Rare event simulation related to financial risks: efficient estimation and sensitivity analysis

Ankush Agarwal, Stefano De Marco, Emmanuel Gobet, Gang Liu

Rare event simulation related to financial risks: efficient estimation and sensitivity analysis

Ankush Agarwal*[1], Stefano De Marco†[2], Emmanuel Gobet‡[2], and Gang Liu§ [2]

[1]Adam Smith Business School, University of Glasgow, University Avenue, G12 8QQ Glasgow, United Kingdom
[2]Centre de Mathématiques Appliquées (CMAP), Ecole Polytechnique and CNRS, Route de Saclay, 91128 Palaiseau Cedex, France

This version: December 28, 2017

Abstract

We propose an adaptive rare event simulation method based on reversible shaking transformations on path space to estimate rare event statistics arising in different financial risk settings which are embedded within a unified framework of isonormal Gaussian process. We provide convergence results for sampling algorithms and our adaptive scheme which are based on Markov chain created using shaking transformations. We also tackle the important problem of calculating sensitivities of rare event statistics with respect to the model parameters by providing general analytical formulas. We demonstrate the strength of our method and application of our results in various numerical examples which cover usual semi-martingale stochastic models (not necessarily Markovian) driven by Brownian motion and, also, models driven by fractional Brownian motion (non semi-martingale) to address various financial risks.

Key words: Rare event, Monte Carlo simulation, Markov chains, ergodic properties, interacting particle systems, Malliavin calculus, sensitivity analysis, fractional Brownian motion, credit default swaps, model misspecification, deep out-of-the-money options

AMS subject classifications (2010): 65C05, 60G15, 65C40, 65C35, 60H07

JEL classification: C63

1 Introduction

During the last thirty years, financial crises and shocks have repeatedly occurred, ranging from the Black Monday in 1987 to the recent Chinese stock market crash in 2015, passing through the financial crisis of 2007-2008 triggered by over-valuated subprime mortgages. As a consequence, banks, insurance companies and regulators are paying more and more attention to the quantification of risk in all its forms - market risk, credit risk, operational risk - and to its management, in particular in the tails and extremes. Since financial institutions often use estimated model parameters, which may include errors due to an inaccurate estimation process, the sensitivities of financial risks with respect to model parameters also provide valuable information to help with the risk management. In this paper, we propose a new adaptive rare event simulation method based on the reversible shaking transformations to estimate rare event statistics associated with various financial risks. For these reversible shaking transformations, we provide error estimates and speed

*Email: ankush.agarwal@glasgow.ac.uk. The author research was conducted while at CMAP, Ecole Polytechnique and is part of the Chair *Financial Risks* of the *Risk Foundation*.

†Email: stefano.demarco@polytechnique.edu. The author research is part of the Chair *Financial Risks* of the *Risk Foundation*.

‡Email: emmanuel.gobet@polytechnique.edu. The author research is part of the Chair *Financial Risks* of the *Risk Foundation* and the *Finance for Energy Market Research Centre*.

§Email: gang.liu1988@gmail.com. The author research is part of the Chair *Financial Risks* of the *Risk Foundation*.

of convergence result in finite dimensions which also allow us to prove the consistency of estimator based on our adaptive scheme. In addition, we provide analytical formulas for the sensitivities of these rare event statistics in a generalized setting. Recently, the reversible shaking transformations have also been combined with Monte Carlo simulation to estimate nested risk statistics in general settings [FGM17].

1.1 Rare Events in Finance

Rare event analysis comes into play when we consider all forms of financial risks related to events in the extreme tails. In the most recent regulation texts, this is sometimes referred to as stress-testing [Bas17]. To fix the convention, hereby, we refer to rare events as the events which have probabilities smaller than 10^{-3} or 10^{-4}. In the following, we list several examples in the literature that cover a wide range of possible financial risk situations where two of them – *model risk* and *credit risk* – are linked to the latest Basel regulations [Bas11].

- The impact of using a misspecified model for hedging financial positions is commonly known as *model risk* [Con06]. It is related to the issue of model robustness as addressed in [EJS98, CM01]. We consider one such problem in Section 3.1 to estimate the model risk.
- A typical problem which arises in managing *credit risk* is to estimate certain default probabilities required for pricing Credit Default Swaps. Such an example inspired from [CFV09, CC10] is considered in Section 3.2.
- Models based on fractional fields are now popular in physics, natural sciences, economics (see [CI13]) and are also used in financial modeling (see [CR98]). In the recent work [GJR14], the authors modeled the volatility of S&P 500 index by fractional Brownian motion (fBM). In Section 3.3, we consider the problem of estimating far-from-the-money implied volatilities (IV) of an underlying process following fSABR model [GJ14].
- Another example of market risk comes from the evaluation of deep out-of-the-money options (see [GT16, FGG$^+$15]). In Section 3.4, we consider options written on a portfolio of assets and estimate their sensitivities with respect to different portfolio and model parameters.

1.2 Numerical Methods in Literature

In all the above examples, crude Monte-Carlo simulations naturally fail to be efficient since the events of interest have small probabilities. Several numerical methods have been therefore developed to sample extreme scenarios. Among them, the Importance Sampling (IS) technique [RK08] transforms the distribution of random variables to be simulated in order to make the event not rare anymore (or less rare). However, it is known that the method relies heavily on the particularity of the model to obtain a feasible transformation. For example, an IS technique has been designed in [GHS00] for computing the 10 days-VaR of a portfolio of options using small-time linearization in the context of lognormal models and in [GT16] to price deep out-of-the-money basket options in lognormal models.

Another approach is based on Interacting Particle System (IPS) (see for example [Del04, DG05, CDLL06]). Usually, it is designed when the rare event is related to the terminal value of a discrete-time Markov chain in $\mathbb{R}^d$. In this approach, a large number of particles representing the underlying state evolve with interactions at every time step until the terminal time where the rare event statistics are evaluated using Feynman-Kac formulas and are computed on the empirical measure associated to these particles. This approach requires to embed the rare event problem into a Markovian setting which causes some difficulties when the model is not Markovian. The method becomes less efficient as the number of time steps gets larger due to the increasing statistical error. For related numerical experiments, see [CFV09, CC10].

Lastly, the splitting techniques initiated in [VV91] decompose the problem of rare event in a sequence of n increasingly rare events (n is of the order of the log-probability of the rare event, thus it is often smaller than 10). Usually, a particle-based approach (similar to IPS) is used to implement the method. Recently in [GL15], the authors have used splitting approach to express rare event statistics as a product of conditional expectations and used tools from ergodic theory to compute these conditional expectations using reversible transformations (also called *shakers*). It leads to Parallel One-Path (POP) method since each conditional statistic is computed in parallel using a single particle evolving over a long path. The authors also combined shakers in path space with the IPS method to provide a new version of the IPS method. Unlike the original IPS method, this approach does not suffer from increasing variance with finer time discretization and turns out to be more suitable for problems involving stochastic processes (without any Markov assumptions).

1.3 Our Contributions

Observe that most of the financial models are often built on Gaussian noise to account for market risk and other financial risks. Some examples include models based on finite dimensional Gaussian variables, or infinite dimensional ones like for standard or fractional Brownian motion and multi-dimensional Brownian Stochastic Differential Equations (SDE). In order to provide a unified treatment, we embed our study in the framework of a general isonormal Gaussian process where the random process is $X := (X(h) : h \in \mathcal{H})$ in association to a real separable Hilbert space $\mathcal{H}$. This setting is also referred as *Gaussian Hilbert space*. In this work, we design rare-event sampling algorithms, adapted to this Gaussian framework, using reversible Gaussian transformations (called *shaker*) with splitting and rejection (POP approach). Hereafter, we refer to these algorithms as POP methods. This general setting facilitates to show the convergence of the reversible Gaussian transformations in infinite dimension using generalized Gebelein inequality [Geb41] for the maximal correlation between Gaussian subspaces [Jan97, Chapter 10]. We also make use of the existing results on the convergence of Metropolis-Hastings sampler to conclude the same in finite dimensions for Gaussian transformations *with rejection* which form the basis of our sampling algorithms. In addition, we provide error estimates and prove the geometric ergodicity of Gaussian transformations *with rejection* which allows us to verify the conditions required to prove the consistency of our adaptive estimator (discussed below). The extension of such theoretical results in infinite dimension is out of reach so far; however, finite-dimensional convergence results are usually sufficient in practice since algorithms are implemented on a computer in finite dimension. The impact of increasing the dimension (by refining the time-monitoring of the processes for instance) is investigated numerically in [ADGL18], where the experiments show that the convergence does not deteriorate as the dimension gets larger.

The rare events statistics are known to be strongly sensitive to the model parameters (see the limit (2.8)). To the best of our knowledge, there are very few contributions on rare event sensitivities, although Basel regulations emphasize more and more the importance of model risks. We refer to [AR99] where such study is handled in the case of compound Poisson process using the score function method coupled with the IS method. Our Gaussian Hilbert framework allows us to elegantly handle the sensitivity analysis of rare event statistics using the tools of Malliavin calculus. We note that in order to derive these results, we do not need any semimartingale models and Itô calculus framework. We will notice (Remark 2.1) that the representation formulas for quantifying the sensitivities in the extremes (see Section 2.1) can be performed at the same time (without extra effort) than the computation of rare events statistics, this is especially attracting.

In POP method, for good numerical performance, it is advantageous to have the conditional probabilities at intermediate levels to be of the same order (for example, in [Lag06] it is argued that the equiprobability choice minimizes the variance of splitting algorithms). However, the appropriate choice of intermediate levels to ensure this condition requires apriori knowledge about the nature of the rare event under consideration. In the absence of such knowledge, choosing appropriate intermediate levels is challenging and requires several pilot algorithm runs to ensure the conditional probabilities requirement. In this work, we propose an adaptive rare event simulation method (complementary to the previous POP methods) where at each intermediate level, except for the last, the conditional probability is fixed to a pre-decided value $p \in (0, 1)$ (typically 10%).

Thus, the intermediate levels are chosen adaptively without making any assumptions on the order of the rare event probability under consideration. We provide the convergence proof for this *adaptive POP method* which requires careful arguments to handle the empirical nature of the intermediate levels. In particular, we use the probabilistic error bounds for the Markov chain based quantile estimators to derive our convergence result.

1.4 Organization

The paper is organized as follows. We define the generic isonormal Gaussian model under study in Section 2. Then we present the splitting principle in Section 2.1: this reduces the rare-event evaluation to computing conditional expectations/probabilities w.r.t. rarer and rarer events. The accounting for sensitivity analysis is performed as well, and we show it fits naturally the splitting principle (Proposition 1). Then, in order to compute the aforementioned conditional expectations, we define the shakers in Subsection 2.2 leading to the POP method. In Section 2.3, we introduce the adaptive rare event simulation method on the path space (adaptive POP method). Section 3 is devoted to applications and experiments in various financial risk problems, namely: model misspecification risk, default probabilities in credit portfolios, estimation of small strike asymptotics in fractional Brownian motion models and parameter sensitivity estimation for deep out-of-the money options. We conclude by summarizing our contributions in Section 4. The proofs are presented in Appendix A.

2 Main Results

We adopt the framework in [Nua06] of an isonormal Gaussian process associated with a general Hilbert space $\mathcal{H}$ (in other words, the framework of Gaussian Hilbert spaces, see [Jan97]). Namely, we assume that $\mathcal{H}$ is a real separable Hilbert space with scalar product $\langle ., . \rangle_{\mathcal{H}}$ and we consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting a centered Gaussian family of scalar random variables $X = (X(h) : h \in \mathcal{H})$ such that

$$\mathbb{E}\left[X(h)X(g)\right] = \langle h, g \rangle_{\mathcal{H}} \quad \text{for any } h, g \in \mathcal{H}.$$

We may refer to X as a path indexed by $h \in \mathcal{H}$. The norm of an element $h \in \mathcal{H}$ is denoted by $\|h\|_{\mathcal{H}}$. The mapping $h \mapsto X(h)$ is linear. Some important examples are as following:

Example 1 (Finite dimensional Gaussian space)**.** *Let $d \in \mathbb{N}^*$, set $\mathcal{H} := \mathbb{R}^d$ and $\langle h, g \rangle_{\mathcal{H}} := \sum_{j=1}^d h_j g_j$ for any $h, g \in \mathbb{R}^d$, denote by $e^i = (\mathbf{1}_{\{j=i\}} : 1 \leq j \leq d)$ the ith element of the canonical basis of $\mathbb{R}^d$. Then $(X(e^1), \ldots, X(e^d))$ is a vector with independent standard Gaussian components.*

Example 2 (Multidimensional Brownian motion (BM))**.** *Let $d \in \mathbb{N}^*$, denote by $\mathcal{H}$ the $\mathbf{L}_2$-space $\mathcal{H} := \mathbf{L}_2(\mathbb{R}^+ \times \{1, \ldots, d\}, \mu)$, where measure μ is the product of Lebesgue measure and the uniform measure which gives mass one to each point $1, \ldots, d$, and set*

$$\langle h, g \rangle_{\mathcal{H}} := \int_{\mathbb{R}^+ \times \{1,\ldots,d\}} h(x)g(x)\mu(\mathrm{d}x) \quad \textit{for any } h, g \in \mathcal{H}.$$

Define

$$X_t^i := X(\mathbf{1}_{[0,t]\times\{i\}}) \quad \textit{for any } t \geq 0, \quad 1 \leq i \leq d.$$

Then the process $(X_t^1, \ldots, X_t^d : t \geq 0)$ is a standard d-dimensional Brownian motion.

Example 3 (Fractional Brownian motion (fBM))**.** *The fBM with Hurst exponent $H \in (0, 1)$ is a $\mathbb{R}$-valued Gaussian process, centered with covariance function*

$$\mathbb{E}(X_t^{(H)} X_s^{(H)}) = \frac{1}{2}\left(t^{2H} + s^{2H} - |t - s|^{2H}\right) := R_H(t, s), \quad \textit{for any } s, t \geq 0.$$

For any fixed $T > 0$, $(X_t^{(H)} : 0 \le t \le T)$ can also be defined within our framework (see [Nua06, Chapter V]). Denote by $\mathcal{H}^0$ the set of step functions on $[0,T]$, and let $\mathcal{H}$ be the Hilbert space defined as the closure of $\mathcal{H}^0$ w.r.t. the scalar product $\langle \mathbf{1}_{[0,t]}, \mathbf{1}_{[0,s]} \rangle_{\mathcal{H}} = R_H(t,s)$. Denote by X the Gaussian process on $\mathcal{H}$ and $(X(\mathbf{1}_{[0,t]}) : 0 \le t \le T)$ defines a fBm $(X_t^{(H)} : 0 \le t \le T)$ with Hurst exponent H.

We can also combine these examples by simultaneously defining, for example, standard BM and fBM. On top of this Gaussian model on $\mathcal{H}$, we can define more sophisticated models frequently used in finance for modeling risk. For the sake of convenience of the reader, here we mention two of them and refer to Section 3 for further developments.

- Local volatility models [Dup94]:

$$\mathrm{d}S_t = b(t, S_t)\mathrm{d}t + \sigma(t, S_t)\mathrm{d}X_t$$

  where X is a standard q-dimensional BM and S stands for the price process of d tradable assets.

- Fractional Brownian Motion (fBM) volatility models [CR98, GJR14]:

$$\mathrm{d}S_t = \mu_t S_t \mathrm{d}t + \sigma_t S_t \mathrm{d}X_t$$

  where S stands for the asset price and the random volatility σ_t is defined through a fractional Brownian motion. To have mean-reverting volatility, we may model σ as a fractional Ornstein-Uhlenbeck process (see, for example, [CKM03, GVZ15]). In Subsection 3.3, we rather consider the fractional SABR model of [GJ14] where the volatility takes the form

$$\sigma_t = \bar{\sigma} \exp\left(-\frac{1}{2}\alpha^2 t^{2H} + \alpha X_t^{(H)}\right), \ t \in [0,T]$$

  where $\bar{\sigma}$ and α are positive parameters, and $X^{(H)}$ is as in Example 3.

2.1 Splitting Method for Rare Event Statistics and Their Sensitivities

In the following, we assume that the probability space at hand $(\Omega, \mathcal{F}, \mathbb{P})$ is such that the σ-field $\mathcal{F}$ is generated by $\{X(h) : h \in \mathcal{H}\}$. For notational simplification, we often identity $\mathcal{H}$ with its orthonormal basis $\mathfrak{b}\mathcal{H} = (\bar{h}_1, \bar{h}_2, \dots)$. To allow great generality, we assume that the rare event is defined through two components, some Rare-event Explanatory Variables (REV) and a level-set function, which are parametrized as follows:

a) We consider random variables taking values in a general metric space $(\mathbf{Z}, \mathcal{Z})$, i.e.

$$Z : \omega \in (\Omega, \mathcal{F}) \mapsto Z(\omega) := \Psi_Z(X(\omega)) \in (\mathbf{Z}, \mathcal{Z}) \tag{2.1}$$

   where Ψ_Z is a measurable mapping from $\mathbb{R}^{\mathcal{H}}$ to $\mathbf{Z}$. The random variable Z stands for REV whose aim is to model the stochasticity of the rare-event.

b) The above REV will be evaluated along a level-set function φ, which completes the definition of the rare event:

$$\varphi : (z, a) \in \mathbf{Z} \times (-\infty, \infty] \mapsto \varphi(z, a) \in [-\infty, \infty).$$

   As we will see, non-positive values of $\varphi(z,a)$ correspond to rare-event scenarios whose probabilities we aim to compute. Furthermore, we assume that for any a, $\varphi(.,a)$ is a measurable map in the first component and that φ is non-increasing w.r.t. the second variable, i.e.

$$\varphi(z,a) \ge \varphi(z,a') \ \text{ for any } -\infty < a \le a' \le \infty \text{ and any } z \in \mathbf{Z}. \tag{2.2}$$

   We take the convention $\varphi(z,\infty) = -\infty$ for any $z \in \mathbf{Z}$. The property (2.2) is crucial for the splitting approach in order to define nested subsets of increasingly rare scenarios (see Equation (2.5) later).

c) The rare event under study is described by the critical paths of Z in set A of the form

$$A := \{z \in \mathbf{Z} : \varphi(z, \bar{a}) \leq 0\}$$

for a given level parameter $\bar{a} \in \mathbb{R}$ such that the probability $\mathbb{P}(Z \in A)$ is small. To avoid a degenerate problem, we assume from now on that $\mathbb{P}(Z \in A) > 0$. To fix ideas, suppose that Z models the capital reserve of a firm which otherwise defaults if its reserve falls below level $\bar{a}$. Then, for $\varphi(z, a) = z - a$, $\mathbb{P}(Z \in A) = \mathbb{P}(Z \leq \bar{a})$ is the default probability for the firm.

d) There is an integrable random variable $\Phi : \Omega \mapsto \mathbb{R}$ modeling the output, for which we wish to evaluate the statistics restricted to the event $\{Z \in A\}$, i.e., to compute

$$\mathbb{E}[\Phi \mathbf{1}_{Z \in A}].$$

We write

$$\Phi : \omega \in (\Omega, \mathcal{F}) \mapsto \Phi(\omega) := \Psi_\Phi(X(\omega)) \in (\mathbb{R}, \mathcal{B}(\mathbb{R})) \tag{2.3}$$

where Ψ_Φ is a measurable mapping from $\mathbb{R}^{\mathcal{H}}$ to $\mathbb{R}$.

