Bayesian Inference for partially observed SDEs Driven by Fractional Brownian Motion

Joseph Dureau
Department of Statistics, London School of Economics
dureau.joseph@gmail.com

Alexandros Beskos
Department of Statistics and Applied Probability,
National University of Singapore
staba@nus.edu.sg

Konstantinos Kalogeropoulos
Department of Statistics, London School of Economics
k.kalogeropoulos@lse.ac.uk

May 7, 2014

Abstract

In this paper we consider continuous-time diffusion models driven by fractional Brownian Motion (fBM), with observations obtained at discrete-time instances. As a prototypical scenario we will give emphasis on a stochastic volatility (SV) model allowing for memory in the volatility increments through an fBM specification. Due to the non-Markovianity of the model and the high-dimensionality of the latent volatility path, estimating posterior expectations is a computationally challenging task. We present novel simulation and re-parameterisation framework based on the Davies and Harte method and use it to construct a Markov chain Monte-Carlo (MCMC) algorithm that allows for computationally efficient parametric Bayesian inference upon application on such models. The algorithm is based on an advanced version of the so-called Hybrid Monte-Carlo (HMC) that allows for increased efficiency when applied on high-dimensional latent variables relevant to the models of interest in this paper. The inferential methodology is examined and illustrated in the SV models, on simulated data as well as real data from the S&P500/VIX time series. Contrary to a long range dependence attribute of the SV process (Hurst parameter $H > 1/2$) many times assumed in the literature, the posterior distribution favours $H < 1/2$ that points towards medium range dependence.
1 Introduction

The presence of memory in stochastic processes is becoming increasingly important in various scientific fields including the analysis of financial and econometric time series. A natural continuous-time modeling framework for processes with memory is offered by using fractional Brownian motion (fBm) as the driving noise in the model specification. This is a zero mean scalar Gaussian process, denoted as $B^H = \{B^H_t; t \geq 0\}$, that is completely specified by its covariance function:

$$
E[ B^H_s B^H_t ] = \frac{1}{2} \left( |t|^{2H} + |s|^{2H} - |t-s|^{2H} \right), \quad 0 \leq s \leq t,
$$

and is parametrised by the Hurst index $H \in (0, 1)$ governing the memory properties of its increments. For $H = 1/2$ we obtain the standard Brownian motion (BM) with independent increments, whereas for $H > 1/2$ we obtain smoother paths, yet of infinite variation, with positively autocorrelated increments that exhibit long range dependence in the sense that their autocorrelations are not absolutely summable. For $H < 1/2$ we obtain rough paths with negatively autocorrelated increments that exhibit medium range dependence. In other words, the autocorrelation of the increments may be summable for the case of $H < 1/2$ but it decays at a much slower rate than exponential which corresponds to a typical short range dependence process.

In this paper we target models where the underlying stochastic process is driven by an fBM and is observed in partial manner. One such example is the volatility of financial time series which is usually observed through the corresponding asset prices. For illustration purposes, we work with this example closely, noting however that the developed framework is applicable for more general classes of partially observed diffusion processes driven by fBM.

The adoption of models with memory for the volatility originates from empirical evidence; see for example [13, 24]. The autocorrelation function of squared returns is often observed to be slowly decaying towards zero, clearly not in an exponential manner that would suggest short range dependence models, nor implying a unit root that would point to integrated processes. In discrete-time an alternative is the long memory stochastic volatility (SV) model introduced by [19, 20], where the log-volatility is a fractional ARIMA process. In continuous-time, the authors of [11] (see [10] for an alternative specification) introduced a fractional SV model that falls within the family of models we examine closely in this paper:

\begin{align*}
\text{(1)} & \quad dS_t = \mu S_t dt + \sigma_S(X_t) S_t dW^H_t ; \\
\text{(2)} & \quad dX_t = b_X(X_t, \zeta) dt + \sigma_X(X_t, \zeta) dB^H_t , \quad t \in [0, \ell] .
\end{align*}

