



HAL
open science

Generative modeling for time series via Schrödinger bridge

Mohamed Hamdouche, Pierre Henry-Labordere, Huyên Pham

► **To cite this version:**

Mohamed Hamdouche, Pierre Henry-Labordere, Huyên Pham. Generative modeling for time series via Schrödinger bridge. Université Paris Cité. 2023. hal-04063041

HAL Id: hal-04063041

<https://hal.science/hal-04063041>

Submitted on 7 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Generative modeling for time series via Schrödinger bridge

Mohamed HAMDUCHE* Pierre HENRY-LABORDERE† Huyên PHAM‡

Abstract

We propose a novel generative model for time series based on Schrödinger bridge (SB) approach. This consists in the entropic interpolation via optimal transport between a reference probability measure on path space and a target measure consistent with the joint data distribution of the time series. The solution is characterized by a stochastic differential equation on finite horizon with a path-dependent drift function, hence respecting the temporal dynamics of the time series distribution. We can estimate the drift function from data samples either by kernel regression methods or with LSTM neural networks, and the simulation of the SB diffusion yields new synthetic data samples of the time series.

The performance of our generative model is evaluated through a series of numerical experiments. First, we test with a toy autoregressive model, a GARCH Model, and the example of fractional Brownian motion, and measure the accuracy of our algorithm with marginal and temporal dependencies metrics. Next, we use our SB generated synthetic samples for the application to deep hedging on real-data sets. Finally, we illustrate the SB approach for generating sequence of images.

Keywords: generative models, time series, Schrödinger bridge, kernel estimation, deep hedging.

1 Introduction

Sequential data appear widely in our society like in video and audio data, and simulation of time series models are used in various industrial applications including clinical predictions [17], and weather forecasts [5]. In the financial industry, simulations of dynamical scenarios are considered in market stress tests, risk measurement, and risk management, e.g. in deep hedging [2]. The design of time series model is a delicate issue, requiring expensive calibration task, and subject to error misspecification and model risk. Therefore, the generation for synthetic samples of time series has gained an increasing attention over the last years, and opens the door in the financial sector for a pure data driven approach in risk management.

*LPSM, Université Paris Cité and Sorbonne Université, and Qube Research and Technologies, hamdouche at lpsm.paris

†Qube Research and Technologies, pierre.henrylaborderre at qube-rt.com

‡LPSM, Université Paris Cité and Sorbonne Université, pham at lpsm.paris; This author is supported by the BNP-PAR Chair “Futures of Quantitative Finance”, and by FiME, Laboratoire de Finance des Marchés de l’Energie, and the “Finance and Sustainable Development” EDF - CACIB Chair

Generative modeling (GM) has become over the last years a successful machine learning task for data synthesis notably in (static) image processing. Several competing methods have been developed and state-of-the-art includes *Likelihood-based models* like energy-based models (EBM) [15], variational auto-encoders (VAE) [14], *Implicit generative models* with the prominent works on generative adversarial network (GAN) [11] and its extensions [1], and recently the new *generation of score-based models* using Langevin dynamics, [21], [22], [12], and diffusion models via Schrödinger bridge, see [13] for the application to a class of stochastic volatility models, and [23], [8]. Generative methods for time series raises challenging issues for learning efficiently the temporal dependencies. Indeed, in order to capture the potentially complex dynamics of variables across time, it is not sufficient to learn the time marginals or even the joint distribution without exploiting the sequential structure. An increasing attention has been paid to these methods in the literature and state-of-the-art generative methods for time series are: *Time series GAN* [27] which combines an unsupervised adversarial loss on real/synthetic data and supervised loss for generating sequential data, *Quant GAN* [25] with an adversarial generator using temporal convolutional networks, *Causal optimal transport COT-GAN* [26] with adversarial generator using the adapted Wasserstein distance for processes, *Conditional loss Euler generator* [20] starting from a diffusion representation time series and minimizing the conditional distance between transition probabilities of real/synthetic samples, *Signature embedding of time series* [9], [18], [3], and *Functional data analysis* with neural SDEs [6].

In this paper, we develop a novel generative model based on Schrödinger bridge approach that captures the temporal dynamics of the time series. This consists in the entropic interpolation via optimal transport between a reference probability measure on path space and a target measure consistent with the joint data distribution of the time series. The solution is characterized by a stochastic differential equation on finite horizon with a path-dependent drift function, called Schrödinger bridge time series (SBTS) diffusion, and the simulation of the SBTS diffusion yields new synthetic data samples of the time series.

Our SB approach differs from related works that have been recently designed for learning marginal (or static) distributions. In [23], the authors perform generative models by solving two SB problems. The paper [8] formulates generative modeling by computing the SB problem between the data and prior distribution. The very recent work [4] proposes momentum SB by considering an additional velocity variable for learning multi marginal distributions. Let us mention also the recent paper [19] that combines SB with h -transform in order to respect aligned data. Instead, our SBTS diffusion interpolates the joint time series distribution starting from an initial deterministic value. Moreover, we propose an alternate method for the estimation of the drift function, which is path-dependent in our case. While [23] uses a logistic regression for estimating the density ratio and then the drift function, which requires additional samples from Gaussian noises, and [8] performs an extension of the Sinkhorn algorithm, we propose a kernel regression method relying solely on data samples, and this turns out to be quite simple, efficient and low-cost computationally. Compared to GAN type methods, the simulation of synthetic samples from SBTS is much faster as it does not require the training of neural networks.

We validate our methodology with several numerical experiments. We first test on some time series models like autoregressive, GARCH models, and also for the fractional Brownian motion with rough paths. The accuracy is measured by some metrics aiming to

capture the temporal dynamics and the correlation structure. We also provide operational metrics of interest for the financial industry by implementing our results on real datasets, and applying to the deep hedging of call options. Finally, we show some numerical illustrations of our SB method in high dimension for the generation of sequential images.

2 Problem formulation

Let μ be the distribution of a time series representing the evolution of some \mathbb{R}^d -valued process of interest (e.g. asset price, claim process, audio/video data, etc), and suppose that one can observe samples of this process at given fixed times of a discrete time grid $\mathcal{T} = \{t_i, i = 1, \dots, N\}$ on $(0, \infty)$. We set $T = t_N$ as the terminal observation horizon. Our goal is to construct a model that generates time series samples according to the unknown target distribution $\mu \in \mathcal{P}((\mathbb{R}^d)^N)$ the set of probability measures on $(\mathbb{R}^d)^N$. For that objective, we propose a dynamic modification of the Schrödinger bridge as follows. Let $\Omega = C([0, T]; \mathbb{R}^d)$ be the space of \mathbb{R}^d -valued continuous functions on $[0, T]$, $X = (X_t)_t$ the canonical process with initial value $X_0 = 0$, and $\mathbb{F} = (\mathcal{F}_t)_t$ the canonical filtration. Denoting by $\mathcal{P}(\Omega)$ as the space of probability measures on Ω , we search for \mathbb{P} (representing the theoretical generative model) in $\mathcal{P}(\Omega)$, close to the Wiener measure \mathbb{W} in the sense of Kullback-Leibler (or relative entropy), and consistent with the observation samples. In other words, we look for a probability measure $\mathbb{P}^* \in \mathcal{P}(\Omega)$ solution to:

$$\mathbb{P}^* \in \arg \min_{\mathbb{P} \in \mathcal{P}_{\mathcal{T}}^{\mu}(\Omega)} H(\mathbb{P}|\mathbb{W}), \quad (2.1)$$

where $\mathcal{P}_{\mathcal{T}}^{\mu}(\Omega)$ is the set of probability measures \mathbb{P} on Ω with joint distribution μ at (t_1, \dots, t_N) , i.e., $\mathbb{P} \circ (X_{t_1}, \dots, X_{t_N})^{-1} = \mu$, and $H(\cdot|\cdot)$ is the relative entropy between two probability measures defined by

$$\mathcal{H}(\mathbb{P}|\mathbb{W}) = \begin{cases} \int \ln \frac{d\mathbb{P}}{d\mathbb{W}} d\mathbb{P}, & \text{if } \mathbb{P} \ll \mathbb{W} \\ \infty, & \text{otherwise.} \end{cases}$$

Denoting by $\mathbb{E}_{\mathbb{P}}$ and $\mathbb{E}_{\mathbb{W}}$ the expectation under \mathbb{P} and \mathbb{W} , we see that $\mathcal{H}(\mathbb{P}|\mathbb{W}) = \mathbb{E}_{\mathbb{P}}[\ln \frac{d\mathbb{P}}{d\mathbb{W}}] = \mathbb{E}_{\mathbb{W}}[\frac{d\mathbb{P}}{d\mathbb{W}} \ln \frac{d\mathbb{P}}{d\mathbb{W}}]$ when $\mathbb{P} \ll \mathbb{W}$. Compared to the classical Schrödinger bridge (SB) (see [16], and the application to generative modeling in [23]), which looks for a probability measure \mathbb{P} that interpolates between an initial probability measure and a target probability distribution at terminal time T , here, we take into account via the constraint in $\mathcal{P}_{\mathcal{T}}^{\mu}$ the temporal dependence of the process observed at sequential times $t_1 < \dots < t_N$, and look for an entropic interpolation of the time series distribution. We call (2.1) the Schrödinger bridge for time series (SBTS) problem.