Our approach is based on the principle of splitting A in a sequence of $n \geq 2$ nested subsets $(A_k)_{k=1}^n$. Consider the level parameters $\bar{a} := a_n < \cdots < a_k < \cdots < a_0 := \infty$ and set

$$A_k := \{z \in \mathbf{Z} : \varphi(z, a_k) \leq 0\}, \tag{2.4}$$

so that, owing to (2.2) we have

$$A := A_n \subset \cdots \subset A_k \subset \ldots A_0 := \mathbf{Z}. \tag{2.5}$$

Note that for describing a given rare-event A, there are many possible couples (level set function φ, level set parameter $\bar{a}$). The choice made by the user has an impact on the performance of the methods (see the example on credit-risk in Subsection 3.2) and it is made according to the knowledge of the model at hand. Later, we will often refer to $(a_k)_{k=1}^n$ as acceptance level parameters. The choice of acceptance levels can be done adaptively which we discuss in Section 2.3. It is clear that the above splitting approach justifies the following decompositions:

$$\mathbb{E}[\Phi \mathbf{1}_{Z \in A}] = \mathbb{E}[\Phi \mathbf{1}_{Z \in A} \mid Z \in A_{n-1}] \prod_{k=1}^{n-1} \mathbb{P}(Z \in A_k \mid Z \in A_{k-1}) \tag{2.6}$$

$$= \mathbb{E}[\Phi \mid Z \in A_n] \prod_{k=1}^{n} \mathbb{P}(Z \in A_k \mid Z \in A_{k-1}). \tag{2.7}$$

Next, we assume that the model at hand depends on a real-valued parameter θ, through the definition of Z and Φ so that $\mathbb{E}[\Phi \mathbf{1}_{Z \in A}]$ should now be written as $\mathbb{E}\left[\Phi^\theta \mathbf{1}_{Z^\theta \in A}\right]$. The sensitivity of the above quantity w.r.t. θ is an important issue to account for because the errors in model calibration and estimation procedures could have a significant impact. This concerns the evaluation of model risk (see, for example, [Con06]). This question is even more delicate when combined with rare-event analysis since it is known that the distribution tails are very sensitive to parameter shocks [AS12]. For example, if $G_\sigma \overset{\mathrm{d}}{=} \mathcal{N}(0, \sigma^2)$ then

$$\lim_{x \to \infty} \frac{\mathbb{P}(G_\sigma \geq x)}{\mathbb{P}(G_{\sigma'} \geq x)} = \begin{cases} 0 & \text{if } 0 < \sigma < \sigma' \\ \infty & \text{if } \sigma > \sigma' > 0 \end{cases}, \tag{2.8}$$

that is a small change of parameters may cause a large change of tail-probabilities.

To quantify the impact of θ on $\mathbb{E}\left[\Phi^\theta \mathbf{1}_{Z^\theta \in A}\right]$, we may evaluate the derivative w.r.t. θ whenever it exists. However, in the rare event setting, as the above expectation is small, most likely its derivative will also be

small. Thus, it makes more sense to evaluate the relative sensitivity defined by

$$\frac{\partial_\theta \mathbb{E}\left[\Phi^\theta \mathbf{1}_{Z^\theta \in A}\right]}{\mathbb{E}\left[\Phi^\theta \mathbf{1}_{Z^\theta \in A}\right]}$$

provided that $\mathbb{E}\left[\Phi^\theta \mathbf{1}_{Z^\theta \in A}\right]$ is differentiable in θ and non zero.

Regarding the computational aspects, the derivative $\partial_\theta \mathbb{E}\left[\Phi^\theta \mathbf{1}_{Z^\theta \in A}\right]$ can be estimated by the resimulation method as follows: Take two values of θ which are close to each other, approximate expectation for each value of θ by Monte-Carlo simulations and form the finite difference as an estimator of the derivative. This is known to be not well suited to the case where the functional inside the expectation is irregular in θ which is typically our case because of the indicator function. A better strategy is to represent the derivative as an expectation (known as the likelihood method in the case of explicit distributions, or based on Integration-By-Parts formula in the Malliavin calculus setting [FLL+99]) and then evaluate it by simulations. This is our approach which we formulate as an assumption.

(IBP) There exists an open set $\Theta \subset \mathbb{R}$ such that $\theta \mapsto \mathbb{E}\left[\Phi^\theta \mathbf{1}_{Z^\theta \in A}\right]$ is differentiable on Θ and for any $\theta \in \Theta$, there is an integrable random variable $\mathcal{I}(Z^\theta, \Phi^\theta)$ such that

$$\partial_\theta \mathbb{E}\left[\Phi^\theta \mathbf{1}_{Z^\theta \in A}\right] = \mathbb{E}\left[\mathcal{I}(Z^\theta, \Phi^\theta)\mathbf{1}_{Z^\theta \in A}\right].$$

Combining this with the splitting approach of Equation (2.7) gives a simple representation of the relative sensitivity.

Proposition 1. *Assume* **(IBP)**. *For any $\theta \in \Theta$ such that $\mathbb{E}\left[\Phi^\theta \mathbf{1}_{Z^\theta \in A}\right] \neq 0$, we have*

$$\frac{\partial_\theta \mathbb{E}\left[\Phi^\theta \mathbf{1}_{Z^\theta \in A}\right]}{\mathbb{E}\left[\Phi^\theta \mathbf{1}_{Z^\theta \in A}\right]} = \frac{\mathbb{E}\left[\mathcal{I}(Z^\theta, \Phi^\theta) \mid Z^\theta \in A\right]}{\mathbb{E}\left[\Phi^\theta \mid Z^\theta \in A\right]}.$$

When we are concerned by the sensitivity of the rare-event probability, it takes the simple form

$$\partial_\theta \left[\log\left(\mathbb{P}\left(Z^\theta \in A\right)\right)\right] := \frac{\partial_\theta \mathbb{P}\left(Z^\theta \in A\right)}{\mathbb{P}\left(Z^\theta \in A\right)} = \mathbb{E}\left[\mathcal{I}(Z^\theta, 1) \mid Z^\theta \in A\right].$$

Remark 2.1. *Observe (for computational perspectives) that the above relative sensitivities take the same form as the first term in (2.7), i.e. a ratio of expectations conditionally to the last level A. Therefore, evaluating $\mathbb{E}\left[\Phi \mathbf{1}_{Z \in A}\right]$ or its relative sensitivity w.r.t. a parameter θ will be achieved with the same computational cost (provided that $\mathcal{I}(Z^\theta, \Phi^\theta)$ is known).*

In full generality on the probabilistic setting, the determination of $\mathcal{I}(Z^\theta, \Phi^\theta)$ is difficult but in our Gaussian noise setting, it can be achieved using the Integration by Parts formula of Malliavin calculus. There are numerous situations where one can obtain such a representation for sensitivities (see [FLL+99, Gob04, GM05, KY09] among others, and [Nua06, Section 6.2] for more references). We establish such a result in the case Z^θ takes values in $\mathbb{R}^d$, and Z^θ, Φ^θ are smooth in θ. Hereafter, we adopt and follow the notation of [Nua06] for the derivative operator D, for the space $\mathbf{D}^{1,2}$ of random variables that are one time Malliavin differentiable with $\mathbf{L}_2$-integrability, and for the divergence operator δ. We say that a family of random variables $(U^\theta : \theta \in \Theta)$ is in $\mathbf{L}_p^{loc}$ $(p \geq 1)$ if for any $\theta \in \Theta$, there is a open set $V_\theta \subset \Theta$ containing θ such that $\sup_{\theta' \in V_\theta} |U^{\theta'}|$ is bounded by a random variable in $\mathbf{L}_p$.

Theorem 2. *Consider $\mathbf{Z} = \mathbb{R}^d$ and let $q > d$. Assume the following conditions:*

(a) *$(\Phi^\theta, \theta \in \Theta)$ is in $\mathbf{L}_2^{loc}$ and Z^θ has a $\mathbf{L}_q$-norm bounded locally uniformly in θ;*

(b) *Φ^θ and Z^θ are continuous and differentiable on Θ and their derivatives $(\dot{\Phi}^\theta, \dot{Z}^\theta : \theta \in \Theta)$ are respectively in $\mathbf{L}_1^{loc}$ and $\mathbf{L}_2^{loc}$;*

(c) *for any $\theta \in \Theta$, $Z^\theta \in \mathbf{D}^{1,2}$ and the Malliavin covariance matrix $\gamma_{Z^\theta} := (\langle D.Z_i^\theta, D.Z_j^\theta \rangle_{\mathcal{H}})_{1 \leq i,j \leq d}$ is invertible a.s.;*

(d) *for any* $\theta \in \Theta$, $\Phi^\theta \sum_{j=1}^d (\gamma_{Z^\theta}^{-1} \dot{Z}^\theta)_j D.Z_j^\theta$ *is in the domain of* δ *and* $\dot{\Phi}^\theta + \delta(\Phi^\theta \sum_{j=1}^d (\gamma_{Z^\theta}^{-1} \dot{Z}^\theta)_j D.Z_j^\theta)$ *has a* $\mathbf{L}_2$*-norm bounded locally uniformly in* θ*;*

(e) *for any* $\theta \in \Theta$ *and any* $i \in \{1, \ldots, d\}$, $\sum_{j=1}^d (\gamma_{Z^\theta}^{-1})_{j,i} D.Z_j^\theta$ *is in the domain of* δ *and* $\delta(\sum_{j=1}^d (\gamma_{Z^\theta}^{-1})_{j,i} D.Z_j^\theta)$ *has a* $\mathbf{L}_q$*-norm bounded locally uniformly in* θ*.*

Then **(IBP)** *is satisfied on* Θ *and*

$$\mathcal{I}(Z^\theta, \Phi^\theta) := \dot{\Phi}^\theta + \delta\left(\Phi^\theta \sum_{j=1}^d (\gamma_{Z^\theta}^{-1} \dot{Z}^\theta)_j D.Z_j^\theta\right).$$

The above theorem together with Proposition 1 and Remark 2.1 provides an important contribution in the literature as the issue of rare event sensitivity has received little attention. Using the result in Theorem 2, analytical sensitivity formulas can be derived in several model settings driven by Gaussian noise. We provide an example in Section 3.4. The proof of Theorem 2 is presented in Appendix A.1.

2.2 Reversible Gaussian Transformations and Parallel-One-Path (POP) Method

In this section, we recall the idea of reversible transformations proposed in [GL15]. We suppose that the state values of X that serve to model the REV Z in (2.1) lie in the path space $\mathbb{R}^{\mathcal{H}}$. Let $\rho := (\rho_h : h \in \mathfrak{b}\mathcal{H}) \in [-1, 1]^{\mathfrak{b}\mathcal{H}}$. Then, the reversible transformations on the Gaussian path X are defined as below:

$$K : \begin{cases} \mathbb{R}^{\mathcal{H}} \times \mathbb{R}^{\mathcal{H}} & \mapsto \mathbb{R}^{\mathcal{H}} \\ (x, x') := (x_h : h \in \mathfrak{b}\mathcal{H},\ x'_h : h \in \mathfrak{b}\mathcal{H}) & \to (\rho_h x_h + \sqrt{1 - \rho_h^2} x'_h : h \in \mathfrak{b}\mathcal{H}). \end{cases} \tag{2.9}$$

Whenever useful, we will write K_ρ to insist on the dependence on the so-called shaking parameter ρ. If $X' = (X'(h) : h \in \mathcal{H})$ is an independent copy of X, we simply denote by $\mathcal{K}$ the random transformation from $\mathbb{R}^{\mathcal{H}} \mapsto \mathbb{R}^{\mathcal{H}}$ as

$$\mathcal{K}(x) = K(x, X').$$

In the stochastic analysis literature, the above parametrized transformation for a constant parameter $\rho_h = \text{constant} \in (0, 1)$ is associated to the Ornstein-Uhlenbeck (or Mehler) semigroup (see [Nua06, Section 1.4]) and simply writes

$$K(x, x') = \rho x + \sqrt{1 - \rho^2} x', \tag{2.10}$$

independently of the choice of the basis $\mathfrak{b}\mathcal{H}$. We call the transformation (2.9) *shaker* and it satisfies the following reversibility property.

Proposition 3 (Reversible shaker)**.** *The following identity holds in distribution:*

$$(X, K(X, X')) \overset{\mathrm{d}}{=} (K(X, X'), X).$$

This type of reversibility property is well-known in the Markov Chain Monte Carlo literature when studying the convergence of Markov chains in large time. Thus, the shaker (2.9) preserves the distribution of X (seen now as a stationary measure) and by iterating the transformations and averaging out the outputs in time, we may obtain a numerical evaluation of related expectations (Birkhoff Law of Large Numbers) which is justified by the following result.

Theorem 4. *Let* $f : \mathbf{Z} \mapsto \mathbb{R}$ *be a measurable function and assume that* $f(Z) \in \mathbf{L}_2$ *where* $Z = \Psi_Z(X)$ *as in* (2.1)*. Define* $X_0 = X, X_{k+1} = K_\rho(X_k, X'_k)$ *and* $Z_k = \Psi_Z(X_k)$ *where the* X'_k *are independent copies of* X*. Then, for* $|\rho|_\infty := \sup_{h \in \mathfrak{b}\mathcal{H}} |\rho_h| < 1$*,*

$$\left| \frac{1}{N} \sum_{k=1}^N f(Z_k) - \mathbb{E}\left[f(Z)\right] \right|_{\mathbf{L}_2}^2 \leq \frac{\mathbb{V}\mathrm{ar}\left(f(Z)\right)}{N} \left(\frac{1 + |\rho|_\infty}{1 - |\rho|_\infty} \right), \quad \forall N \geq 1. \tag{2.11}$$

The proof for the above $\mathbf{L}_2$ convergence result of the Gaussian shaker is given in Appendix A.2.

Our adaptive method is based on shaking *with rejection* transformations on the path space which also form the basis of Parallel-One-Path (POP) method [GL15]. We define the shaking *with rejection* transformation as follows:

$$M_k^K : \begin{cases} \mathbb{R}^{\mathcal{H}} \times \mathbb{R}^{\mathcal{H}} \to \mathbb{R}^{\mathcal{H}}, \\ (x, x') \mapsto K(x, x')\mathbf{1}_{\Psi_Z(K(x,x'))\in A_k} + x\mathbf{1}_{\Psi_Z(K(x,x'))\notin A_k}, \end{cases} \tag{2.12}$$

where (Ψ_Φ, Ψ_Z) are measurable mappings which transform the path X to the rare-event model (Φ, Z) (defined in (2.1) and (2.3)). Recall the definitions of A_k in Equation (2.4). Intuitively, the above transformation states that for starting state $x \in \mathbb{R}^{\mathcal{H}}$ and innovation $x' \in \mathbb{R}^{\mathcal{H}}$, we apply the transformation and keep it as the next state if $\Psi_Z\big(K(x, x')\big) \in A_k$ for $x' \in \mathbb{R}^{\mathcal{H}}$. Otherwise, we restart from the current state x. Furthermore, we define $\mathcal{M}_k^{\mathcal{K}}(.) := M_k^K(., X')$ where X' is the generic isonormal Gaussian path. Due to the reversibility property (Proposition 3), it immediately follows that the conditional distribution $X \mid \Psi_Z(X) \in A_k$ is invariant with respect to the shaking with rejection $\mathcal{M}_k^{\mathcal{K}}(\cdot)$.

Proposition 5. *Let $k \in \{0, 1, \cdots, n-1\}$. The distribution of X conditionally on $\{\Psi_Z(X) \in A_k\}$ is invariant w.r.t. the random transformation $\mathcal{M}_k^{\mathcal{K}}$, i.e. for any bounded measurable $\varphi : \mathbb{R}^{\mathcal{H}} \to \mathbb{R}$ we have*

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

Next, we define a Markov chain based on the transformation $\mathcal{M}_k^{\mathcal{K}}$ which has the above conditional distribution as invariant measure.

Definition 1. *For each $k = 0, \ldots, n-1$, given a starting point $X_{k,0} \in \Psi_Z^{-1}(A_k)$, define*

$$X_{k,i} := \mathcal{M}_k^{\mathcal{K}}(X_{k,i-1}) = M_k^K(X_{k,i-1}, X'_{k,i-1}) \quad \text{for } i \geq 1 \tag{2.13}$$

where $(X'_{k,i})_{i\geq 0}$ is a sequence of independent copies of X' and independent of $X_{k,0}$.

If the above defined Markov chain is ergodic, we have the following approximation which is justified by Theorems 4 and 8:

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

Taking $\varphi(x) = \mathbf{1}_{A_{k+1}}(\Psi_Z(x))$ and $\varphi(x) = \Psi_\Phi(x)$ yields an approximation of $\mathbb{P}\left(\Psi_Z(X) \in A_{k+1} \mid \Psi_Z(X) \in A_k\right)$ and $\mathbb{E}\left[\Psi_\Phi(X) \mid \Psi_Z(X) \in A_k\right]$; therefore owing to (2.6), the product of all the estimators gives an estimation of the rare event probability $\mathbb{P}\left(\Psi_Z(X) \in A\right)$ and the rare event statistics $\mathbb{E}\left[\Psi_\Phi(X)\mathbf{1}_{\Psi_Z(X)\in A}\right]$.
Furthermore, POP method gives a way to automatically initialize each Markov chain and compute the estimator using shaking with rejection transformation:

Step 1. Take $X_{0,0}$ as a copy of X.

Step 2. For $k = 0$, simulate the Markov chain $X_{k,i} = M_k^K(X_{k,i-1}, X'_{k,i-1})$, $0 \leq i \leq N-1$ and calculate

$$p_k^{(N)} = \frac{1}{N}\sum_{k=0}^{N-1} \mathbf{1}_{A_{k+1}}(\Psi_Z(X_{k,i})).$$

Step 3. Set $i_k = \arg\min\{j : \Psi_Z(X_{k,j}) \in A_{k+1}\}$ and $X_{k+1,0} = X_{k,i_k}$. Repeat **Step 2** for $k = 1, \ldots, n-1$.

Output. Compute the probability estimate as

$$p^{(N)} = \prod_{k=0}^{n-1} p_k^{(N)}. \tag{2.14}$$

The convergence of POP method has been proved in [GL15] under the assumptions which are not easily verified. Here, we provide the proof of convergence in finite dimensions without any assumption. The proof is given using a classical result that the occupation measure of an irreducible and stationary Markov chain will converge to its invariant measure. A short proof for such result using the ergodic theorem and martingale techniques can be found in [AG11]. The proof of our result is relegated to Appendix A.3.

Proposition 6. *POP method with $|\rho|_\infty < 1$ converges almost surely in the finite dimensional case $\mathcal{H} = \mathbb{R}^d$.*

In addition, the marginal distribution of each Markov chain converges to its limit, see Appendix A.3 for the proof.

Proposition 7. *Assume the finite dimensional case for $\mathcal{H}$ and $|\rho|_\infty < 1$. For any fixed $k \in \{0, \dots, n-1\}$, denote by $\mathcal{L}(X_{k,N}^{x_{k,0}})$ the law of $X_{k,N}^{x_{k,0}}$ with initialization at a given point $X_{k,0} = x_{k,0} \in \Psi_Z^{-1}(A_k)$, and denote the distribution of X conditionally on $\{\Psi_Z(X) \in A_k\}$ by π_k. Then, for any $x_{k,0} \in \Psi_Z^{-1}(A_k)$ we have*

$$\|\mathcal{L}(X_{k,N}^{x_{k,0}}) - \pi_k\|_{\mathtt{TV}} \to 0$$

as $N \to \infty$, where $\|\cdot\|_{\mathtt{TV}}$ denotes the total variation norm.

Remark 2.2. *The convergence of marginal distributions may have an interesting practical use for stress-testing financial systems. For example, let X denote the financial random environment that a banking system faces, and Z denote the related risk exposure. In order to test the system resilience, regulators usually design some stress test scenario which means imposing a presumably rare event in A on the banking system and then see how the system reacts to this event. Some references on the design of stress test can be found in [Bas17]. In most stress testing designs, regulators artificially construct one or a few elements in A. Using the POP method and Proposition 7, one can sample approximately according to the conditional distributions $X|Z \in A$ and/or $Z|Z \in A$, which gives a more relevant choice of stress test scenarios. Such application of POP method has been explored in [ADGL18].*

Apart from the convergence of POP method stated in Propositions 6 and 7, we are also interested in its error estimates and its speed of convergence. For this purpose, in finite dimensional case, we derive the required results. We will see later that the speed of convergence result will aid us to prove the consistency of adaptive estimator based on shaking with rejection transformations used in POP method.