In the model above $S_t$ and $X_t$ represent the observed asset price and underlying volatility processes respectively and $W^H_t$ is a noise process specified below. The
definition involves also a length \( \ell > 0 \) for the time-period under consideration and the functions \( \sigma_S : \mathbb{R} \mapsto \mathbb{R}, b_X : \mathbb{R} \times \mathbb{R}^p \mapsto \mathbb{R} \) and \( \sigma : \mathbb{R} \times \mathbb{R}^p \mapsto \mathbb{R} \) that are assumed to be known up to the parameters \( \mu \in \mathbb{R} \) and \( \zeta \in \mathbb{R}^p, p \geq 1 \). For example in [11] the log-volatility is set as a fractional OU process.

The increments of the process \( X_t \) inherit the properties of fBM and are also characterised by the parameter \( H \). In contrast with previous literature that restricts attention to models with long range dependence, we are going to consider the extended model that also allows for \( H < 1/2 \). In fact, one of the interesting results in the application of this paper is the evidence pointing towards medium range dependence in the volatility of the S&P500 index. We will also explore other empirical features related to this model, as suggested by the data, targeting estimation of the real world measure. Along these lines we are going to allow for correlation between the increments of \( B^H_t \) and \( W^H_t \) by defining:

\[
dW^H_t = \sqrt{1 - \rho^2} \, dW_t + \rho \, dB^H_t ,
\]

with \( W_t \) being a standard Brownian motion and a parameter \( \rho \in (-1, 1) \). We note however that such a formulation would require extra care if it was to be applied for pricing purposes, to preclude arbitrage opportunities typically associated with fBm; see for example [6]. A simple way to ensure that, is by considering cases where the volatility is an autonomous process, e.g. by setting \( \rho = 0 \), although this may not be supported by empirical evidence; our analysis on the S&P500 data points towards negative values in a similar manner as with diffusion driven SV models. Nevertheless, the presence of arbitrage for the general case \( \rho \neq 0 \) has yet to be fully investigated (to the best of our knowledge).

Despite the fact that the model is formulated in continuous time, the price-process \( S_t \) can only be observed at a finite collection of time instances, \( 0 < t_1 < t_2 < \cdots < t_n \), for some \( n \geq 1 \). Denote the log prices by:

\[
Y_k = \log S_{t_k} , \quad 1 \leq k \leq n , \quad Y = \{Y_1, Y_2, \ldots, Y_n\} ,
\]

from which we aim in drawing inference for the vector of all parameters involved in our general model specification in (1)-(3), \( \theta = (\mu, \rho, \zeta, H) \subset \mathbb{R}^q \), such that \( q = p + 3 \). While there exist various methods for parameter estimation in cases with direct and ideally high frequency observations on the fBM driven SDE (see for example [26]), the available options are limited in the partial observation setting considered here. In [11, 10] the estimation procedure extracts information on the spot volatility based on the quadratic variation of the price process which is then used to estimate \( \theta \). [28] utilises an Euler scheme-type transformation to link the squared increments of observed price process with the volatility, and constructs a wavelet based estimator of \( H \). A common feature of these approaches, as well as with other related ones ([17]), is that the estimation performance depends on the
observation frequency. In this paper we develop a general inferential framework based on data augmentation that can be applied to low frequency data as well.

A first challenge for the statistical inference procedure corresponding to the model in (1)-(3) lies with the unavailability of a closed-form expression for the likelihood function:

$$p(Y | \theta) = \int p(Y | X, \theta) p(dX | \theta).$$

We therefore consider a time discretised version of the infinite-dimensional path $X$, constructed on a regular time grid, denoting by $N \geq 1$ its finite dimension. Note that conditionally on a partition of the latent SV process $X$, the density of $Y$ w.r.t. Lebesque measure, $p(Y | X, \theta)$ under the fractional SV model in (1)-(3), may be well approximated given a large enough $N$ and, e.g., an Euler-type scheme.