Let us now formulate (SBTS) as a stochastic control problem following the well-known connection established for classical (SB) in [7]. Given $\mathbb{P} \in \mathcal{P}(\Omega)$ with finite relative entropy $H(\mathbb{P}|\mathbb{W}) < \infty$, it is known by Girsanov's theorem that one can associate to \mathbb{P} an \mathbb{F} -adapted \mathbb{R}^d -valued process $\alpha = (\alpha_t)$ with finite energy: $\mathbb{E}_{\mathbb{P}}[\int_0^T |\alpha_t|^2 dt] < \infty$ such that

$$\ln \frac{d\mathbb{P}}{d\mathbb{W}} = \int_0^T \alpha_t \cdot dX_t - \frac{1}{2} \int_0^T |\alpha_t|^2 dt, \quad (2.2)$$

and $X_t - \int_0^t \alpha_s ds$, $0 \leq t \leq T$, is a Brownian motion under \mathbb{P} . We then have

$$\mathcal{H}(\mathbb{P}|\mathbb{W}) = \mathbb{E}_{\mathbb{P}} \left[\frac{1}{2} \int_0^T |\alpha_t|^2 dt \right].$$

Therefore, (SBTS) is reformulated equivalently in the language of stochastic control as:

$$\begin{cases} \text{Minimize over } \alpha \in \mathcal{A}, & J(\alpha) = \mathbb{E}_{\mathbb{P}} \left[\frac{1}{2} \int_0^T |\alpha_t|^2 dt \right] \\ \text{subject to } dX_t = \alpha_t dt + dW_t, & X_0 = 0, \quad (X_{t_1}, \dots, X_{t_N}) \overset{\mathbb{P}}{\sim} \mu, \end{cases}$$

where W is a Brownian motion under \mathbb{P} , \mathcal{A} is the set of \mathbb{R}^d -valued \mathbb{F} -adapted processes s.t. $\mathbb{E}_{\mathbb{P}}[\int_0^T |\alpha_t|^2 dt] < \infty$, and $(X_{t_1}, \dots, X_{t_N}) \overset{\mathbb{P}}{\sim} \mu$ is the usual notation for $\mathbb{P} \circ (X_{t_1}, \dots, X_{t_N})^{-1} = \mu$. In the sequel, when there is no ambiguity, we omit the reference on \mathbb{P} in $\mathbb{E} = \mathbb{E}_{\mathbb{P}}$ and $\sim = \overset{\mathbb{P}}{\sim}$. We denote by V_{SBTS} the infimum of this stochastic control problem under joint distribution constraint:

$$V_{SBTS} := \inf_{\alpha \in \mathcal{A}_{\mathcal{T}}^{\mu}} J(\alpha),$$

where $\mathcal{A}_{\mathcal{T}}^{\mu}$ is the set of controls α in \mathcal{A} satisfying $(X_{t_1}, \dots, X_{t_N}) \sim \mu$ with $X_t = \int_0^t \alpha_s ds + W_t$. Our goal is to prove the existence of an optimal control α^* that can be explicitly derived, and then used to generate samples of the time series distribution μ via the probability measure \mathbb{P}^* on Ω , i.e., the simulation of the optimal diffusion process X controlled by the drift α^* .

3 Solution to Schrödinger bridge for time series

Similarly as for the classical Schrödinger bridge problem, we assume that the target distribution μ admits a density with respect to the Lebesgue measure on $(\mathbb{R}^d)^N$, and by misuse of notation, we denote by $\mu(x_1, \dots, x_N)$ this density function. Denote by $\mu_{\mathcal{T}}^W$ the distribution of the Brownian motion on \mathcal{T} , i.e. of $(W_{t_1}, \dots, W_{t_N})$, which admits a density given by (by abuse of language, we use the same notation for the measure and its density)

$$\mu_{\mathcal{T}}^W(x_1, \dots, x_N) = \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi(t_{i+1} - t_i)}} \exp\left(-\frac{|x_{i+1} - x_i|^2}{2(t_{i+1} - t_i)}\right), \quad (3.1)$$

for $(x_1, \dots, x_N) \in (\mathbb{R}^d)^N$ (with the convention that $t_0 = 0$, $x_0 = 0$). The measure μ is absolutely continuous with respect to $\mu_{\mathcal{T}}^W$, and we shall assume that its relative entropy is finite, i.e.,

$$\mathcal{H}(\mu|\mu_{\mathcal{T}}^W) = \int \ln \frac{\mu}{\mu_{\mathcal{T}}^W} d\mu < \infty.$$

The solution to the (SBTS) problem is provided in the following theorem.

Theorem 3.1. *The diffusion process $X_t = \int_0^t \alpha_s^* ds + W_t$, $0 \leq t \leq T$, with α^* defined as*

$$\alpha_t^* = \mathbf{a}^*(t, X_t; (X_{t_i})_{t_i \leq t}), \quad 0 \leq t < T,$$

with $\mathbf{a}^*(t, x; \mathbf{x}_i)$, for $t \in [t_i, t_{i+1})$, $\mathbf{x}_i = (x_1, \dots, x_i) \in (\mathbb{R}^d)^i$, $x \in \mathbb{R}^d$, given by

$$\mathbf{a}^*(t, x; \mathbf{x}_i) = \nabla_x \ln \mathbb{E}_{\mathbb{W}} \left[\frac{\mu}{\mu_{\mathcal{F}}^W} (X_{t_1}, \dots, X_{t_N}) \mid \mathbf{X}_{t_i} = \mathbf{x}_i, X_t = x \right],$$

where we set $\mathbf{X}_{t_i} = (X_{t_1}, \dots, X_{t_i})$, induces a probability measure $\mathbb{P}^* = \frac{\mu}{\mu_{\mathcal{F}}^W} (X_{t_1}, \dots, X_{t_N}) \mathbb{W}$, which solves the Schrödinger bridge time series problem. Moreover, we have

$$V_{SBTS} = \mathcal{H}(\mathbb{P}^* | \mathbb{W}) = \mathcal{H}(\mu | \mu_{\mathcal{F}}^W).$$

Proof. First, observe that $\mathbb{E}_{\mathbb{W}} \left[\frac{\mu}{\mu_{\mathcal{F}}^W} (X_{t_1}, \dots, X_{t_N}) \right] = 1$, and thus one can define a probability measure $\mathbb{P}^* \ll \mathbb{W}$ with density process

$$Z_t = \mathbb{E}_{\mathbb{W}} \left[\frac{d\mathbb{P}^*}{d\mathbb{W}} \mid \mathcal{F}_t \right] = \mathbb{E}_{\mathbb{W}} \left[\frac{\mu}{\mu_{\mathcal{F}}^W} (X_{t_1}, \dots, X_{t_N}) \mid \mathcal{F}_t \right], \quad 0 \leq t \leq T.$$

Notice from the Markov and Gaussian properties of the Brownian motion that for $t \in [t_i, t_{i+1})$, $i = 0, \dots, N-1$, we have $Z_t = h_i(t, X_t; \mathbf{X}_{t_i})$, where for a path $\mathbf{x}_i = (x_1, \dots, x_i) \in (\mathbb{R}^d)^i$, $h_i(\cdot; \mathbf{x}_i)$ is defined on $[t_i, t_{i+1}) \times \mathbb{R}^d$ by

$$h_i(t, x; \mathbf{x}_i) = \mathbb{E}_{Y \sim \mathcal{N}(0, I_d)} \left[\frac{\mu}{\mu_{\mathcal{F}}^W} (\mathbf{x}_i, x + \sqrt{t_{i+1} - t}Y, \dots, x + \sqrt{t_N - t}Y) \right]$$

for $t \in [t_i, t_{i+1})$, $x \in \mathbb{R}^d$, and $\mathbb{E}_{Y \sim \mathcal{N}(0, I_d)}[\cdot]$ is the expectation when Y is distributed according to the Gaussian law $\mathcal{N}(0, I_d)$. Moreover, by the law of conditional expectations, we have

$$h_i(t, x; \mathbf{x}_i) = \mathbb{E}_{\mathbb{W}} \left[h_{i+1}(t_{i+1}, X_{t_{i+1}}; \mathbf{x}_i, X_{t_{i+1}}) \mid X_t = x \right],$$