Theorem 8. *Let the level index $k \geq 0$ be given. Consider the kth Markov chain $(X_{k,i} : i \geq 0)$ with shaking and rejection in $\Psi_Z^{-1}(A_k)$ defined in (2.13), and restricted to the finite d-dimensional case (see Example 1). Assume that $|\rho|_\infty < 1$ and*

$$\sup_{x:\Psi_Z(x)\in A_k} \mathbb{P}\left(K(x, X') \notin A_k\right) = \delta_1 < 1. \tag{2.15}$$

Let $s > 0, q \geq 2$ and set $V(x) := \exp(s\sum_{j=1}^d |x_j|)$. Then, there exist a constant $C > 0$ and a geometric rate $r \in (0,1)$ such that for any measurable function $g : \mathbb{R}^d \to \mathbb{R}$ bounded in $V^{1/q}$-norm (i.e. $\sup_{x\in\mathbb{R}^d} \frac{|g(x)|}{V^{1/q}(x)} < +\infty$) and any initial condition $X_{k,0}$ independent from the future evolution of the chain with $\mathbb{E}[V(X_{k,0})] < +\infty$, we have

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^n g(X_{k,i}) - \mathbb{E}\left[g(X) \mid \Psi_Z(X) \in A_k\right]\right|^q\right] \leq Cn^{-q/2}, \tag{2.16}$$

$$\left|\mathbb{E}\left[g(X_{k,n})\right] - \mathbb{E}\left[g(X) \mid \Psi_Z(X) \in A_k\right]\right| \leq Cr^n, \tag{2.17}$$

for any $n \geq 1$.

See Appendix A.4 for the proof. To get convergence rates (of geometric type as shown in (2.17)), the acceptance rate in the POP algorithm is required to be bounded from below, uniformly in the current point in $\Psi_Z^{-1}(A_k)$, this is the condition (2.15). This condition is related to the geometry of the rare sets, it is application-dependent and it may be difficult to check in practice.

From Theorem 8, it is easy to deduce that $p^{(N)}$ (defined in (2.14)) converges to $\mathbb{P}(Z \in A)$ at rate $\sqrt{N}$, in L_p-norms, details are left to the reader.

2.3 Adaptive Rare Event Simulation Method on Path Space

As mentioned earlier, in the absence of apriori knowledge about the order of the rare event probability, choosing the number of intermediate levels is challenging. To overcome this difficulty, we propose an adaptive algorithm based on the reversible shaking transformation with rejection. For the ease of exposition, let us suppose that Z takes values in $\mathbf{Z} = \mathbb{R}^d$ and that the rare event set is of the form $A = \{z \in \mathbb{R}^d : \varphi(z) \leq \bar{a}\}$ where $\varphi : \mathbb{R}^d \to \mathbb{R}$ is a measurable function and $\bar{a}$ is a given finite threshold. The principle of the adapted algorithm is to set

$$A_k := \{z \in \mathbb{R}^d : \varphi(z) \leq a_k\} \tag{2.18}$$

with online computations of the acceptance level a_k. Notice that this choice of A_k corresponds to the notations (2.4)-(2.5), when defining $\varphi(z, a) := \varphi(z) - a$. To have constant conditional probabilities (as suggested by the optimality results of [Lag06]), we take a_k as the quantile of $V := \varphi(\Psi_Z(X)) = \varphi(Z)$ at level p^k for some fixed $0 < p < 1$. This heuristics guides the following notation and definition.

We denote the p-quantile of the distribution of V as

$$Q_p^1 = F_V^{-1}(p) := \inf\{v \in \mathbb{R} : F_V(v) \geq p\}$$

where $F_V(\cdot)$ is the cumulative distribution function of V. We define the conditional quantile function $g_p(\cdot)$ of V in the following way:

$$g_p(s) := \inf\{v \in \mathbb{R} : \mathbb{P}(V \leq v \mid V \leq s) \geq p\}$$

and also recursively define

$$Q_p^{l+1} := g_p(Q_p^l), l \geq 1.$$

The above formula remains valid for $l = 0$ by setting $Q_p^0 := \infty$. Here, Q_p^k is synonymous with a_k. This is our convention from now on. Moreover, we define

$$r(s) := \mathbb{P}(V \leq \bar{a} \mid V \leq s),$$

then the true rare event probability $\alpha = \mathbb{P}(V \in A)$ can be written in a unique way as

$$\alpha = r(Q_p^{L^*})p^{L^*} \tag{2.19}$$

where $L^* \in \mathbb{N}$ and $r(Q_p^{L^*}) \in (p, 1]$. We are now in a position to define our algorithm with adaptive number of intermediate levels (approximation of L^*):

Initialization. We are given a common initialization point x_0 such that $\varphi(\Psi_Z(x_0)) \leq \bar{a}$. Note that such x_0 can always be obtained by repeatedly using shaking with rejection transformation.

1st Markov chain. Simulate the first N iterations of the Markov chain based on Equation (2.13) with starting state x_0. Then, sort the sample $(V_{N,1}^1, \ldots, V_{N,N}^1)$ in ascending order as $V_{N,(1)}^1 \leq \ldots \leq V_{N,(k)}^1 \leq \ldots \leq V_{N,(N)}^1$ and take $k_p^1 \in \{1, \ldots, N\}$ such that $k_p^1 - 1 < Np \leq k_p^1$. Denote by $\hat{Q}_{N,p}^1 = V_{N,(k_p^1)}^1$, the estimate for Q_p^1 based on N samples.

2nd Markov chain. Start the Markov chain in Equation (2.13) with initial state x_0 and with cascade event set A_1 corresponding to level $\hat{a}_1 := \hat{Q}_{N,p}^1$ (see (2.18)). Again, simulate the first N steps and sort the sample $(V_{N,1}^2, \ldots, V_{N,N}^2)$ in ascending order as $V_{N,(1)}^2 \leq \ldots \leq V_{N,(k)}^2 \leq \ldots \leq V_{N,(N)}^2$. Take k_p^2 such that $k_p^2 - 1 < Np \leq k_p^2$ and denote by $\hat{a}_2 := \hat{Q}_{N,p}^2 = V_{N,(k_p^2)}^2$, the estimate for Q_p^2 based on N samples.

Iteration and stopping. Next, repeat the procedure until the (L_N+1)th step where we have $\hat{Q}_{N,p}^{L_N+1} \leq \bar{a}$ for the first time. The intermediate sets A_k in (2.18) are defined by the acceptance levels $\hat{a}_k = \hat{Q}_{N,p}^k$. Calculate $\hat{r}_N(\hat{Q}_{N,p}^{L_N})$, defined as the proportion of values $(V_{N,1}^{L_N+1}, \ldots, V_{N,N}^{L_N+1})$ which are smaller than $\bar{a}$ with the cascade event set corresponding to acceptance level $\hat{Q}_{N,p}^{L_N}$.

In the case $L_N = 0$, we set by convention $\hat{Q}_{N,p}^0 = \infty$ (similarly to Q_p^0).

Output. Compute the probability estimate as

$$\hat{\alpha}_N := \hat{r}_N(\hat{Q}_{N,p}^{L_N}) p^{L_N}$$

as an approximation of the probability α written in (2.19).

Remark 2.3. *In the following Theorem 9, we assume that the initial points of the above Markov chains are fixed (actually all equal to x_0). The deterministic initialization of Markov chain at each level, indeed, helps to simplify the convergence analysis. However in practice, we could advantageously start the l-th level Markov chain from a point close to the acceptance level, i.e. $X_{l,0}$ equal to the x-configuration of one of the $V_{N,(1)}^l, \ldots, V_{N,(k_p^l)}^l$. The choice of $V_{N,(1)}^l$ is the simplest from algorithmic viewpoint, since we only need to update the smallest $V_{N,i}^l$ (with the corresponding X) during the algorithm run. Besides, we observe only a very small impact of initialization on the numerical results.*

Remark 2.4. *Unlike standard POP method, adaptive algorithm computes each conditional expectation sequentially and thus, it is not amenable to parallelization. In [ADGL18], we propose another variant which combines the adaptive algorithm with POP method to give a partially parallelizable method.*

In order to prove the consistency of estimator $\hat{\alpha}_N$, we make the following assumptions and provide their justification later.

Assumption 1. *The distribution of V admits a density $s \mapsto f(s)$, which is continuous and strictly positive at $s = Q_p^l$ for all $l \in \{1, \ldots, L^*\}$.*

Assumption 2. *For any $q \in (-\infty, \infty]$, let $\hat{g}_{N,p}(s)$ denote the quantile estimator for $g_p(s)$ based on N iterations of the Markov chain based on the rejection set $\{x : \varphi(\Psi_Z(x)) > s\}$. For any $l \in \{0, \ldots, L^*+1\}$, there exists an open interval I_l containing Q_p^l (with the convention $Q_p^0 = \infty$ and $I_0 = \{\infty\}$) and a function $b : I_l \times \mathbb{N}^* \times (0, \infty) \to [0, \infty)$ such that for all $\varepsilon > 0$ and $q \in I_l$*

$$\mathbb{P}\left(|\hat{g}_{N,p}(s) - g_p(s)| > \varepsilon\right) \leq b(s, N, \varepsilon), \text{ and, } \sup_{s \in I_l} b(s, N, \varepsilon) \to 0 \text{ as } N \to \infty.$$

Assumption 3. *For any $s \in (-\infty, \infty]$, let $\hat{r}_N(s)$ denote the mean estimator for $r(s)$ based on N iterations of the Markov chain based on the rejection set $\{x : \varphi(\Psi_Z(x)) > s\}$. For any $l \in \{L^*-1, L^*\} \cap \mathbb{N}$, there exists an open interval J_l containing Q_p^l (with the convention $J_0 = \{\infty\}$) and a function $c : J_l \times \mathbb{N}^* \times (0, \infty) \to [0, \infty)$ such that for all $\varepsilon > 0$ and $s \in J_l$*

$$\mathbb{P}\left(|\hat{r}_N(s) - r(s)| > \varepsilon\right) \leq c(s, N, \varepsilon), \text{ and, } \sup_{s \in J_l} c(s, N, \varepsilon) \to 0 \text{ as } N \to \infty.$$

Theorem 9. *Suppose that Assumptions 1, 2 and 3 hold. Then, $\hat{\alpha}_N$ converges in probability to $\alpha = \mathbb{P}(Z \in A)$ as $N \to \infty$.*

See Appendix A.5 for the proof. Assumption 1 is required for the continuity of $g_p(\cdot)$ and $r(\cdot)$ at quantile levels Q_p^l and is typically assumed to prove convergence of quantile estimators. For example, in [DFJN14], such a condition is required to exhibit probabilistic error bounds for MCMC based quantile estimators. Assumptions 2 and 3 are related to decaying probabilistic error bounds for MCMC based quantile estimators. Such inequalities have been shown to hold in the case of uniformly (see, for example, [GO02, Theorem 2], [DFJN14, Theorem 3], [KLM05, Theorem 1]), geometrically or high order polynomially (see for example, [DFJN14, Theorem 2]) ergodic Markov chains when the starting point of the underlying Markov chain is distributed with the stationary distribution. In Theorem 8, we show that the Markov chain based on shaking

with rejection transformation is geometrically ergodic and thus Assumptions 2 and 3 can be justified. Further note that the above mentioned error bounds decay uniformly to zero as $N \to \infty$ whereas we only require them to decay uniformly over an interval. Even though, we always initialize the Markov chain at hand at a fixed point x_0, we believe that these assumptions are still reasonable because the marginal distribution of Markov chain converges to the stationary distribution (see Proposition 7 and Theorem 8).

3 Applications

In this section, we discuss the application of adaptive rare event algorithm and POP method to various important problems in mathematical finance. The adaptive implementation of classical IPS method has also been presented in [CMFG12, Section 3]) which we generalize by using shaking with rejection transformations to provide an analogue to our adaptive method. The modified version of IPS method has been presented in [GL15]. In order to avoid different terminology, we simply refer to the two adaptive schemes as *adaptive POP* and *adaptive IPS* in this section. In all the following examples, to simplify the exposure and experiments, we assume that the shaking parameter ρ_h is constant w.r.t. h (see (2.9) and (2.10)). We denote this common value[1] by $\rho \in (-1, 1)$.

3.1 Model Misspecification and Robustness

To address the issue of model risk, we consider the Profit&Loss (PL) when the trader uses a Black-Scholes (BS) model to hedge a European call option while the true dynamics of the underlying S is given by a path-dependent volatility model. Let us suppose that there are two volatility levels $\sigma_-, \sigma_+ \in \mathbb{R}_+ \backslash \{0\}$ such that $\sigma_- < \sigma_+$. We propose a discrete-time path-dependent volatility model based on a monitoring period Δ_t (say 1 week) and monitoring dates $t_i = i\Delta_t$, wherein, if the underlying spot price drops below the average of previous four monitored prices, the level of volatility becomes σ_+, otherwise it remains constant at σ_-. The asset price is given as

$$S_t = S_0 \exp\big(-\frac{1}{2}\sigma_-^2 t + \sigma_- W_t\big), \quad t < t_4, \tag{3.1}$$

$$S_t = \begin{cases} S_{t_i} \exp\big(-\frac{1}{2}\sigma_-^2 (t - t_i) + \sigma_-(W_t - W_{t_i})\big) \\ \qquad \text{if } S_{t_i} \geq \frac{1}{4}\sum_{k=1}^4 S_{t_{i-k}} \text{ and } t_i \leq t < t_{i+1}, \\ S_{t_i} \exp\big(-\frac{1}{2}\sigma_+^2 (t - t_i) + \sigma_+(W_t - W_{t_i})\big) \\ \qquad \text{if } S_{t_i} < \frac{1}{4}\sum_{k=1}^4 S_{t_{i-k}} \text{ and } t_i \leq t < t_{i+1}, \end{cases} \qquad \text{when } t \geq t_4.$$

This model corresponds to the usual empirical observation that the underlying volatility is higher when price falls. This is a discrete version of the continuous time model proposed by [Guy14]. Furthermore, we assume the risk-free interest rate to be zero. The resulting model is complete in the sense that any square integrable payoff written on S can be replicated by a self-financing strategy (see [HR98] for complete models with stochastic volatility). The above model is directly written under the risk-neutral measure $\mathbb{P}$.

Meanwhile, we assume that the trader uses a BS model in which the volatility is constant and equal to σ_-. The call option maturity is $T > 0$, and $[0, T]$ is the trading period under consideration. As the trader assumes a BS model, she uses the BS formula to perform delta hedging. For our numerical study, we take $T = 1, n = 50$ Δ_t is s.t. $n\Delta_t = T$) and assume that the trader makes a rebalancing after every period of

[1] The choice of non-constant ρ_h may be guided by the apriori knowledge that some Gaussian components of X may have more or less importance on the rare event occurrence. We do not elaborate further in this direction and leave it for future research.

$5\Delta_t$. At times $t_{j5\Delta_t}, 0 \leq j < 10$, the trader holds δ_j assets, so at the maturity her PL is given by

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

where δ_j is given from the BS-Delta formula with volatility σ_- and spot $S_{5\Delta_t j}$.

Since the realized volatility is higher than the one used for hedging, the trader may underhedge the option (in continuous time hedging, see [EJS98] for precise results) and may incur large losses due to the model risk. Thus, we wish to estimate the probability $\mathbb{P}\left(PL_{\text{trader}} \leq L\right)$. In the model of (3.1), we set $S_0 = K_{\text{strk}} = 10$, $\sigma_- = 0.2$, $\sigma_+ = 0.27$ and take $L = -2.4$. In IPS and POP methods, we deliberately choose 5 intermediate levels as $L_k := kL/5, k = 1, 2, 3, 4, 5$. This choice is made after a set of preliminary runs to ascertain the order of the rare event probability under consideration. As we will see, this choice provides numerical estimates which are close to the estimates from the adaptive algorithms.

The crude Monte Carlo method with 5×10^8 simulations provides a 99% confidence interval for this probability as $[2.93, 3.34] \times 10^{-6}$ with run time of approximately 4000 minutes on a machine with Intel i5-4670 3.40 GHz CPU processor and 16GB memory. The mean estimates and empirical standard deviations of IPS and POP methods and their respective adaptive algorithms using 100 macro-runs are given in Tables 3.1.1-3.1.2. The run time of a single iteration of both IPS and POP methods with fixed shaking parameter is approximately 470 seconds. The adaptive algorithms are performed with parameter $p = 10\%$ for the intermediate conditional probability. From Table 3.1.1, it is clear that POP based estimators provide accurate estimates with a lower standard deviation than IPS based estimators, both schemes being in their non-adaptive versions. In Table 3.1.2, results with adaptive algorithms are compared, here again, the adaptive version corresponding to the POP method yields smaller variances in the estimation.

When comparing standard deviations of Table 3.1.1 and Table 3.1.2, we observe that they are similar (for a given shaker parameter ρ). The reader may think that seemingly adaptive versions do not provide any benefit. One should recall that, the non-adaptive versions require a priori choices of levels (here we choose them by preliminary experiments to have roughly equal conditional probabilities) while with the adaptive version we do not need this kind of a priori knowledge and still we obtain efficient estimators. Actually the advantage really stems from the fully adaptive tuning of levels, which is made possible without deteriorating the variances.

In Figure 3.1.1 (top), we investigate the dependence of the standard deviation (of each conditional probability computed with POP method) w.r.t. the shaking parameter ρ and the level l. We do not report results for $l = 1$ (no rejection) since independent sampling ($\rho = 0$) is obviously the best. We observe that the impact of ρ on the variance is significant: the optimal parameter ρ_l^* minimizing the variance changes from one level to another and ρ_l^* increases with l (the shaking has to become slighter with increasing rarity of the event). These features are easily explained heuristically. Complementary to this, we plot in Figure 3.1.1 (bottom) the rejection rate, which also depends on ρ and l. It appears that ρ_l^* depends very much on l whereas the associated rejection rate remains rather stable and ranges from 60% to 80%. Since we observe a quick explosion of standard deviation when ρ is too close to 1, we recommend to take ρ such that the rejection rate is above 60% rather than below 60%, to be on the safe side when a finer optimization of ρ is not possible. We shall take it as a rule of thumb for further experiments.

Lastly, in Figure 3.1.2 we report statistics on standard deviation and rejection rate for the adaptive algorithm corresponding to POP method. We observe similar features as in Figure 3.1.1. Since different intermediate levels in the adaptive algorithm are correlated, we first run the beginning level to get corresponding value of ρ minimizing the standard deviation. Then, we use this fixed value for the corresponding level in the search of optimal ρ of the next level and so on. We can see in Figure 3.1.2 that with all the values of ρ chosen in this way, the best standard deviation among the final estimators is around 1.5×10^{-7}, which is about 62.5% of standard deviation of the estimator with a constant $\rho = 0.9$ for all the levels.

	IPS			POP		
	mean ($\times 10^{-6}$)	std. ($\times 10^{-7}$)	std./mean	mean ($\times 10^{-6}$)	std. ($\times 10^{-7}$)	std./mean
$\rho = 0.9$	3.10	5.29	0.17	3.13	2.07	0.07
$\rho = 0.7$	3.23	13.3	0.41	3.11	3.98	0.13
$\rho = 0.5$	2.79	25.9	0.93	3.18	8.44	0.27

Table 3.1.1: *Estimators of* $\mathbb{P}(PL_{trader} \leq L)$ *(mean) for* $L = -2.4$ *with empirical standard deviation (std.) for non-adaptive IPS and POP methods based on 100 algorithm macro-runs. Each intermediate level estimator in both methods is based on* $M = N = 10^5$ *simulations.*

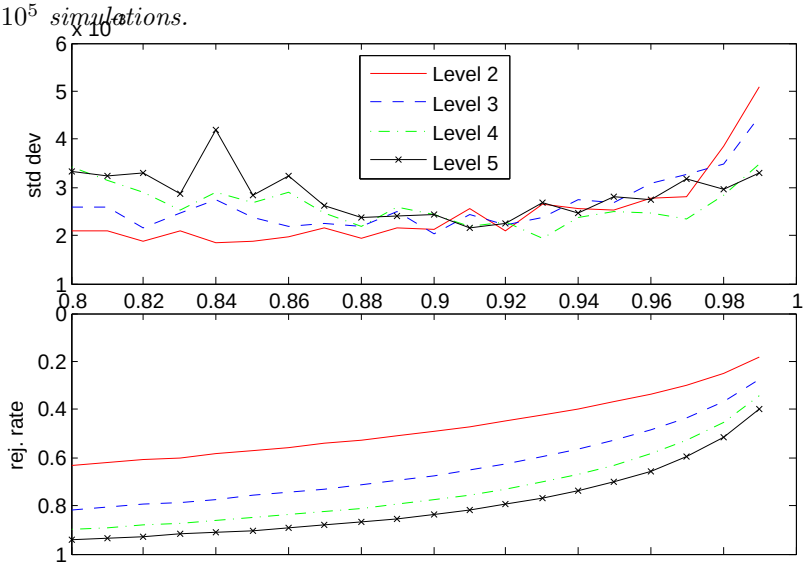


Figure 3.1.1: *POP method, standard deviation (std. dev.) of each conditional probability estimator and corresponding rejection rate (rej. rate), based on 100 macro-runs, for different values of* ρ.