The methodology in [8, 9] partly operates in this setting for a fixed $\theta$ but also relies on calibration against observed option prices. A grid of $H$ values are considered for each of which: (i) the remaining parameters are estimated by historical data, (ii) model option prices are computed via a particle filter on the discretised version of $X$ and a multinomial recombining tree algorithm, and (iii) calibration is used to obtain $H$.

We take a different approach aiming in full Bayesian inference via a Markov Chain Monte Carlo (MCMC) algorithm on the $X, \theta$ space to obtain samples from the joint posterior distribution

$$\Pi(X, \theta | Y) \propto p(Y | X, \theta) p(X | \theta) p(\theta),$$

given a prior $p(\theta)$. If samples from the joint posterior are available, denoted say as $\{X(i), \theta(i)\}_{i \geq 1}$, then variables $\{\theta(i)\}_{i \geq 1}$ are sampled from the marginal posterior $p(\theta | Y)$. Note also that it is essential to construct an MCMC algorithm whose performance does not deteriorate as $N$ gets large, thus giving the opportunity to the user to choose $N$ large enough so that the corresponding approximate posterior is close to the theoretical one, $p(\theta | Y)$. For the standard SV model driven by BM (i.e. $H = 1/2$), such MCMC algorithms are now available (see e.g. [27, 18, 22] and the references therein).

However, major algorithmic design challenges arise when $H \neq 1/2$. First, the costs for naive evaluations of the likelihood are now of $O(N^3)$, as opposed to $O(N)$ for the $H = 1/2$ case, thus prohibiting large values of $N$ that are vital for accurate likelihood evaluations. Second, $H$ and $\zeta$ can be fully identified by a continuous path of the process $X$; see e.g. [26]. That is, the joint distribution of $\{X, H\}$ is degenerate with $p(H | X)$ being a Dirac measure. Thus, to avoid slow mixing issues the algorithmic design will have to decouple this dependence between $X, H$. In the $H = 1/2$ this issue can be tackled by suitable reparametrisation ([27, 18, 22]) or even a particle MCMC algorithm ([2]), but none of these options are available for the case of $H \neq 1/2$. Third, the MCMC algorithms for diffusion models make
intrinsic use of the Markovian structure of $X$, exploiting the fact that conditionally on $Y, \theta$, the SV-path can be split into small ‘blocks’ of time periods with updates on each block involving only computations over it’s associated (small) time period. A critical difference for $H = 1/2$ is that the latent process $X$ is not Markovian any more. Thus, a similar block construction as the above will now be much more expensive: updating a small-size block will now require calculations over the latent path $X$ for the complete time period $[0, \ell]$ under consideration. Therefore, a potentially efficient algorithm should aim at updating as large blocks as possible during its execution. In this paper we address all these issues and offer a general and efficient MCMC algorithm to facilitate Bayesian inference on such models. The first and the second issue are tackled via a novel reparameterisation utilising the Fast Fourier transform (FFT) via a Davies and Harte construction. For the third issue we resort to a Hybrid Monte Carlo (HMC) algorithm ([16, 3, 4]) and in particular develop an advanced version of it that it particularly appropriate when $N$ is big.

The algorithm presented in this paper will have the following characteristics:

a) It’s computational cost per MCMC step will be $O(N \log N)$ - so not much greater than the $O(N)$-costs for the simpler Markovian case $H = 1/2$.

b) It’s mixing time will be $O(1)$ w.r.t. to the number of imputed points $N$; that is, reducing the discretisation error will not worsen the convergence properties of the MCMC algorithm; this will be due to the fact that the algorithm will be well-defined in theory even when considering the complete infinite-dimensional latent path $X = \{X_t; t \in [0, \ell]\}$.

c) It will decouple the full dependence between $X$ and $H$.

d) It will be based on an advanced version of Hybrid Monte-Carlo (HMC), thus will employ Hamiltonian dynamics to allow for making big steps in the state space, while treating big blocks of $X$; in fact in the example applications the whole of the $X$-path on $[0, \ell]$ as well as the parameter vector $\theta$ will be updated simultaneously.