with the convention that $h_N(t_N, x; \mathbf{x}_{N-1}, x) = \frac{\mu}{\mu_{\mathcal{F}}^W} (x_1, \dots, x_{N-1}, x)$. Therefore, for $i = 0, \dots, N-1$, and $\mathbf{x} \in (\mathbb{R}^d)^i$, the function $(t, x) \mapsto h_i(t, x; \mathbf{x}_i)$ is a strictly positive $C^{1,2}([t_i, t_{i+1}) \times \mathbb{R}^d) \cap C^0([t_i, t_{i+1}] \times \mathbb{R}^d)$ classical solution to the heat equation

$$\frac{\partial h_i(\cdot; \mathbf{x}_i)}{\partial t} + \frac{1}{2} \Delta_x h_i(\cdot; \mathbf{x}_i) = 0, \quad \text{on } [t_i, t_{i+1}) \times \mathbb{R}^d,$$

with the terminal condition: $h_i(t_{i+1}, x; \mathbf{x}_i) = h_{i+1}(t_{i+1}, x; \mathbf{x}_i, x)$ (here Δ_x is the Laplacian operator). By applying Itô's formula to the martingale density process Z of \mathbb{P}^* under the Wiener measure \mathbb{W} , we derive

$$\begin{aligned} dZ_t &= \nabla_x h_i(t, X_t; \mathbf{X}_{t_i}) dX_t, \\ &= Z_t \nabla_x \ln h_i(t, X_t; \mathbf{X}_{t_i}) dX_t, \quad t_i \leq t < t_{i+1}, \end{aligned}$$

for $i = 0, \dots, N - 1$. Thus, by defining the process α^* by $\alpha_t^* = \nabla_x \ln h_i(t, X_t; \mathbf{X}_{t_i})$, for $t \in [t_i, t_{i+1})$, $i = 0, \dots, N - 1$, we have

$$\frac{d\mathbb{P}^*}{d\mathbb{W}} = \exp\left(\int_0^T \alpha_t^* dX_t - \frac{1}{2} \int_0^T |\alpha_t^*|^2 dt\right),$$

and by Girsanov's theorem, $X_t - \int_0^t \alpha_s^* ds$ is a Brownian motion under \mathbb{P}^* . On the other hand, by definition of \mathbb{P}^* , and Bayes formula, we have for any bounded measurable function φ on $(\mathbb{R}^d)^N$:

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^*}[\varphi(X_{t_1}, \dots, X_{t_N})] &= \mathbb{E}_{\mathbb{W}}\left[\frac{\mu}{\mu_{\mathcal{T}}^W}(X_{t_1}, \dots, X_{t_N})\varphi(X_{t_1}, \dots, X_{t_N})\right] \\ &= \int \frac{\mu}{\mu_{\mathcal{T}}^W}(x_1, \dots, x_N)\varphi(x_1, \dots, x_N)\mu_{\mathcal{T}}^W(x_1, \dots, x_N)dx_1 \dots dx_N \\ &= \int \varphi(x_1, \dots, x_N)\mu(x_1, \dots, x_N)dx_1 \dots dx_N, \end{aligned}$$

which shows that $(X_{t_1}, \dots, X_{t_N}) \stackrel{\mathbb{P}^*}{\sim} \mu$. Moreover, by noting that

$$J(\alpha^*) = \mathbb{E}_{\mathbb{P}^*}\left[\int_0^T \frac{1}{2}|\alpha_t^*|^2 dt\right] = \mathbb{E}_{\mathbb{P}^*}\left[\ln \frac{d\mathbb{P}^*}{d\mathbb{W}}\right] = \mathcal{H}(\mathbb{P}^*|\mathbb{W}) = \mathcal{H}(\mu|\mu_{\mathcal{T}}^W) < \infty,$$

where we used in the last inequality the fact that $(X_{t_1}, \dots, X_{t_N}) \stackrel{\mathbb{P}^*}{\sim} \mu$, this shows in particular that $\alpha^* \in \mathcal{A}_{\mathcal{T}}^\mu$.

It remains to show that for any $\alpha \in \mathcal{A}_{\mathcal{T}}^\mu$ associated to a probability measure $\mathbb{P} \ll \mathbb{W}$ with density given by (2.2), i.e. $J(\alpha) = \mathcal{H}(\mathbb{P}|\mathbb{W})$, we have

$$J(\alpha) \geq \mathcal{H}(\mu|\mu_{\mathcal{T}}^W). \quad (3.2)$$

For this, we write from Bayes formula and since $W_t = X_t - \int_0^t \alpha_s ds$ is a Brownian motion under \mathbb{P} by Girsanov theorem:

$$\begin{aligned} 1 &= \mathbb{E}_{\mathbb{W}}\left[\frac{\mu}{\mu_{\mathcal{T}}^W}(X_{t_1}, \dots, X_{t_N})\right] \\ &= \mathbb{E}_{\mathbb{P}}\left[\exp\left(\ln \frac{\mu}{\mu_{\mathcal{T}}^W}(X_{t_1}, \dots, X_{t_N}) - \int_0^T \alpha_t dW_t - \frac{1}{2} \int_0^T |\alpha_t|^2 dt\right)\right] \\ &\geq \exp\left(\mathbb{E}_{\mathbb{P}}\left[\ln \frac{\mu}{\mu_{\mathcal{T}}^W}(X_{t_1}, \dots, X_{t_N}) - \int_0^T \alpha_t dW_t - \frac{1}{2} \int_0^T |\alpha_t|^2 dt\right]\right) \\ &= \exp\left(\mathcal{H}(\mu|\mu_{\mathcal{T}}^W) - J(\alpha)\right), \end{aligned}$$

where we use Jensen's inequality, and the fact that $(X_{t_1}, \dots, X_{t_N}) \stackrel{\mathbb{P}}{\sim} \mu$ in the last equality. This proves the required inequality (3.2), and ends the proof. \square

Remark 3.2. *The optimal drift of the Schrödinger bridge time series diffusion is in general path-dependent: it depends at given time t not only on its current state X_t , but also on the past values $\mathbf{X}_{\eta(t)} = (X_{t_1}, \dots, X_{\eta(t)})$, where $\eta(t) = \max\{t_i : t_i \leq t\}$, and we have:*

$$dX_t = \mathbf{a}^*(t, X_t; \mathbf{X}_{\eta(t)})dt + dW_t, \quad 0 \leq t \leq T, \quad X_0 = 0.$$

Moreover, the proof of the above theorem shows that this drift function is explicitly given by

$$\mathbf{a}^*(t, x; \mathbf{x}_i) = \frac{\nabla_x h_i(t, x; \mathbf{x}_i)}{h_i(t, x; \mathbf{x}_i)}, \quad t \in [t_i, t_{i+1}), \mathbf{x}_i \in (\mathbb{R}^d)^i, x \in \mathbb{R}^d, \quad (3.3)$$

for $i = 0, \dots, N-1$, where

$$h_i(t, x; \mathbf{x}_i) = \mathbb{E}_{Y \sim \mathcal{N}(0, I_d)} \left[\rho(\mathbf{x}_i, x + \sqrt{t_{i+1} - t}Y, \dots, x + \sqrt{t_N - t}Y) \right],$$

with $\rho := \frac{\mu}{\mu_T}$ the density ratio.

The following result states an alternate representation of the drift function that will be useful in the next section for estimation.

Proposition 3.3. *For $i = 0, \dots, N-1$, $t \in [t_i, t_{i+1})$, $\mathbf{x}_i = (x_1, \dots, x_i) \in (\mathbb{R}^d)^i$, $x \in \mathbb{R}^d$, we have*

$$\mathbf{a}^*(t, x; \mathbf{x}_i) = \frac{1}{t_{i+1} - t} \frac{\mathbb{E}_\mu \left[(X_{t_{i+1}} - x) F_i(t, x_i, x, X_{t_{i+1}}) \mid \mathbf{X}_{t_i} = \mathbf{x}_i \right]}{\mathbb{E}_\mu \left[F_i(t, x_i, x, X_{t_{i+1}}) \mid \mathbf{X}_{t_i} = \mathbf{x}_i \right]}, \quad (3.4)$$

where

$$F_i(t, x_i, x, x_{i+1}) = \exp \left(-\frac{|x_{i+1} - x|^2}{2(t_{i+1} - t)} + \frac{|x_{i+1} - x_i|^2}{2(t_{i+1} - t_i)} \right),$$

and $\mathbb{E}_\mu[\cdot]$ denotes the expectation under μ .

