


RESEARCH ARTICLE

Signature-based validation of real-world economic scenarios

Hervé Andrès^{1,2} , Alexandre Boumezoued¹ and Benjamin Jourdain²

¹Milliman R&D, Paris, France

²CERMICS, Ecole des Ponts, INRIA, Marne-la-Vallée, France

Corresponding author: Hervé Andrès; Email: hervé.andrès@milliman.com

Received: 17 February 2023; **Revised:** 29 February 2024; **Accepted:** 4 March 2024; **First published online:** 4 April 2024

Keywords: Real-world economic scenarios; economic scenarios validation; insurance; signature; maximum mean distance.

Abstract

Motivated by insurance applications, we propose a new approach for the validation of real-world economic scenarios. This approach is based on the statistical test developed by Chevyrev and Oberhauser ((2022) *Journal of Machine Learning Research*, 23(176), 1–42.) and relies on the notions of signature and maximum mean distance. This test allows to check whether two samples of stochastic processes paths come from the same distribution. Our contribution is to apply this test to a variety of stochastic processes exhibiting different pathwise properties (Hölder regularity, autocorrelation, and regime switches) and which are relevant for the modelling of stock prices and stock volatility as well as of inflation in view of actuarial applications.

1. Introduction

Real-world economic scenarios provide stochastic forecasts of economic variables like interest rates, equity returns, and inflation and are widely used in the insurance sector for a variety of applications including asset and liability management (ALM) studies, strategic asset allocation, computing the solvency capital requirement (SCR) within an internal model or pricing assets or liabilities including a risk premium. Unlike risk-neutral economic scenarios that are used for market consistent pricing, real-world economic scenarios aim at being realistic in view of the historical data and/or expert expectations about future outcomes. In the literature, many real-world models have been studied for applications in insurance. Those applications relate to (i) valuation of insurance products, (ii) hedging strategies for annuity portfolios and (iii) risk calculation for economic capital assessment. On item (i), we can mention the work of Boudreault and Panneton (2009) who study the impact on conditional tail expectation provision of GARCH and regime-switching models calibrated on historical data, and the work of Graf *et al.* (2014) who perform simulations under the real-world probability measure to estimate the risk-return profile of various old-age provision products. On item (ii), Zhu *et al.* (2018) measure the hedging error of several dynamic hedging strategies along real-world scenarios for cash balance pension plans, while Lin and Yang (2020) calculate the value of a large variable annuity portfolio and its hedge using nested simulations (real-world scenarios for the outer simulations and risk-neutral scenarios for the inner simulations). Finally, on item (iii), Hardy *et al.* (2006) compare several real-world models for the equity return process in terms of fitting quality and resulting capital requirements and discuss the problem of the validation of real-world scenarios. Similarly, Otero *et al.* (2012) measure the impact on the Solvency II capital requirements (SCR) when using a regime-switching model in comparison to lognormal, GARCH and E-GARCH models. Floryszczak *et al.* (2019) introduce a simple model for equity returns allowing to avoid over-assessment of the SCR specifically after market disruptions. On the other hand, Asadi and Al Janabi (2020) propose a more complex model for stocks based on ARMA and GARCH processes that

results in a higher SCR than in the Solvency II standard model. This literature shows the importance of real-world economic scenarios in various applications in insurance. We observe that the question of the consistency of the generated real-world scenarios is barely discussed or only from a specific angle such as the model likelihood or the ability of the model to reproduce the 1 in 200 worst shock observed on the market.

In the insurance industry, the assessment of the realism of real-world economic scenarios is often referred to as scenario validation. It allows to verify a posteriori the consistency of a given set of real-world economic scenarios with historical data and/or expert views. As such, it also guides which models can better be used to generate real-world economic scenarios. In the risk-neutral framework, the validation step consists for example in verifying the martingale property of the discounted values along each scenario. In the real-world framework, the most widespread practice is to perform a so-called point-in-time validation. It consists in analyzing the distribution of some variables derived from the generated scenarios (for example, annual log-returns for equity stocks or relative variation for an inflation index) at some specific horizons like one year which is the horizon considered in the Solvency II directive. Generally, this analysis only focuses on the first moments of the one-year distribution as real-world models are often calibrated by a moment-matching approach. The main drawback of this approach is that it only allows to capture properties of the simulated scenarios at some point in time. In particular, the consistency of the paths between $t = 0$ and $t = 1$ year is not studied so that properties like clustering, smoothness and high-order autocorrelation are not captured. Capturing these properties has its importance as their presence or absence in the economic scenarios can have an impact in real practice, such as for example on a strategic asset allocation having a monthly rebalancing frequency or on the SCR calculation when a daily hedging strategy is involved, since the yearly loss distribution will be path-dependent. In this paper, we propose to address this drawback by comparing the distribution of the stochastic process underlying the simulated paths to the distribution of the historical paths. This can be done using a distance between probability measures, called the maximum mean distance (MMD), and a mathematical object, called the signature, allowing to encode a continuous path in an efficient and parsimonious way. Based on these tools, Chevyrev and Oberhauser (2022) designed a statistical test allowing to accept or reject the hypothesis that the distributions of two samples of paths are equal. This test has already been used by Buehler *et al.* (2020) to test whether financial paths generated by a conditional variational auto encoder (CVAE) are close to the historical paths being used to train the CVAE. An alternative way to compare the distributions of two sample of paths is to flatten each sequence of observations into a long vector of length $d \times L$, where L is the length of the sequence of observations and d is the dimension of each observation, and to apply a multi-variate statistical test. However, Chevyrev and Oberhauser have shown that their signature-based test performs overall better (both in terms of statistical power and in terms of computational cost) than standard multi-variate tests on a collection of multidimensional time series data sets. Moreover, this alternative approach requires that each sequence of observations is of the same length which is not a prerequisite in the case of the signature-based test.

Our contribution is to study more deeply this statistical test from a numerical point of view on a variety of stochastic models and to show its practical interest for the validation of real-world stochastic scenarios when this validation is specified as an hypothesis testing problem. First, we present a numerical analysis with synthetic data in order to measure the statistical power of the test, and second, we work with historical data to study the ability of the test to discriminate between several models in practice. Moreover, two constraints are considered in the numerical experiments. The first one is to impose that the distributions of the annual increments are the same in the two compared samples, which is natural since, without this constraint, point-in-time validation methods could distinguish the two samples. Second, in order to mimic the operational process of real-world scenarios validation in insurance, we consider samples of different sizes: the first sample consisting of synthetic or real historical paths is of small size (typically below 50) while the second sample consisting of the simulated scenarios is of greater size (typically around 1000). Our aim is to demonstrate the high statistical power of the test under these constraints. Numerical results are presented for three financial risk drivers, namely, of a stock price, stock volatility

Table 1. Summary of the studied risk factors and associated models in each framework (synthetic data and real historical data).

	Risk factor	Models
Synthetic data	Stock price	Black–Scholes dynamics and rough Heston model
	Stock volatility	Fractional Brownian motion
	Inflation	Regime-switching AR(1) process and Gamma random walk
Historical data	Stock volatility	Ordinary and fractional Ornstein–Uhlenbeck processes
	Inflation	Regime-switching AR(1) process and Gamma random walk

as well as inflation. As for the stock price, we first generate two samples of paths under the widespread Black–Scholes dynamics, each sample corresponding to a specific choice of the (deterministic) volatility structure. We also simulate two samples of paths under both the classic and rough Heston models. As for the stock volatility, the two samples are generated using fractional Brownian motions with different Hurst parameters. Note that the exponential of a fractional Brownian motion, as model for volatility, has been proposed by Gatheral *et al.* (2018) who showed that such a model is consistent with historical data. For the inflation, one sample is generated using a regime-switching $AR(1)$ process and the other sample is generated using a random walk with i.i.d. Gamma noises. Finally, we compare two samples of two-dimensional processes with either independent or correlated coordinates when the first coordinate evolves according to the price in the rough Heston model and the second coordinate evolves according to an $AR(1)$ regime-switching process. Besides these numerical results on simulated paths, we also provide numerical results on real historical data. More specifically, we test historical paths of S&P 500 realized volatility (used as a proxy of spot volatility) against sample paths from a standard Ornstein–Uhlenbeck model on the one hand and against sample paths from a fractional Ornstein–Uhlenbeck model on the other hand. We show that the test allows to reject the former model while the latter is not rejected. Similarly, it allows to reject a random walk model with i.i.d. Gamma noises when applied to US inflation data, while a regime-switching $AR(1)$ process is not rejected. A summary of the studied risk factors and associated models is provided in Table 1.

Remark 1.1. In the literature, the closest work to ours is the one of Bilokon *et al.* (2021). Given a set of paths, they propose to compute the signature of these paths in order to apply the clustering algorithm of Azran and Ghahramani (2006) and to ultimately identify market regimes. The similarity function underlying the clustering algorithm relies in particular on the MMD Using Black–Scholes sample paths corresponding to four different configurations of the drift and volatility parameters, they show the ability of their methodology to correctly cluster the paths and to identify the number of different configurations (i.e. four). Let us point out that our contributions are different from theirs in several ways. First, their objective is to cluster a set of paths in several groups, while our objective is to statistically test whether two sets of paths come from the same probability distribution. Second, our numerical results go beyond the Black–Scholes model and explore more sophisticated models. Moreover, we also provide numerical results on historical data. Finally, our numerical experiments are conducted in a setting closer to our practical applications where the marginal one-year distributions of the two samples are the same or very close, while the frequency of observation of the paths is lower in our case (we consider monthly observations over one year while they consider 100 observations).

The objective of the present article is also to provide a concise introduction to the signature theory that does not require any prerequisite for insurance practitioners. Introduced for the first time by Chen (1957) in the late 50s and then rediscovered in the 90s in the context of rough path theory (Lyons, 1998), the signature is a mapping that allows to characterize deterministic paths up to some equivalence relation (see Theorem B.2. in the Supplementary Material). Chevyrev and Oberhauser (2022) have extended this result to stochastic processes as they have shown that the expected signature of a stochastic process

characterizes its law. The idea to use the signature to address problems in finance is not new although it is quite recent. To our knowledge, Gyurkó and Lyons (2010) are the first in this area. They present a general framework for deriving high order, stable and tractable pathwise approximations of stochastic differential equations relying on the signature and apply their results to the simulation of the Cox–Ingersoll–Ross process. Then, Gyurkó *et al.* (2013) introduced the signature as a way to obtain a faithful transform of financial data streams that is used as a feature of a classification method. Levin *et al.* (2013) use the signature to study the problem of regression where the input and the output variables are paths and illustrate their results by considering the prediction task for AR and ARCH time series models. In the same vein, Cohen *et al.* (2023) address the nowcasting problem – the problem of computing early estimates of economic indicators that are published with a delay such as the GDP – by applying a regression on the signature of frequently published indicators. Ni *et al.* (2021) develop a generative adversarial network (GAN) based on the signature allowing to generate time series that capture the temporal dependence in the training and validation data set both for synthetic and real data. In his PhD thesis, Perez Arribas (2020) shows several applications of the signature in finance including the pricing and hedging of exotic derivatives (see Lyons *et al.*, 2020) or optimal execution problems (see Kalsi *et al.*, 2020 and Cartea *et al.*, 2022). Cuchiero *et al.* (2023b) extend the work of Perez Arribas and develop a new class of asset price models based on the signature of semimartingales allowing to approximate arbitrarily well classical models such as the SABR and the Heston models. Using this new modelling framework, Cuchiero *et al.* (2023a) propose a method to solve the joint S&P 500/VIX calibration problem without adding jumps or rough volatility. Akyildirim *et al.* (2022) introduce a signature-based machine learning algorithm to detect rare or unexpected items in a given data set of time series type. Cuchiero and Möller (2023) employ the signature in the context of stochastic portfolio theory to introduce a novel class of portfolios making in particular several optimization tasks highly tractable. Finally, Bayer *et al.* (2023) propose a new method relying on the signature for solving optimal stopping problems.

The present article is organized as follows: as a preliminary, we introduce in Section 2 the MMD and the signature before describing the statistical test proposed by Chevyrev and Oberhauser (2022). This test is based on these two notions and allows to assess whether two stochastic processes have the same law using finite numbers of their sample paths. Then in Section 3, we study this test from a numerical point of view. We start by studying its power using synthetic data in settings that are realistic in view of insurance applications and then, we apply it to real historical data. We also discuss several challenges related to the numerical implementation of this approach and highlight its domain of validity in terms of the distance between models and the volume of data at hand.

2. From the MMD and the signature to a two-sample test for stochastic processes

In this section, we start by introducing the MMD, which allows to measure how similar two probability measures are. Second, reproducing Kernel Hilbert spaces (RKHS) are presented as they are key to obtain a simple formula for the MMD. Then, we briefly introduce the signature and we show how it allows to construct an RKHS that we can use to make the MMD a metric able to discriminate two probability measures defined on the bounded variation paths quotiented by some equivalence relation. Finally, the statistical test underlying the signature-based validation is introduced. In what follows, \mathcal{X} is a metric space.

2.1. The maximum mean distance

Definition 2.1 (Maximum Mean Distance). *Let \mathcal{G} be a class of functions $f : \mathcal{X} \rightarrow \mathbb{R}$ and μ, ν two Borel probability measures defined on \mathcal{X} . The MMD is defined as*

$$MMD_{\mathcal{G}}(\mu, \nu) = \sup_{f \in \mathcal{G}} \left| \int_{\mathcal{X}} f(x)\mu(dx) - \int_{\mathcal{X}} f(x)\nu(dx) \right|. \tag{2.1}$$

Depending on \mathcal{G} , the MMD is not necessarily a metric (actually, it is a pseudo-metric, that is a metric without the property that two points with zero distance are identical), that is we could have $MMD_{\mathcal{G}}(\mu, \nu) = 0$ for some $\mu \neq \nu$ if the class of functions \mathcal{G} is not rich enough. A sufficiently rich class of functions that makes $MMD_{\mathcal{G}}$ a metric is for example the space of bounded continuous functions on \mathcal{X} equipped with a metric d (Lemma 9.3.2 of Dudley, 2002). A sufficient condition on \mathcal{G} for $MMD_{\mathcal{G}}$ to be a metric is given in Theorem 5 of Gretton *et al.* (2012).

As presented in Definition 2.1, the MMD appears more as a theoretical tool than a practical one since computing this distance seems impossible in practice due to the supremum over a class of functions. However, if this class of function is the unit ball in a reproducing kernel Hilbert space (RKHS), the MMD is much simpler to estimate. Before setting out this result precisely, let us make a quick reminder about Mercer kernels and RKHSs.

Definition 2.2 (Mercer kernel). *A mapping $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is called a Mercer kernel if it is continuous, symmetric and positive semi-definite that is for all finite sets $\{x_1, \dots, x_k\} \subset \mathcal{X}$ and for all $(\alpha_1, \dots, \alpha_k) \in \mathbb{R}^k$, the kernel K satisfies:*

$$\sum_{i=1}^k \sum_{j=1}^k \alpha_i \alpha_j K(x_i, x_j) \geq 0. \tag{2.2}$$

Remark 2.1. In the kernel learning literature, it is common to use the terminology “positive definite” instead of “semi-positive definite” but we prefer the latter one in order to be consistent with the linear algebra standard terminology.

