Study of new rare event simulation schemes and their application to extreme scenario generation

Ankush Agarwal[a,1], Stefano De Marco[a], Emmanuel Gobet[a,1,2], Gang Liu[a]

[a]*Centre de Mathématiques Appliquées (CMAP)*
Ecole Polytechnique, CNRS
Université Paris-Saclay
Route de Saclay, 91128 Palaiseau Cedex, France.

Abstract

This is a companion paper based on our previous work on rare event simulation methods. In this paper, we provide an alternative proof for the ergodicity of shaking transformation in the Gaussian case and propose two variants of the existing methods with comparisons of numerical performance. In numerical tests, we also illustrate the idea of extreme scenario generation based on the convergence of marginal distributions of the underlying Markov chains and show the impact of the discretization of continuous time models on rare event probability estimation.

Keywords: Extreme scenario, rare event, Markov chains, ergodic properties, interacting particle systems
AMS subject classifications (2010): 65C05, 65C40, 37M25, 65C35

Email addresses: `ankush.agarwal@polytechnique.edu` (Ankush Agarwal),
`stefano.demarco@polytechnique.edu` (Stefano De Marco),
`emmanuel.gobet@polytechnique.edu` (Emmanuel Gobet), `gang.liu1988@gmail.com`
(Gang Liu)

1. Introduction

Our aim is to estimate quantities such as $\mathbb{P}(X \in A)$, $\mathbb{E}(\varphi(X)|X \in A)$ and $\mathbb{E}(\varphi(X)1_{X \in A})$ where X is a random variable which takes value in a measurable state space $\mathcal{X}$ and $A \subset \mathcal{X}$ is a small subset. Another problem of interest is to sample from the conditional distribution $X|X \in A$, i.e. the generation of extreme scenarios. These quantities readily appear in the context of reliability for various industrial fields such as communication systems, aircraft safety, nuclear reactors safety, and so forth. Furthermore, they are also related to financial risks, for example model, credit and actuarial risk. Notably, the problem of extreme scenario generation is known as stress-testing in the field of financial risk management. The current methodology is to introduce artificial market specific and idiosyncratic shocks in the model and then to compute the default probabilities/the expected loss. These handpicked scenarios are sometimes simplified and can misrepresent risks in different ways [3] and, in a dynamic setting, can fail to capture the interdependence between various stochastic factors in the market (see Section 4.2 in [6]). In order to study the stability of a financial - say, banking - system, it is essential to generate extreme scenarios given the bank/the system default. In this work (Section 3), we present some examples where this is achieved.

In the case of a rare event, where the probability $\mathbb{P}(X \in A)$ is extremely small, simple Monte Carlo method is inefficient. Alternatively, we can use Importance Sampling methods which can be very efficient but require information about the model and the rare event. The schemes based on Interacting Particle System (IPS) or splitting techniques or ergodic transformations are also well developed. These approaches can be combined. For a detailed description, see [7]. We elaborate here on the splitting and shaking transformation approach designed therein. To implement this approach, we first create a nested sequence of more and more rare events:

$$\mathcal{X} = A_0 \supset \cdots \supset A_k \supset \cdots \supset A_n = A.$$

Then, we can write the quantities above via the following products:

$$\mathbb{P}(X \in A) = \prod_{k=1}^{n} \mathbb{P}(X \in A_k | X \in A_{k-1}), \tag{1.1}$$

$$\mathbb{E}(\varphi(X)\mathbf{1}_{X \in A}) = \mathbb{E}(\varphi(X)|X \in A_n) \prod_{k=1}^{n} \mathbb{P}(X \in A_k | X \in A_{k-1}). \tag{1.2}$$

Thus, instead of estimating the rare event probability directly, we estimate each conditional probability (and expectation) on the right hand side. Two different schemes, Parallel-One-Path (POP) and Interacting Particle System (IPS), using reversible shaking transformation, are proposed in [7] and further studied in [1]. POP is based on the time average of a Markov chain while IPS is based on the space average of a large Markovian particle system. As is shown in the above references, the shaking transformation is the tool which unifies IPS and POP methods. A similar approach to POP has been developed in the engineering sciences community (see [2]), and is called in that context *subset simulation* approach[3]. We emphasize that the shaking transformation designed in [7] avoids using explicit transition kernels and is presented with a simple parametrization. It facilitates quick implementation and makes it easier to tune shaking forces. Our framework can also handle general situations including stochastic processes driven by Brownian motion or Poisson process. Furthermore, we have developed an adaptive version of the POP method and proved its consistency in [1].

In this paper, our aim is threefold:

- We provide efficient variants of POP and IPS methods. The IPS algorithm is revisited by incorporating extra resampling to enhance the independence property between particles and by reducing the size of the particle system to keep the same computational cost. The adaptive POP algorithm is modified to preserve the benefits of both adaptivity and parallelization.