Proof. Fix $i \in \llbracket 0, N-1 \rrbracket$, and $t \in [t_i, t_{i+1})$. From the expression of μ_T^W in (3.1), we have

$$\begin{aligned} & \mathbb{E}_W \left[\frac{\mu}{\mu_T^W}(X_{t_1}, \dots, X_{t_N}) \mid (X_{t_1}, \dots, X_{t_i}) = (x_1, \dots, x_i), X_t = x \right] \\ &= C \int \frac{\mu}{\mu_T^W}(x_1, \dots, x_N) \exp \left(-\frac{|x_{i+1} - x|^2}{2(t_{i+1} - t)} \right) \prod_{j=i+1}^{N-1} \exp \left(-\frac{|x_{j+1} - x_j|^2}{2\Delta t_j} \right) dx_{i+1} \cdots dx_N \\ &= C \int F_i(t, x_i, x, x_{i+1}) \frac{\mu(x_1, \dots, x_N)}{\mu_i(x_1, \dots, x_i)} dx_{i+1} \cdots dx_N = C \mathbb{E}_\mu \left[F_i(t, x_i, x, X_{t_{i+1}}) \mid \mathbf{X}_{t_i} = \mathbf{x}_i \right], \end{aligned} \quad (3.5)$$

where C is a constant varying from line to line and depending only on t and $\mathbf{x}_i = (x_1, \dots, x_i)$, but not on x , and μ_i is the density of $(X_{t_1}, \dots, X_{t_i})$ under μ , i.e.,

$$\mu_i(x_1, \dots, x_i) = \int \mu(x_1, \dots, x_N) dx_{i+1} \cdots dx_N.$$

By plugging the new expression (3.5) into a^* and differentiating with respect to x , we then get

$$a^*(t, x; \mathbf{x}_i) = \frac{1}{t_{i+1} - t} \frac{\mathbb{E}_\mu [(X_{t_{i+1}} - x) F_i(t, x_i, x, X_{t_{i+1}}) | \mathbf{X}_{t_i} = \mathbf{x}_i]}{\mathbb{E}_\mu [F_i(t, x_i, x, X_{t_{i+1}}) | \mathbf{X}_{t_i} = \mathbf{x}_i]}.$$

□

Remark 3.4. *In the case where μ is the distribution arising from a Markov chain, i.e., in the form $\mu(dx_1, \dots, x_N) = \prod_{i=1}^{N-1} \nu_i(x_i, dx_{i+1})$, for some transition kernels ν_i on \mathbb{R}^d , then the conditional expectations in (3.4) will depend on the past values $\mathbf{X}_{t_i} = (X_{t_1}, \dots, X_{t_i})$ only via the last value X_{t_i} .*

4 Generative learning

From Theorem 3.1, we can run an Euler scheme for simulating the Schrödinger bridge diffusion, and then samples of the target distribution μ . For that purpose, we need an accurate estimation of the drift terms, i.e., of the functions a_i^* , for $i = 0, \dots, N - 1$. We propose several estimation methods. In the sequel, for a probability measure ν on $(\mathbb{R}^d)^N$, we denote by $\mathbb{E}_\nu[\cdot]$ the expectation under the distribution ν .

4.1 Drift estimation

Estimation of the density ratio. This method follows the idea in [23]. Denote by $\rho = \frac{\mu}{\mu_T^W}$ the density ratio, and observe that the log-density ratio $\ln \rho$ minimizes over functions r on $(\mathbb{R}^d)^N$ the logistic regression function

$$L_{\text{logistic}}(r) = \mathbb{E}_\mu \left[\ln (1 + \exp(-r(\mathbf{X}))) \right] + \mathbb{E}_{\mu_T^W} \left[\ln (1 + \exp(r(\mathbf{X}))) \right].$$

Therefore, given data samples $\mathbf{X}^{(m)} = (X_{t_1}^{(m)}, \dots, X_{t_N}^{(m)})$, $m = 1, \dots, M$ from μ , and using samples $\mathbf{Y}^{(m)}$ from μ_T^W , we estimate the density ratio ρ by

$$\hat{\rho} = \exp(r_{\hat{\theta}})$$

where $r_{\hat{\theta}}$ is the neural network that minimizes the empirical logistic loss function:

$$\theta \mapsto \frac{1}{M} \sum_{m=1}^M \ln (1 + \exp(-r(\mathbf{X}^{(m)}))) + \ln (1 + \exp(r(\mathbf{Y}^{(m)}))).$$

By writing from (3.3) the Schrödinger drift a^* as

$$a^*(t, x; \mathbf{x}_i) = \frac{\mathbb{E}_{Y \sim \mathcal{N}(0, I_d)} \left[\rho \nabla_x \rho(\mathbf{x}_i, x + \sqrt{t_{i+1} - t} Y, \dots, x + \sqrt{t_N - t} Y) \right]}{\mathbb{E}_{Y \sim \mathcal{N}(0, I_d)} \left[\rho(\mathbf{x}_i, x + \sqrt{t_{i+1} - t} Y, \dots, x + \sqrt{t_N - t} Y) \right]}, \quad (4.1)$$

for $t \in [t_i, t_{i+1})$, we obtain an estimator of a^* by plugging into (4.1) the estimate $\hat{\rho}$, and $r_{\hat{\theta}}$ of ρ , and $\ln \rho$, and then computing the expectation with Monte-Carlo approximations from samples in $\mathcal{N}(0, I_d)$. Notice that this method is very costly as it requires in addition to the training of the neural networks for estimating the density ratio, another Monte-Carlo sampling for estimating finally the drift.

Kernel estimation of the drift. In order to overcome the computational issue of the above estimation method, we propose an alternative approach that relies on the representation of the drift term in Proposition 3.3. Indeed, the key feature of the formula (3.4) is that it involves (conditional) expectations under the target distribution μ , which is amenable to direct estimation using data samples. For the approximation of the conditional expectation, we can then use the classical kernel methods.

From data samples $\mathbf{X}^{(m)} = (X_{t_1}^{(m)}, \dots, X_{t_N}^{(m)})$, $m = 1, \dots, M$ from μ , the Nadaraya-Watson estimator of the drift function is given by

$$\hat{a}(t, x; \mathbf{x}_i) = \frac{1}{t_{i+1} - t} \frac{\sum_{m=1}^M (X_{t_{i+1}}^{(m)} - x) F_i(t, X_{t_i}^{(m)}, x, X_{t_{i+1}}^{(m)}) \prod_{j=1}^i K_h(x_j - X_{t_j}^{(m)})}{\sum_{m=1}^M F_i(t, X_{t_i}^{(m)}, x, X_{t_{i+1}}^{(m)}) \prod_{j=1}^i K_h(x_j - X_{t_j}^{(m)})}, \quad (4.2)$$

for $t \in [t_i, t_{i+1})$, $\mathbf{x}_i \in (\mathbb{R}^d)^i$, $x \in \mathbb{R}^d$, $i = 0, \dots, N - 1$, where K_h is a kernel, i.e., a non-negative real-valued integrable and symmetric function on \mathbb{R}^d , with bandwidth $h > 0$. Common kernel function is the Gaussian density function, but we shall use here for lower time complexity reason, the quartic kernel $K_h(x) = \frac{1}{h} K(\frac{x}{h})$ with

$$K(x) = (1 - |x|^2) 1_{|x| \leq 1}.$$

LSTM network approximation of the path-dependent drift. The conditional expectations in the numerator and denominator of the drift terms can alternately be approximated by neural networks. In order to achieve this, we need a neural network architecture that fits well with the path-dependency of the drift term, i.e., the *a priori* non Markov feature of the data time series distribution μ . We shall then consider a combination of feed-forward and LSTM (Long Short Term Memory) neural network.

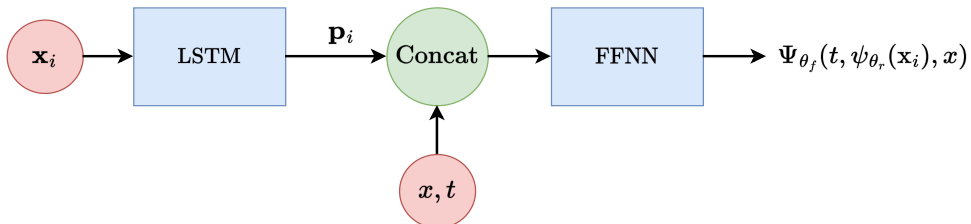


Figure 1: Architecture of the neural network

For $i = 0, \dots, N - 1$, the conditional expectation function in the numerator:

$$\begin{aligned} (t, \mathbf{x}_i, x) &\in [t_i, t_{i+1}) \times (\mathbb{R}^d)^i \times \mathbb{R}^d \\ \mapsto &\mathbb{E}_\mu \left[(X_{t_{i+1}} - X_t) F(t, X_{t_i}, X_t, X_{t_{i+1}}) \mid (\mathbf{X}_{t_i}, X_t) = (\mathbf{x}_i, x) \right], \end{aligned}$$

is approximated by $\Psi_{\theta_f}(t, p_{t_i}, x)$, where $\Psi_{\theta_f} \in NN_{d+k+1, d}$ is a feed-forward neural network with input dimension $1 + k + d$, and output dimension d , and p_{t_i} is an output vector of

dimension k from an LSTM network, i.e., $p_{t_i} = \psi_{\theta_r}(\mathbf{x}_i)$ with $\psi_{\theta_r} \in \text{LSTM}_{i,d,k}$ at time t_i . This neural network is trained by minimizing over the parameter $\theta = (\theta_f, \theta_r)$ the quadratic loss function

$$L(\theta) = \sum_{i=0}^{N-1} \hat{\mathbb{E}} \left| (X_{t_{i+1}} - X) F(\tau, X_{t_i}, X, X_{t_{i+1}}) - \Psi_{\theta_f}(\tau, \psi_{\theta_r}(\mathbf{X}_{t_i}), X) \right|^2.$$

Here $\hat{\mathbb{E}}$ is the empirical loss expectation where $(X_{t_1}, \dots, X_{t_N})$ are sampled from the data distribution μ , τ is sampled according to a uniform law on $[t_i, t_{i+1})$, and X is sampled e.g. from a Gaussian law with mean X_{t_i} for $i = 0, \dots, N-1$. The conditional expectation function in the denominator is similarly approximated.