3.2 Measuring Default Probabilities in Credit Portfolios

In this subsection, we consider a credit portfolio based on asset values of N_0 different firms. Let us suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space where $\{W_1, W_2, \ldots, W_{N_0}, W\}$ are $\mathbb{P}$-standard Brownian motions with constant correlations. We denote by $\{\mathcal{F}_t, t \geq 0\}$ the $\mathbb{P}$-augmentation of the filtration generated by $\{W_1, W_2, \ldots, W_{N_0}, W\}$. As in [CFV09], we assume that the dynamics of asset values is given by the following system of stochastic differential equations

$$\mathrm{d}S_i(t) = rS_i(t)\mathrm{d}t + \sigma(t)S_i(t)\mathrm{d}W_i(t), \quad i = 1, \ldots, N_0, \tag{3.2}$$

where r is the risk-free interest rate, the common stochastic volatility factor $\sigma(t)$ is modeled by a Cox-Ingersoll-Ross model satisfying

$$\mathrm{d}\sigma(t) = \kappa\big(\bar{\sigma} - \sigma(t)\big)\mathrm{d}t + \gamma\sqrt{\sigma(t)}\mathrm{d}W_t, \tag{3.3}$$

where $\kappa, \bar{\sigma}$ and γ are positive constants. Brownian motions are correlated as follows:

$$\mathrm{d}\langle W_i, W_j\rangle_t = \rho^W \mathrm{d}t, i \neq j, \quad \mathrm{d}\langle W_i, W\rangle_t = \rho^\sigma \mathrm{d}t, \quad i = 1, \ldots, N_0.$$

Next, we consider the default boundary for each firm i to be a fixed value $B_i \in \mathbb{R}_+$. The time of default for firm i in the portfolio is defined as

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

The current methods would directly adapt to the case where the default level B_i is replaced by a time-dependent deterministic function.

	Adaptive IPS			Adaptive POP		
	mean ($\times 10^{-6}$)	std. ($\times 10^{-7}$)	std./mean	mean ($\times 10^{-6}$)	std. ($\times 10^{-7}$)	std./mean
$\rho = 0.9$	3.06	4.95	0.16	3.18	2.42	0.08
$\rho = 0.7$	2.98	11.1	0.37	3.10	3.71	0.12
$\rho = 0.5$	2.45	23.6	0.96	3.06	7.27	0.24

Table 3.1.2: *Estimators of $\mathbb{P}(PL_{trader} \leq L)$ (mean) for $L = -2.4$ with empirical standard deviation (std.) for adaptive IPS and POP methods ($p = 10\%$) based on 100 algorithm macro-runs. Each intermediate level estimator in both methods is based on $M = N = 10^5$ simulations.*

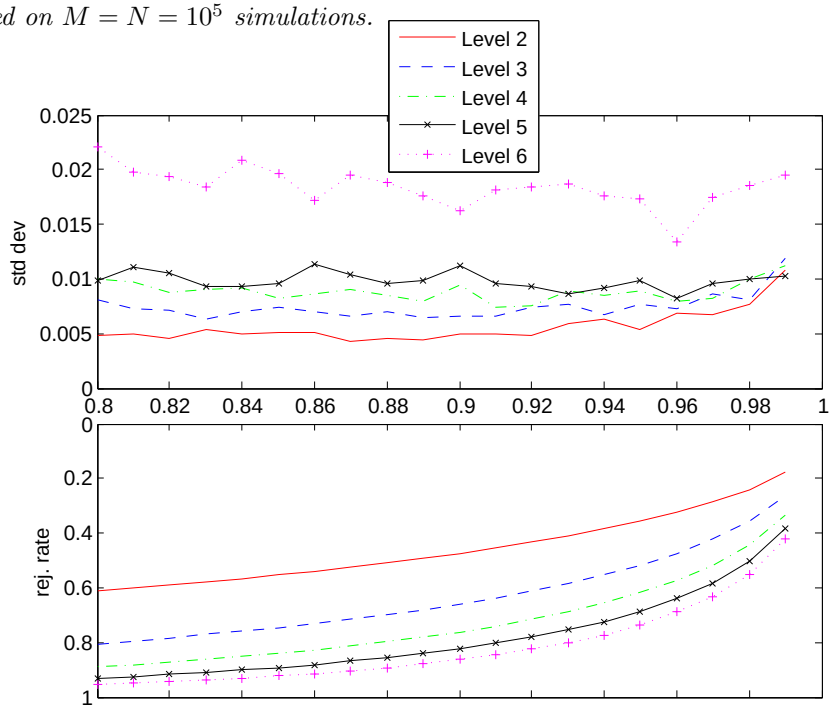


Figure 3.1.2: *Adaptive POP method ($p = 10\%$). Standard deviation (std dev.) and corresponding rejection rate (rej. rate), based on 100 macro-runs, of each quantile estimator $(\hat{Q}^l_{N,p})_{1 \leq l \leq L^*-1}$ and last level occupation measure estimator $\hat{r}_N(\hat{Q}^{L_N}_{N,p})$, for different values of ρ.*

In order to evaluate different tranches in a credit portfolio, we are interested to calculate the probability that at least L defaults occur before T, i.e.

$$\mathrm{P}(L) = \mathbb{P}\left(\sum_{i=1}^{N_0} \mathbf{1}_{\{\tau_i(B_i) \leq T\}} > L\right) = \mathbb{P}\left(\sum_{i=1}^{N_0} \mathbf{1}_{\{\min_t S_i(t) \leq B_i\}} > L\right), \quad 0 < L < N_0.$$

Due to the path dependency of the default scheme and of the stochastic volatility model, it is not clear how to find the optimal change of measure to perform importance sampling to estimate $\mathrm{P}(L)$, which motivates the use of alternative simulation techniques.

A different IPS-based method has been proposed by Carmona et al. [CFV09] in order to compute $\mathrm{P}(L)$ (see also [CC10] for application of this method in other models). We would like to emphasize the main difference between the former IPS approach and our work. The underlying Markov chain for their IPS method is simply the time-discretization of the $(2N_0 + 1)$-dimensional process $(S_i, \min S_i, \sigma, 1 \leq i \leq N_0)$. This poses several difficulties for the authors. Firstly, one needs to exhibit a *good* potential function for the selection of particles which is very delicate because of the high-dimensionality of the problem. Secondly, one needs to choose an appropriate discretization time step Δ_t. This is also intricate, since on the one hand a large number of time steps may help in better selection of the particles in rarer and rarer regions, but on the other hand it slows down the statistical convergence of IPS (the resampling adds noise in the estimation). In our case, we directly consider Markov chains valued on path space, thus avoiding the delicate problem of

choosing the time step Δ_t and the high-dimensional potential function (in our numerical experiments, we have observed that Δ_t has no significant impact on the convergence of our versions of IPS-POP methods when it is small enough. This feature of our algorithms has been investigated in more details in [ADGL18, Section 3.1]). Thus, our approach and results are rather different from those of Carmona et al. [CFV09]. These differences are mainly due to the fact that our method does not require any Markovian assumption on $(S_i, \min S_i, \sigma, 1 \le i \le N_0)$ and could be directly applied to path-dependent models (whenever useful).

In order to express $\mathrm{P}(L)$ in the form of (2.6), we need to create a cascade of decreasing sets $\{A_k\}_{1\le k\le n}$. We define $Z \in \mathbb{R}^{N_0}$ whose i-th component is the minimum of $(S_i(t)/S_i(0))_t$ and we set

$$A_k := \left\{ z \in \mathbb{R}^{N_0} : \sum_{i=1}^{N_0} \mathbf{1}_{\{z_i \times (B_i + \frac{k}{n}(S_i(0) - B_i) \le B_i\}} > L \right\}, \quad 1 \le k \le n,$$

which consists in progressively decreasing the default trigger levels. The nested set condition (2.5) is then fulfilled. Then, we apply POP and IPS methods to compute all the conditional probabilities $\mathbb{P}\left(Z \in A_{k+1} \mid Z \in A_k\right)$.

Remark 3.1. *Another natural way to create the nested sequence of sets is to progressively increase the number of defaults:*

$$\tilde{A}_k := \left\{ z \in \mathbb{R}^{N_0} : \sum_{i=1}^{N_0} \mathbf{1}_{\{z_i S_i(0) \le B_i\}} > \frac{k}{n} L \right\}, \quad 1 \le k \le n.$$

We empirically observe that this choice is in general less accurate. Although we have proven that POP method will eventually converge in all the finite-dimensional cases, how to construct intermediate sets to achieve the best convergence rate remains to be explored.

To perform numerical experiments in the considered model of (3.2)-(3.3), we fix the parameter values as in Table 3.2.1. Further, we fix the total number of firms $N_0 = 125$ and threshold level $B_i = B$ for some

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

Table 3.2.1: *Parameters for credit portfolio model*

$B > 0$. Next, we estimate the default probability $\mathrm{P}(L)$ for different values of L over $T = 1$ with 50 time steps per year in the Euler discretization scheme of Deelstra and Delbaen [DD98]. For $L = 100$ and $B = 36$, the crude Monte Carlo estimator of default probability with 3×10^9 sample paths has a 99% confidence interval as $[4.92, 5.13] \times 10^{-6}$ with run time of approximately 15000 minutes. In Table 3.2.2, we report the results for IPS and POP based estimators for fixed $n = 5$ levels where both the algorithms have run time of approximately 45 seconds for a fixed value of ρ. For different values of the shaking parameter ρ, it is clear that POP based estimator provides more accurate results than IPS method. In Figure 3.2.3, using POP based estimator with fixed number of levels $n = 20$ and 10^4 simulations at each level, we also report $\mathrm{P}(L)$ for different levels of default threshold B based on different values of L. Remarkably, it allows to compute very low probabilities (up to 10^{-24}).

Next, we implement adaptive version corresponding to POP and IPS methods. To estimate $\mathrm{P}(L)$, we fix the conditional probability $\mathbb{P}\left(Z \in A_{k+1} \mid Z \in A_k\right)$ of each, except the last, intermediate level (to be estimated) to $p = 10^{-1}$. In Table 3.2.3, we can see that both IPS and POP based estimates are within the reported confidence interval of the true value for $\rho = 0.9$. However, the corresponding POP based estimator has a lower standard deviation. When comparing with Table 3.2.2, variances are roughly unchanged by using the adaptive scheme, but the advantage of this version is to have a fully simulation-based scheme where we do not need to pre-specify the acceptance threshold levels.

In Figures 3.2.1 and 3.2.2, we report different detailed statistics w.r.t. the level and the shaking parameter (non-adaptive POP method: standard deviation and rejection rate; adaptive POP method: standard

deviation of quantile and occupation measure along with rejection rate). We observe similar behaviors as in the first example of Subsection 3.1. For rare regions (levels $l = 3, 4, 5$), the parameter ρ_l^* minimizing the standard deviation of the l-th conditional probability seems to be associated to rejection rate of 70%. We believe that this (so far empirical) invariance relation between best shaking parameters (for minimal variances) and rejection rate of about $70\% - 80\%$ should give a way to adaptively choose ρ. This will be further investigated in the future. Again we see that for the adaptive POP method, with different values of ρ minimizing standard deviation in each intermediate level, the standard deviation of the final adaptive estimator of the rare event probability is about 60% of that with a constant $\rho = 0.9$.

The above methodology can also be applied directly to better account for the systemic risk and the illiquidity issues, for example, in the settings of [FI14] where inter-bank lending is modeled by a system of coupled diffusion processes in a mean-field regime.

	IPS			POP		
	mean ($\times 10^{-6}$)	std. ($\times 10^{-6}$)	std./mean	mean ($\times 10^{-6}$)	std. ($\times 10^{-6}$)	std./mean
$\rho = 0.9$	5.82	4.37	0.75	5.01	0.80	0.16
$\rho = 0.7$	4.92	1.56	0.32	4.99	1.02	0.20
$\rho = 0.5$	4.79	3.80	0.79	5.02	1.94	0.39

Table 3.2.2: *Estimators of default probability (mean) for $L = 100$ and $B = 36$ with empirical standard deviation (std.) for non-adaptive IPS and POP methods based on 100 algorithm macro-runs. Each intermediate level estimator in both methods is based on $M = N = 10^4$ simulations.*

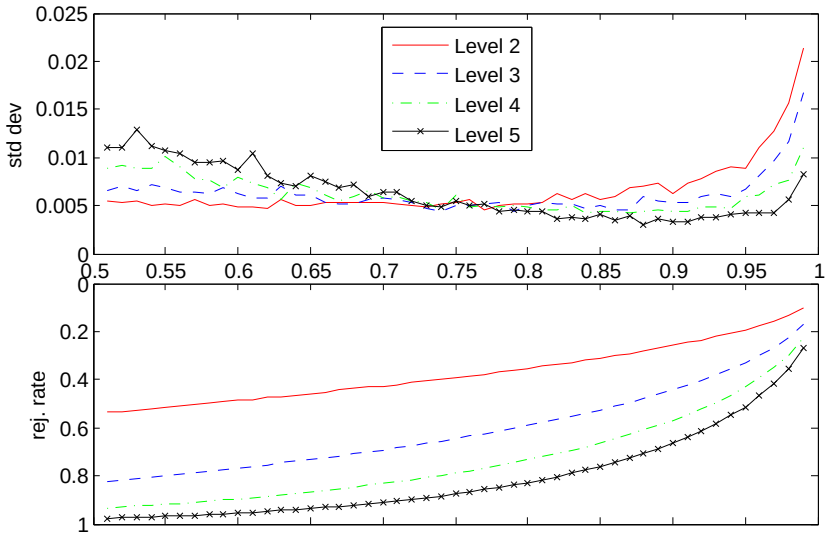


Figure 3.2.1: *POP method, standard deviation (std. dev.) of each conditional probability estimator and corresponding rejection rate (rej. rate), based on 100 macro-runs, for different values of ρ.*

	Adaptive IPS			Adaptive POP		
	mean ($\times 10^{-6}$)	std. ($\times 10^{-6}$)	std./mean	mean ($\times 10^{-6}$)	std. ($\times 10^{-6}$)	std./mean
$\rho = 0.9$	4.93	1.91	0.39	5.16	0.85	0.16
$\rho = 0.7$	5.42	1.58	0.29	4.98	1.02	0.20
$\rho = 0.5$	6.40	5.00	0.78	5.35	2.05	0.38

Table 3.2.3: *Estimators of default probability (mean) for $L = 100$ and $B = 36$ with empirical standard deviation (std.) for adaptive IPS and POP methods ($p = 10\%$) based on 100 algorithm macro-runs. Each intermediate level estimator in both methods is based on $M = N = 10^4$ simulations.*

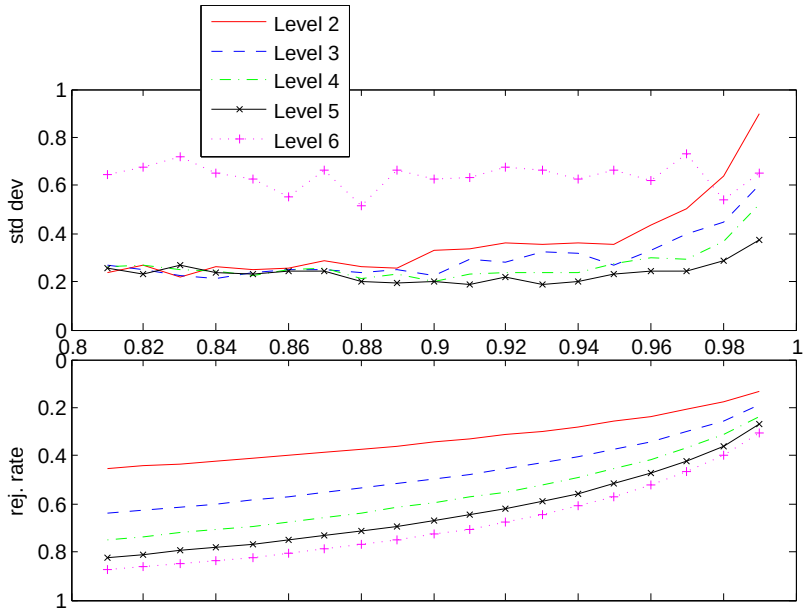
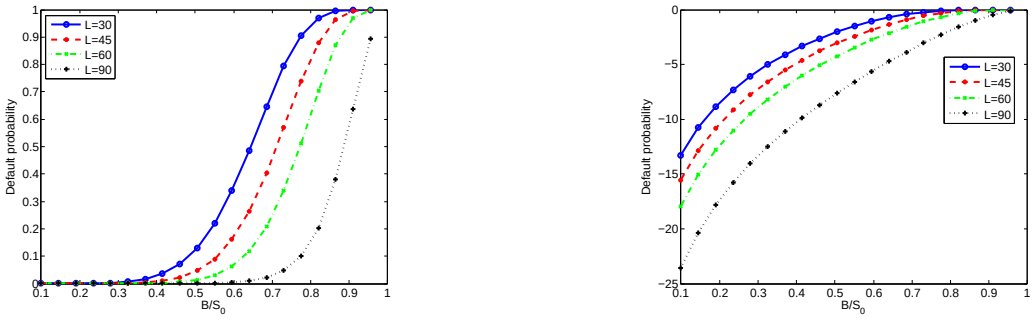


Figure 3.2.2: *Adaptive POP method* ($p = 10\%$). *Standard deviation (std. dev.) and corresponding rejection rate (rej. rate), based on 100 macro-runs, of each quantile estimator* $(\hat{Q}^l_{N,p})_{1\le l\le L^*-1}$ *and last level occupation measure estimator* $\hat{r}_N(\hat{Q}^{L_N}_{N,p})$, *for different values of* ρ. *The std. dev. of occupation measure estimator has been scaled by 10 for easier comparison.*



(a) (b)

Figure 3.2.3: *Plot (a) and log-plot (b) of default probabilities for varying* B/S_0.