- When dealing with stochastic processes, we test the robustness of the two algorithms w.r.t. the dimension of the discretization of the underlying noise (i.e. the number of Brownian increments) in order to address the issue of the curse of dimensionality. As a difference with usual MCMC algorithms where the dimension effect may be significant (see for example [8]), we do not observe any significant decrease of speed of convergence.

- We illustrate the application of these methods for extreme (rare-event) scenario generation.

---

[3]We would like to thank Robert Scheichl for directing our attention to this reference.

2. Parallel-one-path (POP) and interacting particle system (IPS) method

2.1. Shaking transformation

POP and IPS methods are designed to estimate each term on the r.h.s. of (1.1) using a random transformation called *shaking transformation* (a local distribution-preserving perturbation). For a given random variable $X \in \mathcal{X}$, $\mathcal{K}(\cdot)$ is called a shaking transformation for X if $(X, \mathcal{K}(X)) \stackrel{d}{=} (\mathcal{K}(X), X)$. We will consider the case where $\mathcal{K}(X) = K(X, Y)$, where $K(\cdot, \cdot)$ is a deterministic function and Y is another random variable independent of X. Then, we can build a $\mathcal{X}$-valued Markov chain as

$$X_{i+1} = K(X_i, Y_i), \qquad X_0 \stackrel{d}{=} X,$$

where Y_i are independent copies of Y. This Markov chain has the distribution of X as its stationary distribution. Thus, we expect an ergodic theorem to hold, i.e. for any measurable $\Psi : \mathcal{X} \to \mathbb{R}$, setting $Z_i := \Psi(X_i)$ and $Z = \Psi(X)$ we have

$$\frac{1}{N} \sum_{i=1}^{N} Z_i \underset{N \to +\infty}{\to} \mathbb{E}(Z) \, a.s. \tag{2.1}$$

as soon as Z is integrable. The convergence rate of the above approximation depends on X and K. In the Gaussian case, if $X = (X_1, \ldots, X_j, \ldots)$ is an infinite-dimensional centered Gaussian vector, we define

$$K(X, X')_j := \rho_j X_j + \sqrt{1 - \rho_j^2} X_j' \tag{2.2}$$

given an independent copy X' of X and the constants $\rho_j \in [-1, 1]$. Then, we show the following:

Theorem 1 (Theorem 2 [1]). *Assume $Z = \Psi(X)$ is square integrable. If $|\rho| := \sup_j |\rho_j| < 1$, then*

$$\mathbb{E}\left(\left|\frac{1}{N} \sum_{i=1}^{N} Z_i - \mathbb{E}(Z)\right|^2\right) \leq \frac{\mathbb{V}\mathrm{ar}(Z)}{N} \left(\frac{1 + |\rho|}{1 - |\rho|}\right). \tag{2.3}$$

The proof for the above general case can be conducted with the help of the generalized Gebelein inequality. In the following, we provide an elementary

proof in one dimension ($X \overset{d}{=} X' \overset{d}{=} \mathcal{N}(0,1)$) using the properties of Hermite polynomials. Recall that the j-th Hermite polynomial is defined by $H_0 := 1, H_j(x) := \frac{(-1)^j}{j!} e^{\frac{x^2}{2}} \frac{d^j}{dx^j}(e^{-\frac{x^2}{2}}), j \geq 1$, and it satisfies the following properties:

$$H'_j(x) = H_{j-1}(x), \qquad (j+1)H_{j+1}(x) = xH_j(x) - H_{j-1}(x).$$

Lemma 1. *For $\rho \in [-1,1]$, we have $\mathbb{E}\left(H_j(\rho x + \sqrt{1-\rho^2}X')\right) = \rho^j H_j(x)$.*

Proof. Obviously, the result holds for $j = 0$ and $j = 1$. Now suppose it holds for $1, 2, \ldots, j$. We prove it for $j + 1$.

$$(j+1)\mathbb{E}\left(H_{j+1}(\rho x + \sqrt{1-\rho^2}X')\right)$$
$$= \mathbb{E}\left((\rho x + \sqrt{1-\rho^2}X')H_j(\rho x + \sqrt{1-\rho^2}X') - H_{j-1}(\rho x + \sqrt{1-\rho^2}X')\right)$$
$$= \rho x \rho^j H_j(x) + \mathbb{E}\left(\sqrt{1-\rho^2}X'H_j(\rho x + \sqrt{1-\rho^2}X')\right) - \rho^{j-1}H_{j-1}(x).$$