The output of this neural network training yields an approximation:

$$t \in [0, T], \mathbf{x} \in (\mathbb{R}^d)^{\eta(t)}, x \in \mathbb{R}^d \mapsto \hat{\mathbf{a}}(t, x; \mathbf{x}),$$

of the drift term function, which is then used for generating samples $(X_{t_1}, \dots, X_{t_N})$ of μ from the simulation of the diffusion:

$$dX_t = \hat{\mathbf{a}}(t, X_t; \mathbf{X}_{\eta(t)})dt + dW_t, X_0 = 0. \quad (4.3)$$

4.2 Schrödinger bridge time series algorithm

From the estimator $\hat{\mathbf{a}}$ of the drift path-dependent function, we can now simulate the SB SDE (4.3) by an Euler scheme. Let N^π be the number of uniform time steps between two consecutive observation dates t_i and t_{i+1} , for $i = 0, \dots, N-1$, and $t_{k,i}^\pi = t_i + \frac{k}{N^\pi}$, $k = 0, \dots, N^\pi - 1$, the associated time grid. The pseudo-code of the Schrödinger bridge time series (SBTS) algorithm is described in Algorithm 1.

Algorithm 1: SBTS Simulation

Input: data samples of time series $(X_{t_1}^{(m)}, \dots, X_{t_N}^{(m)})$, $m = 1, \dots, M$, and N^π .

Initialization: initial state $x_0 = 0$;

for $i = 0, \dots, N-1$ **do**

Initialize state $y_0 = x_i$;

for $k = 0, \dots, N^\pi - 1$ **do**

Compute $\hat{\mathbf{a}}(t_{k,i}^\pi, y_k; \mathbf{x}_i)$ e.g. by kernel estimator (4.2);

Sample $\varepsilon_k \in \mathcal{N}(0, 1)$ and compute

$$y_{k+1} = y_k + \frac{1}{N^\pi} \hat{\mathbf{a}}(t_{k,i}^\pi, y_k; \mathbf{x}_i) + \frac{1}{\sqrt{N^\pi}} \varepsilon_k,$$

end

Set $x_{i+1} = y_{N^\pi}$.

end

Return: x_1, \dots, x_N

5 Numerical experiments

In this section, we demonstrate the effectiveness of our SBTS algorithm on several examples of time series models, as well as on real data sets for an application to deep hedging. The algorithms are performed on a computer with the following characteristics: Intel(R) i7-7500U CPU @ 2.7GHz, 2 Core(s).

5.1 Evaluation metrics

In addition to visual plot of data vs generator samples path, we use some metrics to evaluate the accuracy of our generators:

- *Marginal metrics* for quantifying how well are the marginal distributions from the generated samples compared to the data ones. These include
 - Classical statistics like mean, 95% and 5% percentiles
 - Kolmogorov-Smirnov test: we compute the p -value, and when $p > \alpha$ (usually 5%), we do not reject the null-hypothesis (generator came from data of reference distribution)
- *Temporal dynamics* metrics for quantifying the ability of the generator to capture the time structure of the time series data. We compute the empirical distribution of the quadratic variation: $\sum_i |X_{t_{i+1}} - X_{t_i}|^2$.
- *Correlation structure* for evaluating the ability of the generator to capture the multi-dimensional structure of the time series. We shall compare the empirical covariance or correlation matrix induced by the generator SBTS and the ones from the data samples.

5.2 Toy autoregressive model of time series

We consider the following toy autoregressive (AR) model:

$$\begin{cases} X_{t_1} &= b + \varepsilon_1, \\ X_{t_2} &= \beta_1 X_{t_1} + \varepsilon_2, \\ X_{t_3} &= \beta_2 X_{t_2} + \sqrt{|X_{t_1}|} + \varepsilon_3, \end{cases}$$

where the noises $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$, $i = 1, \dots, 3$ are mutually independent. The model parameters are $b = 0.7$, $\sigma_1 = 0.1$, $\sigma_2 = \sigma_3 = 0.05$ and $\beta_1 = \beta_2 = -1$.

We use samples of size $M = 1000$ for simulated data of the AR model. The drift of the SBTS diffusion is estimated with a kernel of bandwidth $h = 0.05$, and simulated from euler scheme with $N^\pi = 100$. The runtime for generating 500 paths of SBTS is 8 seconds.

In Figure 2, we plot the empirical distribution of each pair (X_{t_i}, X_{t_j}) from the AR model, and from the generated SBTS. We also show the marginal empirical distributions. Table 1 presents the marginal metrics for the AR model and generator (p -value and percentiles at level 5% and 95%). In Table 2, we give the difference between the empirical correlation from generated samples and AR data samples.

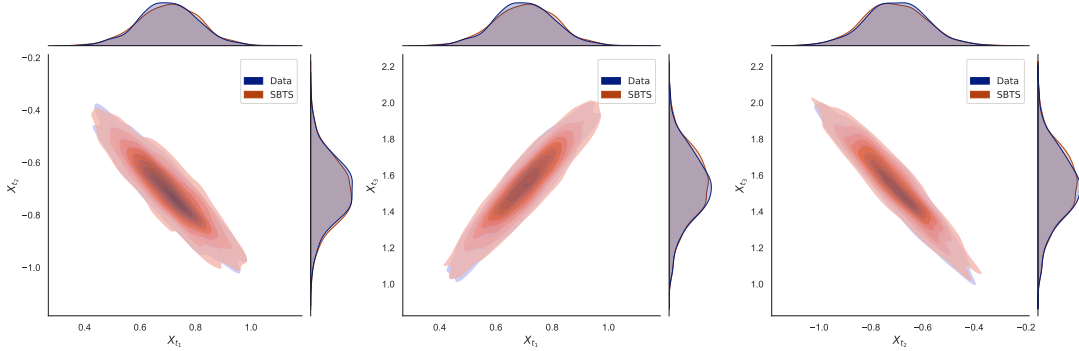


Figure 2: Comparison between the true and generated distribution for each couple (X_{t_i}, X_{t_j}) with $i, j \in \llbracket 1, 3 \rrbracket$ with $i \neq j$

	p-value	q_5	\tilde{q}_5	q_{95}	\tilde{q}_{95}
X_{t_1}	0.98	0.535	0.528	0.855	0.861
X_{t_2}	0.74	-0.873	-0.861	-0.516	-0.514
X_{t_3}	0.90	1.243	1.251	1.808	1.793

Table 1: Marginal metrics for AR model and generator (\tilde{q} for percentile)

	X_{t_1}	X_{t_2}	X_{t_3}
X_{t_1}	0	0.014	-0.01
X_{t_2}	0.014	0	0.013
X_{t_3}	-0.01	0.013	0

Table 2: Difference between empirical correlation from generated samples and reference samples

5.3 GARCH Model

We consider a GARCH model:

$$\begin{cases} X_{t_{i+1}} &= \sigma_{t_{i+1}} \varepsilon_{t_{i+1}} \\ \sigma_{t_{i+1}}^2 &= \alpha_0 + \alpha_1 X_{t_i}^2 + \alpha_2 X_{t_{i-1}}^2, \quad i = 1, \dots, N, \end{cases}$$

with $\alpha_0 = 5$, $\alpha_1 = 0.4$, $\alpha_2 = 0.1$, and the noises $\varepsilon_{t_i} \sim \mathcal{N}(0, 0.1)$, $i = 1, \dots, N$, are i.i.d. The size of the time series is $N = 60$.

The hyperparameters for the training and generation of SBTS are $M = 1000$, $N^\pi = 100$, and a bandwidth for the kernel estimation $h = 0.2$ (larger than for the AR model since by nature the GARCH process is more "volatile"). The runtime for generating 1000 paths is 120 seconds.