We set:

$$\mathcal{H}_0 = \text{span}\{K_x := K(x, \cdot) \mid x \in \mathcal{X}\}. \tag{2.3}$$

With these notations, we have the following theorem due to Moore-Aronszajn (see Theorem 2 of Chapter III in Cucker and Smale, 2002):

Theorem 2.1 (Moore-Aronszajn). *Let K be a Mercer kernel. Then, there exists a unique Hilbert space $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ with scalar product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ satisfying the following conditions:*

- (i) \mathcal{H}_0 is dense in \mathcal{H}
- (ii) For all $f \in \mathcal{H}$, $f(x) = \langle K_x, f \rangle_{\mathcal{H}}$ (reproducing property).

Remark 2.2. The obtained Hilbert space \mathcal{H} is said to be a reproducing kernel Hilbert space (RKHS) whose reproducing kernel is K .

We may now state the main theorem about the MMD.

Theorem 2.2. *Let (\mathcal{H}, K) be a reproducing kernel Hilbert space and let $\mathcal{G} := \{f \in \mathcal{H} \mid \|f\|_{\mathcal{H}} \leq 1\}$. If $\int_{\mathcal{X}} \sqrt{K(x, x)} \mu(dx) < \infty$ and $\int_{\mathcal{X}} \sqrt{K(x, x)} \nu(dx) < \infty$, then for X, X' independent variables distributed according to μ and Y, Y' independent random variables distributed according to ν and such that X and Y are independent, $K(X, X')$, $K(Y, Y')$ and $K(X, Y)$ are integrable and:*

$$MMD_{\mathcal{G}}^2(\mu, \nu) = \mathbb{E}[K(X, X')] + \mathbb{E}[K(Y, Y')] - 2\mathbb{E}[K(X, Y)]. \tag{2.4}$$

The proof of this theorem can be found in Lemma 6 of Gretton *et al.* (2012). A natural question at this stage is that of the choice of the Mercer kernel in order to obtain a metric on the space of probability measures defined on the space of continuous mappings from $[0, T]$ to a finite dimensional vector space or at least on a non-trivial subspace of this space. Chevyrev and Oberhauser (2022) constructed such a Mercer kernel using the signature, which we will now define.

2.2. The signature

This subsection aims at providing a short overview of the signature, more details are given in the Supplementary Material. We call “path” any continuous mapping from some time interval $[0, T]$ to

a finite dimensional vector space E which we equip with a norm $\|\cdot\|_E$. We denote by \otimes the tensor product defined on $E \times E$ and by $E^{\otimes n}$ the tensor space obtained by taking the tensor product of E with itself n times:

$$E^{\otimes n} = \underbrace{E \otimes \cdots \otimes E}_{n \text{ times}}. \tag{2.5}$$

The space in which the signature takes its values is called the space of formal series of tensors. It can be defined as

$$T(E) = \{(\mathbf{t}^n)_{n \geq 0} \mid \forall n \geq 0, \mathbf{t}^n \in E^{\otimes n}\} \tag{2.6}$$

with the convention $E^{\otimes 0} = \mathbb{R}$.

In a nutshell, the signature of a path X is the collection of all iterated integrals of X against itself. In order to be able to define these iterated integrals of X , one needs to make some assumptions about the regularity of X . The simplest framework is to assume that X is of bounded variation.

Definition 2.3 (Bounded variation path). *We say that a path $X : [0, T] \rightarrow E$ is of bounded variation on $[0, T]$ if its total variation*

$$\|X\|_{1,[0,T]} := \sup_{(t_0, \dots, t_r) \in \mathcal{D}} \sum_{i=0}^{r-1} \|X_{t_{i+1}} - X_{t_i}\|_E \tag{2.7}$$

is finite with $\mathcal{D} = \{(t_0, \dots, t_r) \mid r \in \mathbb{N}^*, t_0 = 0 < t_1 < \dots < t_r = T\}$ the set of all subdivisions of $[0, T]$.

Notation 2.1. *We denote by $\mathcal{C}^1([0, T], E)$ the set of bounded variation paths from $[0, T]$ to E .*

Remark 2.3. Intuitively, a bounded variation path on $[0, T]$ is a path whose graph vertical arc length is finite. In fact, if X is a real-valued continuously differentiable path on $[0, T]$, then $\|X\|_{1,[0,T]} = \int_0^T |X'_t| dt$.

We can now define the signature of a bounded variation path.

Definition 2.4 (Signature). *Let $X : [0, T] \rightarrow E$ be a bounded variation path. The signature of X on $[0, T]$ is defined as $S_{[0,T]}(X) = (\mathbf{X}^n)_{n \geq 0}$ where by convention $\mathbf{X}^0 = 1$ and*

$$\mathbf{X}^n = \int_{0 \leq u_1 < u_2 < \dots < u_n \leq T} dX_{u_1} \otimes \cdots \otimes dX_{u_n} \in E^{\otimes n}. \tag{2.8}$$

where the integrals must be understood in the sense of Riemann–Stieljes. We call \mathbf{X}^n the term of order n of the signature and $S_{[0,T]}^N(X) = (\mathbf{X}^n)_{0 \leq n \leq N}$ the truncated signature at order N . Note that when the time interval is clear from the context, we will omit the subscript of S .

Example 2.1. *If X is a one-dimensional bounded variation path, then its signature over $[0, T]$ is very simple as it reduces to the powers of the increment $X_T - X_0$, that is for any $n \geq 0$:*

$$\mathbf{X}^n = \frac{1}{n!} (X_T - X_0)^n. \tag{2.9}$$

The above definition could be extended to less regular paths, namely to paths of finite p -variation with $p < 2$. In this case, the integrals can be defined in the sense of Young (1936). However, if $p \geq 2$, it is no longer possible to define the iterated integrals. Still, it is possible to give a sense to the signature but the definition is much more involved and relies on the rough path theory so we refer the interested reader to Lyons *et al.* (2007).

In this work, we will focus on bounded variation paths where the signature takes its values in the space of finite formal series (as a consequence of Proposition 2.2 of Lyons *et al.*, 2007):

$$T^*(E) := \left\{ \mathbf{t} \in T(E) \mid \|\mathbf{t}\| := \sqrt{\sum_{n \geq 0} \|\mathbf{t}^n\|_{E^{\otimes n}}^2} < \infty \right\} \tag{2.10}$$

where $\|\cdot\|_{E^{\otimes n}}$ is the norm induced by the scalar product $\langle \cdot, \cdot \rangle_{E^{\otimes n}}$ defined on $E^{\otimes n}$ by:

$$\langle x, y \rangle_{E^{\otimes n}} = \sqrt{\sum_{I=(i_1, \dots, i_n) \in \{1, \dots, d\}^n} x_I y_I} \quad \text{for } x, y \in E^{\otimes n} \tag{2.11}$$

with d the dimension of E and x_I (resp. y_I) the coefficient at position I of x (resp. y).

The signature is a powerful tool allowing to encode a path in a hierarchical and efficient manner. In fact, two bounded variation paths having the same signature are equal up to an equivalence relation (the so-called tree-like equivalence, denoted by \sim_t , and defined in Appendix B of the Supplementary Material). In other words, the signature is one-to-one on the space $\mathcal{P}^1([0, T], E) := \mathcal{C}^1([0, T], E) / \sim_t$ of bounded variation paths quotiented by the tree-like equivalence relation. This is presented in a more comprehensive manner in Appendix B of the Supplementary Material. Now, we would like to characterize the law of stochastic processes with bounded variation sample paths using the expected signature, that is the expectation of the signature taken component-wise. In a way, the expected signature is to stochastic processes what the sequence of moments is to random vectors. Thus, in the same way that the sequence of moments characterizes the law of random vectors only if the moments do not grow too fast, we need that the terms of the expected signature do not grow too fast in order to be able to characterize the law of stochastic processes. In order to avoid to have to restrict the study to laws with compact support (as assumed by Fawcett, 2002), Chevyrev and Oberhauser (2022) propose to apply a normalization mapping to the signature ensuring that the norm of the normalized signature is bounded. This property allows them to prove the characterization of the law of a stochastic process by its expected normalized signature (Theorem C.1. in the Supplementary Material). One of the consequences of this result is the following theorem, which makes the connection between the MMD and the signature and represents the main theoretical result underlying the signature-based validation. This result is a particular case of Theorem 30 from Chevyrev and Oberhauser (2022).

Theorem 2.3. *Let E be a Hilbert space and $\langle \cdot, \cdot \rangle$ the scalar product on $T^*(E)$ defined by $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n \geq 0} \langle \mathbf{x}^n, \mathbf{y}^n \rangle_{E^{\otimes n}}$ for all \mathbf{x} and \mathbf{y} in $T^*(E)$. Then the signature kernel defined on $\mathcal{P}^1([0, T], E)$ by $K^{sig}(x, y) = \langle \Phi(x), \Phi(y) \rangle$ where Φ is the normalized signature (see Theorem C.1. in the Supplementary Material), is a Mercer kernel and we denote by \mathcal{H}^{sig} the associated RKHS. Moreover, $MMD_{\mathcal{G}}$ where \mathcal{G} is the unit ball of \mathcal{H}^{sig} is a metric on the space \mathcal{M} defined as:*

$$\mathcal{M} = \left\{ \mu \text{ Borel probability measure defined on } \mathcal{P}^1([0, T], E) \left| \int_{\mathcal{P}^1} \sqrt{K^{sig}(x, x)} \mu(dx) < \infty \right. \right\} \tag{2.12}$$

and we have:

$$MMD_{\mathcal{G}}(\mu, \nu) = \mathbb{E}[K^{sig}(X, X')] + \mathbb{E}[K^{sig}(Y, Y')] - 2\mathbb{E}[K^{sig}(X, Y)] \tag{2.13}$$

where X, X' are independent random variables distributed according to μ and Y, Y' are independent random variables distributed according to ν such that Y is independent from X .

Based on this theorem, Chevyrev and Oberhauser propose a two-sample statistical test that allows to test whether two samples of paths come from the same distribution and that we now introduce.

2.3. A two-sample test for stochastic processes

Assume that we are given a sample (X_1, \dots, X_m) consisting of m independent realizations of a stochastic process of unknown law \mathbb{P}_X and an independent sample (Y_1, \dots, Y_n) consisting of n independent realizations of a stochastic process of unknown law \mathbb{P}_Y . We assume that both processes are in $\mathcal{X} = \mathcal{P}^1([0, T], E)$ almost surely. A natural question is whether $\mathbb{P}_X = \mathbb{P}_Y$. Let us consider the following null and alternative hypotheses:

$$H_0 : \mathbb{P}_X = \mathbb{P}_Y, \quad H_1 : \mathbb{P}_X \neq \mathbb{P}_Y. \tag{2.14}$$

According to Theorem 2.3, we have $MMD_{\mathcal{G}}(\mathbb{P}_X, \mathbb{P}_Y) \neq 0$ under H_1 while $MMD_{\mathcal{G}}(\mathbb{P}_X, \mathbb{P}_Y) = 0$ under H_0 when \mathcal{G} is the unit ball of the reproducing Hilbert space associated to the signature kernel (note that we use the notation K instead of K^{sig} for the signature kernel in this section as there is no ambiguity). Moreover,

$$MMD_{\mathcal{G}}^2(\mathbb{P}_X, \mathbb{P}_Y) = \mathbb{E}[K(X, X')] + \mathbb{E}[K(Y, Y')] - 2\mathbb{E}[K(X, Y)] \tag{2.15}$$

where X, X' are two random variables of law \mathbb{P}_X and Y, Y' are two random variables of law \mathbb{P}_Y with X, Y independent. This suggests to consider the following test statistic:

$$MMD_{m,n}^2(X_1, \dots, X_m, Y_1, \dots, Y_n) := \frac{1}{m(m-1)} \sum_{1 \leq i \neq j \leq m} K(X_i, X_j) + \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} K(Y_i, Y_j) - \frac{2}{mn} \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} K(X_i, Y_j) \tag{2.16}$$

as it is an unbiased estimator of $MMD_{\mathcal{G}}^2(\mathbb{P}_X, \mathbb{P}_Y)$. In the sequel, we omit the dependency on X_1, \dots, X_m and Y_1, \dots, Y_n for notational simplicity. This estimator is even strongly consistent as stated by the following theorem which is an application of the strong law of large numbers for two-sample U -statistics (Sen, 1977).

Theorem 2.4 *Assuming that*

- (i) $\mathbb{E}[\sqrt{K(X, X)}] < \infty$ with X distributed according to \mathbb{P}_X and $\mathbb{E}[\sqrt{K(Y, Y)}] < \infty$ with Y distributed according to \mathbb{P}_Y
- (ii) $\mathbb{E}[|h| \log^+ |h|] < \infty$ where \log^+ is the positive part of the logarithm and

$$h = K(X, X') + K(Y, Y') - \frac{1}{2} (K(X, Y) + K(X, Y') + K(X', Y) + K(X', Y')) \tag{2.17}$$

with X, X' distributed according to \mathbb{P}_X , Y, Y' distributed according to \mathbb{P}_Y and X, X', Y, Y' independent

then:

$$MMD_{m,n}^2 \xrightarrow[m, n \rightarrow +\infty]{a.s.} MMD_{\mathcal{G}}^2(\mathbb{P}_X, \mathbb{P}_Y). \tag{2.18}$$

Under H_1 , $MMD_{\mathcal{G}}^2(\mathbb{P}_X, \mathbb{P}_Y) > 0$ so that $N \times MMD_{m,n}^2 \xrightarrow[m, n \rightarrow +\infty]{a.s.} +\infty$ for $N = m + n$. Thus, we reject the null hypothesis at level α if $MMD_{m,n}^2$ is greater than some threshold c_{α} . In order to determine this threshold, we rely on the asymptotic distribution of $MMD_{m,n}^2$ under H_0 which is due to Gretton *et al.* (2012, Theorem 12).

Theorem 2.5. *Let us define the kernel \tilde{K} by:*

$$\tilde{K}(x, y) = K(x, y) - \mathbb{E}[K(x, X)] - \mathbb{E}[K(X, y)] - \mathbb{E}[K(X, X')] \tag{2.19}$$

where X and X' are i.i.d. samples drawn from \mathbb{P}_X . Assume that:

- (i) $\mathbb{E}[\tilde{K}(X, X')^2] < +\infty$
- (ii) $m/N \rightarrow \rho \in (0, 1)$ as $N = m + n \rightarrow +\infty$.

Under these assumptions, we have:

1. under H_0 :

$$N \times MMD_{m,n}^2 \xrightarrow[N \rightarrow +\infty]{\mathcal{L}} \frac{1}{\rho(1-\rho)} \sum_{\ell=1}^{+\infty} \lambda_{\ell} (G_{\ell}^2 - 1) \tag{2.20}$$

where $(G_\ell)_{\ell \geq 1}$ is an infinite sequence of independent standard normal random variables and the λ_ℓ 's are the eigenvalues of the operator $S_{\hat{K}}$ defined as

$$S_{\hat{K}} : L_2(\mathbb{P}_X) \rightarrow L_2(\mathbb{P}_X) \quad (2.21)$$

$$g \mapsto \mathbb{E}[\hat{K}(\cdot, X)g(X)]$$

with $L_2(\mathbb{P}_X) := \{g : \mathcal{X} \rightarrow \mathbb{R} \mid \mathbb{E}[g(X)^2] < \infty\}$.