The paper is organised as follows. Section 2 presents the adopted reparametrisation, based on the Davies and Harte method for the construction of an fBM, that allows decoupling of $X$ and $H$ while offering simulation of the paths and evaluation of the likelihood with $O(N \log N)$ computations. The target posterior is re-written accordingly for the MCMC algorithm. Section 3 presents the MCMC method, driven by Hamiltonian dynamics, that will allow for mesh-free (i.e. not deteriorating with increasing $N$) rapid mixing of the MCMC method. Section 4 shows the details for the application of the algorithm on the SV model and presents numerical experiments as well as an illustration of the methodology.
to data S&P500/VIX indices. Finally, Section 5 concludes and discusses further directions.

# 2 Davies and Harte Sampling and Reparameterisation

As already mentioned, in practice our MCMC algorithm will consider the driving noise $B^H$ on a user-selected regular grid of $N$ points of mesh-size $\delta = \ell/N$.

## 2.1 fBM Sampling

We will show below the Davies and Harte method for constructing $\{B^H_t; t \in [0, N]\}$ on the grid $\{1, 2 \ldots N\}$. Due to the self-similarity of fBM, we can then obtain fBM on $\{\delta, 2\delta, \ldots, N\delta\}$ by multiplying the originally sampled $N$-dimensional vector with $\delta^H$. We define the following increments on the grid times $0, 1, \ldots, N$:

$$G(k) = B^H_k - B^H_{k-1}, \quad 1 \leq k \leq N,$$

and the $N$-dimensional fractional Gaussian noise (abbreviated to fGn) vector $G$:

$$G_N = \{G(k) : 1 \leq k \leq N\}.$$  

Crucially, for the Davies and Harte methodology fGn is stationary, since for indices $k \geq 0, k_0 > 0$:

$$g(k) := E[G(k_0 + k) G(k_0)]$$

$$= \frac{1}{2} \delta^{2H} (k+1)^{2H} + \frac{1}{2} \delta^{2H} (k-1)^{2H} - \delta^{2H} |k|^{2H}.$$  

For details on the Davies and Harte method for sampling $G_N$, see e.g. [30]. The method describes the simulation $G_N$ via the use of $2N$ iid samples from standard normal distribution $N(0, 1)$. In brief, the algorithm builds up as follows.

We consider the Toeplitz covariance matrix of $G_N$:

$$\Gamma = \begin{pmatrix}
  g(0) & g(1) & \cdots & g(N-2) & g(N-1) \\
  g(1) & g(0) & \cdots & g(N-1) & g(N-2) \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  g(N-1) & g(N-2) & \cdots & g(1) & g(0)
\end{pmatrix}$$

and the matrix:

$$\Gamma^f = \begin{pmatrix}
  0 & g(N-1) & \cdots & g(2) & g(1) \\
  g(N-1) & 0 & \cdots & g(3) & g(2) \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  g(1) & g(2) & \cdots & g(N-1) & 0
\end{pmatrix}.$$
The covariance matrix $\Gamma$ is embedded within the following circular matrix:

$$C = \begin{pmatrix} \Gamma & \Gamma^f \\ \Gamma^f & \Gamma \end{pmatrix},$$

that is, bringing the last element of a row of $C$ at the beginning of that row will give the immediately next row (below, we count the rows and columns of $C$ and other involved matrices from 0 to $2N - 1$ as this simplifies the expressions). Due to the circular property, $C$ allows for a relatively simple and fast eigen-expansion using FFT. In particular, we can write the decomposition:

$$C = PA_H P^*$$

where $i^2 = -1$, $P$ is the unitary $(2N) \times (2N)$-matrix with constituent elements:

$$P_{jk} = \frac{1}{\sqrt{2N}} \exp(-2\pi i \frac{jk}{2N}), \quad 0 \leq j, k \leq 2N - 1,$$

