structural approach. Since the language of continuous-time finance is commonly used in the industry, we choose to first introduce our model in continuous time. We will concentrate on the discrete time version implemented for the purpose of computations in the next subsection.

3.1 Credit Portfolio Model

We consider a portfolio of N firms. N will be typically 125 in the numerical applications presented later in the paper, and 1 for the single-name case presented in the Appendix. The dynamics of the asset values of these N firms are given by the following system of stochastic differential equations:

$$dS_i(t) = rS_i(t)dt + \sigma_i \sigma(t) S_i(t) dW_i(t), \quad i = 1, \dots, N \quad (3.1)$$

where r is the risk-free interest rate, σ_i is an idiosyncratic (non-random) volatility factor, the correlation structure of the driving Wiener processes W_i is given by

$$d\langle W_i, W_{i'} \rangle_t = \rho_{ii'} dt,$$

and the common stochastic volatility factor $\sigma(t)$ is a square-root diffusion satisfying the stochastic differential equation:

$$d\sigma(t) = \kappa(\bar{\sigma} - \sigma(t))dt + \gamma\sqrt{\sigma(t)} dW_t, \quad (3.2)$$

where κ , $\bar{\sigma}$ and γ are positive constants and the Wiener process W_t satisfies

$$d\langle W_i, W \rangle_t = \rho_\sigma dt, \quad i = 1, 2, \dots, N.$$

We impose the condition $\gamma^2 < 2\kappa\bar{\sigma}$ so that $\sigma(t)$ remains positive at all times. Note also that, in contrast to the classical Heston model, the volatility $\sigma(t)$ is a square-root process and not the square volatility. This is not an issue since we are not using explicit formulas for the Heston model which are in fact not available in the correlated multi-dimensional case. Also, for each firm i , we assume the existence of a deterministic boundary $t \mapsto B_i(t)$ and the time of default for firm i is assumed to be given by

$$\tau_i = \inf \{t : S_i(t) \leq B_i(t)\}. \quad (3.3)$$

For the sake of simplicity, we define the portfolio loss function $L(t)$ as the number of defaults prior to time t , i.e.

$$L(t) = \sum_{i=1}^N \mathbf{1}_{\{\tau_i \leq t\}}, \quad t > 0. \quad (3.4)$$

Since the spreads of CDO tranches are derived from the knowledge of a finite number of expectations of the form:

$$\mathbb{E}\{(L(T) - K)^+\},$$

where T is a coupon payment date and K is proportional to the attachment or detachment points of the tranche, we will restrict ourselves to the evaluation of these expectations.

Clearly, the only interesting case is when all of the names in the portfolio are dependent. In [12], the exact distribution of losses is derived for $N = 2$ from the distributions of the hitting times of a pair of correlated Brownian motions. Unfortunately, a tractable general result is not available for $N > 2$, let alone in the case of stochastic volatility! Since the distribution of $L(T)$ is not known in the dependent case, for $N > 2$, one usually relies on approximation methods. Moreover, since N is typically very large (125 in a standard CDO contract), PDE based methods are ruled out.

Instead of computing the spreads of the tranches directly, we compute the probability mass function for $L(T)$, that is we calculate

$$\mathbb{P}(L(T) = k) = p_k(T), \quad k = 0, \dots, N. \quad (3.5)$$

3.2 The Background Markov Chain

We discretize the time variable t of the above continuous time model using a time step Δt , that will be chosen as $\Delta t = (1/20)\text{yr}$ in the numerical experiments reported later in the paper. Notice that we will also need a smaller time step δt . The latter will be chosen to be $\delta t = 10^{-3}\text{yr}$ in our numerical experiments. The Markov chain $\{X_n\}_n$ on which we construct the IPSs used to compute small probabilities is given by:

$$X_n = \left(\sigma(n\Delta t), (S_i(n\Delta t))_{1 \leq i \leq N}, \left(\min_{0 \leq m \leq n} S_i(m\Delta t) \right)_{1 \leq i \leq N} \right), \quad n \geq 0. \quad (3.6)$$

The state space of X_n is $E_n = [0, \infty)^{2N+1}$ so this chain is $(2N+1)$ -dimensional. We assume a constant (i.e. time independent) barrier B_i for each firm $1 \leq i \leq N$, and we define the time τ_i of default of firm i as

$$\tau_i = \min\{n \geq 0; S_i(n\Delta t) \leq B_i\}$$

in analogy with its continuous time version (3.3). In this way, the value of τ_i can be read off the sequence of running minima. Notice also, with this definition of the default times, we do not have to correct for the bias introduced by the discretization of a continuous time boundary crossing problem.