2. under H_1 :

$$N \times MMD_{m,n}^2 \xrightarrow{N \rightarrow +\infty} +\infty \quad (2.22)$$

This theorem indicates that if one wants to have a test with level α , one should take the $1 - \alpha$ quantile of the above asymptotic distribution as rejection threshold. In order to approximate this quantile, we use an approach proposed by Gretton *et al.* (2009) that aims at estimating the eigenvalues in Theorem 2.5 using the empirical Gram matrix spectrum. This approach relies on the following theorem (Theorem 1 of Gretton *et al.*, 2009).

Theorem 2.6. *Let $(\lambda_\ell)_{\ell \geq 1}$ be the eigenvalues defined in Theorem 2.5 and $(G_\ell)_{\ell \geq 1}$ be a sequence of i.i.d. standard normal variables. For $N = m + n$, we define the centred Gram matrix $\hat{A} = HAH$ where $A = (K(Z_i, Z_j))_{1 \leq i, j \leq N}$ ($Z_i = X_i$ if $i \leq m$ and $Z_i = Y_{i-m}$ if $i > m$) and $H = I_N - \frac{1}{N} \mathbf{1}\mathbf{1}^T$. If $\sum_{\ell=1}^{+\infty} \sqrt{\lambda_\ell} < \infty$ and $m/N \rightarrow \rho \in (0, 1)$ as $N \rightarrow +\infty$, then under H_0 :*

$$\frac{1}{\hat{\rho}(1 - \hat{\rho})} \sum_{\ell=1}^{+\infty} \frac{v_\ell}{N} (G_\ell^2 - 1) \xrightarrow{N \rightarrow +\infty} \frac{1}{\rho(1 - \rho)} \sum_{\ell=1}^{+\infty} \lambda_\ell (G_\ell^2 - 1) \quad (2.23)$$

where $\hat{\rho} = m/N$ and the v_ℓ 's are the eigenvalues of \hat{A} .

Therefore, we can approximate the asymptotic distribution in Theorem 2.5 by

$$\frac{1}{\hat{\rho}(1 - \hat{\rho})} \sum_{\ell=1}^R \frac{v_\ell}{N} (G_\ell^2 - 1) \quad (2.24)$$

with R is the truncation order and $v_1 > v_2 > \dots > v_R$ are the R first eigenvalues of \hat{A} in decreasing order. A rejection threshold is then obtained by simulating several realizations of the above random variable and then computing their empirical quantile at level $1 - \alpha$.

3. Implementation and numerical results

The objective of this section is to show the practical interest of the two-sample test described in the previous section for the validation of real-world economic scenarios. In the sequel, we refer to the two-sample test as the signature-based validation test. As a preliminary, we discuss the challenges implied by the practical implementation of the signature-based validation test.

3.1. Practical implementation of the signature-based validation test

3.1.1. Signature of a finite number of observations

In practice, the historical paths of the risk drivers under interest are not observed continuously, and the data are only available on a discrete time grid with steps that are generally not very small (one day for the realized volatility data and one month for the inflation considered in Section 3.3). One has to embed these observations into a continuous path in order to be able to compute the signature and *a fortiori* the MMD with the model that we want to test. The two most popular embeddings in the literature are the linear and the rectilinear embeddings. The former one consists in a plain linear interpolation of the observations, while the latter consists in an interpolation using only parallel shifts with respect to the x and y -axis. In the sequel, we will only use the linear embedding as this choice does not seem to have

a material impact on the information contained in the obtained signature as shown by the comparative study led by Fermanian (section 4.2 of Fermanian, 2021). Even if several of the models considered below have paths with unbounded variation, we discretize them with a rather crude step (consistent with insurance practice) and then we compute the signatures of the continuous paths with bounded variation deduced by the linear interpolation embedding.

3.1.2. Extracting information of a one-dimensional process

Remember that if X is a one-dimensional bounded variation path, then its signature over $[0, T]$ is equal to the powers of the increment $X_T - X_0$. As a consequence, finer information than the global increment about the evolution of X on $[0, T]$ is lost. In practice, X is often of dimension 1. Indeed, the validation of real-world economic scenarios is generally performed separately for each risk driver and not all together as a first step and only then the validation is performed on a more global level (e.g. through the analysis of the copula among risk drivers). For example, when validating risk drivers such as equity or inflation, X represents an equity or inflation index which is one-dimensional. Moreover, even for multidimensional risk drivers such as interest rates, practitioners tend to work on the marginal distributions (e.g. they focus separately on each maturity for interest rates) so that the validation is again one-dimensional. In order to nonetheless be able to capture finer information about the evolution of X on $[0, T]$, one can apply a transformation to X to recover a multi-dimensional path. The two most widely used transformations are the time transformation and the lead-lag transformation. The time transformation consists in considering the two-dimensional path $\hat{X}_t : t \mapsto (t, X_t)$ instead of $t \mapsto X_t$. The lead-lag transformation has been introduced by Gyurkó *et al.* (2013) in order to capture the quadratic variation of a path in the signature. Let X be a real-valued stochastic process and $0 = t_0 < t_{1/2} < t_1 < \dots < t_{N-1/2} < t_N = T$ be a partition of $[0, T]$. The lead-lag transformation of X on the partition $(t_{i/2})_{i=0, \dots, 2N}$ is the two-dimensional path $t \mapsto (X_t^{lead}, X_t^{lag})$ defined on $[0, T]$ where:

1. the lead process $t \mapsto X_t^{lead}$ is the linear interpolation of the points $(X_{t_{i/2}}^{lead})_{i=0, \dots, 2N}$ with:

$$X_{t_{i/2}}^{lead} = \begin{cases} X_{t_j} & \text{if } i = 2j \\ X_{t_{j+1}} & \text{if } i = 2j + 1 \end{cases} \tag{3.1}$$

2. the lag process $t \mapsto X_t^{lag}$ is the linear interpolation of the points $(X_{t_{i/2}}^{lag})_{i=0, \dots, 2N}$ with:

$$X_{t_{i/2}}^{lag} = \begin{cases} X_{t_j} & \text{if } i = 2j \\ X_{t_j} & \text{if } i = 2j + 1 \end{cases} \tag{3.2}$$

Illustrations of the lead and lag paths as well as of the lead-lag transformation are provided in Figure 1. Note that by summing the area of each rectangle formed by the lead-lag transformation between two orange dots in Figure 1(b), one recovers the sum of the squares of the increments of X which is exactly the quadratic variation of X over the partition $(t_i)_{i=0, \dots, N}$. The link between the signature and the quadratic variation is more formally stated in Proposition A.1. of the Supplementary Material.

Remark 3.1. The choice of the dates $(t_{i+1/2})_{i=0, \dots, N-1}$ such that $t_i < t_{i+1/2} < t_{i+1}$ can be arbitrary since the signature is invariant by time reparametrization (see Proposition B.1. in the Supplementary Material).

A third transformation can be constructed from the time and the lead-lag transformations. Indeed, given a finite set of observations $(X_t)_{t=0, \dots, N}$, one can consider the three-dimensional path $t \mapsto (t, X_t^{lead}, X_t^{lag})$. We call this transformation the time lead-lag transformation. Finally, the cumulative lead-lag transformation is the two-dimensional path $t \mapsto (\tilde{X}_t^{lead}, \tilde{X}_t^{lag})$ where \tilde{X}_t^{lead} (resp. \tilde{X}_t^{lag}) is the lead (resp. lag) transformation of the points $(\tilde{X}_t)_{t=0, \dots, N+1}$ with:

$$\tilde{X}_t = \begin{cases} 0 & \text{for } i = 0 \\ \sum_{k=0}^{i-1} X_{t_k} & \text{for } i = 1, \dots, N + 1. \end{cases} \tag{3.3}$$

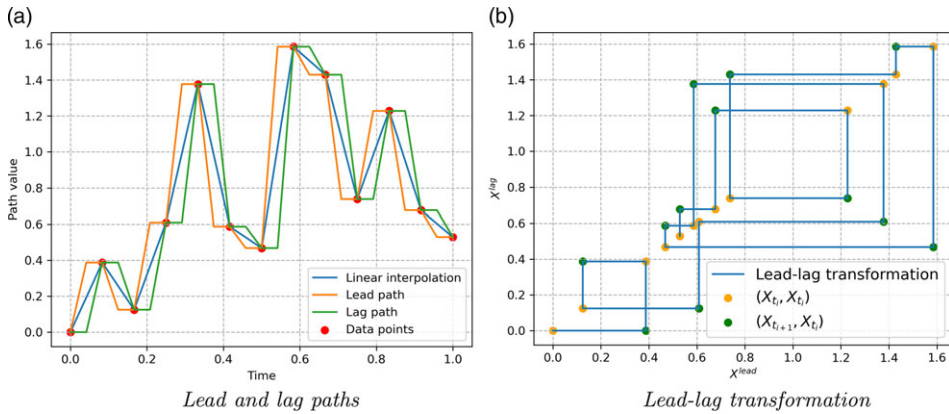


Figure 1. Lead-lag illustrations.

This transformation has been introduced by Chevyrev and Kormilitzin (2016) because its signature is related to the statistical moments of the initial path X . More details on this point are provided in Remark A.1. in the Supplementary Material.

3.1.3. Numerical computation of the signature and the MMD

The numerical computation of the signature is performed using the `iisignature` Python package (version 0.24) of Reizenstein and Graham (2020). The `signatory` Python package of Kidger and Lyons (2020) could also be used for faster computations. Because the signature is an infinite object, we compute in practice only the truncated signature up to some specified order R . The influence of the truncation order on the statistical power of the test will be discussed in Section 3.2. Note that we will focus on truncation orders below 8 as there is not much information beyond this order given that we work with a limited number of observations of each path which implies that the approximation of high-order iterated integrals will rely on very few points.

3.2. Analysis of the statistical power on synthetic data

In this subsection, we apply the signature-based validation on simulated data, that is the two samples of stochastic processes are numerically simulated. Keeping in mind insurance applications, the two-sample test is structured as follows:

- Each path is obtained by a linear interpolation from a set of 13 equally spaced observations of the stochastic process under study. The first observation (i.e. the initial value of each path) is the same across all paths. These 13 observations represent monthly observations over a period of one year. In insurance practice, computational time constraints around the asset and liability models generally limit the simulation frequency to a monthly time step. The period of one year is justified by the fact that one needs to split the historical path under study into several shorter paths to get a test sample of size greater than 1. Because the number of historical data points is limited (about 30 years of data for the major equity indices), a split frequency of one year appears reasonable given the monthly observation frequency.
- The two samples are assumed to be of different sizes (i.e. $m \neq n$ with the notations of Section 2.3). Several sizes m of the first sample will be tested while the size n of the second sample is always set to 1000. The first sample representing historical paths, we will mainly consider small values of m as m will in practice be equal to the number of years of available data (considering a split of the historical path in one-year length paths as discussed above). For the second sample which consists in simulated paths (for example by an Economic Scenario Generator),

we take 1000 simulated paths as it corresponds to a lower bound of the number of scenarios typically used by insurers. Numerical tests (not presented here) have also been performed with a sample size of 5000 instead of 1000, but the results were essentially the same.

- As we aim to explore the capability of the two-sample test to capture properties of the paths that cannot be captured by looking at their marginal distribution at some dates, we impose that the distributions of the increment over $[0, T]$ of the two compared stochastic processes are the same. In other words, we only compare stochastic processes $(X_t)_{0 \leq t \leq T}$ and $(Y_t)_{0 \leq t \leq T}$ satisfying $X_T - X_0 \stackrel{L}{=} Y_T - Y_0$ with $T = 1$ year. This constraint is motivated by the fact that two models that do not have the same marginal one-year distribution are already discriminated by the current point-in-time validation methods. Moreover, it is a common practice in the insurance industry to calibrate the real-world models by minimizing the distance between model and historical moments so that the model marginal distribution is often close to the historical marginal distribution. Because of this constraint, we will remove the first-order term of the signature in our estimation of the MMD because it is equal to the global increment $X_T - X_0$ and it does not provide useful statistical information.

In order to measure the ability of the signature-based validation to distinguish two different samples of paths, we compute the statistical power of the underlying test, which is the probability to correctly reject the null hypothesis under H_1 , by simulating 1000 times two samples of sizes m and n respectively and counting the number of times that the null hypothesis (the stochastic processes underlying the two samples are the same) is rejected. The rejection threshold is obtained using the empirical Gram matrix spectrum as described in Section 2.3. First, we generate a sample of size m and a sample of size n under H_0 to compute the eigenvalues of the matrix A in Theorem 2.6. Then, we keep the 20 first eigenvalues in decreasing order and we perform 10,000 simulations of the random variable in Equation (2.24) whose distribution approximates the MMD asymptotic distribution under H_0 . The rejection threshold is obtained as the empirical quantile of level 99% of these samples. For each experiment presented in the sequel, we also simulate 1000 times two samples of sizes m and n under H_0 and we count the number of times that the null hypothesis is rejected with this rejection threshold, which gives us the type I error. This step allows us to verify that the computed rejection threshold provides indeed a test of level 99% in all experiments. As we obtain a type I error around 1% in all numerical experiments, we conclude to the accuracy of the computed rejection threshold.

We will now present numerical results for three stochastic processes: the fractional Brownian motion, the Black–Scholes dynamics and the rough Heston model. Two time-series models will also be considered: a regime-switching $AR(1)$ process and a random walk with i.i.d. Gamma noises. Finally, we study a two-dimensional process where one component evolves according to the rough Heston dynamics and the other one evolves according to a regime-switching $AR(1)$ process. The choice of this catalogue of models is both motivated by

1. the fact that the models are either widely used or realistic for the simulation of specific risk drivers (equity volatility, equity prices and inflation) and
2. the fact that the models allow to illustrate different pathwise properties such as Hölder regularity, volatility, autocorrelation or regime switches.