In Figure 3, we plot four sample paths of the GARCH process to be compared with four sample paths of the SBTS. Figure 4 represents samples plot of the joint distribution between $(X_{t_1}$ and the terminal value $X_{t_N})$ of the time series. Figure 5 provides some metrics to evaluate the performance of SBTS. On the left, we represent the p -value for

each of the marginals of the generated SBTS. On the right, we compute for each marginal index $i = 1, \dots, N = 60$, the difference between ρ_i and $\hat{\rho}_i$ where ρ_i (resp. $\hat{\rho}_i$) is the sum over j of the empirical correlation between X_i and X_{t_j} from GARCH (resp. generated SBTS), and plot its mean and standard deviation.

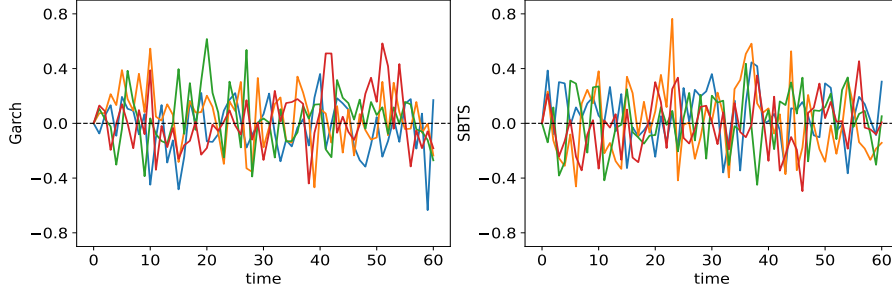


Figure 3: Samples path of reference GARCH (left) and generator SBTS (right)

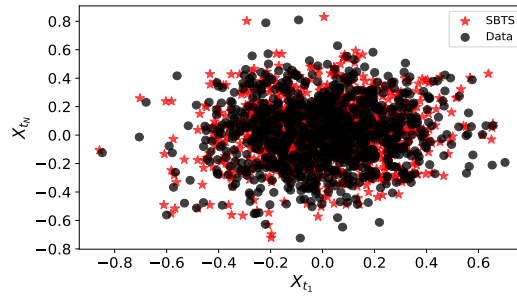


Figure 4: Samples plot of the joint distribution (X_{t_1}, X_{t_N})

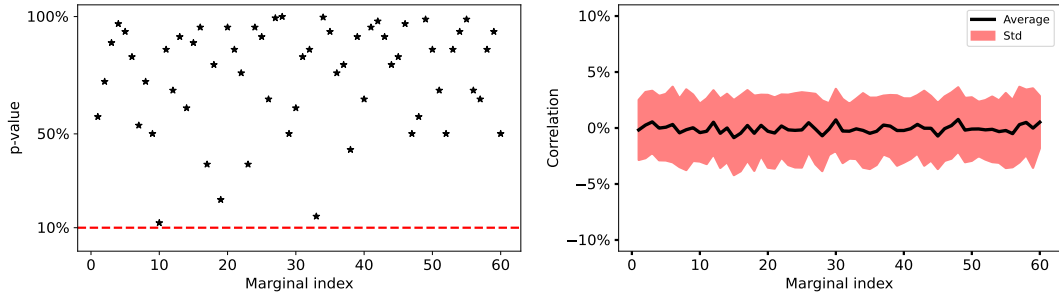


Figure 5: *Left*: p -value for the marginals X_{t_i} . *Right*: Difference between the term-by-term empirical correlation from generated samples and reference samples.

5.4 Fractional Brownian Motion

We consider a fractional Brownian motion (FBM) with Hurst index H that measures the roughness of this Gaussian process. We plot in Figure 6 four samples path of FBM with $H = 0.2$, and samples paths generated by SBTS. The generator is trained with $M =$

1000 sample paths, and the hyperparameters used for the simulation are $N^\pi = 100$, with bandwidth $h = 0.05$ for the kernel estimation of the Schrödinger drift. The runtime for 1000 paths is 120 seconds.

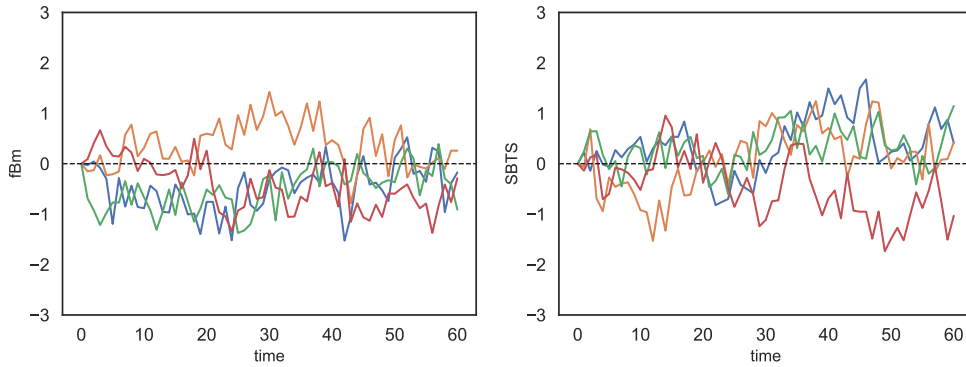


Figure 6: Samples path of reference FBM (left) and generator SBTS (right)

Figure 7 represents the covariance matrix of $(X_{t_1}, \dots, X_{t_N})$ for $N = 60$ of the FBM and of the generated SBTS, while we plot in Figure 8 the empirical distribution of the quadratic variation $\sum_{i=0}^{N-1} |X_{t_{i+1}} - X_{t_i}|^2$ for the FBM and the SBTS.

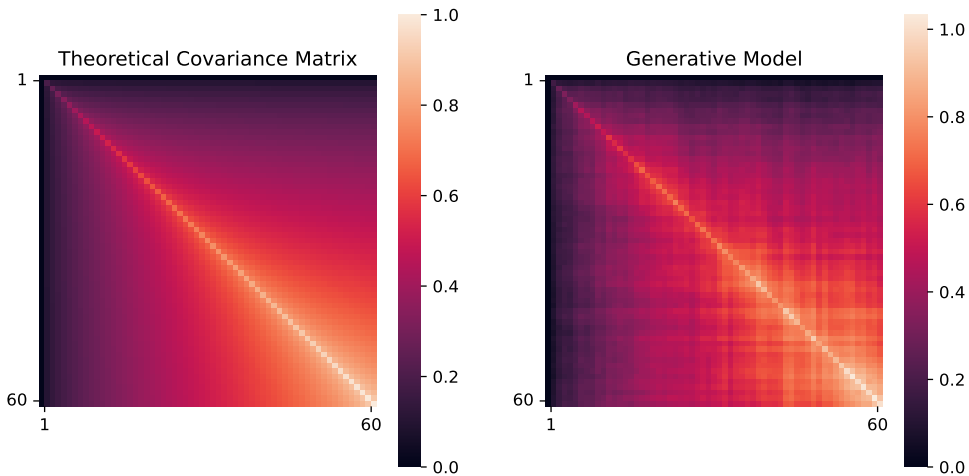


Figure 7: Covariance matrix for reference FBM and SBTS

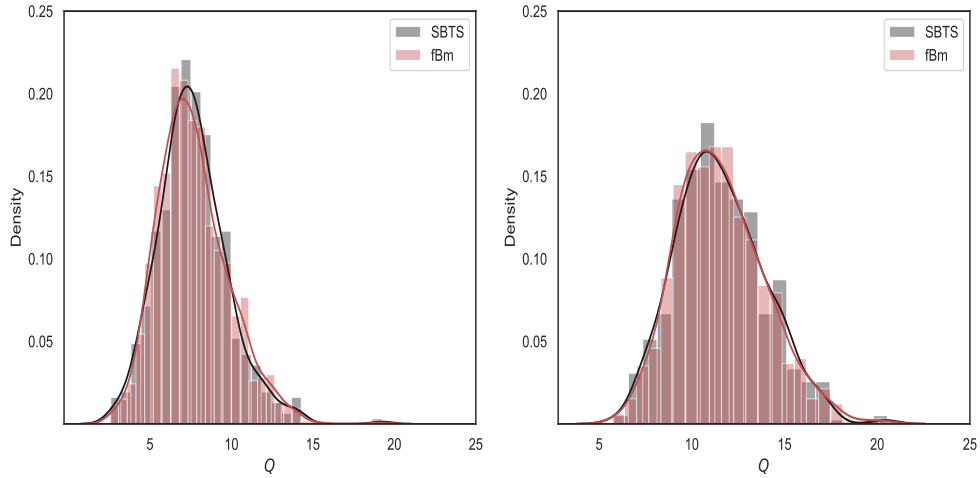


Figure 8: Quadratic variation distribution for $N = 30$ (left), $N = 60$ (right) and $T = t_N = 1$

Finally, we provide estimate of the Hurst index from our generated SBTS with the standard estimator (see e.g. [10]) given by:

$$\hat{H} = \frac{1}{2} \left[1 - \frac{\log \left(\sum_{i=0}^{N-1} |X_{t_{i+1}} - X_{t_i}|^2 \right)}{\log N} \right].$$

For $N = 60$, we get: $\hat{H} = 0.2016$, Std = 0.004.