It is easy to prove that for a smooth function h with polynomial growth, we have $\mathbb{E}(h'(X')) = \mathbb{E}(X'h(X'))$. Then take $h(x') := H_j(\rho x + \sqrt{1-\rho^2}x')$ and write

$$\mathbb{E}\left(\sqrt{1-\rho^2}X'H_j(\rho x + \sqrt{1-\rho^2}X')\right) = (1-\rho^2)\mathbb{E}\left(H'_j(\rho x + \sqrt{1-\rho^2}X')\right)$$
$$= (1-\rho^2)\rho^{j-1}H_{j-1}(x),$$

which allows to conclude. $\qquad \square$

Proof of Theorem 1 (dimension 1). We consider the case $X \overset{d}{=} \mathcal{N}(0,1)$. We denote the single shaking parameter by ρ and define $T_\rho\Psi(X) := \mathbb{E}'(\Psi(K(X, X')))$ $= \mathbb{E}\left(\Psi(\rho X + \sqrt{1-\rho^2}X')\right)$ where the expectation is taken only w.r.t. X'. Since Hermite polynomials form a complete orthonormal system in $L^2(X)$, there exist $(a_j \in \mathbb{R}, j \geq 1)$ such that $Z = \Psi(X) = \mathbb{E}(\Psi(X)) + \sum_{j \geq 1} a_j H_j(X)$. Then, the above lemma gives

$$T_\rho(\Psi(X)) = \mathbb{E}(\Psi(X)) + \sum_{j \geq 1} \rho^j a_j H_j(X);$$

$$\mathbb{E}\left(\left|T_\rho(\Psi(X)) - \mathbb{E}(\Psi(X))\right|^2\right) \leq \sum_{j \geq 1} \rho^{2j} a_j^2 \leq |\rho|^2 \mathbb{V}\mathrm{ar}(\Psi(X)).$$

5

It readily follows that, for $i < i'$,

$$\mathbb{C}\mathrm{ov}\left(Z_i, Z_{i'}\right) = \mathbb{E}\left(\left(Z_i - \mathbb{E}\left(Z\right)\right)\left(T_{\rho^{i'-i}}(\Psi(X_i)) - \mathbb{E}\left(Z\right)\right)\right) \le |\rho|^{i'-i}\mathbb{V}\mathrm{ar}\left(Z\right).$$

This allows us to get the upper bound (2.3), since the l.h.s. of (2.3) is equal to $\frac{1}{N^2}\sum_{1 \le i, i' \le N} \mathbb{C}\mathrm{ov}\left(Z_i, Z_{i'}\right)$. $\qquad\square$

Using the shaking transformation, we define a second random transformation, called *shaking with rejection*, which allows to formulate the POP and IPS methods. For a fixed $k \in \{0, \ldots, n\}$, we define

$$\mathcal{M}_k^{\mathcal{K}}(X) := \mathcal{K}(X)\mathbf{1}_{\mathcal{K}(X) \in A_k} + X\mathbf{1}_{\mathcal{K}(X) \notin A_k}. \tag{2.4}$$

As earlier, we write $\mathcal{M}_k^{\mathcal{K}}(X) = M_k^K(X, Y)$ when $\mathcal{K}(X) = K(X, Y)$ and Y is another random variable independent of X.

2.2. POP method

The POP method is designed to estimate each conditional probability $\mathbb{P}(X \in A_k | X \in A_{k-1})$ separately. The key observation for this technique is that the conditional distribution $X \mid X \in A_k$ is invariant with respect to the shaking with rejection $\mathcal{M}_k^{\mathcal{K}}(\cdot)$.

Proposition 2 ([7, Proposition 2.2]). *Let $k \in \{0, 1, \cdots, n\}$. For any bounded measurable φ, we have*

$$\mathbb{E}\left(\varphi(\mathcal{M}_k^{\mathcal{K}}(X)) \mid X \in A_k\right) = \mathbb{E}\left(\varphi(X) \mid X \in A_k\right). \tag{2.5}$$

This proposition enables us to define a Markov chain which has $X|X \in A_k$ as its invariant measure.

Definition 1. *For each $k = 0, \ldots, n$, given a starting point $X_{k,0}$, define*

$$X_{k,i} := \mathcal{M}_k^{\mathcal{K}}(X_{k,i-1}) = M_k^K(X_{k,i-1}, Y_{k,i-1}) \quad \text{for } i \ge 1 \tag{2.6}$$

where $(Y_{k,i})_{i \ge 0}$ is a sequence of independent copies of Y, independent of $X_{k,0}$.