3.2.1. The fractional Brownian motion

The fractional Brownian motion (fBm) is a generalization of the standard Brownian motion that, outside this standard case, is neither a semimartingale nor a Markov process and whose sample paths can be more or less regular than those of the standard Brownian motion. More precisely, it is the unique centred Gaussian process $(B_t^H)_{t \geq 0}$ whose covariance function is given by:

$$\mathbb{E} [B_s^H B_t^H] = \frac{1}{2} (s^{2H} + t^{2H} - (s - t)^{2H}) \quad \forall s, t \geq 0 \tag{3.4}$$

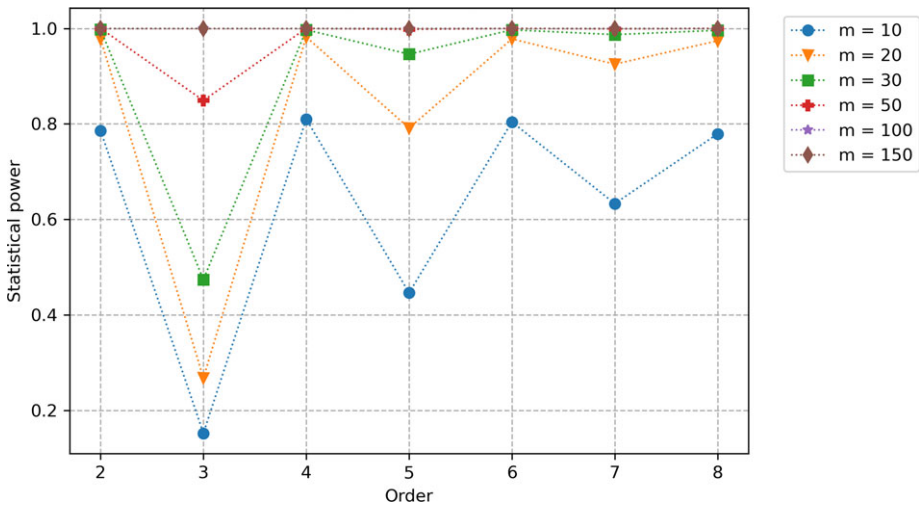


Figure 2. Statistical power (as a function of the truncation order and the size of the first sample) of the signature-based validation test when comparing fBm paths with $H = 0.1$ against fBm paths with $H' = 0.2$. Note that the lead-lag transformation is applied.

where $H \in (0, 1)$ is called the Hurst parameter. Taking $H = 1/2$, we recover the standard Brownian motion. The fBm exhibits two interesting pathwise properties.

1. The fBm sample paths are $H - \epsilon$ Hölder for all $\epsilon > 0$. Thus, when $H < 1/2$, the fractional Brownian motion sample paths are rougher than those of the standard Brownian motion and when $H > 1/2$, they are smoother.
2. The increments are correlated: if $s < t < u < v$, then $\mathbb{E}[(B_t^H - B_s^H)(B_v^H - B_u^H)]$ is positive if $H > 1/2$ and negative if $H < 1/2$ since $x \mapsto x^{2H}$ is convex if $H > 1/2$ and concave otherwise.

One of the main motivations for studying this process is the work of Gatheral *et al.* (2018) which shows that the historical volatility of many financial indices essentially behaves as a fBm with a Hurst parameter around 10%.

In the following numerical experiments, we will compare samples from a fBm with Hurst parameter H and samples from a fBm with a different Hurst parameter H' . One can easily check that B_1^H has the same distribution than $B_1^{H'}$ since B_1^H and $B_1^{H'}$ are both standard normal variables. Thus, the constraint that both samples have the same one-year marginal distribution (see the introduction of Section 3.2) is satisfied. Note that a variance rescaling should be performed if one considers a horizon that is different from 1 year. We start with a comparison of fBm paths having a Hurst parameter $H = 0.1$ with fBm paths having a Hurst parameter $H = 0.2$ using the lead-lag transformation. In Figure 2, we plot the statistical power as a function of the truncation order R for different values of the first sample size m (we recall that the size of the second sample is fixed to 1000). We observe that even with small sample sizes, we already obtain a power close to 1 at order 2. Note that the power does not increase with the order but decreases at odd orders when the sample size is smaller than 50. This can be explained by the fact that the odd-order terms of the signature of the lead-lagged fBm are linear combinations of monomials in $B_{t_1}^H, \dots, B_{t_n}^H$ that are of odd degree. Since (B_t^H) is a centred Gaussian process, the expectation of these terms are zero no matter the value of H . As a consequence, the contribution of odd-order terms of the signature to the MMD is the same under H_0 and under H_1 . If we conduct the same experiment for $H = 0.1$ versus $H' = 0.5$ (corresponding to the standard Brownian motion), we obtain cumulated powers greater than 99% for all tested orders and sample sizes (even $m = 10$) even if the power of the odd orders is small (below 45%). This is very promising as it shows that the signature-based validation allows distinguishing

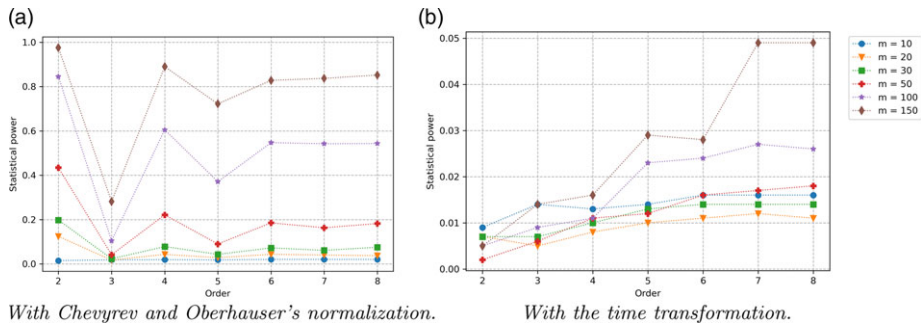


Figure 3. Statistical power (as a function of the truncation order and the size of the first sample) of the signature-based validation test when comparing fBm paths with $H = 0.1$ against fBm paths with $H' = 0.2$.

very accurately rough fBm paths (with a Hurst parameter in the range of those estimated by Gatheral *et al.*, 2018) from standard Brownian motion paths even with small sample sizes.

Note that in these numerical experiments, we have not used any tensor normalization while it is a key ingredient in Theorem 2.3. This is motivated by the fact that the power is much worse when we use Cheyrev and Oberhauser’s normalization (see Example 4 in Cheyrev and Oberhauser, 2022) as one can see on Figure 3(a). These lower powers can be understood as a consequence of the fact that the normalization is specific to each path. So the normalization can bring the distribution of $S(\hat{B}^{H'})$ closer to the one of $S(\hat{B}^H)$ than without normalization so that it is harder to distinguish them at fixed sample size. Moreover, if the normalization constant λ is smaller than 1 (which we observe numerically), the high-order terms of the signature become close to zero and their contribution to the MMD is not material. For $H = 0.1$ versus $H' = 0.5$, we observed that the powers remain very close to 100%.

Also note that the lead-lag transformation is key for this model as replacing it by the time transformation (see Section 3.1.2) results in much lower statistical powers, see Figure 3(b). This observation is consistent with the previous study from Fermanian (2021) which concluded that the lead-lag transformation is the best choice in a learning context.

Before moving to the Black–Scholes dynamics, we present results of the test when the signature is replaced by the log-signature. The log-signature is a more parsimonious – though equivalent – representation of paths than the signature as it contains more zeros. A formal definition of the log-signature and more insights can be found in the Supplementary Material. Although no information is lost by the log-signature, it is not clear whether the MMD is still a metric when the signature is replaced by the log-signature in the kernel (see Remark C.1. in the Supplementary Material). Numerically, the log-signature shows satisfying powers for $H = 0.1$ versus $H' = 0.2$, especially when the truncation order is 2 (see Figure 4). One can remark in particular that the power decreases with the order. This observation likely results from the $1/n$ factor appearing in the log-signature formula (Equation (A.14) of the Supplementary Material) which makes high-order terms of the log-signature small so that the even-order terms no longer compensate the odd-order terms.

3.2.2. The Black–Scholes dynamics

In the well-known Black–Scholes model (Black and Scholes, 1973), the evolution of the stock price $(S_t)_{t \geq 0}$ is modelled using the following dynamics:

$$dS_t = \mu S_t dt + \gamma(t) S_t dW_t, \quad S_0 = s_0 \tag{3.5}$$

where $\mu \in \mathbb{R}$, γ is a deterministic function of time and $(W_t)_{t \geq 0}$ is a standard Brownian motion. Because of its simplicity, this model is still widespread in the insurance industry, in particular for the modelling of equity and real estate indices.

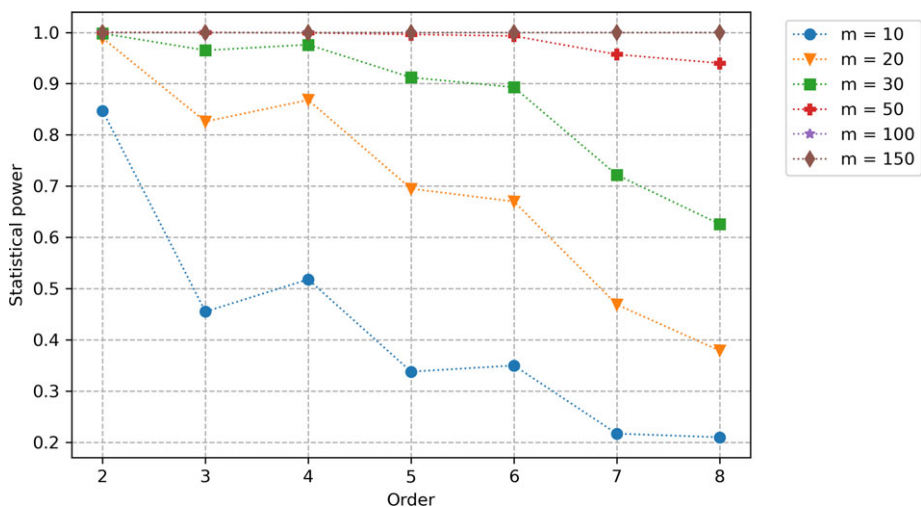


Figure 4. Statistical power (as a function of the truncation order and the size of the first sample) of the log-signature-based validation test when comparing fBm paths with $H = 0.1$ against fBm paths with $H' = 0.2$. Note that the lead-lag transformation is applied.

Like for the fractional Brownian motion, we want to compare two parametrizations of this model that share the same one-year marginal distribution. For this purpose, we consider a Black–Scholes dynamics (BSd) with drift μ and constant volatility σ and a BSd with the same drift μ but with a deterministic volatility $\gamma(t)$ satisfying $\int_0^1 \gamma^2(s)ds = \sigma^2$ which guarantees that the one-year marginal distribution constraint is met. For the sake of simplicity, we take a piecewise constant volatility with $\gamma(t) = \gamma_1$ if $t \in [0, 1/2)$ and $\gamma(t) = \gamma_2$ if $t \in [1/2, 1]$. In this setting, the objective of the test is no longer to distinguish two stochastic processes with different regularity but two stochastic processes with different volatility which is *a priori* more difficult since the volatility is not directly observable in practice. When $\mu = 0$, the first two terms of the signature of the lead-lag transformation have the same asymptotic distribution in the two parametrizations as the time step converges to 0. This is explained in Example A.3. in the Supplementary Material. We conjecture that this result extends to the full signature so that the two models cannot be distinguished using the signature.

We start by comparing BSd paths with $\mu = 0.05$ and $\sigma = 0.2$ and BSd paths with $\mu = 0.05$, $\gamma_1 = 0$ and $\gamma_2 = \sqrt{2}\sigma$ using the lead-lag transformation. We consider a zero volatility on half of the time interval in order to obtain very different paths in the two samples. Despite this extreme parametrization, we obtain very low powers (barely above 2% even for a sample size of $m = 500$). It seems that the constraint of same quadratic variation in both samples makes the signatures from sample 1 too close from those of sample 2.

In order to improve the power of the test, we can consider another data transformation which allows capturing information about the initial one-dimensional path in a different manner. We observed that the time lead-lag transformation (see Section 3.1.2) allowed to better distinguish the signatures from the two parametrizations given above. However, the order of magnitude of the differentiating coefficients of the signature (i.e. the coefficients of the signature that are materially different between the two samples) was significantly smaller than the one of non-differentiating coefficients so that the former were hidden by the latter when computing $MMD_{m,n}$. To address this issue, we applied a rescaling to all coefficients of the signature to make sure they are all of the same order of magnitude. Concretely, given two samples $\mathcal{S}_1 = \{X_1, \dots, X_m\}$ and $\mathcal{S}_2 = \{Y_1, \dots, Y_n\}$ of d -dimensional paths, the rescaling is performed as follows.

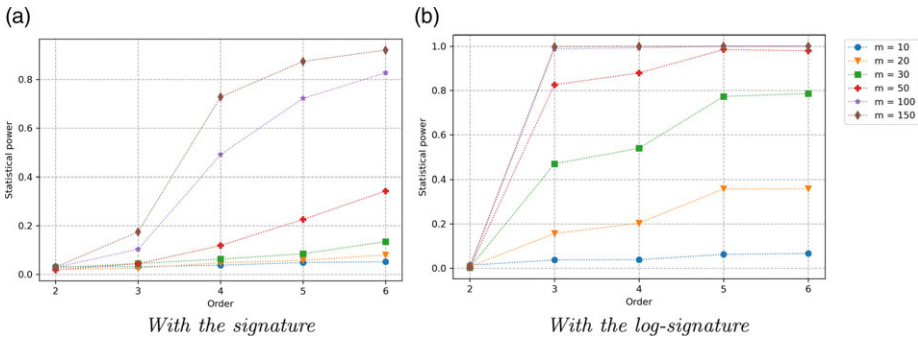


Figure 5. Statistical power (as a function of the truncation order and the size of the first sample) of the validation test when comparing constant volatility BSd paths against piecewise constant BSd paths using the time lead-lag transformation with the rescaling procedure.

1. For all $i \in \{1, \dots, m\}$ and for all $j \in \{1, \dots, n\}$, compute $S(X_i)$ and $S(Y_j)$.
2. For all $\ell \in \{1, \dots, R\}$ and for all $I = (i_1, \dots, i_\ell) \in \{1, \dots, d\}^\ell$, compute

$$M_i^\ell = \max \left(\max_{i=1, \dots, m} |X_{i,I}^\ell|, \max_{j=1, \dots, n} |Y_{j,I}^\ell| \right) \tag{3.6}$$

where $X_{i,I}^\ell$ (resp. $Y_{j,I}^\ell$) is the coefficient at position I of the ℓ -th term of the signature of X_i (resp. Y_j).

3. For all $i \in \{1, \dots, m\}$, for all $j \in \{1, \dots, n\}$, for all $\ell \in \{1, \dots, R\}$ and for all $I = (i_1, \dots, i_\ell) \in \{1, \dots, d\}^\ell$, compute the rescaled signature as

$$\hat{X}_{i,I}^\ell = \frac{X_{i,I}^\ell}{M_i^\ell} \quad \text{and} \quad \hat{Y}_{j,I}^\ell = \frac{Y_{j,I}^\ell}{M_j^\ell}. \tag{3.7}$$

This procedure guarantees that all coefficients of the signature lie within $[-1, 1]$. Using this normalization for the time lead-lag rescaling, the power of the test is significantly better than with the plain lead-lag transformation, as shown in Figure 5(a). In Figure 5(b), we show that the power can be further improved by considering the log-signature instead of the signature. Note that the increase of the power starts at order 3 which makes sense since order 2 only allows to capture the quadratic variation over $[0, T]$ (which is the same in the two parametrizations) while order 3 allows to capture the evolution of the quadratic variation over time. Alternatively, one can consider, instead of the time lead-lag transformation, the cumulative lead-lag transformation (see Equation (3.3)) on the log-paths (i.e. on $\log S_t$), which provides even better statistical powers as shown in Figure 6.