The very form of the resampling distribution (2.5) shows that in order to have more simulation paths realizing rare events corresponding to unusually high numbers of defaults, an obvious strategy is to choose a set of potential functions becoming larger as the likelihood of default increases. Indeed, the resampling step will select paths with high Gibbs resampling weights, and the paths with small weights will have a greater chance of not being selected, and hence disappear. For the purpose of our numerical experiments we choose a parameter $\alpha > 0$, and we define the multiplicative potential functions G_p by:

$$G_p^\alpha(Y_p) = \exp[-\alpha(V(X_p) - V(X_{p-1}))], \quad (3.7)$$

where

$$V(X_p) = \sum_{i=1}^N \log(\min_{0 \leq m \leq p} S_i(m\Delta t)).$$

We shall drop the superscript α when the dependence upon this free parameter is irrelevant. Notice that

$$\begin{aligned} G_p^\alpha(Y_p) &= \exp[-\alpha(V(X_p) - V(X_{p-1}))] \\ &= \exp\left[-\alpha \sum_{i=1}^N \log \frac{\min_{0 \leq m \leq p} S_i(m\Delta t)}{\min_{0 \leq m \leq p-1} S_i(m\Delta t)}\right], \end{aligned}$$

where the above logarithm is obviously less than or equal to zero. Clearly, different choices of α give different distributions for the resampling weights, and as a result, we expect that different choices of α will give different sets of loss levels k for which the probability $\mathbb{P}(L(t) = k)$ can be computed by IPS as a positive number. For a given value of k , contrary to a plain Monte Carlo computation, the IPS algorithm produces enough sample paths with k losses for the estimation procedure to be acceptable if we choose α appropriately. In the numerical computations reported below, we use an idea which could be traced back to [5], at least in an implicit form, and which was used systematically in [3]. Instead of choosing α and getting reasonable estimates of $\mathbb{P}(L(t) = k)$ for some values of k depending upon α , we reverse the procedure, and for each k , we pick the best α . Note that in the single-name case presented in the appendix, since we can compare the variances of the IPS and MC estimators over the range of k 's, we can afford to use the standard approach fixing α first.

Finally, it is worth noticing that because of the special form (3.7) of the resampling weights,

1. we only need to keep track of (X_{p-1}, X_p) instead of the full history $Y_p = (X_0, X_1, \dots, X_p)$ thereby minimizing the storage space needed for the implementation of the algorithm;
2. $\prod_{1 \leq k < p} G_k^\alpha(Y_k) = \exp[-\alpha(V(X_{p-1}) - V(X_0))]$, thereby providing a significant simplification of the computations.

3.3 Detailed IPS Algorithm

We divide the time interval $[0, T]$ into n equal intervals $[(p-1)T/n, pT/n]$, $p = 1, 2, \dots, n$. These are the times we stop and perform the selection step. We introduce the chain $(X_p)_{0 \leq p \leq n} = (\tilde{X}_{pT/n})_{0 \leq p \leq n}$ and the whole history of the chain is denoted by $Y_p = (X_0, \dots, X_p)$.

Since it is not possible to sample directly from the distribution of $(X_p)_{0 \leq p \leq n}$ for $N > 2$, we will have to apply an Euler scheme during the mutation stage; we let δt denote the sufficiently small time step used. In general δt will be chosen so that $\delta t \ll \Delta t = T/n$.

Our algorithm is built with the weight function defined in equation (3.7). As mentioned earlier, because of the special form of the resampling weights, instead of working with the entire histories Y_p , we need only to keep track of X_p and its *parent* X_{p-1} . We introduce a special notation, say \hat{W}_p , for the “parent” of X_p . So for all practical purposes, instead of being an entire historical path, $y_p = (x_0, x_1, \dots, x_p)$, for implementation purposes, a particle will only be a couple $y_p = (w_p, x_p)$.

Initialization. We start with M identical copies, $\hat{X}_0^{(j)}$, $1 \leq j \leq M$, of the initial condition X_0 . That is,

$$\hat{X}_0^{(j)} = (\sigma(0), (S_1(0), \dots, S_N(0)), (S_1(0), \dots, S_N(0))), \quad 1 \leq j \leq M.$$

and we define their parents by $\hat{W}_0^{(j)} = \hat{X}_0^{(j)}$. In this way we have our initial set of M particles $(\hat{W}_0^{(j)}, \hat{X}_0^{(j)})$, $1 \leq j \leq M$.

Now suppose that at time p , we have a set of M particles $(\hat{W}_p^{(j)}, \hat{X}_p^{(j)})$, $1 \leq j \leq M$.

Selection Stage. We compute the normalizing constant:

$$\hat{\eta}_p^M = \frac{1}{M} \sum_{j=1}^M \exp \left[-\alpha \left(V(\hat{X}_p^{(j)}) - V(\hat{W}_p^{(j)}) \right) \right]. \quad (3.8)$$

Then, we choose independently M particles according to the empirical distribution

$$\eta_p^M(d\check{W}, d\check{X}) = \frac{1}{M \hat{\eta}_p^M} \sum_{j=1}^M \exp \left[-\alpha \left(V(\hat{X}_p^{(j)}) - V(\hat{W}_p^{(j)}) \right) \right] \times \delta_{(\hat{W}_p^{(j)}, \hat{X}_p^{(j)})}(d\check{W}, d\check{X}). \quad (3.9)$$

The particles that are selected are denoted $(\check{W}_p^{(j)}, \check{X}_p^{(j)})$.