5.5 Application to deep hedging on real-data sets

In this paragraph, we use generated time series for applications to risk management, and notably the pricing of derivatives and the computation of associated hedging strategies via deep hedging approach. We use samples of historical data for generating by SBTS new synthetic time series samples. We then compute deep hedging strategies that are trained from these synthetic samples, and we compare with the PnL and the replication error based on historical dataset. The general backtest procedure is illustrated in Figure 9.

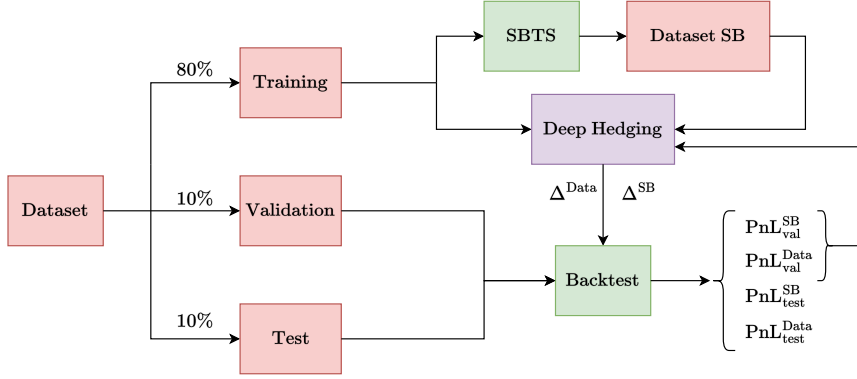


Figure 9: Procedure of backtest for deep hedging

We consider stock price S from the company *Apple* with data (ticker is AAPL) from January 1, 2010 to January 30, 2020, and produce $M = 2500$ samples of $N = 60$ successive days, with a sliding window. The hyperparameters for the generation of SBTS synthetic samples are $N^\pi = 100$, bandwidth $h = 0.05$.

We plot in Figure 10 four sample paths of the SBTS diffusion to be compared with the real ones from *Apple*. We illustrate the excess of kurtosis of the real data by plotting in Figure 11 the tail distribution for the return $R_{t_i} = \frac{S_{t_{i+1}}}{S_{t_i}} - 1$: x in log-scale $\mapsto \mathbb{P}[|R| \geq x]$, and found that the excess of kurtosis of real data is 1.96, to be compared with the one from SBTS, and equal to 2.34. Figure 12 represents the empirical distribution of the *Apple* time series data vs SBTS, while Figure 13 shows their covariance matrices.

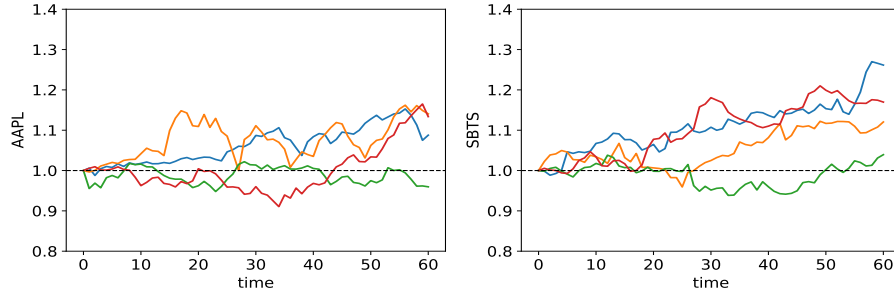


Figure 10: Four paths generated by Schrodinger bridge (Right) vs real ones (Left)

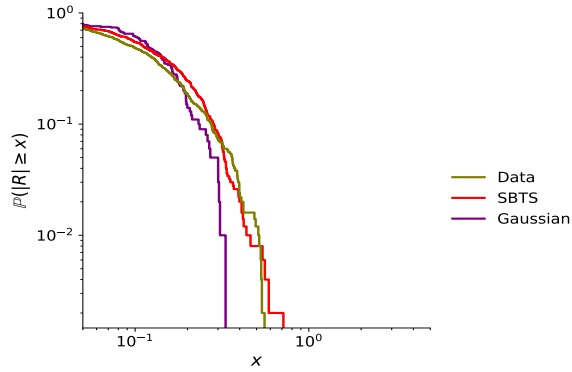


Figure 11: Plot of tail distribution for the return: x in log-scale $\mapsto \mathbb{P}[|R| \geq x]$. Excess of kurtosis for real-data = 1.96, for generated SBTS = 2.34

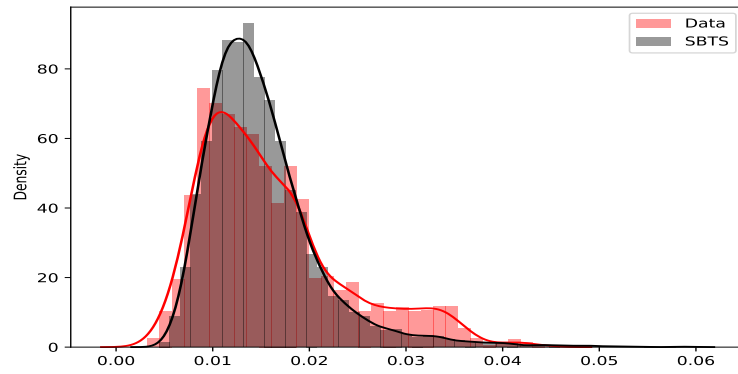


Figure 12: Comparison of quadratic variation distribution

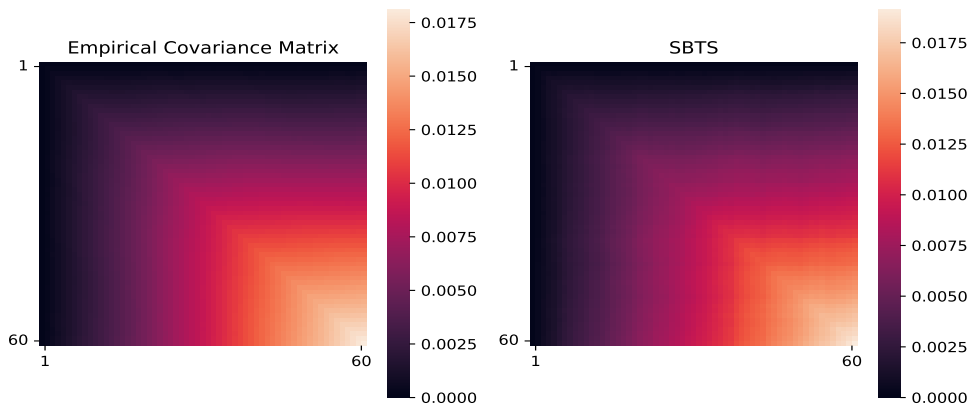


Figure 13: Covariance matrix for real-data and generative SBTS

The synthetic time series generated by SBTS is now used for the deep hedging of ATM call option $g(S_T) = (S_T - S_0)_+$, i.e., by minimizing over the initial capital p (premium) and

the parameters of the neural network Δ the (empirical) loss function, called replication error:

$$\mathbb{E}|\text{PnL}^{\text{P},\Delta}|^2, \quad \text{with} \quad \text{PnL}^{\text{P},\Delta} = p + \sum_{i=0}^{N-1} \Delta(t_i, S_{t_i})(S_{t_{i+1}} - S_{t_i}) - g(S_T).$$

We then compare with the deep hedging on historical data by looking at the PnL and replication errors. The historical data set of Apple is split in the chronological order, namely training data set from 01/01/2007 to 31/12/2017, validation data set from 01/01/2018 to 31/12/2018, and test set from 01/01/2019 to 30/01/2020. As pointed out in [24], it is important not to break the time structure as it may lead to an overestimation of the model performance.

In Figure 14, we plot the empirical distribution of the PnL with deep hedging obtained from real data vs SBTS, and backtested on the validation and test sets. It appears that the PnL from SBTS has a smaller variance (hence smaller replication error), and yields less extreme values, i.e. outside the zero value, than the PnL from real data. This is also quantified in Table 3 where we note that the premium obtained from SBTS is higher than the one from real data, which means that one is more conservative with SBTS by charging a higher premium.

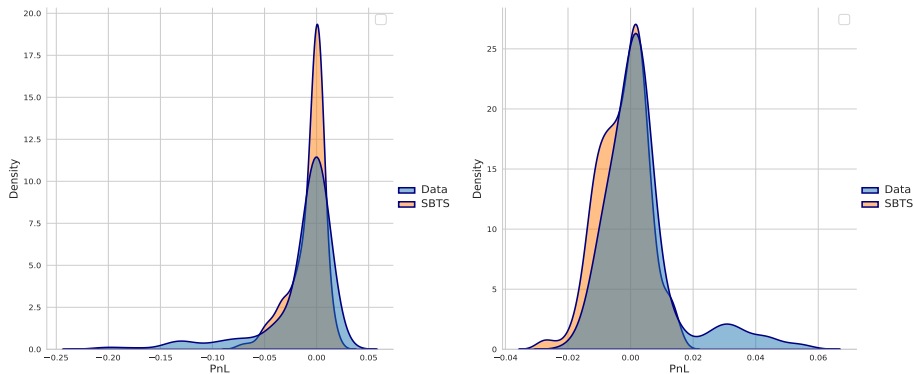


Figure 14: Deep hedging PnL distribution with backtest from validation set (*left*) and test set (*right*).