and $P^* = (\overline{p}_{kl})$ is it’s complex transpose. The diagonal matrix:

$$\Lambda_H = \text{diag}\{\lambda_0, \lambda_1, \ldots, \lambda_{2N-1}\}$$

is constructed via the eigenvalues:

$$\lambda_k = \sum_{j=0}^{2N-1} c_{0,j} \exp(-2\pi i \frac{jk}{2N}), \quad 0 \leq k \leq 2N - 1,$$

where $c_{0,j}$ denotes the corresponding elements from the 0-row of matrix $C$. Using FFT, the matrix $\Lambda_H$ can be obtained from the $c_{0,j}$’s with $O(N \log N)$ calculations. One can now easily obtain the square-root matrix:

$$C^{1/2} = PA_H^{1/2} P^*.$$

So, the sampling algorithm involves simulating $Z_0 \sim N(0, I_{2N})$ and retrieving the first $N$ elements of $\sqrt{C}Z_0 = P\sqrt{\Lambda_H} P^* Z_0$. where the two calculations that involve multiplications with $P$ or $P^*$ can also be carried out with $O(N \log N)$ costs using FFT. As noted in [30], a simple way to avoid the $O(N \log N)$ calculations of $P^*Z$ is to simulate the distribution of $W = P^*Z_0$ as follows:

- Simulate independently $W_0, W_N \sim N(0, 1)$.
- Simulate independently $V, V' \sim N(0, I_{N-1})$.
- Set $W_j = \frac{1}{\sqrt{2}}(V_j + iV'_j)$ and $W_{2N-j} = \frac{1}{\sqrt{2}}(V_j - iV'_j)$, for $1 \leq j \leq N - 1$. 

7
Thus, to perform the sampling task, one equivalently needs to return:

$$P \Lambda_{H}^{1/2} Q Z$$

for $Z \sim N(0, I_{2N})$ and the matrix:

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix},$$

for the $N \times N$ sub-matrices: $Q_{11} = \text{diag}\{1, 1/\sqrt{2}, 1/\sqrt{2}, \ldots, 1/\sqrt{2}\}$; $Q_{12} = \{q_{ij}\}$ with $q_{i,-1} = 1/\sqrt{2}$ for $1 \leq i \leq N - 1$, otherwise $q_{ij} = 0$; $Q_{12} = \{q_{ij}\}$ with $q_{i,-1} = 1/\sqrt{2}$ for $1 \leq i \leq N - 1$, otherwise $q_{ij} = 0$; finally, $Q_{22} = \text{diag}_{\text{inv}} = \{1, -i/\sqrt{2}, -i/\sqrt{2}, \ldots, -i/\sqrt{2}\}$, where diag_{inv} denotes a matrix with non-zero entries in the inverse diagonal.

The complete algorithm for generating $G_N$ is shown in Table 1.

---

**Simulation of $G_N$:**

(i) Sample $Z \sim N(0, I_{2N})$.

(ii) Calculate $Z' = P \sqrt{\Lambda_{H}} Q Z$.

(iii) Return the first $N$ elements of $Z'$.

---

Table 1: The algorithm for sampling the fGN vector $G_N$ defined via (5), (6). It’s computational cost is $O(N \log N)$ due to using FFT.

2.2 Reparameterisation

The algorithm in Table 1 gives rise to a linear mapping, say $F_H$, for generating $G_N$ from $2N$ iid standard Gaussian variables, $Z = (Z_i)_{i=1}^{2N}$, that is:

$$Z \mapsto F_H(Z) = G_N.$$  

Thus, the latent variable principle described in the Introduction of Section 1 will be implemented using the iid Gaussian vector $Z$ (which is a-priori independent from $H$) rather than the SV-path $X$. Indeed, we will work with the joint posterior distribution of $(Z, \theta)$, possessing a density w.r.t. $\otimes_{i=1}^{2N} N(0, 1) \times \text{Leb}_q$, i.e. the product of $2N$ standard Gaussian laws, times the $q$-dimensional Lebesgue measure. That is, we have that the posterior distribution $\Pi_N$ for $(Z, \theta)$ is specified as follows:

$$d\Pi_N \left( Z, \theta \mid Y \right) \propto p(\theta)p_N(Y \mid Z, \theta).$$

(9)
Subscript $N$ in the probability distribution or density expressions in (9) and in the sequel emphasizes the finite-dimensional approximations due to using an $N$-dimensional proxy for the theoretical infinite dimensional SV path $X$. We follow an Euler scheme to give rise to this approximation as shown in the analytical expressions below and in the Appendix.