Therefore, if the above defined Markov chain is ergodic, from Birkhoff's ergodic theorem, we have the following approximation:

$$\mathbb{E}\left(\varphi(X)|X \in A_k\right) \approx \frac{1}{N}\sum_{i=0}^{N-1}\varphi(X_{k,i}), \quad \text{as } N \to +\infty. \tag{2.7}$$

6

Setting $\varphi \equiv \mathbf{1}_{A_{k+1}}$ for $k \leq n-1$ yields an approximation of $\mathbb{P}\left(X \in A_{k+1} | X \in A_k\right)$ and the product of all the estimators provides an estimate of the rare event probability $\mathbb{P}\left(X \in A\right)$ thanks to (1.1). Notice that the POP method is especially suitable for parallel implementation and it is very economical in terms of memory consumption.

Application to stress testing. Another application of the above approach is to generate samples for $X \mid X \in A$ by using the last level Markov chain $(X_{n,i})_{i \geq 0}$. We know that for an aperiodic ergodic Markov chain, its marginal distribution will converge to its unique stationary distribution. Thus, if we run the Markov chain $X_{n,\cdot}$ with enough iterations, its marginal distribution will be close to the conditional distribution of $X \mid X \in A$. In other words, for N large, $(X_{n,N}, X_{n,2N}, \ldots, X_{n,lN})$ will constitute a sample of size l of (almost independent) random variables with distribution (approximately equal to) $X \mid X \in A$. This can be used to build extreme scenarios, for example, in stress testing. As recalled in the introduction, this is an important question in risk assessment. This application is illustrated later in Section 3.

Adaptive POP. A suitable choice of nested subsets $(A_k)_{0 \leq k \leq n}$ is essential for an efficient application of the POP method. If any of the conditional probabilities $\mathbb{P}(X \in A_k \mid X \in A_{k-1})$ is small, we again fall in the case of rare event. Ideally, we would wish to have all the conditional probabilities to be of the same order but this is not possible without apriori knowledge. In order to overcome this difficulty, an adaptive version of the POP method is developed in [1] together with a proof of its convergence. This adaptive version finds good nested subsets using the quantile estimators of each intermediate level Markov chain. However the subsets (A_k) are determined sequentially which rules out a parallel implementation. In subsection 2.4.1 we propose a variant of this adaptive POP method which allows to recover a fully parallel version of the algorithm.

2.3. IPS method

With the aim of estimating the rare event probability $\mathbb{P}\left(X \in A\right)$, we define a Markov chain as follows:

Definition 2. $(X_i)_{0 \leq i \leq n-1}$ *is a $\mathcal{X}$-valued Markov chain defined as*

$$X_0 \stackrel{\mathrm{d}}{=} X, \qquad X_i := \mathcal{M}_i^{\mathcal{K}}(X_{i-1}) = M_i^K(X_{i-1}, Y_{i-1}) \quad \text{for } 1 \leq i \leq n-1, \ (2.8)$$

where $(Y_i)_{0 \le i \le n-2}$ is a sequence of independent copies of Y, independent of X_0.

With the above definition we obtain the following result.

Theorem 3 ([7]). *For any bounded measurable function $\varphi : \mathcal{X} \to \mathbb{R}$, we have*

$$\mathbb{E}\left(\varphi(X)\mathbf{1}_{X \in A}\right) = \mathbb{E}\left(\varphi(X_{n-1})\prod_{i=0}^{n-1}\mathbf{1}_{A_{i+1}}(X_i)\right). \tag{2.9}$$

The interpretation of rare event probability as an expectation related to a Markov chain, i.e. in a Feynman-Kac measure way, enables the use of the IPS method. The steps of this algorithm are described below in Algorithm 1 with $J = 1$. The idea behind this procedure is to imitate the evolutionary selection and mutation process via the interaction of a large number of particles (see, for example, [5]). As seen from Algorithm 1, each generation is used to approximate the conditional distribution $X \mid X \in A_k$.

2.4. Variants of POP and IPS methods

2.4.1. POP method with level refinement

As shown through the numerical examples in [7], the POP method is efficient when provided with good choices of nested subsets, which is possible thanks to the aforementioned level adaptive version. However, the advantage of parallelization is lost. Here, we propose a variant with level refinement which allows to recover the possibility of parallelization.

To begin, we fix a threshold value q (for example $q = 0.05$) for each conditional probability under which the estimation using one Markov chain is considered insufficient, and we arbitrarily choose some nested subset A_k (such as an equi-distant partition of the entire space according to some criteria function). Then, we run one Markov chain as defined in Definition 1 at each level in parallel to estimate all the conditional probabilities $\mathbb{P}(X \in A_k \mid X \in A_{k-1})$. Next, we check if the estimator $p_k^{(N)}$ is larger than q and accept the estimate for each level for which it is true. Otherwise, if $p_{k_0}^{(N)} < q$ for some k_0, we change the algorithm to the adaptive scheme as in [1] to estimate $\mathbb{P}(X \in A_{k_0} \mid X \in A_{k_0-1})$. Notice that the Markov chain already used for level k_0 need not be simulated again. Another scheme is to put new nested subsets between A_{k_0} and A_{k_0-1} and run the POP method again in

parallel. Finally, the product of all the estimators provides an estimate of the rare event probability.