Mutation Stage. The stochastic volatility $\sigma(t)$ and the correlation between Brownian motions prevent us from knowing the transition probability of X_n in closed form and perform the mutation in one step. We need Monte Carlo simulations based on an approximation scheme. We choose a plain Euler scheme to make our life easier. We fix a time step $\delta t \ll \Delta t$ (as already mentioned, we will choose $\delta t = 10^{-3}$ in the numerical experiments reported below). For each of the selected particles, $(\check{W}_p^{(j)}, \check{X}_p^{(j)})$, we apply an Euler scheme from time t_p to time t_{p+1} with step size δt to each $\check{X}_p^{(j)}$ so that $\check{X}_p^{(j)}$ becomes $\hat{X}_{p+1}^{(j)}$. We then set $\hat{W}_{p+1}^{(j)} = \check{X}_p^{(j)}$. It should be noted that each of the particles are evolved independently, and that the true dynamics of X_p , given by the discretization of (3.1,3.2), is applied rather than some other measure. It is this fact that separates IPS from importance sampling.

Conclusion. At maturity, i.e. at time n such that $n\Delta t = T$, we tally the total number of losses for each of the M particles by computing the function f_n defined by

$$f(\hat{X}_n^{(j)}) = \sum_{i=1}^N \mathbf{1}_{\{X_n^{(j)}(N+1+i) \leq B_i\}},$$

where we use the last N component of X_n defined in (3.6). The estimator $\hat{p}_k^M(T)$ of $\mathbb{P}(L(T) = k) = p_k(T)$ is then given by

$$\hat{p}_k^M(T) = \left[\frac{1}{M} \sum_{j=1}^M \mathbf{1}_{\{f(\hat{X}_n^{(j)})=k\}} \exp[\alpha(V(\hat{W}_n^{(j)}) - V(\hat{X}_0))] \right] \times \left[\prod_{p=0}^{n-1} \hat{\eta}_p^M \right]. \quad (3.10)$$

As we explained earlier, this estimator is unbiased in the sense that $\mathbb{E}[\hat{p}_k^M(T)] = p_k(T)$.

4 Numerical Results

In this section we report on numerical experiments with an implementation of the IPS procedure described in Section 2 with the stochastic volatility model described in Section 3.

4.1 Parameters of the Numerical Experiments

All the computations reported in this section were done for a homogeneous portfolio of $N = 125$ names. The following table gives the parameters used for the stochastic volatility dynamics (3.2) of the process $\sigma(t)$:

σ_0	κ	$\bar{\sigma}$	γ	ρ_σ
0.4	3.5	0.4	0.7	-0.06

For the sake of simplicity (in part justified by the homogeneity of the portfolio), we assume that the starting points of the N firm values are the same and that

the thresholds giving the default barriers are also the same. The parameters of the dynamics (3.1) of the asset values of the N firms used in our numerical experiments are given by:

$S_i(0)$	r	ρ	B_i	N_{sel}	δt
90	0.06	0.1	36	20	10^{-3}

We also included in the above table the number N_{sel} of selections per year used in the IPS algorithm (essentially the reciprocal of the time step Δt used in the text earlier), as well as the time step δt for the Euler steps needed for the simulation of one mutation step.

The left pane of Figure 4.1 gives a line plot of the probability distributions of the number of loss $L(T)$ for maturities $T = 1, 2, 3, 4, 5$. Since the probability of small losses is overwhelmingly high for $T = 1$, it is difficult to see the details of these distributions because the scale on the vertical scale is determined by the shortest maturity, we reproduced in the right pane of Figure 4.1, the loss distributions with maturities $T = 2, 3, 4, 5$ in their own scale. These plots were obtained from the parameters given in the tables above with $M = 10^5$ Monte Carlo samples with 10^3 Euler time steps per year.

Figure 4.2 gives the line graphs of the probability *densities* for maturities $T = 1$ and $T = 5$. It shows clearly that for $T = 5$ years, there are enough losses for the Monte Carlo simulation procedure to estimate $\mathbb{P}\{L(T) = k\}$ for values of k going up to $k = 120$, and that the increase in variance of the estimator is not dramatic when one switch from $M = 10^5$ samples to 5×10^4 samples. However, Monte Carlo estimates are more troublesome for $T = 1$ year. Indeed, the estimates are pretty wildly oscillating for values of k above $k = 70$, totally unreliable for values of k above $k = 80$, and one sees a significant increase in variance when one halves the number of samples from 10^5 .

4.2 Performance of the IPS Algorithm

In this subsection, we compare the results of our implementation of the IPS algorithm to the results obtained with plain Monte Carlo simulations, as used above to determine a benchmark estimate of the loss distributions. As already noticed, the major shortcoming of Monte Carlo computations is the large variance of the estimator, and the fact that, to estimate the probability of the rare events in the tails of the distributions, one needs a prohibitively large number of simulations, e.g. we used $M = 10^5$ samples in the computations reported above, but despite this large number, all the estimates of the probabilities $\mathbb{P}\{L(T) = k\}$ for $k \geq 80$ are typically 0, especially for $T = 1$ to which we restrict ourselves in this section.