3.3 Fractional Brownian Motion for Modelling Volatility

The fractional Brownian motion (fBM) $(B^{(H)}_t)_{t\in\mathbb{R}}$with Hurst exponent $H \in (0,1)$ was defined in Example 3. For $H \neq 1/2$, it is well known that $B^{(H)}$ is not a semimartingale. In order to represent fBM, we make use of the Mandelbrot and van Ness representation of $B^{(H)}$ as an integral w.r.t. a standard Brownian motion B:

$$B^{(H)}_t = C_H\Big[\int_{-\infty}^{t}\big[(t-s)^{H-\frac{1}{2}} - (-s)_+^{H-\frac{1}{2}}\big]\mathrm{d}B_s \quad \text{with} \quad C_H = \sqrt{\frac{2H\Gamma(3/2-H)}{\Gamma(H+1/2)\Gamma(2-2H)}}.$$

Recently, Gatheral and co-authors [GJR14] have successfully employed fBM to model the market observed volatility of stock indexes. In addition, classical Heston model [Hes93] has also been redefined with fBM in order to evaluate derivative prices more accurately (see, for example, [ER16, ER17]). In order to demonstrate the application of POP method for models which are not necessarily based on semimartingales, we consider the fractional SABR (fSABR) model proposed by Gatheral et al. [GJ14]. In fSABR, the underlying asset

	$\alpha = 0.5$				$\alpha = 1.0$			
ρ^{BZ}	$H = 0.15$	$H = 0.25$	$H = 0.75$	$H = 0.9$	$H = 0.15$	$H = 0.25$	$H = 0.75$	$H = 0.9$
-0.3	2.6133	2.5515	2.8058	2.9753	0.8251	0.8267	0.9211	0.9632
-0.5	2.4222	2.3823	2.6733	2.8715	0.7905	0.7913	0.8950	0.9449
-0.7	2.2593	2.2042	2.5465	2.7918	0.7597	0.7591	0.8686	0.9277
-0.9	2.1235	2.0653	2.4339	2.6919	0.7325	0.7297	0.8449	0.9113

Table 3.3.1: *Estimates of critical negative moment $\tilde{q}$ in fSABR model (3.4) using POP method.*

dynamics are given by

$$\frac{\mathrm{d}S_t}{S_t} = \sigma_t \mathrm{d}Z_t, \qquad \sigma_t = \bar{\sigma} \exp\left(-\frac{1}{2}\alpha^2 t^{2H} + \alpha B_t^{(H)}\right), \tag{3.4}$$

where Z_t is a standard Brownian motions with instantaneous correlation ρ^{BZ} with B_t (i.e. $\mathrm{d}\langle B, Z\rangle_t = \rho^{BZ}\mathrm{d}t$). Under the model (3.4), we use POP method to estimate the small-strike tail asymptotic slope of implied variance

$$\beta_L := \limsup_{x \to -\infty} \frac{I^2(x)T}{|x|} \tag{3.5}$$

where $I(x)$ is the BS implied volatility of a Vanilla option on S with log-moneyness $x = \log K/S_0$ and maturity T. The estimate of the slope can be, in turn, used to obtain estimate of the critical negative moment $\tilde{q} := \sup\{q : \mathbb{E}[S_T^{-q}] < \infty\}$ from the well-known moment formula [Lee04, Theorem 3.4]

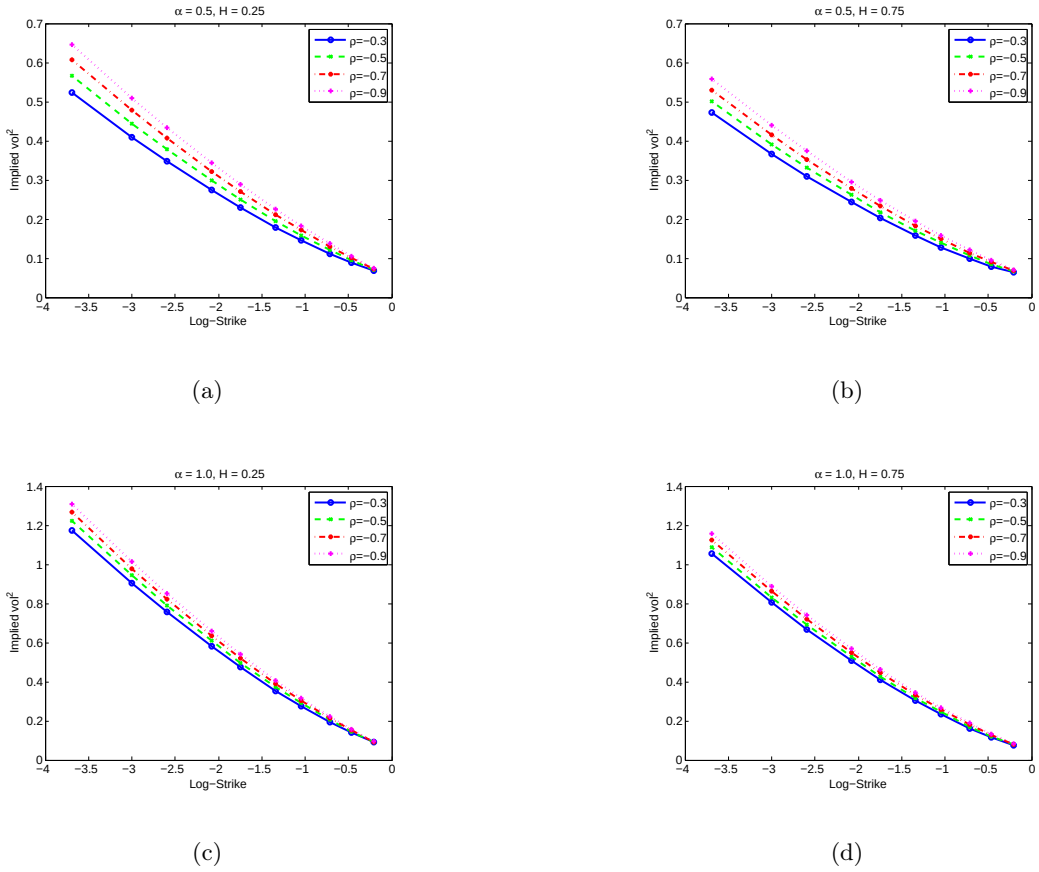
$$\tilde{q} = 1/2\beta_L + \beta_L/8 - 1/2. \tag{3.6}$$

We work with the following parameter values: $S_0 = 40, \bar{\sigma} = 0.235, r = 0, T = 1.0$ and $\alpha = 0.5, 1.0$. We use intermediate levels at $[32.5, 25, 19.5, 14, 10.5, 7, 5, 3, 2, 1]$ in the POP method (shaking parameter value $= 0.9$) with 10^5 simulations[2] at each level in order to estimate the implied volatility at different values of the log-moneyness. The output values are based on 100 independent algorithm macro-runs. We observe on Figure 3.3.1 that the squared implied volatilities seemingly behave linearly for large negative values of the log-moneyness, which suggests that the limit superior in (3.5) is presumably a limit (see Remark 3.2 below for a related discussion).

In light of (3.5), we could use the most extreme value of the implied variance $I^2(x_{\min})$ (corresponding to $x_{\min} = -3.75$ in Figure 3.3.1) in order to evaluate β_L. Instead of doing so, we compute the slope β_L by linear interpolation of the two most extreme implied variances $I^2(x_{\min})$ and $I^2(x_{\min} + \Delta x)$. We observe that following one or the other procedure has no significant impact on the results. This yields the estimates of $\tilde{q}$ in Table 3.3.1. From our numerical results, we can observe that $\tilde{q}$ increases with the value of the correlation ρ^{BZ} in the model. Conversely, $\tilde{q}$ decreases with the value of the parameter α. There is no global monotonicity appearing from the relationship between $\tilde{q}$ and value of $H \in (0, 1)$. On the other hand, one does see (as expected) the emergence of two different regimes for $H < 1/2$ and $H > 1/2$. These observations suggest that it is possible - at least in theory - to calibrate the value of one of these model parameters from extreme implied volatility estimates, for example by using POP method. Moreover, the plots in 3.3.1 indicate a 'tilting' effect of the correlation parameter ρ^{BZ} on the whole smile curve, analogous to that in standard stochastic volatility models based on Brownian motion. This indicates that under the fractional model (3.4), too, the appropriate value of the correlation parameter can be reasonably inferred from market implied volatilities by observing the slopes of the left- and right hand sides of the smile.

Remark 3.2. *While the formulas (3.5)-(3.6) always hold when β_L is defined via a limit superior, it is interesting to notice that there is a (large) class of models for which the limit superior can actually be updated to a true limit, thus providing the full asymptotic equivalence $I^2(x)T \sim |x|$ as $x \to -\infty$. This*

[2]We exactly simulate the skeleton of Z, B and B^H (with a step length of $T/100$) as a correlated Gaussian vector since the covariance matrix of this vector can be computed explicitly.



(a) (b)

(c) (d)

Figure 3.3.1: *Squared implied volatility as a function of log-strike in the fSABR model* (3.4).

class is fully characterized in Gulisashvili [Gul12, Theorem 3.5]. Recall that a positive measurable function f defined on some neighborhood of infinity is said to be regularly varying with index $\alpha \in \mathbb{R}$ if for every $\lambda > 0$, $\frac{f(\lambda x)}{f(x)} \to \lambda^\alpha$ as $x \to \infty$. Furthermore, the class of Pareto-type functions is introduced in [Gul12]. Let g be positive measurable functions defined on $(0, c)$ for some $c > 0$: if there exist two functions g_1 and g_2 that are regularly varying with index α and such that $g_1(x^{-1}) \leq g(x) \leq g_2(x^{-1})$ for all $0 < x < c$, then we say that the function g is of weak Pareto-type near zero with index α.
Gulisashvili [Gul12] proves the following: under the assumption $0 < \tilde{q} < \infty$, the asymptotic formula

$$\lim_{x \to -\infty} \frac{I^2(x)T}{|x|} = \beta_L \tag{3.7}$$

holds if and only if the following condition is satisfied:

i) The put price function $P(K) = \mathbb{E}[(K - S_T)^+]$, $K > 0$, is of weak Pareto type near zero with index $\alpha_1 = -\tilde{q} - 1$.

It is possible to relate the property i) in a more direct way to the distribution of the stock price: condition i) holds for the put price if one of the following two conditions is satisfied:

ii) The cdf of the stock price $F(K) = \mathbb{P}(S_T \leq K)$ is of weak Pareto type near zero with index $\alpha_2 = -\tilde{q}$.

iii) The density $p_T(\cdot)$ of the stock price S_T (if it exists) is of weak Pareto type near zero with index $\alpha_3 = -\tilde{q} + 1$.

The implication iii) $\Rightarrow$ i) is proven in [Gul12], Theorem 3.11. The implication ii) $\Rightarrow$ i) can be proven following the lines of the proofs of Theorems 3.11 and 3.7 in [Gul12].

Figure 3.3.1 suggests that squared implied volatilities behave asymptotically linearly with log-moneyness, and we can therefore conjecture that equation (3.7) holds for the fSABR model (3.4). An analysis of the cdf or the density function of the stock price, as performed in [GVZ15] for a class of models with Gaussian self-similar stochastic volatility, would allow to show that properties ii) and iii) hold true in the fSABR model. We leave such kind of investigation for future research.

3.4 Estimating Sensitivities for Out-of-the-Money Options

In this example, we consider a d-dimensional Black-Scholes model in which the asset price vector $S = (S^1, S^2, \dots, S^d)$ is given as

$$\frac{\mathrm{d}S_t^i}{S_t^i} = \mu^i \mathrm{d}t + \sigma^i \mathrm{d}(LW_t)^i \tag{3.8}$$

where $\sigma^i > 0$ for all $i = 1, \dots, d$, W is an d-dimensional standard Brownian motion, and L is the symmetric square root of a d-dimensional correlation matrix C, so that $LL^* = C$ (here $L = L^*$). Hereafter we assume that the matrix C (therefore L) is invertible. Denoting by Z^i the log of S^i, one has

$$Z_T^i = Z_0^i + \left(\mu^i - \frac{1}{2}(\sigma^i)^2\right) T + \sigma^i (LW_T)^i \tag{3.9}$$

with $Z_0^i = \log(S_0^i)$. Equation (3.8) allows to model separately the individual volatility σ^i of each asset and the correlation between the driving Brownian factors. The introduction of a volatility smile on each asset can be achieved simply by switching from constant to local volatility functions $\sigma^i(t, \cdot)$ (which can be separately calibrated to option data on each asset).

We consider a digital-style payoff written on a generalized basket, whose financial evaluation is defined by

$$\mathcal{P} := \mathbb{P}\left(\varphi(Z_T, \bar{a}) \geq 0\right)$$

where

$$\varphi(z, \bar{a}) := \sum_{i=1}^{d} \varepsilon_i p_i e^{z^i} - \bar{a} \tag{3.10}$$

with $p_i > 0, \varepsilon_i \in \{-1, 1\}$ and $\bar{a} \in \mathbb{R}$. This setting can cover the situation of risk management of an insurance contract (when each asset evolves with its own drift coefficient μ^i), and of course the pricing of a digital option on the basket, which corresponds to set $\mu^i = r$, where r is a risk-free interest rate. We are interested in computing the sensitivities of $\mathcal{P}$ with respect to different model parameters, such as

- p_i in order to assess the influence of the individual weights, possibly in order to reweight the portfolio and lower the risk,
- σ^i in order to quantify the impact of individual volatilities on the tails of the basket,
- $C_{i,j} = (LL^*)_{i,j}$ for $i < j$, in order to study the effect of pair-wise correlations on the product.

In order to obtain explicit sensitivity formulas, we apply Theorem 2 with $\Phi^\theta = 1$, $Z^\theta = Z_T$ in the context of multidimensional Brownian motion (Example 2), where θ plays the role of one the model parameters or payoff parameters above. A direct computation shows

$$(D_t Z_T)_{i,j} = \sigma^i L_{i,j} \mathbf{1}_{t \leq T} = \Sigma_{i,j} \mathbf{1}_{t \leq T} \qquad \gamma_{Z_T} = T\, \Sigma\Sigma^*,$$

where we denote Σ the matrix $\Sigma = \mathrm{diag}(\sigma) L$, where $\mathrm{diag}(\sigma)_{i,j} = \sigma^i \delta_{i,j}$. Under our assumption, the matrices Σ and γ_{Z_T} are invertible.

In what follows, we denote $A_{i,\cdot}$ (respectively $A_{\cdot,i}$) the i-th row (respectively i-th column) of the matrix A.

Sensitivity w.r.t. p_i. In view of (3.10) we have $\partial_{p_i}\mathcal{P} = \partial_{Z_0^i}\mathcal{P}\frac{1}{p_i}$ and it suffices to compute sensitivity w.r.t. $\theta = Z_0^i$. Clearly $\partial_{Z_0^i} Z_T = e^i$ where e^i is the i-th element of the canonical basis of $\mathbb{R}^d$, therefore the weight $\mathcal{I}(Z^\theta, 1)$ in Theorem 2 becomes

$$\begin{aligned}\mathcal{I}(Z^\theta,1) &= \delta\left(\sum_{j=1}^d \left(\gamma_{Z_T}^{-1}\partial_{Z_0^i} Z_T\right)_j D.Z_T^j\right)\\ &= \delta\left(\sum_{j=1}^d (\gamma_{Z_T}^{-1})_{j,i}(\Sigma \mathbf{1}_{[0,T]})_{j,.}\right) = \delta\left(\sum_{j=1}^d (\gamma_{Z_T}^{-1})_{i,j}(\Sigma \mathbf{1}_{[0,T]})_{j,.}\right)\\ &= \frac{1}{T}\delta\left((\Sigma\Sigma^*)^{-1}\Sigma \mathbf{1}_{[0,T]})_{i,.}\right) = \frac{1}{T}\delta\left(((\Sigma^*)^{-1})_{i,.}\mathbf{1}_{[0,T]}\right) = \frac{1}{T}\Sigma^{-1}e^i \cdot W_T.\end{aligned}$$

The computation of the sensitivities with respect to σ^i and $C_{i,j}$ involves quantities of the form $\delta((AW_T)^i \times u^* \mathbf{1}_{[0,T]}(\cdot))$, where A is a $d \times d$ matrix and u a (constant) vector in $\mathbb{R}^d$. We will therefore make use of the following formula

$$\delta((AW_T)^i u^* \mathbf{1}_{[0,T]}(\cdot)) = (AW_T)^i u \cdot W_T - T\,(Au)^i \quad . \tag{3.11}$$

Equation (3.11) can be proven using the identity $\delta(F\,U.) = F\delta(U.) - \langle DF, U\rangle$ which holds for $U \in dom(\delta)$ and $F \in \mathbf{D}^{1,2}$, where we denote $\langle V, U\rangle = \sum_{j=1}^d \int_0^T V_t^j U_t^j \mathrm{d}t$.

Sensitivity w.r.t. $\theta = \sigma^i$. We have $\partial_{\sigma^i} Z_T = (-\sigma^i T + (LW_T)^i)e^i$. Since $\partial_{\sigma^i} Z_T$ and e^i are collinear, the computations are very similar to the previous ones, and we obtain

$$\begin{aligned}\mathcal{I}(Z^\theta,1) &= \delta\left(\sum_{j=1}^d \left(\gamma_{Z_T}^{-1}\partial_{\sigma^i} Z_T\right)_j D.Z_T^j\right)\\ &= \frac{1}{T}\delta\left((-\sigma^i T + (LW_T)^i)((\Sigma^*)^{-1})_{i,.}\mathbf{1}_{[0,T]}\right)\\ &= -\sigma^i\delta\left(((\Sigma^*)^{-1})_{i,.}\mathbf{1}_{[0,T]}\right) + \frac{1}{T}\delta\left((LW_T)^i((\Sigma^*)^{-1})_{i,.}\mathbf{1}_{[0,T]}\right)\\ &= \Sigma^{-1}e^i \cdot W_T\left(-\sigma^i + \frac{1}{T}(LW_T)^i\right) - (L\Sigma^{-1})_{i,i}\end{aligned}$$

where we have applied the identity (3.11) with $A = L$ and $u^* = ((\Sigma^*)^{-1})_{i,.}$ in the last step.

Sensitivity w.r.t. $\theta = C_{i,j}$, $i < j$. We wish to take partial derivatives of functions defined on the set of correlation matrices $\mathcal{C} = \left\{(C_{i,j})_{i,j} : C \in \mathcal{S}_{\geq 0}^d, C_{i,i} = 1, C \text{ invertible}\right\}$ with respect to each of the entries $C_{i,j}$, $i < j$, where $\mathcal{S}_{\geq 0}^d$ denotes the set of symmetric and positive matrices. This is possible under the invertibility assumption because, given a matrix $C \in \mathcal{C}$ and fixed $i < j$, the whole set $\{C_\varepsilon := C + \varepsilon e^{i,j} + \varepsilon e^{j,i}, \varepsilon \in \mathbb{R}\}$ is contained in $\mathcal{C}$ for ε small enough, where $e^{i,j}$ denotes the matrix such that $(e^{i,j})_{i,j} = 1$ and with zero entries elsewhere.[3] We set $\dot{C} := \partial_\varepsilon C_\varepsilon|_{\varepsilon=0} = e^{i,j} + e^{j,i}$.
In particular, for the symmetric square root $L = \sqrt{C}$, the partial derivative $\dot{L} := \partial_{C_{i,j}} L$ solves the Sylvester equation [Hig08, p.58]

$$\dot{L}\,L + L\,\dot{L} = \dot{C} = e^{i,j} + e^{j,i}.$$

[3]The matrices C_ε are clearly symmetric and satisfy $(C_\varepsilon)_{i,i} = 1$. The invertibility of C_ε for ε small enough follows from the continuity of the smallest eigenvalue $\lambda_{\min}$ from $\mathcal{S}_{\geq 0}^d$ into $\mathbb{R}$, $A \mapsto \lambda_{\min}(A)$ (with respect to, say, the topology induced by the Hilbert-Schmidt norm), see [HJ90, Hoffman and Wielandt's theorem, p.368].