We also considered a slight variation of the BSd that has autocorrelation. Let $(t_i)_{0 \leq i \leq N}$ be an equally spaced partition of $[0, 1]$ with $\Delta t = t_{i+1} - t_i = 1/N$. The autocorrelated discretized BSd $(S_t^C)_{0 \leq t \leq N}$ is defined as follows:

$$\begin{cases} S_{t_{i+1}}^C = S_{t_i}^C \exp \left(\left(\mu^C - \frac{\gamma_{t_i}^2}{2} \right) \Delta t + \gamma_{t_i} \sqrt{\Delta t} G_{i+1} \right), & i \in \{0, \dots, N-1\} \\ S_0^C = s_0 \end{cases} \tag{3.8}$$

where $(G_i)_{1 \leq i \leq N}$ is a sequence of standard normal random variables satisfying:

$$\text{Cov}(G_i, G_j) = \begin{cases} 1 & \text{if } i = j \\ \rho & \text{if } |i - j| = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

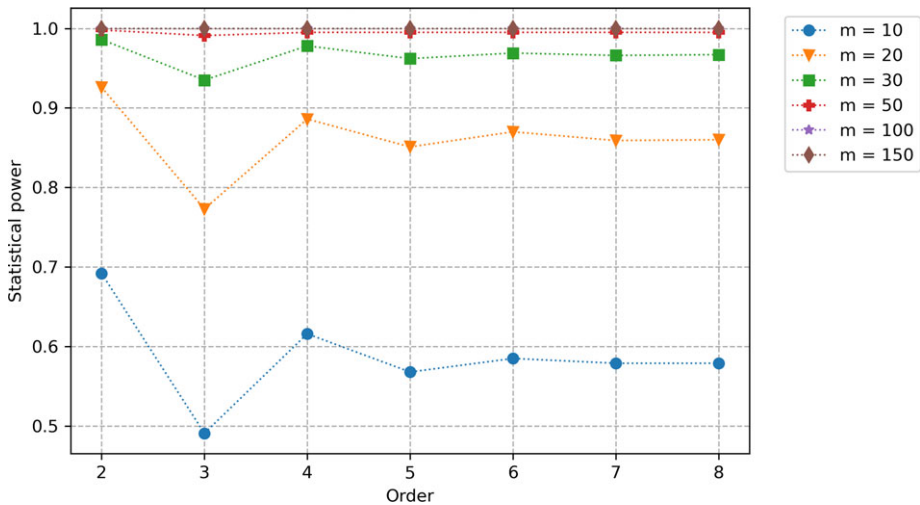


Figure 6. Statistical power (as a function of the truncation order and the size of the first sample) of the signature-based validation test when comparing constant volatility BSd paths against piecewise constant BSd paths using the cumulative lead-lag transformation on the log-paths.

The covariance matrix of (G_1, \dots, G_N) is positive definite when $\rho \in \left(-\frac{1}{2 \cos(\frac{\pi}{N+1})}, \frac{1}{2 \cos(\frac{\pi}{N+1})}\right)$. Indeed, the covariance matrix is a tridiagonal Toeplitz matrix so its eigenvalues are given by (p. 59 in Smith, 1985):

$$\lambda_k = 1 + 2\rho \cos\left(\frac{k\pi}{N+1}\right), \quad k = 1, \dots, N. \tag{3.10}$$

In our framework, we have $N = 12$ and we can check that $[-0.5, 0.5] \subset \left(-\frac{1}{2 \cos(\frac{\pi}{N+1})}, \frac{1}{2 \cos(\frac{\pi}{N+1})}\right)$. In Figure 7, we compare BSd paths with $\mu = 0.05$ and $\sigma = 0.2$ and autocorrelated BSd paths with correlation $\rho \in \{-0.5, -0.4, \dots, 0.5\}$ and a piecewise constant volatility like in the previous setting but with $\gamma_1 = \sigma/\sqrt{2}$ and (γ_2, μ^C) chosen such that S_1^C has the same distribution as $s_0 e^{\mu - \sigma^2/2 + \sigma W_1}$. Here, the first sample size is fixed to $m = 30$. While it was not possible to distinguish BSd paths with different volatility functions using the lead-lag transformation, we observe that the introduction of autocorrelation makes the distinction again possible even with a small sample size. More precisely, except if $\rho \in \{-0.1, 0, 0.1, 0.2\}$, we obtain a power greater than 90% at order 2. We note, however, a decrease of the power with the truncation order as it appears that apart from the term of order 2, all the other terms of the signature are very close between the two samples.

3.2.3. The rough Heston model

First introduced as the limit of a microscopic model of prices on financial markets by El Euch *et al.* (2018), the rough Heston model can be seen as a fractional extension of the classic Heston model. Indeed, its dynamics can be written as follows:

$$\begin{aligned} dS_t &= S_t \sqrt{V_t} dW_t \\ V_t &= V_0 + \int_0^t \frac{(t-s)^{H-1/2}}{\Gamma(H+1/2)} \left((\theta - \lambda V_s) ds + \sigma \sqrt{V_s} dB_s \right) \end{aligned} \tag{3.11}$$

where Γ is the gamma function, $S_0, V_0, \theta, \lambda, \sigma > 0$, $H \in (0, 1/2]$ and $\rho \in [-1, 1]$ is the correlation between the Brownian motions W and B . Note that for $H = 1/2$, V is a Cox–Ingersoll–Ross (CIR)

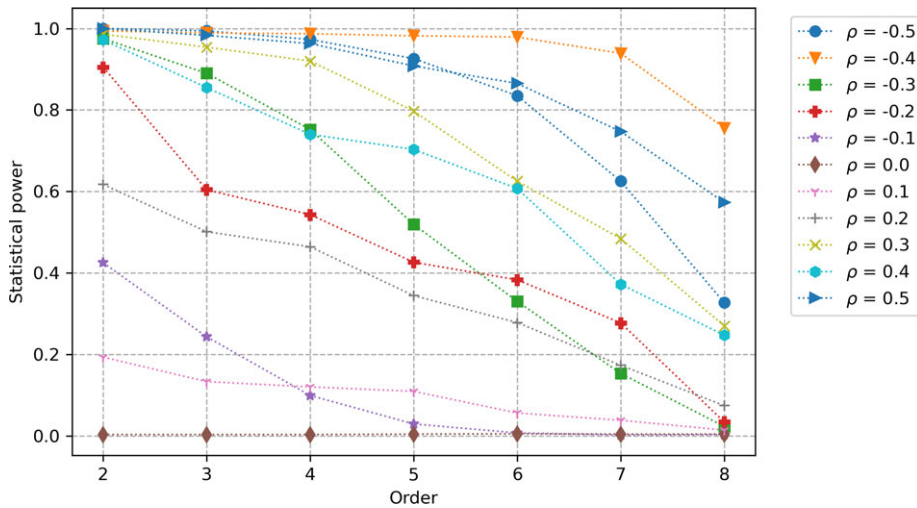


Figure 7. Statistical power (as a function of the truncation order and the correlation of the autocorrelated BSd paths) of the log-signature-based validation test when comparing constant volatility BSd paths against autocorrelated BSd paths using the lead-lag transformation with the rescaling procedure. The size of the first sample is fixed to $m = 30$.

process so that we recover the dynamics of the classic Heston model. Although mainly studied for applications under the risk-neutral probability (pricing and hedging options), the rough Heston model has also been used under the real-world probability in the context of portfolio selection (Han and Wong, 2021 and Abi Jaber *et al.*, 2021) and optimal reinsurance (Ma *et al.*, 2023).

The purpose of this section is to compare price paths (i.e. paths of S with the above notation) from the rough Heston model to price paths from the classic Heston model. In particular, let us insist on the fact that we do not include the simulations of the volatility process V in the two samples on which the signature-based validation test is applied because this process is not observable in practice. The spirit of this comparison is quite similar to the one described in the previous section: we evaluate the ability of the test to distinguish two stochastic processes having different volatilities. However, the present comparison is more challenging than the one involving the Black–Scholes dynamics since the difference between the two volatilities is more subtle: they differ essentially in their Hölder regularity but the overall evolution is close. We illustrate this point in Figure 8. Note in particular that the two price paths are very close. These simulations as well as the following ones rely on the approximate discretization scheme of Alfonsi and Kebaier (2024) for the rough Heston model and the explicit scheme $E(0)$ of Alfonsi (2005) for the CIR in the classic Heston model.

We start with a comparison between rough Heston paths having a Hurst parameter $H = 0.1$ and classic Heston paths. The parameters $S_0, V_0, \theta, \lambda, \sigma$ and ρ are identical between the two models and the retained values are those presented in the caption of Figure 8. Although this configuration does not imply that both models have exactly the same marginal one-year distributions, the two distributions are very close as shown in Figure 9. Besides, a two-sample Kolmogorov–Smirnov test (applied on 1000 annual log-returns from each model) yields a p -value of 0.31, hence the hypothesis that the two distributions are the same cannot be rejected at standard levels. As a first step, we simulate paths with a time step $\Delta = 1/120$ over a one-year horizon and then we extract every tenth values to obtain monthly trajectories. The use of this finer discretization for the simulation aims at limiting the discretization error. In this setting, we obtain very low powers (below 2%) be it with the lead-lag transformation, the time lead-lag transformation or the cumulative lead-lag transformations. Besides, neither the use of the log-signature instead of the signature nor the rescaling procedure described in the previous section leads to any improvement. Given that the price evolution is extremely close in both models (as shown in Figure 8), the obtained

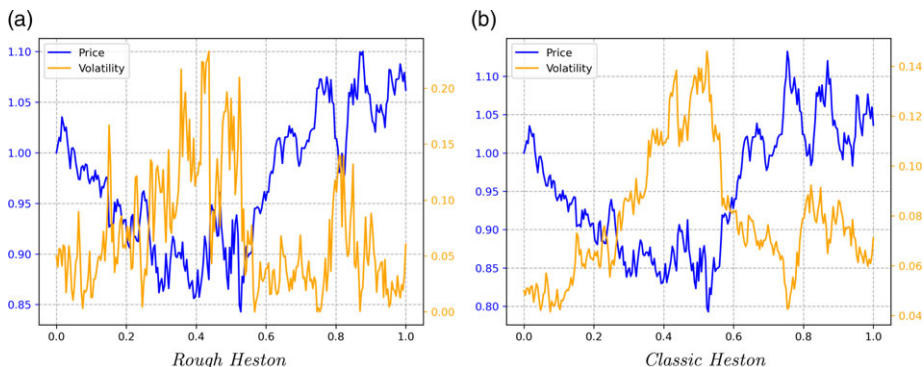


Figure 8. Comparison between a sample path of (S_t, V_t) from the rough Heston model with Hurst parameter $H=0.1$ and a sample path from the classic Heston model. We took $S_0 = 1, V_0 = 0.05, \theta = 0.05, \lambda = 0.3, \sigma = 0.3$ and $\rho = -0.7$ in both cases. Moreover, both paths are sampled using the same Gaussian random variables so that they look similar.

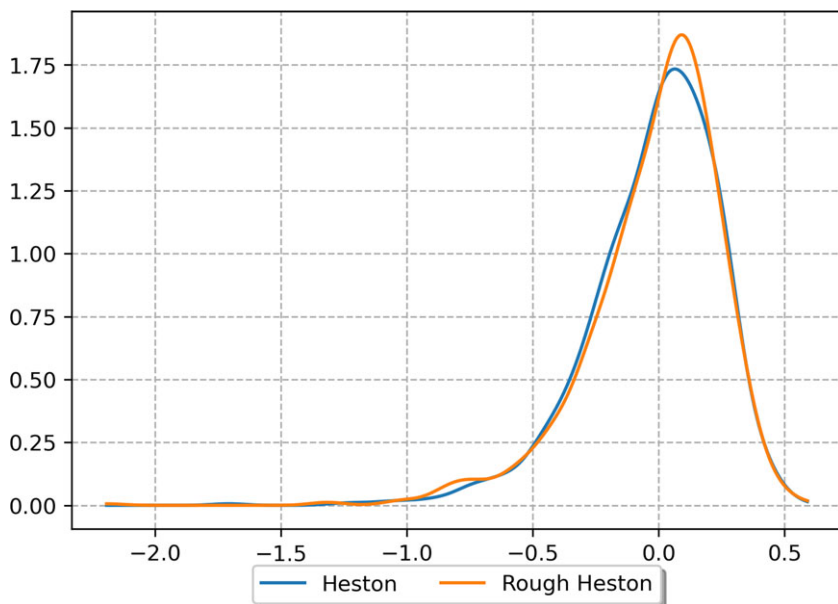


Figure 9. Kernel-density estimates of annual log-returns using Gaussian kernels. The annual log-returns are obtained by simulating 1000 one-year paths with a monthly frequency of each model.

powers are no surprise. In order to be able to distinguish between the two models, one has to extract somehow the volatility from the price path. However, this is illusory with monthly observations of the price over a period of one year. Therefore, as a second step, we propose to consider daily simulations of the price over one year and to compute a monthly realized volatility as

$$RV_{k\Delta} = \sqrt{\sum_{\ell=0}^{p-1} \left(\log \frac{S_{(k-1)\Delta+(\ell+1)\Delta'}}{S_{(k-1)\Delta+\ell\Delta'}} \right)^2} \tag{3.12}$$

where $k \in \{1, \dots, 12\}, \Delta = 1/12, \Delta' = 1/252$ and $p = 21$ ($252 = 21 \times 12$ is the average number of business days in a year). Now, if we apply the lead-lag transformation to the obtained monthly realized

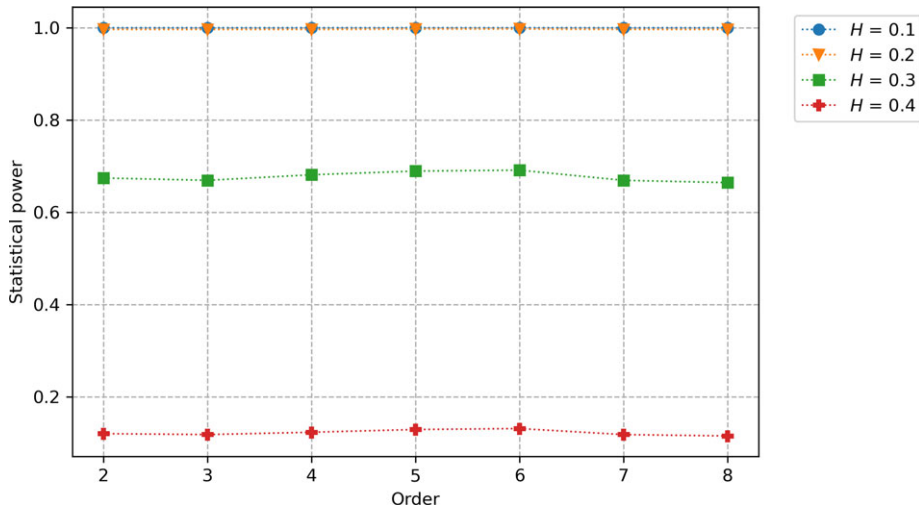


Figure 10. Statistical power of the log-signature-based validation test when comparing rough Heston paths to classic Heston paths with the lead-lag transformation on the monthly realized volatility paths extracted from the daily prices paths. The size of the first sample is fixed to $m = 30$.

volatility paths, we obtain statistical powers that are close to 1 at all truncation orders and for all sample sizes $m \in \{10, 20, 30, 50, 100, 150\}$ with the signature-based validation test. Note that these results are consistent with those obtained when comparing fBm paths with Hurst parameter $H = 0.1$ and standard Brownian motion paths. To complete this numerical study, we also test rough Heston paths with $H \in \{0.2, 0.3, 0.4\}$ against classic Heston paths. The resulting statistical powers are reported in Figure 10 for $m = 30$ and using the log-signature (which slightly improves the powers compared to the signature as for the fBm). As expected, the powers decrease when H gets closer to 0.5 (corresponding to the classic Heston), but we still have very good powers for $H = 0.2$. For $H = 0.3$ and $H = 0.4$, one needs to work with samples of greater size to achieve a statistical power above 95% ($m \geq 100$ for $H = 0.3$ and $m \geq 600$ for $H = 0.4$).