Figure 4.3 illustrates the differences between estimates of the loss distribution obtained by a standard Monte Carlo method and by the IPS algorithm described in the previous sections. The distributions are plotted on a loga-

rithmic scale in order to emphasize the differences which would not appear otherwise. Several comments are in order.

Note that there is a dramatic decrease in the number of simulation samples, from $M = 10^5$ to $M = 200$ in the present situation. However, in all honesty, the computational budget of each simulation is higher for the IPS because of the $N_{selec} = 20$ selection steps per year. Still, the saving is enormous because of the drastic variance reduction of the IPS algorithm.

It is easy to see that for each value of k (or for each small range of contiguous values of k) it is possible to choose a form of the multiplicative potential V and a value of α (and hence a set of weight functions G_i 's) so that the IPS algorithm estimates the value(s) of $\mathbb{P}\{L(T) = k\}$ without bias, and with a significantly smaller number of simulation samples. This fact applies as well to the computation of one expectation $\mathbb{E}\{f_n(Y_n)\}$ if f_n is highly concentrated on a *small set of paths*. This fact was advocated by Del Moral and Garnier in their original discussion of the rare events of the Gaussian random walk, and despite the fact that our model is significantly more involved and with higher dimension, it is confirmed in our situation.

However, the conclusion we draw from our numerical experiments, and perfectly captured by Figure 4.3, is that this positive result cannot be extended to a large set of probabilities or expectations without paying a price. Indeed, if one tries to restrict the IPS to a single weight potential function V , and do the fine tuning by varying *one single* parameter, the number α in our case, it is very difficult to estimate the entire distribution at once without a bias. Even though our claim is only based on numerical circumstantial evidence, we believe that it is true and we hope to address it with elements of solutions in a further investigation.

As explained earlier (see also the bullet points in the next subsection) we estimate the entire distribution at once (as opposed to one probability at a time) by running the IPS simultaneously for several values of the parameter α . Figure 4.4 gives a surface plot of the actual numbers of simulated sample paths of the IPS against the number of losses they produce at maturity $T = 1$, and the value of α used in the weights of the selection steps of the IPS. This surface plot clearly shows the expected pattern: for each value of α , only a small number of values of k will be reached by the paths of the IPS. Moreover, this range of values of k moves upward as α increases. As mentioned earlier, the idea using this *loss map* to relate k and α is borrowed from [3].

4.3 Practical Remarks

The following remarks emphasize the robustness and stability of the results from the IPS algorithm.

- We varied the number N_{sel} of selections/mutations (between and including $N_{sel} = 15$ and $N_{sel} = 35$) and for a given set of weight functions, we did not see significant difference.

- Concerning the role played by the choice of the weights functions G_i , or in our case the weight potential function V , we tried several variations on the same theme, and whether we replace the logarithm of the running minimum by similar functions such as the double logarithm or the square root of the logarithm, or even if we replace the running minimum by the actual level of the value of the firm, and any of these functions, the numerical results are qualitatively the same!
- Clearly, for a given weight potential function V , varying α changes the range of k reached by the historical paths re-sampled during the selection steps. So for a given function f_n , one can demonstrate numerically that there is a set of values of the parameter α reducing the variance of the estimator of $\mathbb{E}\{f_n(Y_n)\}$. However, since we chose to compute the entire loss distribution at a given maturity, we used the method of [3] according to which one runs IPS for a wide range of α and for each $k \in \{0, 1, \dots, N\}$, estimates the loss probability $\mathbb{P}\{L(T) = k\}$ using the IPS estimator given by the best value of α , say $\alpha(k)$, for a simple criterion.

5 Conclusion

In this paper, we demonstrated how one can use the IPS approach of Del Moral and Garnier [5] for the computation of rare events probabilities to the field of credit risk. We chose a structural first passage model with stochastic volatility for the purpose of illustration. We showed that, even for realistic portfolios of $N = 125$ names, for an appropriate choice of weight function, the IPS algorithm can produce loss probabilities which would require a significantly larger number of Monte Carlo simulations (and hence higher computing time and budget) than traditional Monte Carlo methods. This is in sharp contrast with standard importance sampling procedures plagued with instability issues in the stochastic volatility models considered in this paper. Explicit formulas for the asymptotic variance of the IPS estimator are derived for the single-name case in the Appendix where more single names results are gathered. Some of these single name results, together with informal intuitive arguments can be used in the multi-name case to guess reasonable values of the free parameters of the weights driving the selection stage of the IPS algorithm.

Appendix

A The Single-Name Constant Volatility Case

A.1 Default Probabilities

In this subsection, we analyze the variance of the estimator (3.10) in the single name case. So from now on, we restrict ourselves to $N = 1$ and to constant deterministic volatility $\sigma; \sigma(t) \equiv \sigma$. Also, as before we work with a time independent barrier, say B . Using IPS, we

compute the probability of default before maturity T :

$$P^B(0, T) = \mathbb{P}(\min_{u \leq T} S(u) \leq B) = \mathbb{E}[\mathbf{1}_{\min_{u \leq T} S(u) \leq B}].$$

We use the classical explicit formula for $P^B(0, T)$ as a benchmark:

$$P^B(0, T) = 1 - \left(\mathcal{N}(d_2^+) - \left(\frac{S_0}{B} \right)^p \mathcal{N}(d_2^-) \right), \quad (\text{A.1})$$

with \mathcal{N} denoting the standard normal $\mathcal{N}(0, 1)$ cumulative distribution function and

$$d_2^\pm = \frac{\pm \ln\left(\frac{S_0}{B}\right) + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}},$$

$$p = 1 - \frac{2r}{\sigma^2}.$$

We are only interested in values of B that make the above event rare.