From (3.9), we derive $\partial_{C_{i,j}} Z_T = \partial_{C_{i,j}}[\mathrm{diag}(\sigma) L W_T] = \mathrm{diag}(\sigma)\dot{L} W_T$. This yields

$$\begin{aligned}
\mathcal{I}(Z^\theta, 1) &= \delta\Big(\sum_{l=1}^{d}\Big(\gamma_{Z_T}^{-1}\partial_{C_{i,j}} Z_T\Big)_l D.Z_T^l\Big) \\
&= \frac{1}{T}\delta\Big(\sum_{l=1}^{d}\Big((\Sigma\Sigma^*)^{-1}\mathrm{diag}(\sigma)\dot{L}W_T\Big)_l \Sigma_{l,\cdot}\mathbf{1}_{[0,T]}\Big) \\
&= \frac{1}{T}\sum_{l=1}^{d}((\Sigma\Sigma^*)^{-1}\mathrm{diag}(\sigma)\dot{L}W_T)_l\ \Sigma^*_{\cdot,l}\cdot W_T - \sum_{l=1}^{d}((\Sigma\Sigma^*)^{-1}\mathrm{diag}(\sigma)\dot{L}\Sigma^*_{\cdot,l})_l
\end{aligned}$$

(using (3.11) with $A = (\Sigma\Sigma^*)^{-1}\mathrm{diag}(\sigma)\dot{L}$ and $u^* = \Sigma_{l,\cdot}$)

$$\begin{aligned}
&= \frac{1}{T} W_T \cdot (\Sigma^*(\Sigma\Sigma^*)^{-1}\mathrm{diag}(\sigma)\dot{L}W_T) - \mathrm{Tr}((\Sigma\Sigma^*)^{-1}\mathrm{diag}(\sigma)\dot{L}\Sigma^*) \\
&= \frac{1}{T} W_T \cdot L^{-1}\dot{L}W_T - \mathrm{Tr}(L^{-1}\dot{L}).
\end{aligned}$$

Since $W_T \cdot L^{-1}\dot{L}W_T$ is a scalar, it is equal to its transpose $W_T \cdot \dot{L}L^{-1}W_T$, and thus to its average $\frac{1}{2}W_T \cdot (L^{-1}\dot{L} + \dot{L}L^{-1})W_T$. Similarly, $\mathrm{Tr}(L^{-1}\dot{L}) = \frac{1}{2}\mathrm{Tr}(L^{-1}\dot{L} + \dot{L}L^{-1})$. We claim that

$$L^{-1}\dot{L} + \dot{L}L^{-1} = L^{-1}\dot{C}L^{-1}, \tag{3.12}$$

which gives the final representation

$$\begin{aligned}
\mathcal{I}(Z^\theta, 1) &= \frac{1}{2T} W_T \cdot (L^{-1}(e^{i,j} + e^{j,i})L^{-1})W_T - \frac{1}{2}\mathrm{Tr}(L^{-1}(e^{i,j} + e^{j,i})L^{-1}) \\
&= \frac{1}{T}(L^{-1}W_T)^i (L^{-1}W_T)^j - (C^{-1})_{i,j}
\end{aligned}$$

where the final step follows after a few standard manipulations.
In order to prove (3.12), proceed as follows: starting from the derivative of C^{-1}, we obtain

$$-C^{-1}\dot{C}C^{-1} = \partial_\varepsilon C_\varepsilon^{-1}|_{\varepsilon=0} = \partial_\varepsilon L_\varepsilon^{-2}|_{\varepsilon=0} = \partial_\varepsilon L_\varepsilon^{-1}|_{\varepsilon=0} L^{-1} + L^{-1}\partial_\varepsilon L_\varepsilon^{-1}|_{\varepsilon=0} = -L^{-1}\dot{L}L^{-2} - L^{-2}\dot{L}L^{-1}.$$

Now multiplying by L on the left and right we obtain (3.12).

We estimate relative model sensitivity using POP method in the setting of (3.8). It is important to observe that the relative sensitivity can be directly evaluated by the POP method using the ratio of two time-average approximations of $\mathcal{I}(Z^\theta, \Phi^\theta)$ and Φ^θ respectively, along only one Markov chain defined by applying shaking with rejection with respect to $Z^\theta \in A$, see Remark 2.1. The computations at intermediate levels are unnecessary which very much simplifies the numerical evaluation. We consider a two-dimensional example which is similar to the example discussed in [GT16]. We take interest rate $r = \mu^i = 0.01$ and for the other parameters $K = 100$, $T = 1$, $\sigma^1 = 0.25$, $\sigma^2 = 0.225$, correlation parameter $C_{1,2} = 0.9$, $p_1 = 10$, $S_0^1 = 10$, $p_2 = 5$, $S_0^2 = 20$ and estimate the sensitivities of the rare event statistics $\mathbb{E}(K - p_1 S_T^1 - p_2 S_T^2)_+$ with respect to p_1, σ^1 and $C_{1,2}$. Observe that we choose $\sigma^2 = C_{1,2}\sigma^1$, which corresponds to the critical case described in [GT16, Theorem 1] where the asymptotics of the density of the basket undergoes a change of regime. It is thus arguably delicate to obtain a tractable analytical approximation via the derivation of the density.

In Table 3.4.1, we compare the results of finite difference method using simple Monte Carlo with common random numbers [GRZ84] to those of the POP method with the number of simulations as indicated. Here, the rare event probability $\mathbb{P}(p_1 S_T^1 + p_2 S_T^2 \le K)$ is around 1.7×10^{-3}. We deliberately choose such an example in order to show the application of POP method in "not-so-rare" situations. Actually, when the rare event probability becomes smaller, the performance of POP method is considerably improved with respect to the simple Monte Carlo. For 10^6 paths, the run time of MC method is 64 seconds and for POP method with shaking parameter $\rho = 0.85$ is 80 seconds.

	Relative sensitivity w.r.t.		
	p_1	σ^1	$C_{1,2}$
POP method (10^6) (mean/std)	-0.7155 (0.0046)	24.0078 (0.1760)	3.1058 (0.0253)
Finite difference (10^6) (mean/std)	-0.7120 (0.0157)	23.9252 (0.4838)	3.0866 (0.1128)
Finite difference (10^9) (99% conf. interval)	$(-0.7155, -0.7129)$	$(23.9285, 24.0108)$	$(3.0801, 3.0990)$

Table 3.4.1: *Estimates of relative sensitivity w.r.t. different model parameters.*

4 Conclusion

In this work, we introduced an adaptive algorithm to efficiently compute statistics of rare events using reversible shaking transformations. We established various convergence results which provides theoretical foundation for consistency of our algorithm. We embedded the underlying model in a setting of Hilbert Gaussian spaces which allows to easily analyze the sensitivities with respect to the model parameters which is an important concern in risk management. With the help of various examples, we demonstrated that this Gaussian space viewpoint is sufficiently large to encompass

- many important models in financial engineering and stochastic finance, ranging from Markovian to path-dependent models passing through fractional Brownian motion models,
- many risk management issues, such as model risk, credit risk, market risk.

Finally, our numerical experiments show better performance of the POP method and the adaptive algorithm compared to the modified IPS method and its adaptive implementation.

References

[ADGL18] A. Agarwal, S. De Marco, E. Gobet, and G. Liu. Study of new rare event simulation schemes and their application to extreme scenario generation. *Mathematics and Computers in Simulation*, 143(Supplement C):89–98, 2018.

[AG11] S. Asmussen and P.W. Glynn. A new proof of convergence of MCMC via the ergodic theorem. *Statistics & Probability Letters*, 81(10):1482–1485, 2011.

[AR99] S. Asmussen and R.Y. Rubinstein. Sensitivity analysis of insurance risk models via simulation. *Management Science*, 45(8):1125–1141, 1999.

[AS12] C. Alexander and J-M. Sarabia. Quantile uncertainty and value-at-risk. *Risk Analysis: An International Journal*, 32(8):1293–1308, 2012.

[Bas11] Basel Committee on Banking Supervision. Basel III: A global regulatory framework for more resilient banks and banking systems. https://www.bis.org/publ/bcbs189.pdf, December 2011.

[Bas17] Basel Committee on Banking Supervision. Stress testing principles. http://www.bis.org/bcbs/publ/d428.pdf, December 2017.

[Bil95] P. Billingsley. *Probability and measure*. John Wiley & Sons, 1995.

[CC10] R. Carmona and S. Crépey. Particle methods for the estimation of credit portfolio loss distributions. *Int. J. Theor. Appl. Finance*, 13(4):577–602, 2010.

[CDLL06] F. Cérou, P. Del Moral, F. Le Gland, and P. Lezaud. Genetic genealogical models in rare event analysis. *ALEA Lat. Am. J. Probab. Math. Stat.*, 1:181–203, 2006.

[CFV09] R. Carmona, J.P. Fouque, and D. Vestal. Interacting particle systems for the computation of rare credit portfolio losses. *Finance and Stochastics*, 13(4):613–633, 2009.

[CI13] S. Cohen and J. Istas. *Fractional fields and applications*. Springer, 2013.

[CKM03] P. Cheridito, H. Kawaguchi, and M. Maejima. Fractional Ornstein-Uhlenbeck processes. *Electronic Journal of Probability*, 8(3):14, 2003.

[CM01] P. Carr and D. Madan. Towards a theory of volatility trading. *Option Pricing, Interest Rates and Risk Management, Handbook in Mathematical Finance*, pages 458–476, 2001.

[CMFG12] F. Cérou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte Carlo for rare event estimation. *Statistics and Computing*, 22(3):795–808, 2012.

[Con06] R. Cont. Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance*, 16(3):519–547, 2006.

[CR98] F. Comte and E. Renault. Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4):291–323, 1998.

[DD98] G. Deelstra and F. Delbaen. Convergence of discretized stochastic (interest rate) processes with stochastic drift term. *Applied stochastic models and data analysis*, 14(1):77–84, 1998.

[Del04] P. Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer, New-York, 2004.

[DFJN14] C.R. Doss, J.M. Flegal, G.L. Jones, and R.C. Neath. Markov chain Monte Carlo estimation of quantiles. *Electronic Journal of Statistics*, 8(2):2448–2478, 2014.

[DG05] P. Del Moral and J. Garnier. Genealogical particle analysis of rare events. *Annals of Applied Probability*, 15:2496–2534, 2005.

[Die90] J. Dieudonné. *Eléments d'analyse*. Jacques Gabay, Paris, 1990. Tome 1, (chapitres I à XI), Fondements de l'Analyse moderne.

[Dup94] B. Dupire. Pricing with a smile. *Risk*, 7(1):18–20, 1994.

[EJS98] N. El Karoui, M. Jeanblanc-Picqué, and S.E. Shreve. Robustness of the Black and Scholes formula. *Mathematical Finance*, 8(2):93–126, 1998.

[ER16] O. El Euch and M. Rosenbaum. The characteristic function of rough Heston models. *arXiv preprint arXiv:1609.02108, to appear in Mathematical Finance*, 2016.

[ER17] O. El Euch and M. Rosenbaum. Perfect hedging in rough Heston models. *arXiv preprint arXiv:1703.05049*, 2017.

[FGG+15] P.K. Friz, J. Gatheral, A. Gulisashvili, A. Jacquier, and J. Teichmann, editors. *Large Deviations and Asymptotic Methods in Finance*. Springer Proceedings in Mathematics and Statistics. Springer-Verlag, Berlin, 2015.

[FGM17] G. Fort, E. Gobet, and E. Moulines. MCMC design-based non-parametric regression for rare-event. Application to nested risk computations. *Monte Carlo Methods and Applications*, 23(1):21–42, 2017.

[FI14] J-P. Fouque and T. Ichiba. Stability in a model of inter-bank lending. *SIAM Journal on Financial Mathematics*, 4(1):784–803, 2014.

[FLL+99] E. Fournié, J.M. Lasry, J. Lebuchoux, P.L. Lions, and N. Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance and Stochastics*, 3(4):391–412, 1999.

[FM03] G. Fort and E. Moulines. Convergence of the Monte Carlo expectation maximization for curved exponential families. *Annals of Statistics*, pages 1220–1259, 2003.

[Geb41] H. Gebelein. Das statistische Problem der Korrelation als Variations-und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 21(6):364–379, 1941.

[GHS00] P. Glasserman, P. Heidelberger, and P. Shahabuddin. Variance reduction techniques for estimating Value-at-Risk. *Management Science*, 46:1349–1364, 2000.

[GJ14] J. Gatheral and T. Jaisson. Fractional volatility models. *Presentation at Bloomberg seminar,* http://mfe.baruch.cuny.edu/wp-content/uploads/2012/09/FractionalVolatility2014BBQ.pdf, June 2014.

[GJR14] J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *Preprint available at SSRN 2509457, to appear in Quantitative Finance*, 2014.

[GL15] E. Gobet and G. Liu. Rare event simulation using reversible shaking transformations. *SIAM Scientific Computing*, 37(5):A2295–A2316, 2015.

[GM05] E. Gobet and R. Munos. Sensitivity analysis using Itô-Malliavin calculus and martingales. Application to stochastic control problem. *SIAM Journal of Control and Optimization*, 43:5:1676–1713, 2005.

[GO02] P.W. Glynn and D. Ormoneit. Hoeffding's inequality for uniformly ergodic markov chains. *Statistics & probability letters*, 56(2):143–146, 2002.

[Gob04] E. Gobet. Revisiting the Greeks for European and American options. In J. Akahori, S. Ogawa, and S. Watanabe, editors, *Stochastic processes and applications to mathematical finance*, pages 53–71. World Scientific, 2004.

[GRZ84] S. Gal, R.Y. Rubinstein, and A. Ziv. On the optimality and efficiency of common random numbers. *Mathematics and Computers in Simulation*, 26(6):502–512, 1984.

[GT16] A. Gulisashvili and P. Tankov. Tail behavior of sums and differences of log-normal random variables. *Bernoulli*, 22(1):444–493, 2016.

[Gul12] A. Gulisashvili. Asymptotic equivalence in Lee's moment formulas for the implied volatility, asset price models without moment explosions, and Piterbarg's conjecture. *International Journal of Theorical and Applied Finance*, 15:1–34, 2012.

[Guy14] J. Guyon. Path-dependent volatility. *Risk magazine*, october 2014.

[GVZ15] A. Gulisashvili, F. Viens, and X. Zhang. Extreme-strike asymptotics for general gaussian stochastic volatility models. ArXiv preprint, arXiv:1502.05442, 2015.

[Hes93] S.L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2):327–343, 1993.

[Hig08] N.J. Higham. *Functions of matrices: theory and computation*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2008.

[HJ90] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge Univ. Press, Cambridge, 1990.

[HR98] D.G. Hobson and L.C.G. Rogers. Complete models with stochastic volatility. *Mathematical Finance*, 8(1):27–48, 1998.

[Jan97] S. Janson. *Gaussian Hilbert spaces*, volume 129. Cambridge university press, 1997.

[KLM05] I. Kontoyiannis, L.A. Lastras-Montano, and S.P. Meyn. Relative entropy and exponential deviation bounds for general markov chains. In *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on*, pages 1563–1567. IEEE, 2005.

[KY09] A. Kohatsu-Higa and K. Yasuda. Estimating multidimensional density functions using the Malliavin-Thalmaier formula. *SIAM Journal on Numerical Analysis*, 47:1546–1575, 2009.

[Lag06] A. Lagnoux. Rare event simulation. *Probability in the Engineering and Informational Sciences*, 20(1):45–66, 2006.

[Lee04] R.W. Lee. The moment formula for implied volatility at extreme strikes. *Mathematical Finance*, 14(3):469–480, 2004.

[MT09] S.P. Meyn and R.L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009.

[Nua06] D. Nualart. *Malliavin calculus and related topics*. Springer Verlag, second edition, 2006. (with corrections on the webpage of the author).

[RK08] R.Y. Rubinstein and D.P. Kroese. *Simulation and the Monte Carlo method*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2008.

[Shi04] I. Shigekawa. *Stochastic analysis*, volume 224. American Mathematical Soc., 2004.

[VV91] M. Villén-Altamirano and J. Villén-Altamirano. RESTART: a method for accelerating rare event simulations. In *Proceedings of the 13th International Teletraffic Congress, Copenhagen*, pages 71–76, 1991.

A Proofs

A.1 Proof of Theorem 2

The proof follows a standard routine inspired by [FLL+99, Gob04, GM05, KY09] but it requires a specific careful analysis because of the indicator function. Firstly, properly mollify the indicator function $z \to \mathbf{1}_{\varphi(z,\bar{a})\leq 0}$. Secondly, compute the derivative of the expectation for the mollified function, then, integrate by parts and take the limit w.r.t. the mollified parameter. Mollifying and passing to the limit is the critical part. In [KY09, Section 6], it has been done for functions which are almost everywhere continuous. Here we do not impose such restrictions.

Step 1. Let us define the measure $\bar{\mu}(\mathrm{d}z) = (1+|z|)^{-q}\mathrm{d}z$ on $\mathbb{R}^d$ with q as in the statement and as $q > d$, this is a finite measure. Since $\mathbf{1}_A$ is in $\mathbf{L}_4(\mu)$, there is a sequence $(\xi_k)_{k\in\mathbb{N}}$ of smooth functions with compact support, such that

$$\int_{\mathbb{R}^d} |\mathbf{1}_{z\in A} - \xi_k(z)|^4 (1+|z|)^{-q}\mathrm{d}z \underset{k\to\infty}{\longrightarrow} 0. \tag{A.1}$$

W.l.o.g. we assume that $0 \leq \xi_k \leq 1$. Now, define

$$u_k(\theta) := \mathbb{E}\left[\Phi^\theta \xi_k(Z^\theta)\right], \qquad u(\theta) := \mathbb{E}\left[\Phi^\theta \mathbf{1}_{Z^\theta\in A}\right],$$
$$v_k(\theta) := \mathbb{E}\left[\mathcal{I}(Z^\theta,\Phi^\theta)\xi_k(Z^\theta)\right], \qquad v(\theta) := \mathbb{E}\left[\mathcal{I}(Z^\theta,\Phi^\theta)\mathbf{1}_{Z^\theta\in A}\right].$$

Going forward, we shall establish three results. Firstly, $u_k(\theta) \underset{k\to\infty}{\longrightarrow} u(\theta)$ for any $\theta \in \Theta$, then, $u_k'(\theta) = v_k(\theta)$ for any $\theta \in \Theta$, and finally, v_k converges to v locally uniformly on Θ. By [Die90, Statement (8.6.4) Chap. VIII], this proves that u is differentiable on Θ and its derivative is v.

Step 2: Proof of $u_k'(\theta) = v_k(\theta)$. We can show

$$u_k'(\theta) = \partial_\theta \mathbb{E}\left[\Phi^\theta \xi_k(Z^\theta)\right] = \mathbb{E}\left[\dot{\Phi}^\theta \xi_k(Z^\theta)\right] + \mathbb{E}\left[\Phi^\theta \sum_{i=1}^d \partial_{z_i}\xi_k(Z^\theta)\dot{Z}_i^\theta\right],$$

from the dominated convergence theorem using the boundedness of $\xi_k, \nabla\xi_k$ and the uniform controls in the assumptions (a)-(b). Further, by the chain rule property, $\xi_k(Z^\theta) \in \mathbf{D}^{1,2}$ with $D.[\xi_k(Z^\theta)] = \sum_{i=1}^d \partial_{z_i}\xi_k(Z^\theta) D.Z_i^\theta$. Moreover by definition of δ as the adjoint operator of D, we have

$$\begin{aligned} v_k(\theta) &= \mathbb{E}\left[\dot{\Phi}^\theta \xi_k(Z^\theta) + \langle \sum_{i=1}^d \partial_{z_i}\xi_k(Z^\theta) D.Z_i^\theta, \Phi^\theta \sum_{j=1}^d (\gamma_{Z^\theta}^{-1}\dot{Z}^\theta)_j D.Z_j^\theta\rangle_{\mathcal{H}}\right] \\ &= \mathbb{E}\left[\dot{\Phi}^\theta \xi_k(Z^\theta) + \Phi^\theta \sum_{i=1}^d \partial_{z_i}\xi_k(Z^\theta) \sum_{j=1}^d (\gamma_{Z^\theta})_{i,j}(\gamma_{Z^\theta}^{-1}\dot{Z}^\theta)_j\right] \\ &= \mathbb{E}\left[\dot{\Phi}^\theta \xi_k(Z^\theta) + \Phi^\theta \sum_{i=1}^d \partial_{z_i}\xi_k(Z^\theta)\dot{Z}_i^\theta\right] = u_k'(\theta). \end{aligned}$$

Step 3: Proof of $(u_k, v_k) \underset{k\to\infty}{\longrightarrow} (u, v)$ locally uniformly on Θ. Assume for a while the $\mathbf{L}_2$-convergence

$$\mathbb{E}\left[|\xi_k(Z^\theta) - \mathbf{1}_{Z^\theta\in A}|^2\right] \underset{k\to\infty}{\longrightarrow} 0 \quad \text{locally uniformly in } \theta \in \Theta. \tag{A.2}$$

Then from above and (d), we deduce that for any $\theta \in \Theta$, there is an open set $V \subset \theta$ such that

$$|v_k(\theta) - v(\theta)| \le \sup_{\theta\in V}\left|\dot{\Phi}^\theta + \delta(\Phi^\theta \sum_{j=1}^d (\gamma_{Z^\theta}^{-1}\dot{Z}^\theta)_j D.Z_j^\theta)\right|_{\mathbf{L}_2} \sup_{\theta\in V}\left|\xi_k(Z^\theta) - \mathbf{1}_{Z^\theta\in A}\right|_{\mathbf{L}_2} \underset{k\to\infty}{\longrightarrow} 0.$$

The same arguments apply for $u_k - u$. Consequently, it remains to justify (A.2).