		Training Set		Validation Set		Test Set	
	Premium	Mean	Std	Mean	Std	Mean	Std
Data	0.0415	0.0008	0.0098	-0.0154	0.0371	0.003	0.012
SBTS	0.0471	0.0004	0.0109	-0.0075	0.0164	-0.0024	0.0076

Table 3: Mean of PnL and its Std (replication error).

6 Further tests in high dimension

In this section, we illustrate how our SB approach can be used for generating samples in very high dimension.

We use a data set of images from MNIST with training size $M = 10000$. We first start with handwritten digital numbers, and plot in Figure 15 the static images from real MNIST data set and the ones generated by SBTS. The number of pixels is 28×28 , and the runtime for generating 16 iamges is equal to 120 seconds.

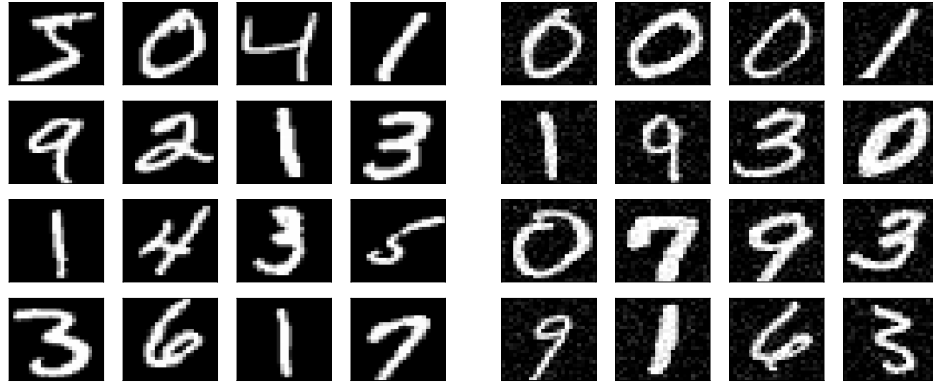


Figure 15: *Left:* MNIST samples. *Right:* Generated SBTS samples.

Next, in Figures 16 and 17, we plot sequential images sampled from real data set, and compare with the sequence generated by SBTS algorithm. The number of pixels is 14×14 , and the runtime for generating 100 paths of sequential images is equal to 9 minutes.

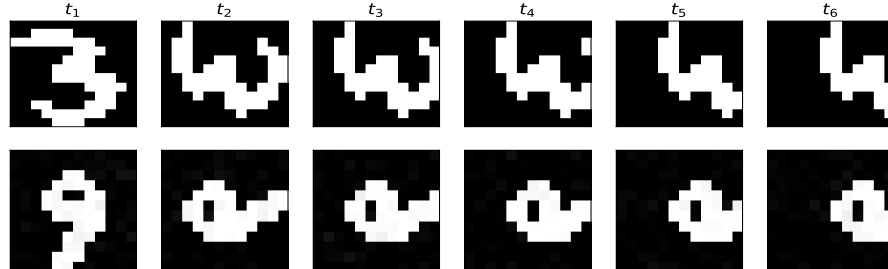


Figure 16: *Top:* a time series sampled from real distribution. *Bottom:* generated time series via SBTS

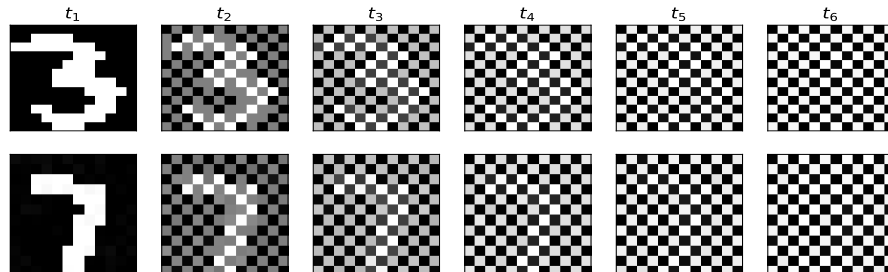


Figure 17: *Top:* a time series sampled from real distribution. *Bottom:* generated time series via SBTS

References

- [1] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *ICML*, 2017.
- [2] H. Buehler, L. Gonon, J. Teichmann, and B. Wood. Deep hedging. *Quantitative Finance*, 19(8):1271–1291, 2019.
- [3] H. Buehler, B. Horvath, T. Lyons, I. Perez-Arribas, and B. Wood. A data-driven market simulator for small data environments. *SSRN 3632431*, 2020.
- [4] T. Chen, G-H. Liu, M. Tao, and E. Theodorou. Deep momentum multi-marginal Schrödinger bridge. *arXiv:2303.0175*, 2023.
- [5] Y. Chen, Y. Wang, D. Kirschen, and B. Zhang. Model-free renewable scenario generation using generative adversarial networks. *IEEE Transactions on power systems*, 33(3):3265–3275, 2018.
- [6] V. Choudhary, S. Jaimungal, and M. Bergeron. FuNVol: A Multi-Asset Implied Volatility Market Simulator using Functional Principal Components and Neural SDEs. *arXiv:2303.00859*, 2023.
- [7] P. Dai Pra. A stochastic control approach to reciprocal diffusion process. *Applied Mathematics and Optimization*, 23(1):313–329, 1991.
- [8] V. De Bortoli, J. Thornton, J. Heng, and A. Doucet. Diffusion Schrödinger bridge with applications to score-based generative modeling. *ArXiv: 2106.01357*, 2021.
- [9] A. Fermanian. Embedding and learning with signatures. *arXiv:1911.13211*, 2019.
- [10] J. Gairing, P. Imkeller, R. Shevchenko, and C. Tudor. Hurst index estimation in stochastic differential equations driven by fractional brownian motion. *Journal of Theoretical Probability*, 33:1691–1714, 2020.
- [11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Faley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.
- [12] F. Guth, S. Coste, V. De Bortoli, and S. Mallat. Wavelet score-based generative modeling. *arXiv: 2208.05003*, 2022.
- [13] P. Henry-Labordere. From (martingale) Schrodinger bridges to a new class of stochastic volatility model. *SSRN.3353270*, 2019.
- [14] D. Kingma and M. Welling. Auto-encoding variational Bayes. In *ICLR*, 2014.
- [15] Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, and F. Huang. A tutorial on energy-based modeling. *Predicting structured data*, 1, 2006.
- [16] C. Leonard. A survey of the schrodinger problem and some of its connections with optimal transport. *Dynamical Systems*, 34(4):1533–1574, 2014.

- [17] X. Lyu, S. Hueser, S.L. Hyland, G. Zerveas, and G. Raetsch. Improving clinical predictions through unsupervised time series representation learning. *arXiv:1812.00490*, 2018.
- [18] H. Ni, L. Szpruch, M. Wiese, S. Liao, and B. Xiao. Conditional Sig-Wasserstein GANs for time series generation. *arXiv:2006.05421*, 2020.
- [19] V. Ram Somnath, M. Pariset, Y-P. Hsieh, M.R. Martinez, A. Krause, and C. Bunne. Aligned diffusion Schrödinger bridges. *Arxiv:2302.11419*, 2023.
- [20] C. Remlinger, J. Mikael, and R. Elie. Conditional versus adversarial Euler-based generators for time series. *arXiv:2102.05313*, 2021.
- [21] Y. Song and S. Ermon. Generative modeling by estimating gradients of the distribution. In *NIPS*, pages 11918–11930, 2019.
- [22] Y. Song, J. Sohl-Dickstein, D. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on learning Representation*, 2021.
- [23] G. Wang, Y. Jiao, Q. Xu, Y. Wang, and C. Yang. Deep generative learning via schrödinger bridge. *arXiv:2106.10410*.
- [24] W. Wang and J. Ruf. A note on spurious model selection. *Quantitative Finance*, 22(10):1797–2000, 2022.
- [25] M. Wiese, R. Knobloch, R. Korn, and Kretschmer. Quant gans: deep generation of financial time series. *Quantitative Finance*, 20(9):1419–1440, 2020.
- [26] T. Xu, W. Li, M. Munn, and B. Acciaio. COT-GAN: Generative Sequential Data via Causal Optimal Transport. In *NeurIPS*, 2020.
- [27] J. Yoon, D. Jarrett, and Van der Schaar. Time-series generative adversarial networks. In *NeurIPS*, 2019.