We remark that it is a standard result that the variance associated with the traditional Monte Carlo method for computing $P^B(0, T)$ is $P^B(0, T)(1 - P^B(0, T))$. We also remark that for a single name the Markov chain $(X_p)_{0 \leq p \leq n}$ defined in Section 3.2 simplifies to

$$X_p = (S(t_p), \min_{u \leq t_p} S(u)).$$

Then, following the IPS setup described in Section 2, the rare event probability $P^B(0, T)$ admits the Feynman-Kac representation:

$$P^B(0, T) = \gamma_n(L_n^{(B)}(1)),$$

where $L_n^{(B)}(1)$ is given by the weighted indicator function defined for any path $y_n = (x_0, \dots, x_n) \in F_n$ by

$$\begin{aligned} L_n^{(B)}(1)(y_n) &= L_n^{(B)}(1)(x_0, \dots, x_n) = \mathbf{1}_{\{\min_{u \leq T} S(u) \leq B\}} \prod_{1 \leq p < n} G_p^-(x_0, \dots, x_p) \\ &= \mathbf{1}_{\{\min_{u \leq T} S(u) \leq B\}} e^{\alpha(V(x_{n-1}) - V(x_0))} \\ &= \mathbf{1}_{\{\min_{u \leq T} S(u) \leq B\}} e^{\alpha(\log(\min_{u \leq t_{n-1}} S(u)/S_0))}, \end{aligned}$$

for our choice of multiplicative potential function (3.7). Also, notice that we have $\|L_n^{(B)}(1)(y_n)\| \leq 1$ since $\log(\min_{u \leq t_{n-1}} S(u)/S_0) \leq 0$ and $\alpha > 0$ by assumption. Next, following the IPS selection-mutation algorithm outlined in Section 3.3, we form the estimator

$$P_M^B(0, T) = \gamma_n^M(L_n^{(B)}(1)) = \eta_n^M(L_n^{(B)}(1)) \prod_{1 \leq p < n} \eta_p^M(G_p). \quad (\text{A.2})$$

By Lemma 1, $P_M^B(0, T)$ is an unbiased consistent estimator of $P^B(0, T)$. While many estimators are unbiased, the key to determining the efficiency of our estimator is to study its variance and prove a central limit theorem.

A.2 Variance Analysis

In the present situation we can prove:

Theorem 1. *The estimator $P_B^M(0, T)$ given in equation (A.2) is unbiased, and it satisfies the central limit theorem*

$$\sqrt{M}\mathbb{E}[P_B^M(0, T) - P^B(0, T)] \xrightarrow{M \rightarrow \infty} N(0, \sigma_n^B(\alpha)^2),$$

with the asymptotic variance

$$\begin{aligned} & \sigma_n^B(\alpha)^2 \\ &= \sum_{p=1}^n \left[\mathbb{E} \left\{ e^{-\alpha \log(\min_{u \leq t_{p-1}} S(u))} \right\} \times \mathbb{E} \left\{ P_{B,p,n}^2 e^{\alpha \log(\min_{u \leq t_{p-1}} S(u))} \right\} - P^B(0, T)^2 \right], \end{aligned} \tag{A.3}$$

where $P_{B,p,n}$ is the collection of functions defined by

$$P_{B,p,n}(x) = \mathbb{E} \left\{ \mathbf{1}_{\min_{t \leq T} S(t) \leq B} | X_p = x \right\},$$

and $P^B(0, T)$ is given by (A.1).

The proof follows directly by applying Theorem 2.3 in [5] with the weight function that we have defined in (3.7).

In the constant volatility single-name case, the asymptotic variance $\sigma_n^B(\alpha)^2$ can be obtained explicitly in terms of double and triple integrals with respect to explicit densities. This will be used in our comparison of variances for IPS and pure Monte Carlo in the following section. The details of these explicit formulas are given in the Ph.D. Dissertation [11].

As shown numerically in the next section the variance for IPS is of order p^2 with $p = P^B(0, T)$ (small in the regime of interest), in contrast to being of order p for the direct Monte Carlo simulation. This is indeed a very significant variance reduction in the regime p small, as already observed in [5], in a different context.

A.3 Numerical Results

In this subsection, we compute the probability of default for different values of the barrier comparing IPS to the standard Monte Carlo method. Notice that, in both cases, we implemented the continuity correction for the barrier level described in [2] to account for the fact that we are using a discrete approximation to the continuous barrier for both IPS and Monte Carlo. For the different values of the barrier we use, we can calculate the exact probability of default from equation (A.1).