Under the assumption (e), we have the integration by parts formula at order 1 (derived as in the proof of Step 2), i.e. for any smooth function ζ with compact support and any $i \in \{1, \dots, d\}$,

$$\mathbb{E}\left[\partial_{z_i}\zeta(Z^\theta)\right] = \mathbb{E}\left[\zeta(Z^\theta)\delta\left(\sum_{j=1}^d (\gamma_{Z^\theta}^{-1})_{j,i} D.Z_j^\theta\right)\right].$$

Therefore, from [Shi04, Theorem 5.4] the distribution of Z^θ has a continuous density $p_{Z^\theta}(.)$ w.r.t. the Lebesgue measure, which is uniformly bounded by a function depending only on the $\mathbf{L}_q$-norms of $\delta\left(\sum_{j=1}^d (\gamma_{Z^\theta}^{-1})_{j,i} D.Z_j^\theta\right), 1 \le i \le d$. In view of (a)-(e), we deduce that for any $\theta \in \Theta$, there is a neighborhood $V \subset \Theta$ of θ such that $\sup_{\theta'\in V} |p_{Z^{\theta'}}|_\infty := C_V < \infty$, and

$$\begin{aligned} \mathbb{E}\left[|\xi_k(Z^{\theta'}) - \mathbf{1}_{Z^{\theta'}\in A}|^2\right] &\le \left(\mathbb{E}\left[|\xi_k(Z^{\theta'}) - \mathbf{1}_{Z^{\theta'}\in A}|^4 (1 + |Z^{\theta'}|)^{-q}\right]\right)^{1/2} \left(\mathbb{E}\left[(1 + |Z^{\theta'}|)^q\right]\right)^{1/2} \\ &\le \left(\int_{\mathbb{R}^d} |\xi_k(z) - \mathbf{1}_{z\in A}|^4 (1 + |z|)^{-q} C_V \mathrm{d}z\right)^{1/2} \sup_{\theta'\in V}\left(\mathbb{E}\left[(1 + |Z^{\theta'}|)^q\right]\right)^{1/2}. \end{aligned}$$

Owing to (A.1), the above converges to 0 as $k \to \infty$, uniformly w.r.t. $\theta' \in V$, and (A.2) is proved. □

A.2 Proof of Theorem 4

Denote by e_N the l.h.s. of the above inequality. We have

$$e_N = \frac{1}{N^2}\left[\sum_{1\le k\le N} \mathbb{V}\mathrm{ar}\left(f(Z_k)\right) + 2\sum_{1\le k<l\le N} \mathbb{C}\mathrm{ov}\left(f(Z_k), f(Z_l)\right)\right].$$

By the reversible shaker property, Z_k and Z have the same law, thus $\sum_{k=1}^N \mathbb{V}\mathrm{ar}\left(f(Z_k)\right) = N\mathbb{V}\mathrm{ar}\left(f(Z)\right)$. On the other hand, for $l > k$, we have

$$|\mathbb{C}\mathrm{ov}\left(f(Z_k), f(Z_l)\right)| \le \rho_{X_k,X_l}\sqrt{\mathbb{V}\mathrm{ar}\left(f(Z_k)\right)}\sqrt{\mathbb{V}\mathrm{ar}\left(f(Z_l)\right)} = \rho_{X_k,X_l}\mathbb{V}\mathrm{ar}\left(f(Z)\right)$$

where ρ_{X_k,X_l} is the so-called *Renyi maximal correlation coefficient* between X_k and X_l, i.e. the supremum of the correlation between a function g_k of X_k and a function g_l of X_l, the supremum being taken over all functions (g_k, g_l) with squared integrability properties. We claim that

$$\rho_{X_k,X_l} \le |\rho|_\infty^{l-k}. \tag{A.3}$$

The proof is provided at the end. With (A.3) at hand, we deduce

$$\Big|\sum_{1\le k<l\le N} \mathbb{C}\mathrm{ov}\left(f(Z_k), f(Z_l)\right)\Big| \le N\frac{|\rho|_\infty}{1-|\rho|_\infty}\mathbb{V}\mathrm{ar}\left(f(Z)\right).$$

Finally, we get

$$e_N \le \frac{\mathbb{V}\mathrm{ar}\left(f(Z)\right)}{N}\left[1 + 2\frac{|\rho|_\infty}{1-|\rho|_\infty}\right],$$

which finishes the proof of (2.11).

It remains to justify (A.3). This is a consequence of [Jan97, Theorem 10.11]. Indeed, assume without loss of generality that $k = 1$ (for notational convenience). Now define a Gaussian Hilbert space $\mathcal{G}$ for all the variables from shaker iteration $k = 1$ to $l > 1$. For this, set $\mathfrak{H} := \{\mathfrak{h} = (h_1, \dots, h_l) \in \mathcal{H}^l\}$: endowed with the scalar product $\langle \mathfrak{h}, \mathfrak{g}\rangle_{\mathfrak{H}} = \sum_{i=1}^l \langle h_i, g_i\rangle_{\mathcal{H}}$, $\mathfrak{H}$ is a Hilbert space to which we associate the Gaussian process $\mathfrak{X} = \{\mathfrak{X}(\mathfrak{h}) : \mathfrak{h} \in \mathfrak{H}\}$. Let $\mathcal{G}$ denote the Gaussian Hilbert space spanned by $\{\mathfrak{X}(\mathfrak{h}) : \mathfrak{h} \in \mathfrak{H}\}$.

In view of (2.9) we observe that (X_1, X_l) can be realized jointly as follows:

$$X_1 = \Big\{\mathfrak{X}(\mathfrak{h}) : \mathfrak{h} = (h, 0, \dots, 0), h \in \mathfrak{b}\mathcal{H}\Big\},$$
$$X_l = \Big\{\mathfrak{X}(\mathfrak{h}) : \mathfrak{h} = (\rho_h^{l-1}h, \rho_h^{l-2}\sqrt{1-\rho_h^2}h, \cdots, \sqrt{1-\rho_h^2}h), h \in \mathfrak{b}\mathcal{H}\Big\}.$$

Let $\mathcal{G}_1$ denote the Gaussian subspace spanned by $\{\mathfrak{X}(\mathfrak{h}) : \mathfrak{h} = (h, 0, \dots, 0), h \in \mathcal{H}\}$ and similarly for $\mathcal{G}_l$. Then, [Jan97, Theorem 10.11] states that ρ_{X_1,X_l} is equal to the norm of the operator $P_{\mathcal{G}_l,\mathcal{G}_1}$ which is defined as the orthogonal projection of $\mathcal{G}$ onto $\mathcal{G}_l$ and then restricted to $\mathcal{G}_1$. Thus, from [Jan97, Definition 10.6], we get $\rho_{X_k,X_l} = \|P_{\mathcal{G}_l,\mathcal{G}_1}\| = \sup_{X\in\mathcal{G}_1, Y\in\mathcal{G}_l} |\mathrm{Corr}(X, Y)| \le |\rho|_\infty^{l-1}$. The proof of (A.3) is complete. □

Remark A.1. *The above result is of standalone interest and provides an important first step towards proving the* $\mathbf{L}_2$ *convergence of shaking transformation with rejection in infinite dimension.*

A.3 Proof of Propositions 6 and 7

A.3.1 Proof of Propositions 6

The following theorem provides an important result for our proof.

Theorem 10 ([AG11]). *Let S be a measurable space. Assume that X is an S-valued Markov chain, starting at a given x, which is η-irreducible for some measure η and has a stationary distribution π. Then, for $f : S \to \mathbb{R}_+$*

$$\frac{1}{N}\sum_{j=0}^{N-1} f(X_j) \to \int_S f(y)\pi(\mathrm{d}y) \quad \mathbb{P}_x - a.s. \tag{A.4}$$

as $N \to \infty$ for π-a.a. $x \in S$. If in addition, the one step transition kernel can be written either as

$$\mathbb{P}_x(X_1 \in dy) = q(x,y)\eta(dy)$$

or as

$$\mathbb{P}_x(X_1 \in dy) = (1 - a(x))\delta_x(dy) + a(x,y)q(x,y)\eta(dy) \tag{A.5}$$

with $a(x), a(x,y) > 0$ for each $x, y \in S$, then X is a positive recurrent Harris chain and the convergence (A.4) holds for all $x \in S$.

Using the above results, we can show that the POP method with $|\rho|_\infty < 1$ converges almost surely in all the finite dimensional cases (i.e. $S = \mathcal{H} = \mathbb{R}^d$ of Example 1). Firstly, we explain how the transformation with shaking and rejection at level k (defined in (2.12)) can be interpreted in the form (A.5), which is well-known as *Metropolis-Hastings sampler*. Unlike usual Metropolis-Hastings sampler where explicit transition densities and acceptance functions are used, we use implicit transition densities and acceptance functions. Namely, in the case of shaking for standard d-dimensional normal variable, i.e. $K(x, X') = (\rho_i x_i + \sqrt{1-\rho_i^2} X'_i)_{1\leq i\leq d}$ with i.i.d. standard Gaussian variables $(X'_i)_{1\leq i\leq d}$, the measure η can be taken as the Lebesgue measure on $\mathbb{R}^d$ and the transition density is given by

$$q(x,y) = \exp\left(-\sum_{i=1}^{d} \frac{|y_i - \rho_i x_i|^2}{2(1-\rho_i^2)}\right)(2\pi)^{-q/2}\prod_{i=1}^{d}(1-\rho_i^2)^{-1/2}$$

(here we use $\sup_{1\leq i\leq d}|\rho_i| = |\rho|_\infty < 1$). Then the acceptance function corresponds to $a(x,y) = 1_{\Psi_Z(y)\in A_k}$ and the local mean acceptance rate to $a(x) = \int_{\mathbb{R}^d} a(x,y)q(x,y)\mathrm{d}y$.

Secondly, the assumption $a(x) > 0$ writes in our rare event setting as $\mathbb{P}(\Psi_Z(K(x,X')) \in A_k) > 0$ for any x s.t. $\Psi_Z(x) \in A_k$. This inequality holds true since we assume $0 < \mathbb{P}(Z \in A) \leq \mathbb{P}(\Psi_Z(X) \in A_k)$, i.e. $\Psi_Z^{-1}(A_k)$ has a strictly positive Lebesgue measure.

Thirdly, notice that the existence of a stationary distribution has been shown previously (Proposition 5) and we can easily see that the Markov chain in POP method is η-irreducible, due to the strictly positive transition density p. Therefore, from Theorem 10 we easily deduce the almost sure convergence at each level, using the *random* initialization described in algorithm of POP method. □

A.3.2 Proof of Proposition 7

This follows from the well-known result for positive Harris recurrent Markov chain which states that if the chain is in addition aperiodic, then its marginal distribution converges to its stationary distribution (see for example [MT09, Theorem 13.0.1]). The existence of an implicit positive transition density in finite dimension ensures that our Markov chain is aperiodic and the result of Proposition 7 holds. □
Note that in Theorem 8, under stronger conditions (i.e. the acceptance rate is bounded away from 0, see (2.15)), the convergence is geometric.

A.4 Proof of Theorem 8

To alleviate notation, we set $\mathcal{A}_k = \Psi_Z^{-1}(A_k)$, we write π_k for the conditional distribution of $X \mid X \in \mathcal{A}_k$ and last, we remove all the indices k from the mathematical objects under study. The transition kernel of

the Markov chain $(X_i : i \geq 0)$ is denoted by $Q_{\texttt{POP}}$, in view of the definition of the algorithm it is defined as follows: for any measurable positive function f and any $x \in \mathcal{A}$, we have

$$Q_{\texttt{POP}} f(x) := \int_{\mathcal{A}} q(x,z) f(z) \mathrm{d}z + f(x) \int_{\mathcal{A}^c} q(x,z) \mathrm{d}z, \tag{A.6}$$
$$\text{with } q(x,z) = \exp\left(- \sum_{j=1}^{d} \frac{|z_j - \rho_j x_j|^2}{2(1-\rho_j^2)} \right) (2\pi)^{-q/2} \prod_{j=1}^{d} (1-\rho_j^2)^{-1/2}.$$

The density q is the density of the proposal distribution.

The proof of (2.16) goes through an application of [FM03, Proposition 2], restated below in our context.

Proposition 11. *Let $Q_{\texttt{POP}}$ be a phi-irreducible aperiodic transition kernel on $\mathcal{A}$ and let $\mathcal{C} \subset \mathcal{A}$ be an accessible petite set. Assume there exist $\delta \in (0,1)$, $b \in \mathbb{R}$ and a measurable $V : \mathcal{A} \to [1, +\infty)$, bounded on $\mathcal{C}$, such that*

$$Q_{\texttt{POP}} V(x) \leq \delta V(x) + b \mathbf{1}_{x \in \mathcal{C}}, \qquad x \in \mathcal{A}. \tag{A.7}$$

Let $p \geq 2$ and assume that $\mathbb{E}[V(X_0)] < +\infty$. Then there exists a finite positive constant C such that for any g bounded in $V^{1/p}$-norm, we have

$$\mathbb{E}\left[\left| \sum_{i=1}^{n} g(X_i) - \pi(g) \right|^p \right] \leq C \left(\sup_{\mathcal{A}} \frac{|g|^p}{V} \right) n^{p/2}, \qquad \forall n \geq 1.$$

Observe that if the assumptions of the above proposition hold, one can also apply [MT09, Theorem 15.0.1], which leads to the inequality (2.17).

Proof of the drift condition (A.7). In view of (A.6), we start to bound the first term on the RHS. Using the exponential form for V and simple inequalities, we get

$$\begin{aligned}
\int_{\mathcal{A}} q(x,z) V(z) \mathrm{d}z &\leq V(x) \int_{\mathbb{R}^q} q(x,z) \frac{V(z)}{V(x)} \mathrm{d}z \\
&= V(x) \prod_{j=1}^{d} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi(1-\rho_j^2)}} \exp\left(- \frac{(z_j - \rho_j x_j)^2}{2(1-\rho_j^2)} \right) e^{s|z_j| - s|x_j|} \mathrm{d}z_j \\
&\leq V(x) \prod_{j=1}^{d} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi(1-\rho_j^2)}} \exp\left(- \frac{\tilde{z}_j^2}{2(1-\rho_j^2)} + s|\tilde{z}_j| \right) e^{-s(1-|\rho_j|)|x_j|} \mathrm{d}\tilde{z}_j \\
&:= V(x) C(\rho_1, \ldots, \rho_d) e^{-s \sum_{j=1}^{d} (1-|\rho_j|)|x_j|}.
\end{aligned}$$

Since $|\rho|_\infty < 1$, for any $|x| > R$ with R large enough, we get

$$\int_{\mathcal{A}} q(x,z) V(z) \mathrm{d}z \leq V(x) \frac{(1-\delta_1)}{2}, \qquad \forall |x| > R. \tag{A.8}$$

On the other hand, for $|x| \leq R$, the above left hand side is bounded (since continuous on a compact set), therefore

$$\sup_{x : |x| \leq R} \int_{\mathcal{A}} q(x,z) V(z) \mathrm{d}z = C_{(\text{A.9})} < +\infty. \tag{A.9}$$

To complete (A.7), we set

$$\mathcal{C} = \{ x \in \mathcal{A} : V(x) \leq L \}, \qquad \text{with} \qquad L \geq \frac{4 C_{(\text{A.9})}}{1-\delta_1} \vee \tilde{R}$$

with $\tilde{R}$ large enough so that $\mathcal{C} \cap \mathcal{A} \neq \emptyset$. Then, plugging (2.15), (A.8) and (A.9) into (A.6), we get

$$Q_{\texttt{POP}} V(x) \leq \frac{(1+\delta_1)}{2} V(x) + C_{(A.9)}, \qquad \forall x \in \mathcal{A}. \tag{A.10}$$

Now, for $x \in \mathcal{A} \cap \mathcal{C}^c$, by definition of $\mathcal{C}$ with the threshold L, we have

$$Q_{\texttt{POP}} V(x) \leq \frac{(3+\delta_1)}{4} V(x) + \frac{(\delta_1 - 1)}{4} V(x) + C_{(A.9)} \leq \frac{(3+\delta_1)}{4} V(x).$$

This, together with (A.10) and $V \geq 1$ proves

$$Q_{\texttt{POP}} V(x) \leq \frac{(3+\delta_1)}{4} V(x) + C_{(A.9)} \mathbf{1}_{x \in \mathcal{C}},$$

i.e. the announced inequality (A.7) with $\delta := \frac{(3+\delta_1)}{4}$ and $b := C_{(A.9)}$.

Verification of the other assumptions of Proposition 11.

- $Q_{\texttt{POP}}$ is Lebesgue-irreducible [MT09, Proposition 4.2.1 (ii)], since for any measurable set $A \subset \mathcal{A}$ with strictly positive Lebesgue measure, we have

$$Q_{\texttt{POP}}(x, A) \geq \int_A q(x, z) \mathrm{d}z > 0, \qquad \forall x \in \mathcal{A}.$$

- The set $\mathcal{C}$ is a ν_1-small set (see [MT09, Section 5.2]), hence a petite set [MT09, Proposition 5.5.3]: indeed, for any $x \in \mathcal{C}$ and any measurable set $A \subset \mathcal{A}$, we have

$$Q_{\texttt{POP}}(x, A) \geq \int_A \inf_{x \in \mathcal{C}} q(x, z) \mathrm{d}z := \nu(A),$$

  and the measure ν is non-zero since q continuous and $\mathcal{C}$ is compact.

- $Q_{\texttt{POP}}$ is aperiodic: in fact, it is strongly aperiodic as a consequence of the existence of the ν_1-small set $\mathcal{C}$ with $\nu(\mathcal{C}) > 0$ and irreducibility, see [MT09, p. 114].

- The set $\mathcal{C}$ is accessible (see [MT09, p. 86]) since from any starting point $x \in \mathcal{A}$, the probability of hitting $\mathcal{C}$ in finite time is strictly positive: indeed, it bounded from below as

$$\mathbb{P}_x \left(\tau_{\mathcal{C}} < +\infty \right) \geq Q_{\texttt{POP}}(x, \mathcal{C}) \geq \int_{\mathcal{A} \cap \mathcal{C}} q(x, z) \mathrm{d}z > 0.$$

□

A.5 Proof of Theorem 9

To begin, we first recall the following classical equivalence result which will be used in our proof.

Theorem 12 (Theorem 20.5 [Bil95])**.** *A necessary and sufficient condition for sequence of random variables $X_n \xrightarrow{P} X$ is that each subsequence $\{X_{n_k}\}$ contains a further subsequence $\{X_{n_{k(i)}}\}$ such that $X_{n_{k(i)}} \to X$ almost surely as $i \to \infty$.*

We split the proof in several steps.

Lemma 1. *There exists a subsequence $\{N_{k(i)}\}_{i \geq 1}$ such that for any $l \in \{1, \ldots, L^* + 1\}$ and any $\epsilon > 0$, we have*

$$\sum_{i \geq 1} \mathbb{P} \left(|\hat{Q}^l_{N_{k(i)}, p} - Q^l_p| > \varepsilon \right) < \infty. \tag{A.11}$$

Thus, $\hat{Q}^l_{N_{k(i)}, p}$ converges to Q^l_p almost surely as $i \to \infty$.