3.2.4. Time series models

Let us denote by I_t the value of an index (e.g. the Consumer Price Index) at time t and by $X_t = \log \frac{I_t}{I_{t-\Delta}}$ the log-return of this index over a time interval of size Δ . The first model under study assumes that the log-returns evolve as a regime-switching AR(1) process:

$$X_{t+\Delta} = \mu_{S_{t+\Delta}} + \phi_{S_{t+\Delta}}(X_t - \mu_{S_{t+\Delta}}) + \sigma_{S_{t+\Delta}} \epsilon_{t+\Delta} \tag{3.13}$$

where X_0 is fixed and S is a time-homogeneous discrete Markov chain with $K \geq 2$ states and whose transition matrix is denoted by P . The noises $(\epsilon_t)_{t \geq 0}$ are assumed to be i.i.d. standard normal variables. The second model under study assumes that the log-returns are i.i.d. non-centred Gamma noises whose shift, shape and scale parameters are, respectively, denoted by γ , α and β . In the following, we refer to this model as the Gamma random walk. The choice of these two models will be further motivated in Section 3.3.2.

Note that for the regime-switching AR(1) process, the annual log-return $\log \frac{I}{I_0}$ is distributed according to a Gaussian mixture while for the Gamma random walk, the annual log-return is Gamma distributed. Therefore, in order to still have close distributions for the annual log-return between the two models, we choose the parameters of the models so that the three first moments of the annual log-return are the same in both models. This is made possible by the fact that we can compute all moments in both cases and the fact that the parameters γ , α and β of the Gamma noises ϵ_t can be explicitly written as

Table 2. Parameters of the regime-switching AR(1) process and the Gamma random walk. The parameters of the former are inspired from parameters calibrated on real inflation data that we present later while the parameters of the latter are obtained by moment-matching as described above.

	X_0	μ_1	μ_2	ϕ_1	ϕ_2	σ_1	σ_2	P
Regime-switching AR(1)	0	0.002	0.006	0.45	0.6	0.0025	0.004	$\begin{pmatrix} 0.95 & 0.05 \\ 0.1 & 0.9 \end{pmatrix}$
	γ	α	β					
Gamma random walk	-0.6880	0.4734	1.4534					

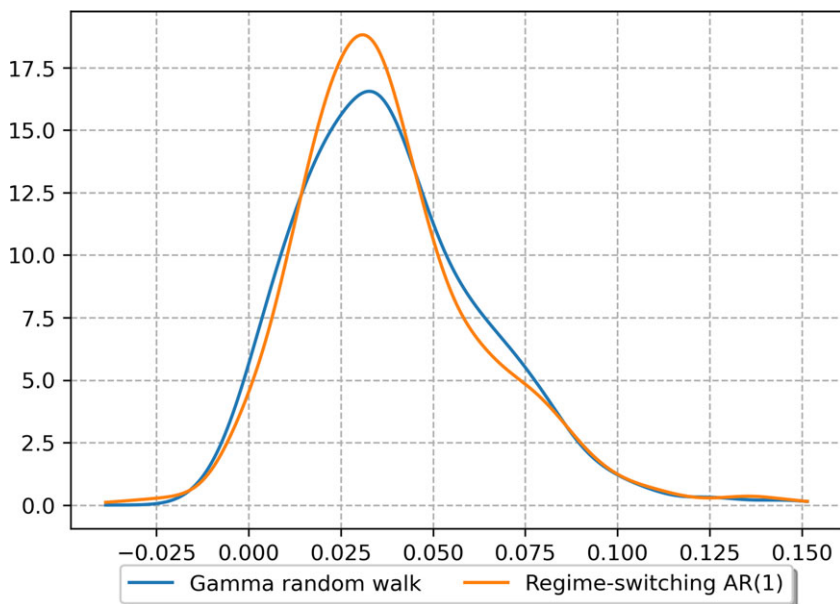


Figure 11. Kernel-density estimates of annual log-returns using Gaussian kernels. The annual log-returns are obtained by simulating 1000 one-year monthly paths of each model.

a function of the three first moments of ϵ_t (see section 17.7 of Johnson *et al.*, 1994). Therefore, given $X_0, (\mu_i)_{1 \leq i \leq K}, (\phi_i)_{1 \leq i \leq K}, (\sigma_i)_{1 \leq i \leq K}$ and P , we can find γ, α and β such that the three first moments of the annual log-return are matched. In Figure 11, the distributions of the annual log-return are compared for the parameters reported in Table 2. We observe that they are close which is confirmed by a two-sample Kolmogorov–Smirnov test (applied on 1000 simulated one-year monthly paths of each model) yielding a p -value of 0.50.

In Figure 12, we plot the statistical power of the two-sample test using the lead-lag transformation and the log-signature for an AR(1) process with two regimes and for a Gamma random walk. The parameters of both models are given in Table 2. Note that the regime at $t = 0$ is sampled from the stationary distribution of the Markov chain and that we work again on the log-paths as this representation improves slightly the results. We obtain statistical powers that are very close to 1 at any order for a sample size greater than $m = 30$. This third case shows that the signature-based validation test is still powerful when working with time-series models and allows to distinguish between paths exhibiting changes of regimes over

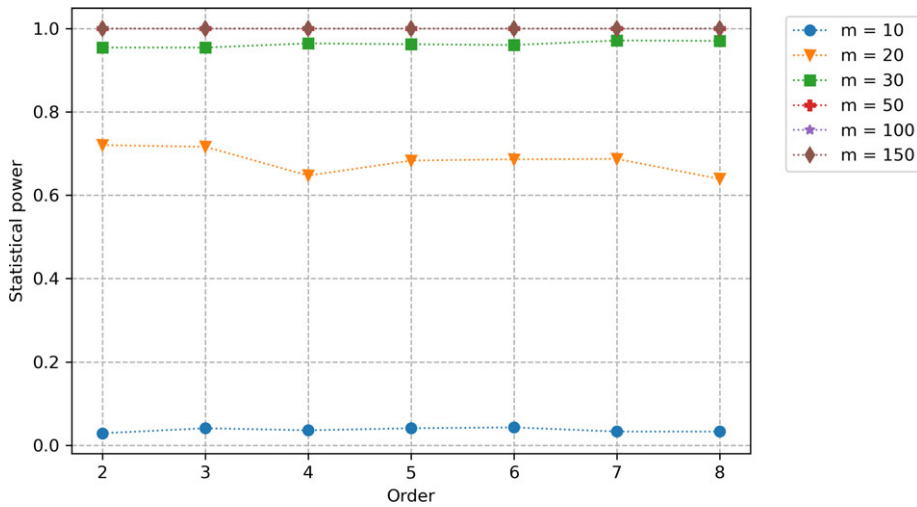


Figure 12. Statistical power (as a function of the truncation order and the size of the first sample) of the log-signature-based validation test when comparing regime-switching AR(1) log-paths against Gamma random walk log-paths using the lead-lag transformation.

time and first-order autocorrelation from paths with i.i.d. log-returns. Note that we have performed the same experiment with a regime-switching random walk (i.e. $\phi_1 = \phi_2 = 0$) instead of a regime-switching AR(1) process, and we obtained statistical powers above 88% for a sample size greater than $m = 50$ which shows that the first-order autocorrelation component is not necessary to distinguish the two models.

3.2.5. A two-dimensional process

In the four preceding examples, we compare the paths of one-dimensional processes but in practice economic scenarios consist in joint simulations of several correlated risk drivers. In this subsection, we propose to study whether the signature-based validation test also allows to capture the dependence structure between two models. To this end, we consider the following two samples.

1. The first sample consists of paths of the two-dimensional process $Z = (S, I)$ where (S, V) is solution of the rough Heston SDE (3.11) and the log-returns of I are modelled by a regime-switching AR(1) process. The Brownian motions W and B in Equation (3.11) are assumed to be independent from the noises ϵ in Equation (3.13). Besides, the parameters of each model are those presented in the caption of Figure 8 and in Table 2, respectively.
2. The second sample consists of paths of the exact same two-dimensional process except that W is correlated with ϵ (Gaussian dependence). In this case, the process is denoted by \tilde{Z} .

Note that the marginal distributions of the vectors Z and \tilde{Z} are exactly the same making again the two samples very similar.

In this multidimensional setting, it is *a priori* no longer necessary to perform a transformation on the simulated paths to extract information beyond the increments over $[0, T]$. However, starting with a correlation of 0.5, the measured statistical power of the plain signature-based validation test is below 2%. The rescaling procedure helps to improve materially these results as we obtain powers above 60% for $m = 150$. In order to achieve higher powers, especially for lower sample sizes, we considered the signature of the sequence of the log-returns of S and I in the test. This is motivated by the fact that the log-returns are the quantities that are uncorrelated for Z and correlated for \tilde{Z} . This modification allowed to considerably increase the powers for all sample sizes. We also considered the lead-lag

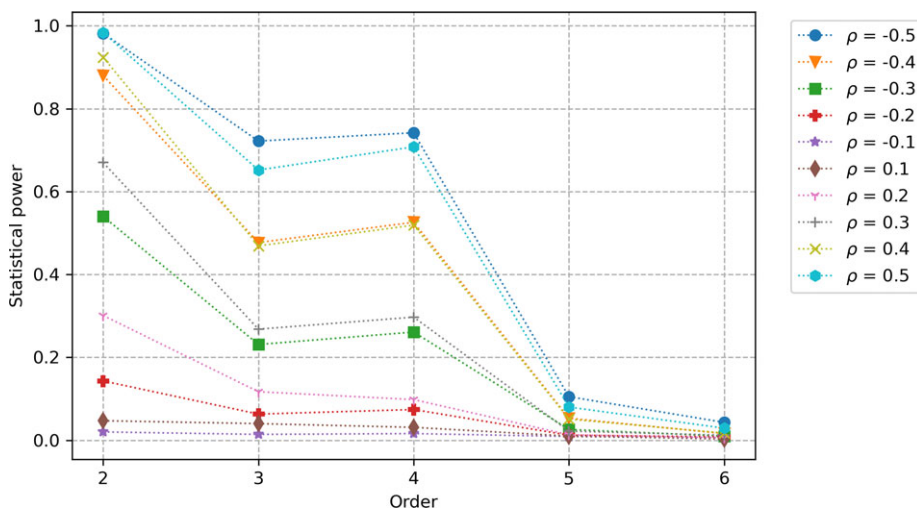


Figure 13. Statistical power (as a function of the truncation order and the correlation between the two components of \tilde{Z}) of the signature-based validation test when comparing paths of Z against paths of \tilde{Z} . The lead-lag transformation with the rescaling procedure is applied on the log-returns of each component of Z and \tilde{Z} . Note that the size of the first sample is fixed to $m = 30$.

transformation of the log-returns of S and I and we further improved the numerical results. In particular, we obtain a statistical power above 98% for a sample size of $m = 30$. In Figure 13, we show the obtained powers in this configuration for values of the correlation between W and ϵ in the range $\{-0.5, -0.4, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.4, 0.5\}$ when the size of the first sample is fixed to $m = 30$. When the absolute value of the correlation is greater or equal to 0.4, the statistical power is above 88%. Thus, we conclude that the signature-based validation test can also be powerful in a multidimensional setting. Note that the powers are decreasing with the truncation order. This seems to be mainly explained by the fact that the terms of odd order of the signature are essentially the same between the two samples (which can be seen as a consequence of the fact that correlation is captured through quadratic terms) so that the differentiating terms of even order are hidden.

3.3. Application to historical data

The purpose of this subsection is to show that the signature-based validation test is able to discriminate between stochastic models calibrated on historical data. This is of practical interest for the validation task of real-world economic scenarios, but more generally it is of interest for academics or practitioners that would like to compare a new model to existing ones based on a criteria of consistency with historical data. The following two subsections present numerical results on a realized volatility data set and on an inflation data set, respectively.

3.3.1. Realized volatility data

We consider the daily realized estimates of the S&P 500 from January 2000 to January 2022 obtained from Heber *et al.* (2009). On the log-volatilities derived from this data set (illustrated in Figure 14(a)), we calibrate an ordinary Ornstein-Uhlenbeck (OU) process and a fractional Ornstein-Uhlenbeck (FOU) process whose dynamics is recalled below:

$$dY_t = \alpha(\theta - Y_t)dt + \sigma dW_t^H \tag{3.14}$$

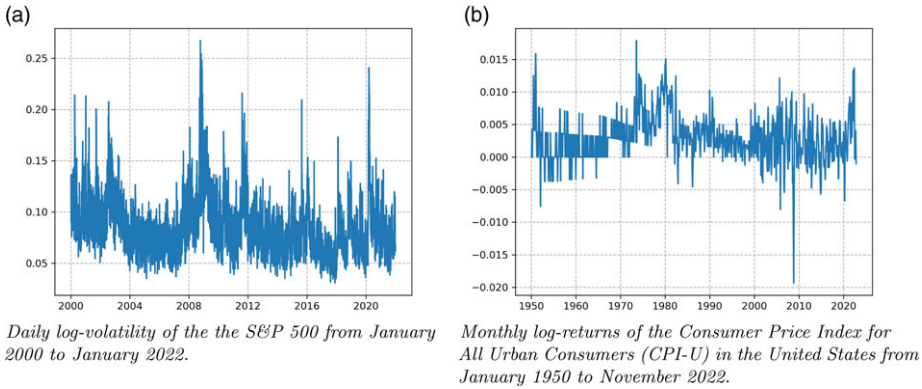


Figure 14. Data sets illustrations. Note on the right figure that between 1950 and the beginning of 70s, the smaller precision in the measurement of the inflation index leads to monthly log-returns that seem to oscillate between some fixed values.

where $(W_t^H)_{t \geq 0}$ is a fractional Brownian motion with Hurst parameter H . The OU process (corresponding to $H = 1/2$ in Equation 3.14) is calibrated using the maximum likelihood estimator of (α, θ, σ) , while the FOU process is calibrated using a two-step method. First, the Hurst parameter is estimated using the approach of Gatheral *et al.* (2018). This approach relies on the following scaling property of the fractional Brownian motion:

$$\mathbb{E} [|W_{t+\Delta}^H - W_t^H|^q] = \Delta^{qH} \mathbb{E}[|G|^q] \tag{3.15}$$

where G is a standard normal random variable. Thus, using the daily log-volatilities $(y_{k\Delta})_{k=0, \dots, N}$, we estimate the parameters ξ_q for several values of q (we took $q \in \{0.5, 1, 1.5, 2, 3\}$ as Gatheral *et al.*, 2018) in the following linear regression:

$$\log \left(\frac{1}{\lfloor \frac{N\Delta'}{\Delta'} \rfloor} \sum_{k=1}^{\lfloor \frac{N\Delta'}{\Delta'} \rfloor} |y_{k\Delta'} - y_{(k-1)\Delta'}|^q \right) = \beta_q + \xi_q \log(\Delta') + \epsilon_{\Delta'} \tag{3.16}$$

where $\Delta' \in \{\Delta, 2\Delta, \dots, p\Delta\}$ (we took $\Delta = 1/252$ and $p = 252$), β_q is the intercept and $\epsilon_{\Delta'}$ is the noise term. The slopes ξ_q of these regressions are then themselves regressed against q . The slope of this regression is our estimate \hat{H} of the Hurst parameter. Alternatively, one could consider the procedures presented in Cont and Das (2022) and Cont and Das (2023). Second, the mean-reversion speed α , the mean-reversion level θ and the volatility of volatility σ are estimated using the method-of-moments estimators from Wang *et al.* (2023):

$$\begin{aligned} \hat{\theta} &= \frac{1}{N} \sum_{k=1}^N y_{k\Delta}, \\ \hat{\alpha} &= \left(\frac{N^2 \hat{\sigma}^2 \hat{H} \Gamma(2\hat{H})}{N \sum_{k=1}^N y_{k\Delta}^2 - (\sum_{k=1}^N y_{k\Delta})^2} \right)^{1/(2\hat{H})}, \\ \hat{\sigma} &= \sqrt{\frac{\sum_{k=1}^{N-2} (y_{(k+2)\Delta} - 2y_{(k+1)\Delta} + y_{k\Delta})^2}{N(4 - 2^{2\hat{H}})\Delta^{2\hat{H}}}}. \end{aligned} \tag{3.17}$$

The calibrated parameters are reported in Table 3.