2.4.2. IPS method with extra resampling

Initialization:
For given integers J and M, set $M' = \lfloor \frac{M}{J} \rfloor$;
Draw $(X_0^{(M',m)}, m = 1, \cdots, M')$ which are i.i.d. copies of X ;
$p_0^{(M')} = \frac{1}{M'} \sum_{m=1}^{M'} \mathbf{1}_{A_1}(X_0^{(M',m)})$;
for $i = 0$ until $n - 2$ **do**
$\quad$ **for** $m = 1$ until M' **do**
$\quad\quad$ *Selection step:*
$\quad\quad$ **if** $X_i^{(M',m)} \in A_{i+1}$ **then**
$\quad\quad\quad$ $\hat{X}_{i,0}^{(M',m)} = X_i^{(M',m)}$;
$\quad\quad$ **else**
$\quad\quad\quad$ Pick $\hat{X}_{i,0}^{(M',m)}$ uniformly in $\{X_i^{(M',m)} \in A_{i+1}\}$ and independently of everything else;
$\quad\quad$ **end**
$\quad\quad$ *Mutation step:*
$\quad\quad$ **for** $j = 1$ until J **do**
$\quad\quad\quad$ $\hat{X}_{i,j}^{(M',m)} = M_{i+1}^K(\hat{X}_{i,j-1}^{(M',m)}, Y_{i,j-1}^{(m)})$ where $Y_{i,j}^{(m)}$ are i.i.d copies of Y;
$\quad\quad$ **end**
$\quad\quad$ $X_{i+1}^{(M',m)} = \hat{X}_{i,J}^{(M',m)}$;
$\quad$ **end**
$\quad$ $p_{i+1}^{(M')} = \frac{1}{M'} \sum_{m=1}^{M'} \mathbf{1}_{A_{i+2}}(X_{i+1}^{(M',m)})$;
end
Result: $p^{(M')} = \prod_{i=0}^{n-1} p_i^{(M')}$

Algorithm 1: IPS method with extra resampling and reduced size for computing $\mathbb{P}(X \in A)$

An intrinsic feature of the IPS method is that different generations of particles are correlated with each other. Thus, if the empirical measure is inaccurate at the first generation, it is likely to be inaccurate at the following generation, which amplifies the variance of the final estimator. Here, we propose a way to reduce this dependency between generations. As previously seen,

the shaking with rejection transformation is invariant with respect to the conditional distribution of X. Hence, if we apply several iterations of the transformation to obtain the next particles generation, the distribution of the system will be less influenced by the previous state. In order to keep the same computational cost, we reduce the size of the particle system. Thus, we run the proposed version of IPS algorithm with particle size $\lfloor \frac{M}{J} \rfloor$ and the transformation $\mathcal{M}_k^{\mathcal{K}}(X)$ applied J times at each time step. Regarding the convergence analysis, we can show that for a given J, the convergence still holds as $\frac{M}{J}$ goes to infinity. As we see in the numerical experiments, this variant of the IPS method has a better performance, compared to the standard algorithm (Subsection 2.3) without extra resampling ($J = 1$).

Another variant of the above algorithm can be obtained by using the intermediate states during the extra resampling at each level $\big(\hat{X}_{i,j}^{(M',m)}; j = 1, 2, \ldots, J-1\big)$ to contribute to the conditional probability estimator $\big(p_{i+1}^{(M')}; i = 0, 1, \ldots, n-2\big)$. In Table 3.2, we observe that this algorithm has similar standard deviation as the extra resampling variant of the IPS method.[4]

3. Numerical experiments

In this section, we provide new and complementary numerical tests on two examples chosen from [1]. For completeness, each example is recalled briefly.

3.1. Credit portfolio default probability

We first consider an example of a credit portfolio based on the asset values of N_0 different firms $(S_1(t), \ldots, S_{N_0}(t) : 0 \leq t \leq T)$. Each S_i follows a Heston model with the same CIR volatility $(\sigma(t) : 0 \leq t \leq T)$ and with the same parameters, given by

N_0	$S_i(0)$	r	ρ^W	$\sigma(0)$	κ	$\bar{\sigma}$	γ	ρ^σ	T
125	90	0.06	0.10	0.4	3.5	0.4	0.7	-0.06	1

The parameter r stands for the drift of S_i, γ for the volatility of volatility, $\bar{\sigma}$ for the long-term volatility, κ for its mean-reverting rate; the correlation

---

[4]We would like to thank Arnaud Guyader and Tony Lelievre for suggesting this variant.