Proof. We proceed by induction on l. Indeed, for $l = 1$ with empty rejection ($Q_p^0 = \infty$), we have

$$\mathbb{P}\left(|\hat{Q}^1_{N,p} - Q^1_p| > \varepsilon\right) \leq b(N, \varepsilon),$$

where $b(N, \varepsilon) = \lim_{s\to\infty} b(s, N, \varepsilon)$. As $b(N, \varepsilon) \to 0$ due to Assumption 2, we can find a subsequence $N_{k_1(i)}$ such that $\sum_{i\geq 1} b(N_{k_1(i)}, \varepsilon) < \infty$. This ensures that (A.11) holds for $l = 1$. Now suppose that there exists a subsequence $\{N_{k'(i)}\}_{i\geq 1}$ such that (A.11) is true for $\{1, \dots, l\}$ and let us prove it for $l + 1$. We have on $\{\hat{Q}^l_{N_{k'(i)},p} \in I_l\}$ by Assumption 2

$$\mathbb{P}\left(|\hat{Q}^{l+1}_{N_{k'(i)},p} - g_p(\hat{Q}^l_{N_{k(i)},p})| > \varepsilon | \hat{Q}^l_{N_{k'(i)},p}\right) \leq b(\hat{Q}^l_{N_{k'(i)},p}, N_{k'(i)}, \varepsilon).$$

The term on the right hand side above is bounded by $\sup_{s\in I_l} b(s, N_{k'(i)}, \varepsilon)$ and we can find a further subsequence $\{N_{k(i)}\}_{i\geq 1}$ such that on $\{\hat{Q}^l_{N_{k(i)},p} \in I_l\}$, we have

$$\sum_{i\geq 1} \mathbb{P}\left(|\hat{Q}^{l+1}_{N_{k(i)},p} - g_p(\hat{Q}^l_{N_{k(i)},p})| > \varepsilon \mid \hat{Q}^l_{N_{k(i)},p}\right) \leq \sum_{i\geq 1} \sup_{s\in I_l} b(s, N_{k(i)}, \varepsilon) < \infty. \tag{A.12}$$

Then, on the subsequence $\{N_{k(i)}\}_{i\geq 1}$ chosen as above, we consider

$$\begin{aligned}
\mathbb{P}\left(|\hat{Q}^{l+1}_{N_{k(i)},p} - Q^{l+1}_p| > \varepsilon\right) &\leq \mathbb{P}\left(|\hat{Q}^{l+1}_{N_{k(i)},p} - g_p(\hat{Q}^l_{N_{k(i)},p})| > \varepsilon/2\right) + \mathbb{P}\left(|g_p(\hat{Q}^l_{N_{k(i)},p}) - g_p(Q^l_p)| > \varepsilon/2\right) \\
&\leq \mathbb{P}\left(|\hat{Q}^{l+1}_{N_{k(i)},p} - g_p(\hat{Q}^l_{N_{k(i)},p})| > \varepsilon/2, \hat{Q}^l_{N_{k(i)},p} \in I_l\right) \\
&+ \mathbb{P}\left(\hat{Q}^l_{N_{k(i)},p} \notin I_l\right) + \mathbb{P}\left(|g_p(\hat{Q}^l_{N_{k(i)},p}) - g_p(Q^l_p)| > \varepsilon/2\right) \\
&:= \texttt{I} \ + \texttt{II} \ + \texttt{III}.
\end{aligned} \tag{A.13}$$

The term `I` in the right hand side of (A.13) is handled by the result (A.12) such that we have

$$\sum_{i\geq 1} \mathbb{P}\left(|\hat{Q}^{l+1}_{N_{k(i)},p} - g_p(\hat{Q}^l_{N_{k(i)},p})| > \varepsilon/2, \hat{Q}^l_{N_{k(i)},p} \in I_l\right) < \infty. \tag{A.14}$$

Furthermore, Assumption 1 implies that the function $g_p(s)$ is continuous at $s = Q^l_p$. This combined with the induction hypothesis at level l implies that the series with general terms given by `II` and `III` converge similarly to (A.14). Therefore, (A.11) is proved for $l + 1$ and the result follows. □

Corollary 1. *When $\frac{\log\alpha}{\log p}$ is not an integer, i.e. $Q_p^{L^*+1} < \bar{a} < Q_p^{L^*}$, there exists a subsequence $\{N_{k(i)}\}_{i\geq 1}$ such that*

$$\mathbb{P}\left(L_{N_{k(i)}} = L^* \text{ for } i \text{ large enough }\right) = 1.$$

Proof. This is a direct consequence from Lemma 1. □

Lemma 2. *Assume $L^* \neq 0$. When $\frac{\log\alpha}{\log p}$ is not an integer, there exists a subsequence $\{N_{k(i)}\}_{i\geq 1}$ such that for any $\varepsilon > 0$, we have*

$$\sum_{i\geq 1} \mathbb{P}\left(|\hat{Q}^{L_{N_{k(i)}}}_{N_{k(i)},p} - Q^{L^*}_p| > \epsilon\right) < \infty.$$

Proof. We work with the subsequence $\{N_{k(i)}\}_{i\geq 1}$ given by Lemma 1 and make a trivial decomposition:

$$\mathbb{P}\left(|\hat{Q}^{L_{N_{k(i)}}}_{N_{k(i)},p} - Q^{L^*}_p| > \epsilon\right) = \mathbb{P}\left(\hat{Q}^{L_{N_{k(i)}}}_{N_{k(i)},p} - Q^{L^*}_p > \epsilon\right) + \mathbb{P}\left(Q^{L^*}_p - \hat{Q}^{L_{N_{k(i)}}}_{N_{k(i)},p} > \epsilon\right). \tag{A.15}$$

Recall, that $\hat{Q}_{N_{k(i)},p}^{L_{N_{k(i)}}+1}$ is the first quantile estimation which lies below $\bar{a}$. In the first term in r.h.s of Equation (A.15), if $\hat{Q}_{N_{k(i)},p}^{L_{N_{k(i)}}} > Q_p^{L^*} + \epsilon$ and $\hat{Q}_{N_{k(i)},p}^{L_{N_{k(i)}}+1} \leq \bar{a}$, then there is no $\hat{Q}_{N_{k(i)},p}^{l}$ which lies in the interval $]Q_p^{L^*} - \delta, Q_p^{L^*} + \delta[$ with $\delta = \min\{\epsilon, Q_p^{L^*} - \bar{a}\} > 0$. So $\{\hat{Q}_{N_{k(i)},p}^{L_{N_{k(i)}}} - Q_p^{L^*} > \epsilon\}$ implies $\{|\hat{Q}_{N_{k(i)},p}^{L^*} - Q_p^{L^*}| > \delta\}$ and we have

$$\mathbb{P}\left(\hat{Q}_{N_{k(i)},p}^{L_{N_{k(i)}}} - Q_p^{L^*} > \epsilon\right) \leq \mathbb{P}\left(|\hat{Q}_{N_{k(i)},p}^{L^*} - Q_p^{L^*}| > \delta\right).$$

Next, we make another decomposition:

$$\mathbb{P}\left(Q_p^{L^*} - \hat{Q}_{N_{k(i)},p}^{L_{N_{k(i)}}} > \epsilon\right) \leq \mathbb{P}\left(Q_p^{L^*} - \hat{Q}_{N_{k(i)},p}^{L_{N_{k(i)}}} > \epsilon, |\hat{Q}_{N_{k(i)},p}^{L^*} - Q_p^{L^*}| \leq \epsilon\right) + \mathbb{P}\left(|\hat{Q}_{N_{k(i)},p}^{L^*} - Q_p^{L^*}| > \epsilon\right).$$

On the joint event in the first probability in the above r.h.s. inequality, we must have $\hat{Q}_{N_{k(i)},p}^{L_{N_{k(i)}}} < \hat{Q}_{N_{k(i)},p}^{L^*}$, and consequently $\hat{Q}_{N_{k(i)},p}^{L^*+1} > \bar{a}$ (by definition of $\hat{Q}_{N_{k(i)},p}^{L_{N_{k(i)}}}$ as the last quantile estimation above $\bar{a}$). Thus, it follows that

$$\mathbb{P}\left(Q_p^{L^*} - \hat{Q}_{N_{k(i)},p}^{L_{N_{k(i)}}} > \epsilon, |\hat{Q}_{N_{k(i)},p}^{L^*} - Q_p^{L^*}| \leq \epsilon\right) \leq \mathbb{P}\left(|\hat{Q}_{N_{k(i)},p}^{L^*+1} - Q_p^{L^*+1}| > \bar{a} - Q_p^{L^*+1}\right).$$

We are able to conclude the proof by collecting the above results and using Lemma 1 with $l = L^*$, $l = L^* + 1$ and various $\epsilon > 0$. □

Lemma 3. *When $\frac{\log \alpha}{\log p}$ is not an integer, there exists a subsequence $\{N_{k(i)}\}_{i\geq 1}$ such that for any $\epsilon > 0$ we have*

$$\sum_{i\geq 1} \mathbb{P}\left(|\hat{r}_{N_{k(i)}}(\hat{Q}_{N_{k(i)},p}^{L_{N_{k(i)}}}) - r(Q_p^{L^*})| > \epsilon\right) < \infty.$$

Consequently, $\hat{r}_{N_{k(i)}}(\hat{Q}_{N_{k(i)},p}^{L_{N_{k(i)}}})$ converges to $r(Q_p^{L^*})$ almost surely as $i \to \infty$.

Proof. Assume first that $L^* \geq 1$. Then, we have the existence of a subsequence $\{N_{k(i)}\}_{i\geq 1}$ such that the result in Lemma 2 holds. For this subsequence on $\{\hat{Q}_{N_{k(i)},p}^{L_{N_{k(i)}}} \in J_{L^*}\}$, due to Assumption 3, we have

$$\mathbb{P}\left(\left|\hat{r}_{N_{k(i)}}(\hat{Q}_{N_{k(i)},p}^{L_{N_{k(i)}}}) - r(\hat{Q}_{N_{k(i)},p}^{L_{N_{k(i)}}})\right| > \varepsilon | \hat{Q}_{N_{k(i)},p}^{L_{N_{k(i)}}}\right) \leq c(\hat{Q}_{N_{k(i)},p}^{L_{N_{k(i)}}}, N_{k(i)}, \varepsilon).$$

The term in the above r.h.s. is bounded by $\sup_{s\in J_{L^*}} c(s, N_{k(i)}, \varepsilon)$ and we can find a further subsequence $\{N_{\tilde{k}(i)}\}_{i\geq 1}$ such that on $\{\hat{Q}_{N_{\tilde{k}(i)},p}^{L_{N_{\tilde{k}(i)}}} \in J_{L^*}\}$, we have

$$\sum_{i\geq 1} \mathbb{P}\left(\left|\hat{r}_{N_{\tilde{k}(i)}}(\hat{Q}_{N_{\tilde{k}(i)},p}^{L_{N_{\tilde{k}(i)}}}) - r(\hat{Q}_{N_{\tilde{k}(i)},p}^{L_{N_{\tilde{k}(i)}}})\right| > \varepsilon | \hat{Q}_{N_{\tilde{k}(i)},p}^{L_{N_{\tilde{k}(i)}}}\right) \leq \sum_{i\geq 1} \sup_{s\in J_{L^*}} c(s, N_{\tilde{k}(i)}, \varepsilon) < \infty. \tag{A.16}$$

Thus, we use the subsequence $\{N_{\tilde{k}(i)}\}_{i\geq 1}$ and make the following decomposition

$$\begin{aligned}
&\mathbb{P}\left(\left|\hat{r}_{N_{\tilde{k}(i)}}(\hat{Q}_{N_{\tilde{k}(i)},p}^{L_{N_{\tilde{k}(i)}}}) - r(Q_p^{L^*})\right| > \varepsilon\right) \\
&\leq \mathbb{P}\left(\left|\hat{r}_{N_{\tilde{k}(i)}}(\hat{Q}_{N_{\tilde{k}(i)},p}^{L_{N_{\tilde{k}(i)}}}) - r(\hat{Q}_{N_{\tilde{k}(i)},p}^{L_{N_{\tilde{k}(i)}}})\right| > \varepsilon/2\right) + \mathbb{P}\left(\left|r(\hat{Q}_{N_{\tilde{k}(i)},p}^{L_{N_{\tilde{k}(i)}}}) - r(Q_p^{L^*})\right| > \varepsilon/2\right) \\
&\leq \mathbb{P}\left(\left|\hat{r}_{N_{\tilde{k}(i)}}(\hat{Q}_{N_{\tilde{k}(i)},p}^{L_{N_{\tilde{k}(i)}}}) - r(\hat{Q}_{N_{\tilde{k}(i)},p}^{L_{N_{\tilde{k}(i)}}})\right| > \varepsilon/2, \hat{Q}_{N_{\tilde{k}(i)},p}^{L_{N_{\tilde{k}(i)}}} \in J_{L^*}\right) \\
&+ \mathbb{P}\left(\hat{Q}_{N_{\tilde{k}(i)},p}^{L_{N_{\tilde{k}(i)}}} \notin J_{L^*}\right) + \mathbb{P}\left(\left|r(\hat{Q}_{N_{\tilde{k}(i)},p}^{L_{N_{\tilde{k}(i)}}}) - r(Q_p^{L^*})\right| > \varepsilon/2\right).
\end{aligned} \tag{A.17}$$

The first term in the right hand side of (A.17) is handled by the result (A.16) such that we have

$$\sum_{i\geq 1}\mathbb{P}\left(\left|\hat{r}_{N_{\tilde{k}(i)}}(\hat{Q}^{L_{N_{\tilde{k}(i)}}}_{N_{\tilde{k}(i)},p})-r(\hat{Q}^{L_{N_{\tilde{k}(i)}}}_{N_{\tilde{k}(i)},p})\right|>\varepsilon/2,\hat{Q}^{L_{N_{\tilde{k}(i)}}}_{N_{\tilde{k}(i)},p}\in J_{L^*}\right)<\infty.$$

The second term in the r.h.s. of (A.17) also gives a convergent series in view of Lemma 2 and the last term is handled like the second term by noting that $r(s)$ is continuous at $s=Q_p^{L^*}$ (Assumption 1). Now consider the case $L^*=0$ and write

$$\mathbb{P}\left(\left|\hat{r}_N(\hat{Q}^{L_N}_{N,p})-r(Q_p^{L^*})\right|>\varepsilon\right)\leq\mathbb{P}\left(\left|\hat{r}_N(\hat{Q}^{0}_{N,p})-r(Q_p^{0})\right|>\varepsilon,L_N=0\right)+\mathbb{P}\left(L_N\neq 0\right).$$

The existence of a subsequence such that the series formed by the first probability term in the above r.h.s. converges over this subsequence directly follows from Assumption 3. Moreover, by definition of L_N and since $L^*=0$, $\{L_N\neq 0\}\subset\{\hat{Q}^1_{N,p}>\bar{a}\}\subset\{|\hat{Q}^1_{N,p}-Q^1_p|>\bar{a}-Q^1_p>0\}$: we conclude by Lemma 1 with $l=1$. □

Proof of Theorem 9, when $\log\alpha/\log p$ is not an integer. We use the results from Corollary 1 and Lemma 3 to conclude from Theorem 12 that estimator $\hat{\alpha}_N$ is consistent. □

Next, we prove the convergence when $\log\alpha/\log p$ is an integer. This case needs to be dealt with separately as we no longer have convergence of L_N to L^*. When $\alpha=p^{L^*}$, the estimator can be expressed as

$$\hat{\alpha}_N=1_{\{L_N=L^*-1\}}\hat{r}_N(\hat{Q}^{L^*-1}_{N,p})p^{L^*-1}+1_{\{L_N=L^*\}}\hat{r}_N(\hat{Q}^{L^*}_{N,p})p^{L^*}+1_{\{L_N\notin\{L^*-1,L^*\}\}}\hat{r}_N(\hat{Q}^{L_N}_{N,p})p^{L_N}.$$

Then, the error of our estimator is given as:

$$\begin{aligned}\hat{\alpha}_N-p^{L^*}=&1_{\{L_N=L^*-1\}}\left(\hat{r}_N(\hat{Q}^{L^*-1}_{N,p})-p\right)p^{L^*-1}\\&+1_{\{L_N=L^*\}}\left(\hat{r}_N(\hat{Q}^{L^*}_{N,p})-1\right)p^{L^*}\\&+1_{\{L_N\notin\{L^*-1,L^*\}\}}\left(\hat{r}_N(\hat{Q}^{L_N}_{N,p})p^{L_N}-p^{L^*}\right).\end{aligned}\tag{A.18}$$

Lemma 4. *If $\alpha=p^{L^*}$, there exists a subsequence $\{N_{k(i)}\}_{i\geq 1}$ such that*

$$\mathbb{P}\left(L_{N_{k(i)}}\in\{L^*-1,L^*\},\ \textit{for } i \textit{ large enough}\right)=1.$$

Proof. Note that $Q_p^{L^*+1}<\bar{a}$ always holds. From Lemma 1, we have a subsequence $\{N_{k(i)}\}_{i\geq 1}$ such that $\hat{Q}^{L^*+1}_{N_{k(i)},p}$ converges almost surely to $Q_p^{L^*+1}$. By the definition of L_N we have $L_{N_{k(i)}}+1\leq L^*+1$ as $i\to\infty$. Similarly, provided that $L^*>1$, $\hat{Q}^{L^*-1}_{N_{k(i)},p}$ converges almost surely to $Q_p^{L^*-1}>\bar{a}$, thus $L_{N_{k(i)}}\geq L^*-1$ as $i\to\infty$. □

Lemma 5. *For $l\in\{L^*-1,L^*\}\cap\mathbb{N}$, there exists a subsequence $\{N_{k(i)}\}_{i\geq 1}$ such that for any $\epsilon>0$, we have*

$$\sum_{i\geq 1}\mathbb{P}\left(|\hat{r}_{N_{k(i)}}(\hat{Q}^{l}_{N_{k(i)},p})-r(Q^l_p)|>\epsilon\right)<\infty.$$

Thus for such l, $\hat{r}_{N_{k(i)}}(\hat{Q}^{l}_{N_{k(i)},p})$ converges to $r(Q^l_p)$ almost surely as $i\to\infty$.

Proof. Similar to (A.17), we write for a subsequence $\{N_{k(i)}\}_{i\geq 1}$ which satisfies the result in Lemma 1 and 3,

$$\begin{aligned}\mathbb{P}\left(\left|\hat{r}_{N_{k(i)}}(\hat{Q}^{l}_{N_{k(i)},p})-r(Q^l_p)\right|>\varepsilon\right)\leq&\mathbb{P}\left(\left|\hat{r}_{N_{k(i)}}(\hat{Q}^{l}_{N_{k(i)},p})-r(\hat{Q}^l_{N_{k(i)},p})\right|>\varepsilon/2,\hat{Q}^l_{N_{k(i)},p}\in J_l\right)\\&+\mathbb{P}\left(\hat{Q}^l_{N_{k(i)},p}\notin J_l\right)+\mathbb{P}\left(\left|r(\hat{Q}^l_{N_{k(i)},p})-r(Q^l_p)\right|>\varepsilon/2\right).\end{aligned}$$

If $l = 0$, the two last probabilities on the above r.h.s. are 0, since $\hat{Q}^0_{N_{k(i)},p} = Q^0_p = \infty$, while the first probability forms a convergent series on a subsequence in view of Assumption 3.
If $l > 0$, we argue as in the proof of Lemma 1, by finding a common subsequence in view of Assumption 3, the result in Lemma 1 and the local continuity of $r(\cdot)$. □

Proof of Theorem 9, when $\log\alpha/\log p$ is an integer. In the r.h.s. of Equation (A.18), we consider the subsequence $\{N_{k(i)}\}_{i\geq 1}$ which satisfies the result in Lemma 5 for $l = L^* - 1$ and $l = L^*$, such that we get $\hat{r}_{N_{k(i)}}(\hat{Q}^{L^*-1}_{N_{k(i)},p}) - p = \hat{r}_{N_{k(i)}}(\hat{Q}^{L^*-1}_{N_{k(i)},p}) - r(Q^{L^*-1}_p)$ converges to zero almost surely and that $\hat{r}_{N_{k(i)}}(\hat{Q}^{L^*}_{N_{k(i)},p}) - 1 = \hat{r}_{N_{k(i)}}(\hat{Q}^{L^*}_{N_{k(i)},p}) - r(Q^{L^*}_p)$ converges to zero almost surely, respectively. Finally, we apply Lemma 4 and Theorem 12 to complete the proof. □