Using the calibrated parameters and considering $Y_0 = \theta$, we simulate 10,000 one-year paths with a monthly frequency for both models and we check that the simulated annual log-returns are close to the

Table 3. Parameters calibrated on daily realized log-volatilities of the S&P 500. Note that the decrease in the α value between the ordinary and the fractional Ornstein-Uhlenbeck models results from the rough noise in latter model ($H < 1/2$) that already captures a part of the negative autocorrelation.

(a) Ornstein-Uhlenbeck parameters			(b) Fractional Ornstein-Uhlenbeck parameters			
θ	α	σ	H	θ	α	σ
-5.0132	90.7993	8.2209	0.0916	-5.0131	0.2383	0.7876

Table 4. p -value of the two-sample Kolmogorov–Smirnov test applied to historical annual log-returns and simulated annual log-returns for each model. The historical annual log-returns are computed using a one-year window which is moved with a monthly step for the realized volatility data set and with a quarterly step for the inflation data set.

(a) Realized volatility data set		(b) Inflation data set	
OU	FOU	GRW	RSAR(1)
0.0852	0.6042	0.1227	0.1302

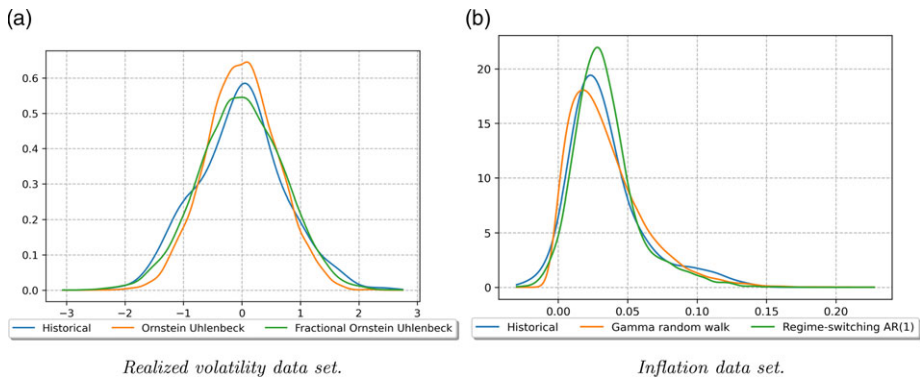


Figure 15. Kernel-density estimates of annual log-returns using Gaussian kernels. The historical annual log-returns are computed using a one-year window which is moved with a monthly step for the realized volatility data set and with a quarterly step for the inflation data set.

historical ones both graphically and using the two-sample Kolmogorov–Smirnov test. This preliminary verification allows to determine whether it would be possible to reject one of the models based on the marginal distributions. The plotted densities in Figure 15(a) appear reasonably close, and the hypothesis of same distribution is not rejected by the Kolmogorov–Smirnov test (see Table 4a) at standard levels (note, however, that the p -value of the OU paths is below 10%).

Now, we want to compare the paths simulated by each model to historical data using the signature-based validation test. To this end, we implement the following steps.

1. In order to be able to apply the signature-based validation test, we construct one-year “historical paths” with a monthly frequency from the monthly observations $(y_{k\bar{\Delta}})_{k=0,\dots,\bar{N}}$ of the realized log-volatility (we keep the last value of each month) as follows: we split the observations $(y_{k\bar{\Delta}})_{k=0,\dots,\bar{N}}$ of the realized log-volatility into $m = \lfloor \bar{N}/12 \rfloor$ one-year historical paths. The i th path consists of $y_{12i\bar{\Delta}}, y_{(12i+1)\bar{\Delta}}, \dots, y_{12(i+1)\bar{\Delta}}$. Here, $m = 22$ which lies in the range studied in the previous subsection.

Table 5. *p*-value of the signature-based validation test applied to historical paths and simulated paths for each model.

(a) Realized volatility data set		(b) Inflation data set	
OU	FOU	GRW	RSAR(1)
0.0008	0.2214	0.0000	0.6919

2. The lead-lag transformation is applied to these historical paths and to 1000 simulated sample paths of each calibrated model.
3. For both models, we test the transformed historical paths against 1000 transformed simulated sample paths using the signature-based validation test with the signature truncated at order 4. Note that the term of order 1 of the signature is removed because it is not very informative given that the increments in both models have close distributions.
4. The *p*-value of each test, obtained by computing $1 - \hat{F}_{H_0}(M\hat{M}D_{m,n}^2)$ where \hat{F}_{H_0} is the empirical cumulative distribution function of $MMD_{m,n}^2$ under H_0 and $M\hat{M}D_{m,n}^2$ is the test statistic value, is reported in Table 5a.

We observe that the OU model is rejected at any standard level while the FOU model is not. This result shows using a different method than Gatheral *et al.* (2018) that a rough volatility model is consistent with historical data.

3.3.2. Inflation data

The second data set that we consider contains monthly observations of the Consumer Price Index for All Urban Consumers (CPI-U) in the United States from January 1950 to November 2022. It is obtained from the U.S. Bureau of Labor Statistics (2023). As pointed out in Section 3.3.2. of the survey article of Petropoulos *et al.* (2022), the state-of-the-art inflation forecasting models are the dynamic stochastic general equilibrium (DSGE) models which rely on the description of the behaviour of the main economic agents (households, firms, governments and central banks) and their interactions through macro- and micro-economic concepts derived from economic theory. However, this class of models requires also GDP data, interest rates data, expert judgments, etc. which is beyond the scope of this study. The literature on inflation models requiring only inflation data is rather limited and we are not aware of a comprehensive survey paper comparing the performance of several models. We decide to calibrate a Gamma random walk (GRM) and a regime-switching *AR*(1) (RSAR(1)) process (see specifications in Section 3.2.4) on the CPI-U log-returns since:

1. Simple models such as random walks models are widely used in the insurance industry in particular for the computation of the SCRs. The choice of a Gamma distribution allows to capture the positive asymmetry in the distribution of the inflation rate (see Figure 15(b)).
2. A regime-switching model is very natural given that the history of inflation is a succession of periods of low and high inflation of varying lengths. The idea to use such models is not new as it was first proposed for US inflation by Evans and Wachtel (1993). We follow Amisano and Fagan (2013) in the use of an regime-switching *AR*(1) process except that we do not make the transition probabilities depend on money growth.

The GRM is calibrated by matching the three first moments of the historical annual log-returns while the RSAR(1) process is calibrated by log-likelihood maximization. The calibrated parameters are reported in Table 6.

Similarly to the previous section, we simulate 10,000 one-year paths with a monthly frequency for both calibrated models and we check that the distributions of the annual log-returns (i.e. the annual

Table 6. Parameters calibrated on the log-returns of the CPI-U. We refer to Section 3.2.4 for the definitions of these parameters.

(a) Gamma random walk parameters			(b) Regime-switching AR(1) parameters						
γ	α	β	μ_1	μ_2	ϕ_1	ϕ_2	σ_1	σ_2	P
-0.00047	0.18208	0.01832	0.0022	0.0062	0.4622	0.5774	0.0025	0.0042	$\begin{pmatrix} 0.9870 & 0.0130 \\ 0.0637 & 0.9363 \end{pmatrix}$

inflation rates) are close to the historical ones. The comparison of the empirical densities (see Figure 15b) does not reveal significant deviations from the historical density and the hypothesis of same distribution is not rejected by the Kolmogorov–Smirnov test (see Table 4b) for both models. Note that the simulated paths all start from 0, and the initial regime for the RSAR(1) process is sampled from the stationary distribution of the Markov chain.

Again, we complete this preliminary verification with the signature-based validation test. We implement the same steps described in the previous section (here, we obtain $m = 72$ historical paths) except that we work with the log-signature which lead to higher statistical powers on synthetic data. The obtained p -values (see Table 5b) show that the GRM model is rejected at any standard level while the RSAR(1) process is not. This is particularly interesting given that the p -values of the two-sample Kolmogorov–Smirnov test reported in Table 4b are very close. Moreover, this result is in line with the empirical observation that the inflation dynamics exhibits roughly two regimes in the inflation data set (see Figure 14): a regime of low inflation (e.g. between 1982 and 2021) and a regime of high inflation (e.g. between 1972 and 1982).

3.4. Numerical results summary

In this section, we have first presented the statistical power of the signature-based validation test in five different settings. In each of these settings, we have shown that high statistical powers can be achieved even in a small sample configuration and with a constraint on the closeness of the compared paths by using the following levers:

1. several representations of the original paths can be considered (log-paths, realized volatility, log-returns);
2. a transformation (lead-lag, time lead-lag, cumulative lead-lag) lifting the chosen representation of the paths to higher dimensional paths can be applied before taking the signature ;
3. several truncation orders can be tested;
4. the signature and the log-signature can be compared;
5. a rescaling of the terms of the signature can be applied.

The combinations of these levers resulting in the highest statistical power over the tested sample sizes in the five settings are presented in Table 7. Note that items 2, 4 and 5 are part of the *generalized signature method* introduced by Morrill *et al.* (2020) that aims at providing a unifying framework for the use of the signature as a feature in machine learning. A natural question at this stage is how to choose the path representation, transformation, truncation order, whether to use the signature or log-signature and whether to use a rescaling or not in a new setting that is not studied here. Unfortunately, we have not been able to find a general rule especially because it does not seem possible to relate these choices to the properties of the models under study except in some cases that we have exhibited above (e.g. even truncation order should be used for centred Gaussian processes). Despite this, our numerous experiments suggest some heuristics:

- the path representation should be chosen to “isolate” a property of the paths allowing to discriminate between models (e.g. extracting the realized volatility is key to distinguish rough Heston paths from classic Heston paths).

Table 7. Configurations of the two-sample test leading to the best statistical power.

Setting	Representation	Transformation	Order	Signature type	Rescaling
Fractional Brownian motion (Section 3.2.1)	Original paths	Lead-lag	2	Log-signature	No
Black-Scholes dynamics (Section 3.2.2)	Log-paths	Cumulative lead-lag	2	Signature	No
Rough Heston (Section 3.2.3)	Realized volatility	Lead-lag	No material influence	Log-signature	No
Time series models (Section 3.2.4)	Log-paths	Lead-lag	No material influence	Log-signature	No
Two-dimensional process (Section 3.2.5)	Log-returns	Lead-lag	2	Signature	Yes

- Truncating the signature at order 2 or 3 is generally the best choice to maximize the statistical power.
- The rescaling procedure is recommended when the dimension of the paths is greater or equal to 3 since the risk of a discriminating coefficient of the signature being “overwhelmed” by other coefficients is higher.

When no prior knowledge on the data is available, a possible strategy for the practitioners is the following: use the test for every combination of the levers above and reject the null hypothesis if a majority of the tests is rejected. Applying this strategy for the two historical data sets leads to the same conclusions about the four studied models: the ordinary Ornstein-Uhlenbeck and the Gamma random walk are rejected while the fractional Ornstein-Uhlenbeck and the regime-switching $AR(1)$ process are not.

Second, we have shown that the signature-based validation test allows to reject some models, while others are not rejected despite the fact that all produce annual log-returns that are reasonably close to the historical ones. In particular, a two-sample Kolmogorov–Smirnov test is not able to distinguish the distribution of the annual log-returns in each model from the historical distribution. As such, the current point-in-time validation approach would not have rejected the ordinary Ornstein-Uhlenbeck model and the Gamma random walk. This demonstrates that the signature-based validation test is a promising tool to validate real-world models.

4. Concluding remarks

We propose a new approach for the validation of real-world economic scenarios motivated by insurance applications. This approach relies on the formulation of the problem of validating real-world economic scenarios as a two-sample hypothesis testing problem where the first sample consists of historical paths, the second sample consists of simulated paths of a given real-world stochastic model and the null hypothesis is that the two samples come from the same distribution. For this purpose, we use the statistical test developed by Chevyrev and Oberhauser (2022) which precisely allows to check whether two samples of stochastic processes paths come from the same distribution. It relies on the notions of signature and MMD which are presented in this article. Our contribution is to study this test from a numerical point of view in settings that are relevant for applications. More specifically, we start by measuring the statistical power of the test on synthetic data under two practical constraints: first, the marginal one-year distributions of the compared samples are equal or very close so that point-in-time validation methods

are unable to distinguish the two samples and second, one sample is assumed to be of small size (below 50) while the other is of larger size (1000). To this end, we apply the test to three stochastic processes in continuous time, namely the fractional Brownian motion (fBm), the Black-Scholes dynamics (BSd) and the rough Heston model, and two time series models, namely a regime-switching $AR(1)$ process and a random walk with i.i.d. Gamma increments. The test is also applied to a two-dimensional process combining the price in the rough Heston model and a regime-switching $AR(1)$ process. The numerical experiments have highlighted the need to configure the test specifically for each stochastic process to achieve a good statistical power. In particular, the path representation (original paths, log-paths, realized volatility or log-returns), the path transformation (lead-lag, time lead-lag or cumulative lead-lag), the truncation order, the signature type (signature or log-signature) and the rescaling are key ingredients to be adjusted for each model. For example, the test achieves statistical powers that are close to one in the following settings which illustrate three different risk factors (stock volatility, stock price and inflation, respectively):

- fBm paths with Hurst parameter $H = 0.1$ against fBm paths with Hurst parameter $H' = 0.2$ using the lead-lag transformation and the log-signature;
- BSd paths with constant volatility against BSd paths with piecewise constant volatility using the time lead-lag transformation and the log-signature with a proper rescaling or using the cumulative lead-lag transformation on the log-paths along with the signature;
- paths of the rough Heston model with Hurst parameter $H \leq 0.2$ against paths of the classic Heston model using the lead-lag transformation on the realized volatility extracted from the price paths along with the signature;
- paths of a regime-switching $AR(1)$ process against paths of a random walk with i.i.d. Gamma increments using the lead-lag transformation on the log-paths along with the log-signature;
- paths of a two-dimensional process where the two coordinates are independent against paths of the same two-dimensional process where the two coordinates are correlated with correlation (in absolute value) above 40% using the lead-lag transformation on the log-returns along with the signature and a proper rescaling.