$\times 10^{-6}$	$J=1$	$J=2$	$J=3$	$J=4$	$J=5$	$J=6$	$J=7$	$J=8$	$J=9$	$J=10$	POP
$\rho=0.9$	2.50	1.21	1.04	1.01	1.03	1.06	1.11	1.19	1.31	1.37	0.80
$\rho=0.7$	1.74	1.40	1.29	1.25	1.20	1.29	1.28	1.33	1.37	1.43	1.02
$\rho=0.5$	4.46	3.69	3.56	3.11	3.18	2.90	2.67	2.73	2.63	2.61	1.94

Table 3.1: Empirical standard deviation of IPS estimators of default probability for $L = 100$ and $B = 36$ based on 1000 algorithm macro-runs, with $M = 10^4$ and particle system size equal to $M' = \lfloor \frac{M}{J} \rfloor$. The last column is the empirical standard deviation of the POP method using $n = M = 10^4$ iterations at each level.

between the Brownian motions driving S_i and S_j is ρ^W and the correlation between each asset and its volatility is ρ^σ.

Our aim is to calculate the probability that at least L defaults occur before T

$$\mathrm{P}(L) = \mathbb{P}\left(\sum_{i=1}^{N_0} \mathbf{1}_{\{\tau_i(B) \leq T\}} > L\right), \quad 0 < L < N_0, \tag{3.1}$$

where $\tau_i(B)$ is the default time for the firm i with default level B:

$$\tau_i(B) := \inf\big\{t \geq 0 : S_i(t) \leq B\big\}.$$

Different IPS-based methods have been proposed in [4] and [1] to compute $\mathrm{P}(L)$ using a dynamic and a static point of view, respectively. We will continue with the static point view, i.e. applying shaking transformation on the path space. In order to express $\mathrm{P}(L)$ in the form of (1.1), we need to create a cascade of decreasing sets $(A_k)_{1 \leq k \leq n}$. We set

$$A_k := \left\{\sum_{i=1}^{N_0} \mathbf{1}_{\{\tau_i(S_i(0) - (S_i(0) - B_i)\frac{k}{n}) \leq T\}} > L\right\}, \quad 1 \leq k \leq n.$$

We estimate $\mathrm{P}(L)$ for different values of L with $N_{\text{step}} = 50$ time steps in the Euler discretization scheme for $(S_1, \ldots, S_{N_0}, \sigma)$ as explained in [1]. The random variable X modeling the rare event is Gaussian, associated to the N_{step} increments of the $N_0 + 1$ Brownian motions and we use the Gaussian shaker (2.2) with $\rho_j = \text{Cst}$.

For $L = 100$ and $B = 36$, a 99% confidence interval $[4.92, 5.13] \times 10^{-6}$ is given in [1] as reference. In Table 3.1, we report the results for IPS estimators with different number J of shaking transformation iterations applied at each time for fixed $n = 5$ levels.

$\times 10^{-6}$	$J=2$	$J=4$	$J=5$	$J=8$	$J=10$	$J=12$	$J=15$	POP
$\rho=0.9$	1.46	1.14	1.09	1.00	1.00	1.02	1.02	0.80

Table 3.2: Empirical standard deviation of IPS estimator based on the variant of Algorithm 1 of default probability for $L=100$ and $B=36$ based on 1000 algorithm macro-runs, with $M=10^4$ and particle system size equal to $M'=\lfloor\frac{M}{J}\rfloor$.

As we can see from Table 3.1, the IPS method with fewer particles but extra resampling at each step have better performance compared to the case without resampling. Heuristically, a good choice is $J=4$. However, even after extra resampling, the POP method yields smaller standard deviation.

Typically, in the classical IPS method such as in [4], the choice of appropriate discretization step (Δ_t) in the Euler scheme is quite delicate. On one hand, small Δ_t reduces the bias but on the other hand, small Δ_t adds statistical errors because the number of selection/mutation steps increases in the IPS method used there. In our version of the IPS and POP methods, the underlying random variable X lies in a large-dimensional space $\mathcal{X}=\mathbb{R}^{(N_0+1)N_{\text{step}}}$. Therefore it is important to assess how the large dimension affects the statistical errors of the methods (like in MCMC sampler, for example, see [8]). The important point is that we use reversible transformation directly in the path space. This suggests that our methods are less sensitive to time-discretization. We investigate this problem in Figure 3.1, where we study the impact of N_{step} in the POP method with $n=5$ levels. We report the numerical results only for the POP method, as the qualitative phenomenon for the IPS method is the same. The graphs show that the choice of N_{step}, when large, has no significant impact on the optimal value of the shaking parameter ρ, which corresponds to the minimum standard deviation of the conditional probability estimator at a given level. Additionally, both the standard deviation and the rejection rate are quite insensitive to N_{step}. These are important advantages of the methods studied in this work. This allows to decouple the problem of bias reduction and the control of statistical convergence.