The following are the parameters we used for both IPS and Monte Carlo.

r	σ	S_0	δt	T	n (# of mutations in IPS)	M
.06	.25	80	.001	1	20	20000

The number of simulations M is the same for IPS and Monte Carlo, and from an empirical investigation, we chose $\alpha = 18.5$ in the IPS method (note that $18.5/125$ is within the range of α 's used in Section 4 in the case of 125 names). The results are shown in Figure A.1.

Indeed probabilities of order 10^{-14} will be irrelevant in the context of default probabilities but the user can see that IPS is capturing the rare events probabilities for the single name case whereas traditional Monte Carlo is not able to capture these values below 10^{-4} .

In Figure A.2 we show how the variance decreases with the barrier level, and therefore with the default probability, for Monte Carlo and IPS. In the IPS case the variance is obtained empirically and using the integral formulas derived in the [11]. We deduce that the variance for IPS decreases as p^2 (p is the default probability), as opposed to p in the case of Monte Carlo simulation.

Each Monte Carlo and IPS simulation gives an estimate of the probability of default (whose theoretical value does not depend on the method) as well as an estimate of the standard deviation of the estimator (whose theoretical does depend on the method). Therefore, it is instructive from a practical point of view to compare the two methods by comparing the empirical ratios of their standard deviation to the probability of default for each method. If $p(B)$ is the probability of default for a certain barrier level B , then the standard deviation, $p_2(B)$, for traditional Monte Carlo is given by,

$$p_2^{\text{Monte Carlo}}(B) = \sqrt{p(B)} \times \sqrt{(1-p(B))},$$

and the theoretical ratio for Monte Carlo is given by

$$\frac{p_2^{\text{Monte Carlo}}(B)}{p(B)} = \frac{\sqrt{(1-p(B))}}{\sqrt{p(B)}},$$

which can be computed using (A.1). For IPS, the corresponding ratio is

$$\frac{p_2^{\text{IPS}}(B)}{p(B)} = \frac{\sigma_n^B(\alpha)}{p(B)},$$

where $\sigma_n^B(\alpha)$ is given in Theorem 1. It is computed using the formula given in [11].

In Figure A.3 one sees that there are specific regimes where it is more efficient to use IPS as opposed to traditional Monte Carlo for certain values of the barrier level (below $.65 \times S_0$). This is to be expected since IPS is well suited to rare event probabilities whereas Monte Carlo is not.

References

1. A. BASSAMBOO AND S. JAIN, *Efficient importance sampling for reduced form models in credit risk*, in WSC '06: Proceedings of the 38th conference on Winter simulation, Winter Simulation Conference, 2006, pp. 741–748.
2. M. BROADIE, P. GLASSERMAN, AND S. KOU, *A continuity correction for discrete barrier options*, *Mathematical Finance*, 7 (1997), pp. 325–349.
3. R. CARMONA AND S. CREPEY, *Importance sampling and interacting particle systems for the estimation of markovian credit portfolios loss distributions*, tech. rep., July 2008.
4. P. DEL MORAL, *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*, Springer-Verlag, 2004.
5. P. DEL MORAL AND J. GARNIER, *Genealogical particle analysis of rare events*, *Annals of Applied Probability*, 15 (2005), pp. 2496–2534.
6. J. FOUQUE, R. SIRCAR, AND K. SOLNA, *Stochastic volatility effects on defaultable bonds*, *Applied Mathematical Finance*, 13 (2006), pp. 215–244.
7. J. FOUQUE, B. WIGNALL, AND X. ZHOU, *Modeling correlated defaults: First passage model under stochastic volatility*, *Journal of Computational Finance*, 11 (2008).
8. K. GIESECKE, *in preparation*, tech. rep., 2008.
9. P. GLASSERMAN, *Monte Carlo Methods in Financial Engineering*, Springer-Verlag, 2004.
10. P. GLASSERMAN AND J. LI, *Importance sampling for portfolio credit risk*, *Management Science*, 51 (2005), pp. 1643–1656.
11. D. VESTAL, *Interacting particle systems for pricing credit derivatives*, Ph.D. Dissertation, University of California Santa Barbara, 2008.
12. C. ZHOU, *An analysis of default correlations and multiple defaults*, *Review of Financial Studies*, 14 (2001), pp. 555–76.