In addition to these numerical experiments on synthetic data, we show that the test also performs well on historical data since it rejects some models, whereas others are not rejected even if the distributions of the annual log-increments are very close in all the models. For example, we show that the fractional Ornstein-Uhlenbeck model with Hurst parameter around 0.1 is consistent with historical log-volatility of the S&P 500 while the ordinary Ornstein-Uhlenbeck model is not, which is another piece of evidence that volatility is rough (Gatheral *et al.*, 2018). These results indicate that this test represents a promising validation tool for real-world scenarios in a practical framework motivated by insurance applications. More broadly, the test appears as a universal tool for academics and practitioners that would like to challenge a new model against historical data.

Acknowledgments. The authors are grateful to the anonymous referees for their valuable comments.

Competing interests. The authors declare none.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/asb.2024.12>

References

- Abi Jaber, E., Miller, E. and Pham, H. (2021) Markowitz portfolio selection for multivariate affine and quadratic Volterra models. *SIAM Journal on Financial Mathematics*, **12**(1), 369–409. ISSN 1945-497X.
- Akyildirim, E., Gambarara, M., Teichmann, J. and Zhou, S. (2022) Applications of signature methods to market anomaly detection. *arXiv preprint arXiv:2201.02441*.

- Alfonsi, A. (2005) On the discretization schemes for the CIR (and Bessel squared) processes. *Monte Carlo Methods and Applications*, **11**(4), 355–384. ISSN 0929-9629,1569-3961.
- Alfonsi, A. and Kebaier, A. (2024) Approximation of Stochastic Volterra Equations with kernels of completely monotone type. *Mathematics of Computation*, **93**(346), 643–677. ISSN 0025-5718,1088-6842.
- Amisano, G. and Fagan, G. (2013) Money growth and inflation: A regime switching approach. *Journal of International Money and Finance*, **33**, 118–145.
- Asadi, S. and Al Janabi, M.A.M. (2020) Measuring market and credit risk under Solvency II: Evaluation of the standard technique versus internal models for stock and bond markets. *European Actuarial Journal*, **10**(2), 425–456. ISSN 2190-9733.
- Azran, A. and Ghahramani, Z. (2006) A new approach to data driven clustering. *Proceedings of the 23rd International Conference on Machine Learning*, pp. 57–64.
- Bayer, C., Hager, P.P., Riedel, S. and Schoenmakers, J. (2023) Optimal stopping with signatures. *Annals of Applied Probability*, **33**(1), 238–273. ISSN 1050-5164,2168-8737.
- Bilokon, P., Jacquier, A. and McIndoe, C. (2021) Market regime classification with signatures. *arXiv preprint arXiv:2107.00066*.
- Black, F. and Scholes, M. (1973) The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**(3), 637–654. ISSN 0022-3808.
- Boudreault, M. and Panneton, C.-M. (2009) Multivariate models of equity returns for investment guarantees valuation. *North American Actuarial Journal*, **13**(1), 36–53. ISSN 1092-0277.
- Buehler, H., Horvath, B., Lyons, T., Perez Arribas, I. and Wood, B. (2020) Generating financial markets with signatures. Available at SSRN 3657366.
- Cartea, A., Pérez Arribas, I. and Sánchez-Betancourt, L. (2022) Double-execution strategies using path signatures. *SIAM Journal on Financial Mathematics*, **13**(4), 1379–1417. ISSN 1945-497X.
- Chen, K.-T. (1957) Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula. *Annals of Mathematics* (2), **65**, 163–178. ISSN 0003-486X.
- Chevyrev, I. and Kormilitzin, A. (2016) A primer on the signature method in machine learning. *arXiv preprint arXiv:1603.03788*.
- Chevyrev, I. and Oberhauser, H. (2022) Signature moments to characterize laws of stochastic processes. *Journal of Machine Learning Research*, **23**(176), 1–42.
- Cohen, S.N., Lui, S., Malpass, W., Mantoan, G., Nesheim, L., de Paula, A., Reeves, A., Scott, C., Small, E. and Yang, L. (2023) Nowcasting with signature methods. *arXiv preprint arXiv:2305.10256*.
- Cont, R. and Das, P. (2022) Rough volatility: Fact or artefact? *arXiv preprint arXiv:2203.13820*.
- Cont, R. and Das, P. (2023) Quadratic variation and quadratic roughness. *Bernoulli*, **29**(1), 496–522.
- Cuchiero, C., Gazzani, G., Möller, J. and Svaluto-Ferro, S. (2023a) Joint calibration to spx and vix options with signature-based models. *arXiv preprint arXiv:2301.13235*.
- Cuchiero, C., Gazzani, G. and Svaluto-Ferro, S. (2023b) Signature-based models: Theory and calibration. *SIAM Journal on Financial Mathematics*, **14**(3), 910–957. ISSN 1945-497X.
- Cuchiero, C. and Möller, J. (2023) Signature methods in stochastic portfolio theory. *arXiv preprint arXiv:2310.02322*.
- Cucker, F. and Smale, S. (2002) On the mathematical foundations of learning. *Bulletin (New Series) of the American Mathematical Society*, **39**(1), 1–49. ISSN 10273-0979.
- Dudley, R.M. (2002) *Real Analysis and Probability*, Cambridge Studies in Advanced Mathematics, Vol. 74. Cambridge: Cambridge University Press. ISBN 0-521-00754-2. Revised reprint of the 1989 original.
- El Euch, O., Fukasawa, M. and Rosenbaum, M. (2018) The microstructural foundations of leverage effect and rough volatility. *Finance and Stochastics*, **22**(2), 241–280. ISSN 0949-2984,1432-1122.
- Evans, M. and Wachtel, P. (1993) Inflation regimes and the sources of inflation uncertainty. *Journal of Money, Credit and Banking*, **25**(3), 475–511.
- Fawcett, T. (2002) *Problems in stochastic analysis: Connections between rough paths and non-commutative harmonic analysis*. Ph.D. Thesis, University of Oxford.
- Fermanian, A. (2021) Embedding and learning with signatures. *Computational Statistics & Data Analysis*, **157**, Paper No. 107148, 23. ISSN 0167-9473.
- Floryszczak, A., Lévy Véhel, J. and Majri, M. (2019) A conditional equity risk model for regulatory assessment. *ASTIN Bulletin*, **49**(1), 217–242. ISSN 0515-0361.
- Gatheral, J., Jaisson, T. and Rosenbaum, M. (2018) Volatility is rough. *Quantitative Finance*, **18**(6), 933–949. ISSN 1469-7688.
- Graf, S., Haertel, L., Kling, A. and Ruß, J. (2014) The impact of inflation risk on financial planning and risk-return profiles. *Astin Bulletin*, **44**(2), 335–365. ISSN 0515-0361.
- Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B. and Smola, A. (2012) A kernel two-sample test. *Journal of Machine Learning Research*, **13**, 723–773. ISSN 1532-4435.
- Gretton, A., Fukumizu, K., Harchaoui, Z. and Sriperumbudur, B.K. (2009) A fast, consistent kernel two-sample test. *In Advances in Neural Information Processing Systems*, **22**.
- Gyurkó, L.G. and Lyons, T. (2010) Rough paths based numerical algorithms in computational finance. *In Mathematics in Finance. Contemporary Mathematics*, Vol. 515, pp. 17–46. Providence, RI: American Mathematical Society.
- Gyurkó, L.G., Lyons, T., Kontkowski, M. and Field, J. (2013) Extracting information from the signature of a financial data stream. *arXiv preprint arXiv:1307.7244*.
- Han, B. and Wong, H.Y. (2021) Mean-variance portfolio selection under Volterra Heston model. *Applied Mathematics and Optimization*, **84**(1), 683–710. ISSN 0095-4616,1432-0606.
- Hardy, M.R., Freeland, R.K. and Till, M.C. (2006) Validation of long-term equity return models for equity-linked guarantees. *North American Actuarial Journal*, **10**(4), 28–47. ISSN 1092-0277.

- Heber, G., Lunde, A., Shephard, N. and Sheppard, K. (2009) Oxford-man institute's realized library. *Version 0.1, Oxford&Man Institute, University of Oxford*.
- Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994) *Continuous Univariate Distributions*, 2nd ed. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, Vol. 1. New York: John Wiley & Sons, Inc. ISBN 0-471-58495-9. A Wiley-Interscience Publication.
- Kalsi, J., Lyons, T. and Arribas, I.P. (2020) Optimal execution with rough path signatures. *SIAM Journal on Financial Mathematics*, **11**(2), 470–493. ISSN 1945-497X.
- Kidger, P. and Lyons, T. (2020) Signatory: differentiable computations of the signature and logsignature transforms, on both CPU and GPU. *International Conference on Learning Representations*.
- Levin, D., Lyons, T. and Ni, H. (2013) Learning from the past, predicting the statistics for the future, learning an evolving system. arXiv preprint [arXiv:1309.0260](https://arxiv.org/abs/1309.0260).
- Lin, X.S. and Yang, S. (2020) Efficient dynamic hedging for large variable annuity portfolios with multiple underlying assets. *Astin Bulletin*, **50**(3), 913–957. ISSN 0515-0361.
- Lyons, T. (1998) Differential equations driven by rough signals. *Revista Matemática Iberoamericana*, **14**(2), 215–310. ISSN 0213-2230.
- Lyons, T., Caruana, M. and Lévy, T. (2007) *Differential Equations Driven by Rough Paths*. Lecture Notes in Mathematics, Vol. 1908. Berlin: Springer. ISBN 978-3-540-71284-8; 3-540-71284-4. Lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6–24, 2004, With an introduction concerning the Summer School by Jean Picard.
- Lyons, T., Nejad, S. and Perez Arribas, I. (2020) Non-parametric pricing and hedging of exotic derivatives. *Applied Mathematical Finance*, **27**(6), 457–494. ISSN 1350-486X, 1466-4313.
- Ma, J., Lu, Z. and Chen, D. (2023) Optimal reinsurance-investment with loss aversion under rough Heston model. *Quantitative Finance*, **23**(1), 95–109. ISSN 1469-7688, 1469-7696.
- Morrill, J., Fermanian, A., Kidger, P. and Lyons, T. (2020) A generalised signature method for multivariate time series feature extraction. arXiv preprint [arXiv:2006.00873](https://arxiv.org/abs/2006.00873).
- Ni, H., Szpruch, L., Sabate-Vidales, M., Xiao, B., Wiese, M. and Liao, S. (2021) Sig-wasserstein gans for time series generation. *Proceedings of the Second ACM International Conference on AI in Finance*, pp. 1–8.
- Otero, L., Durán, P., Fernández, S. and Vivel, M. (2012) Estimating insurers capital requirements through markov switching models in the solvency ii framework. *International Research Journal of Finance and Economics*, **86**, 20–38.
- Perez Arribas, I. (2020) *Signatures in machine learning and finance*. Ph.D. Thesis, University of Oxford.
- Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M.Z., Barrow, D.K., Taieb, S.B., Bergmeir, C., Bessa, R.J., Bijak, J., Boylan, J.E., Browell, J., Carnevale, C., Castle, J.L., Cirillo, P., Clements, M.P., Cordeiro, C., Oliveira, F.L.C., De Baets, S., Dokumentov, A., Ellison, J., Fiszeder, P., Franses, P.H., Frazier, D.T., Gilliland, M., Gönlü, M.S., Goodwin, P., Grossi, L., Grushka-Cockayne, Y., Guidolin, M., Guidolin, M., Gunther, U., Guo, X., Guseo, R., Harvey, N., Hendry, D.F., Hollyman, R., Januschowski, T., Jeon, J., Jose, V.R.R., Kang, Y., Koehler, A.B., Kolassa, S., Kourentzes, N., Leva, S., Li, F., Litsiou, K., Makridakis, S., Martin, G.M., Martinez, A.B., Meeran, S., Modis, T., Nikolopoulos, K., Önköl, D., Paccagnini, A., Panagiotelis, A., Panapakidis, I., Pavia, J.M., Pedio, M., Pedregal, D.J., Pinson, P., Ramos, P., Rapach, D.E., Reade, J.J., Rostami-Tabar, B., Rubaszek, M., Sermpinis, G., Shang, H.L., Spiliotis, E., Syntetos, A.A., Talagala, D., Talagala, T.S., Tashman, L., Thomakos, D., Thorarindottir, T., Todini, E., Arenas, J.R.T., Wang, X., Winkler, R.L., Yusupova, A. and Ziel, F. (2022) Forecasting: Theory and practice. *International Journal of Forecasting*, **38**(3), 705–871.
- Reizenstein, J.F. and Graham, B. (2020) Algorithm 1004: The iisignature library: Efficient calculation of iterated-integral signatures and log signatures. *ACM Transactions on Mathematical Software*, **46**(1), Art. 8, 21. ISSN 0098-3500, 1557-7295.
- Sen, P.K. (1977) Almost sure convergence of generalized U -statistics. *Annals of Probability*, **5**(2), 287–290. ISSN 0091-1798.
- Smith, G.D. (1985) *Numerical Solution of Partial Differential Equations*, 3rd ed. Oxford Applied Mathematics and Computing Science Series. New York: The Clarendon Press, Oxford University Press. ISBN 0-19-859641-3; 0-19-859650-2. Finite difference methods.
- US Bureau of Labor Statistics. (2023) Consumer price index for all urban consumers, U.S. city average, all items. <https://www.bls.gov/cpi/data.htm>. Accessed on: 2023-01-27.
- Wang, X., Xiao, W. and Yu, J. (2023) Modeling and forecasting realized volatility with the fractional ornstein–uhlenbeck process. *Journal of Econometrics*, **232**(2), 389–415.
- Young, L.C. (1936) An inequality of the Hölder type, connected with Stieltjes integration. *Acta Mathematica*, **67**(1), 251–282. ISSN 0001-5962.
- Zhu, X., Hardy, M.R. and Saunders, D. (2018) Dynamic hedging strategies for cash balance pension plans. *Astin Bulletin*, **48**(3), 1245–1275. ISSN 0515-0361.