Stress-testing. In Figure 3.2, we exhibit extreme scenarios for the asset price of firm 1 and its volatility, in the situation of various defaults (level 1 and last level, $n=5$, $L=100$) with two different values of B. Each scenario is obtained by using 10^4 iterations of shaking with rejection (with $\rho=0.9$). For this, we have used the first and last-level Markov chain $X_{1,.}$ and $X_{5,.}$, respectively, to get samples from the distributions $X\mid X\in A_1$ and
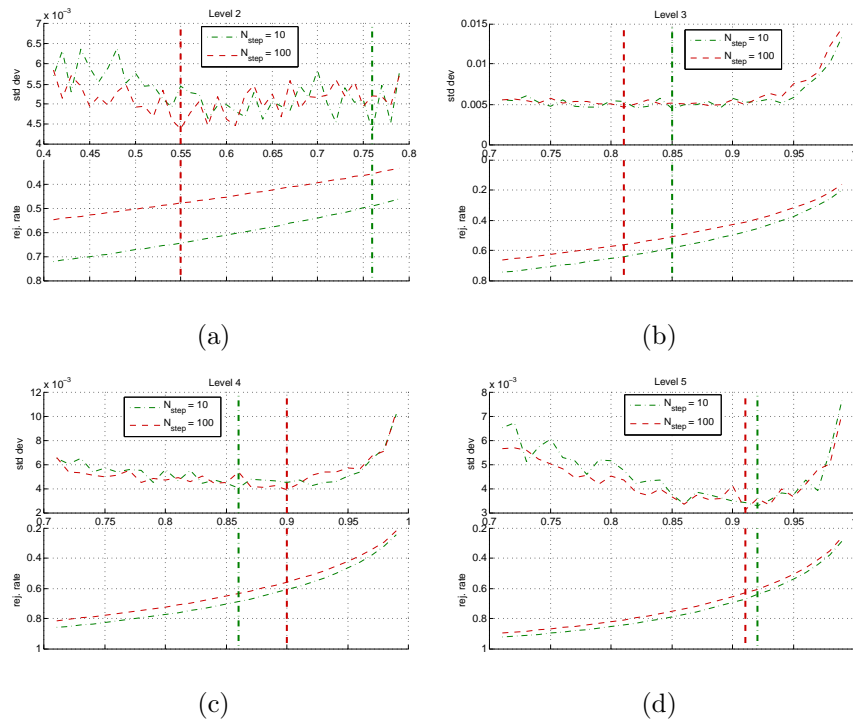
Figure 3.1: The impact of the number of discretization times (N_{step}) on the optimal shaking parameter (ρ) at each level of the POP method.

$X \mid X \in A$ respectively, as explained in Subsection 2.2. We believe that such tools can be efficiently exploited by regulators and risk managers for a better risk assessment.

3.2. Model misspecification in financial engineering and extreme scenarios

We consider the Profit&Loss (PL) when a trader uses a Black-Scholes (BS) model with volatility $\sigma_- > 0$ in order to hedge a European call option, while the true dynamics of the underlying S is given by a path-dependent volatility model with volatility switching between σ_- and $\sigma_+ > \sigma_-$. We propose a discrete-time version of the model based on a monitoring period Δ_t (say 1 week) and monitoring dates $t_i = i\Delta_t$: if the underlying spot price drops below the average of previous four monitored prices, the level of volatility becomes σ_+, otherwise it sticks to σ_- (see [1] for more details). This model corresponds to the usual empirical observation that the underlying volatility
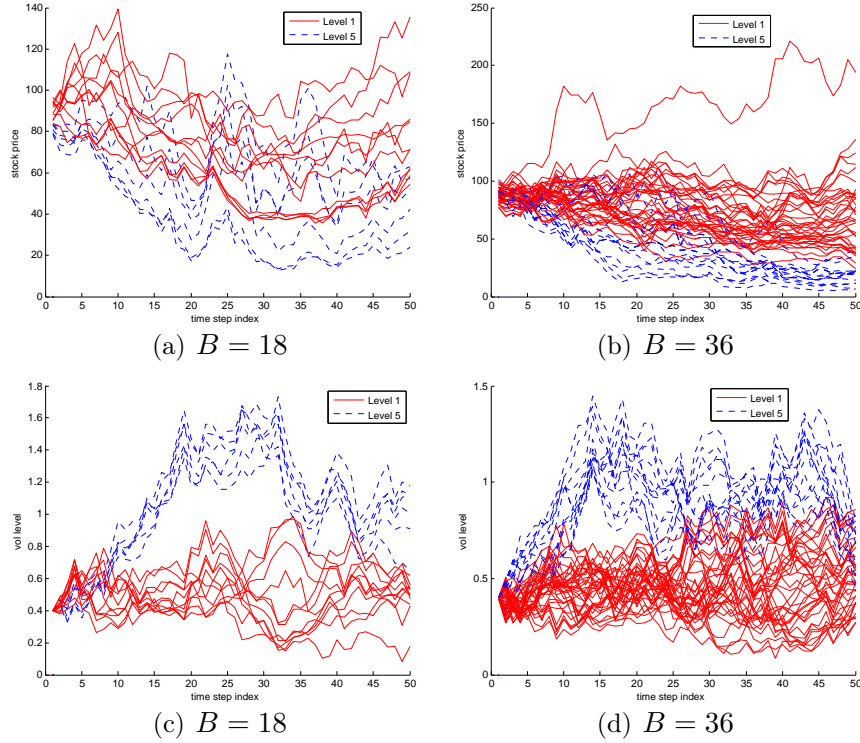
13