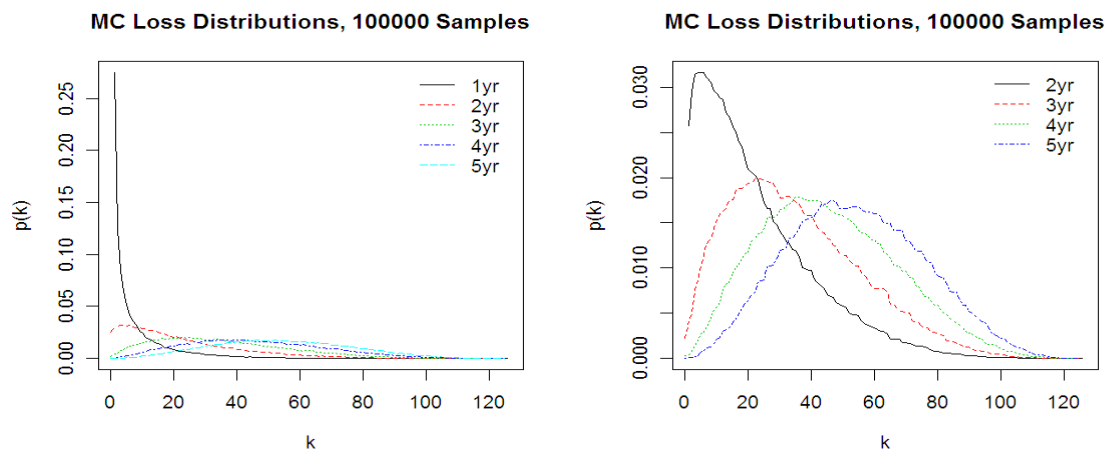


Fig. 4.1 Loss distributions estimated from $M = 10^5$ Monte Carlo samples, of a homogeneous portfolio of $N = 125$ names for maturities $T = 1, 2, 3, 4, 5$. The parameters used are given in the tables above.

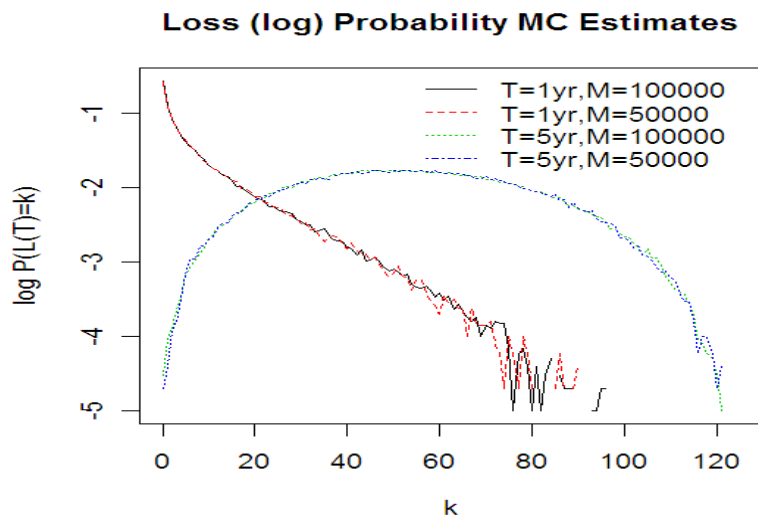


Fig. 4.2 Comparison of the estimates of the loss distribution (log-scale) given by plain Monte Carlo with $M = 10^5$ samples and $M = 50000$ samples for maturities $T = 1$ yr and $T = 5$ yr. The portfolio is homogeneous, has $N = 125$ names and the parameters are given in the tables in the text.

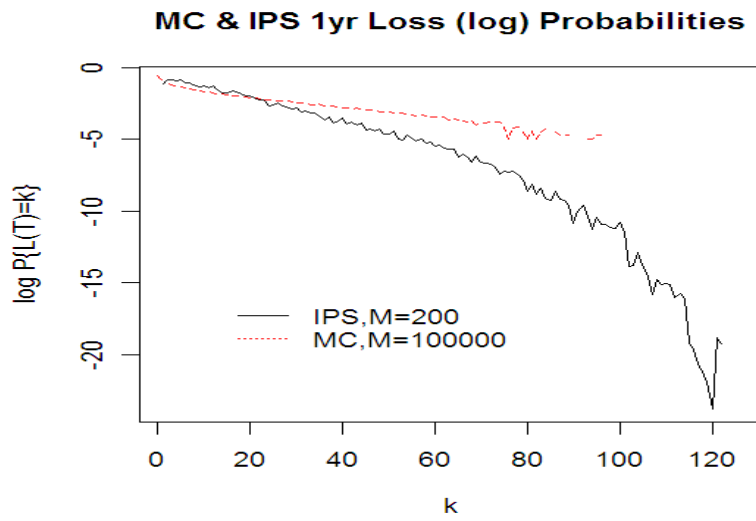


Fig. 4.3 Comparison of the estimates of the loss distribution (log-scale) given by plain Monte Carlo with $M = 10^5$ samples and IPS with $M = 200$ samples. The portfolio is homogeneous, has $N = 125$ names and the parameters are given in the tables in the text.

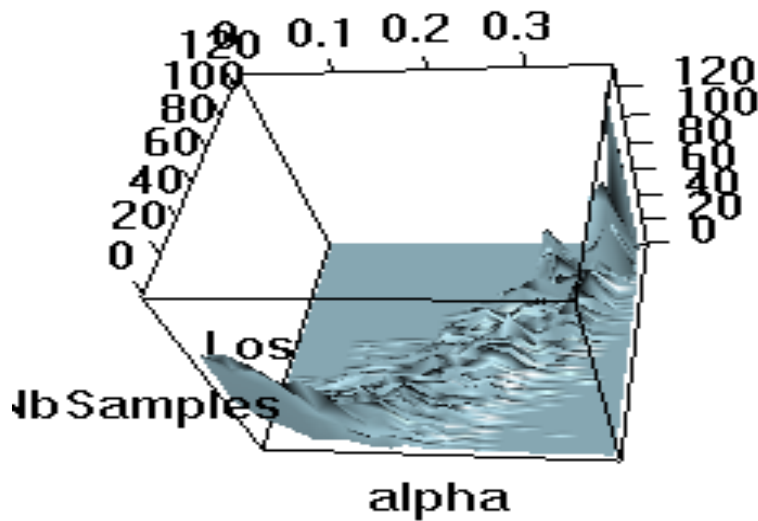


Fig. 4.4 Surface plot of the numbers of samples against the number k of losses at maturity $T = 1$ and the value of the parameter α .