Figure 3.2: Sample paths for the asset price of firm 1 at Level 1 and Level 5 in the POP method and the respective volatility sample paths.

is higher when price falls.

The call option maturity is T, and $[0, T]$ is the trading period under consideration. For our numerical experiment, we take $T = 1$, $\Delta_t = T/50$ and we assume that the trader's rebalancing period is $5\Delta_t$. At times $t_{5\Delta_t j}, 0 \leq j < 10$, the trader holds δ_j assets where δ_j is given from the BS-Delta with volatility σ_- and spot $S_{5\Delta_t j}$. Hence at the maturity the trader's PL is given by

$$PL_{\text{trader}} := \mathbb{E}_{\text{trader}}\big[(S_T - K_{\text{strk}})_+\big] + \sum_{j=0}^{9} \delta_j(S_{5\Delta_t(j+1)} - S_{5\Delta_t j}) - (S_T - K_{\text{strk}})_+.$$

Because of the mismatch between the realized model and the hedging model, the trader may incur large losses. Large loss probability $\mathbb{P}(A)$ with $A = \{PL_{\text{trader}} \leq L\}$ has been estimated using the POP and IPS methods in [1]. Here we are more interested to know what are the typical scenarios which
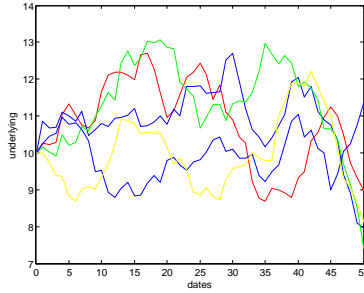
14

Figure 3.3: Typical paths of the underlying stock price which lead to large hedging loss

generate large losses. We set $S_0 = K_{\text{strk}} = 10$, $\sigma_- = 0.2$, $\sigma_+ = 0.27$ and take $L = -2.4$. Applying the principle of subsection 2.2 on stress-testing, we can get samples from the distribution $X \mid X \in A$; 5 typical scenarios are reported in Figure 3.3. Each scenario is obtained using 10^4 iterations of shaking with rejection (with $\rho = 0.9$). As intuitively expected, typical extreme scenarios exhibit large fluctuations of S.

References

[1] A. Agarwal, S. De Marco, E. Gobet, G. Liu, Rare event simulation related to financial risks: efficient estimation and sensitivity analysis. Preprint available at https://hal-polytechnique.archives-ouvertes.fr/hal-01219616 (2015).

[2] S-K. Au, J.L. Beck, Estimation of small failure probabilities in high dimensions by subset simulation. Probab. Eng. Mech. 16(4) (2001) 263–277.

[3] T. Breuer, I. Csiszar, Systematic stress tests with entropic plausibility constraints. J. Bank. & Finance 37(5) (2013) 1552-1559.

[4] R. Carmona, J.P. Fouque, D. Vestal, Interacting particle systems for the computation of rare credit portfolio losses. Finance Stoch. 13(4) (2009) 613–633.

[5] P. Del Moral, Feynman-Kac formulae: Genealogical and Interacting Particle Systems with Applications. Springer, New-York, 2004.

[6] European Central Bank, EU banks' liquidity stress testing and contingency funding plans ECB Publications http://www.ecb.europa.eu/pub/pdf/other/eubanksliquiditystresstesting200811en.pdf (2008).

[7] E. Gobet, G. Liu, Rare event simulation using reversible shaking transformations. SIAM Sci. Comput. 37(5) (2015) A2295–A2316.

[8] N.S. Pillai, A.M. Stuart, A.H. Thiery, Optimal scaling and diffusion limits for the Langevin algorithm in high dimensions. Ann. Appl. Probab. 22(6) (2012) 2320-2356.