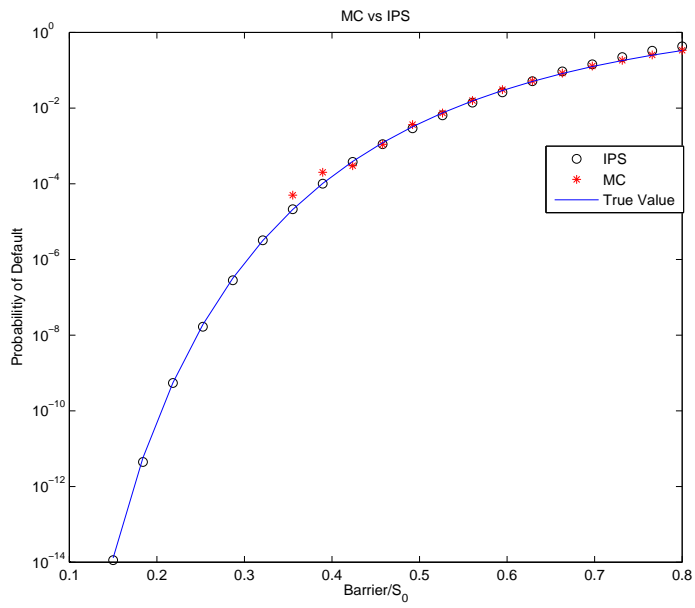


Fig. A.1 Default probabilities for different barrier levels for IPS and Monte Carlo

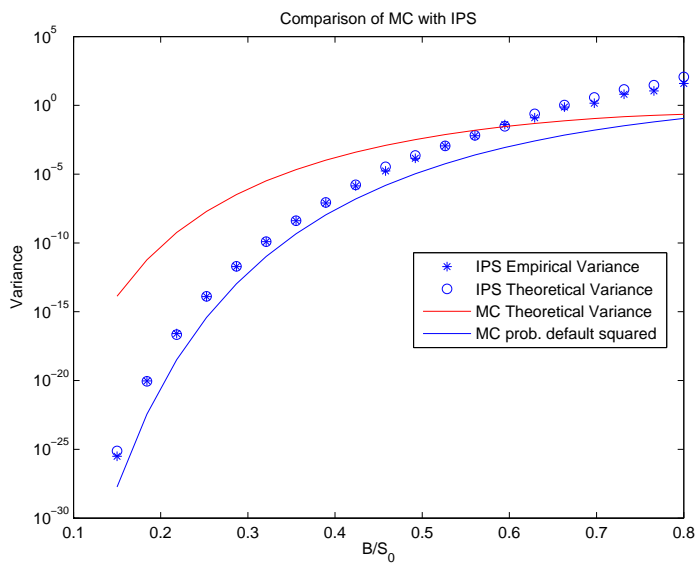


Fig. A.2 Variances for different barrier levels for IPS and Monte Carlo

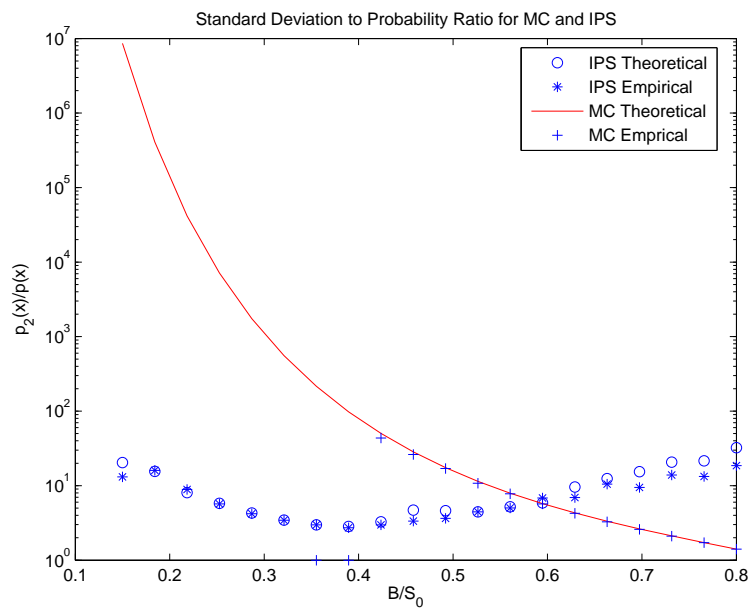


Fig. A.3 Standard deviation-to-probability ratio for Monte Carlo and IPS