We are interested in the expression of the likelihood $p_N(Y \mid Z, \theta)$ in the context of the SV-model of interest given in (1)-(3). It is useful to re-express the model in terms of $U_t = \log S_t$, so that after applying Itô’s formula:

$$
\begin{align*}
    dU_t &= (\mu - \sigma S(X_t)^2/2) dt + \sigma S(X_t) \left( \sqrt{1 - \rho^2} dW_t + \rho dB_H^t \right), \\
    dX_t &= b(X_t, \theta) dt + \sigma X_t dB_H^t.
\end{align*}
$$

Conditionally on $(Z, \theta)$, thus also on the driving $B_H$, the log-process $U$ does satisfy the Markov property. From the specification of the model, we have that:

$$
Y_k \mid Y_{k-1}, B_H, \theta \sim N(m_k(B_H, \theta), V_k(B_H, \theta)),
$$

for $1 \leq k \leq n$ and $Y_0 \equiv U_0$, with mean and variance parameters:

$$
\begin{align*}
    m_k(B_H, \theta) &= Y_{k-1} + \int_{t_{k-1}}^{t_k} (\mu - \sigma S(X_t)^2/2) dt + \rho \int_{t_{k-1}}^{t_k} \sigma S(X_t) dB_H^t; \\
    V_k(B_H, \theta) &= (1 - \rho^2) \int_{t_{k-1}}^{t_k} \sigma S(X_t)^2 dt.
\end{align*}
$$

From (10), it is trivial to write down the expression for $p(Y \mid B_H, \theta)$. In practice, all continuous-time paths will be replaced by $N$-dimensional vectors over the discrete-time instances $\{i\delta\}_{i=1}^N$. Thus, recalling the deterministic mappings:

$$
Z \mapsto G_N \mapsto (B_H^1, B_H^2, \ldots, B_H^N),
$$

arising in the Davies and Harte method, the expression for $p(Y \mid B_H, \theta)$ implied by (10), after taking products over $k = 1, 2, \ldots, n$, in continuous-time will provide an expression for $p_N(Y \mid Z, \theta)$ in discrete-time. The analytical calculation of $p_N(Y \mid Z, \theta)$ is provided in the Appendix. To avoid cumbersome expressions, we will henceforth assume the following expression for the target distribution:

$$
\Pi_N(Z, \theta) \propto e^{-\frac{1}{2} \langle Z, Z \rangle - \Phi(Z, \theta)}
$$

(11)

where, following (9), we have defined:

$$
\Phi(Z, \theta) = -\log p(\theta) - \log p_N(Y \mid Z, \theta).
$$

(12)

In the next Section we describe an efficient MCMC sampler tailored to the sampling task of the high-dimensional target posterior in (11).
3 An Efficient MCMC Sampler

We will be using an HMC sampler for sampling from the joint posterior distribution of $Z, \theta$ in (11). The standard version of HMC was introduced in [14]. Here, we will call upon an advanced version, tailored to the structure of the high-dimensional target distributions we are dealing with in our context. The method is closely related to the algorithms developed in [3, 4] for efficient sampling of distributions defined as a change of measure from Gaussian laws in infinite dimensions. The works in [3, 4] illustrate the mesh-free mixing property of the algorithms in the context of distributions of diffusion paths driven by BM. In the context of this paper, the algorithm will be extended to also take under consideration the involved parameters and the different set-up with a product of standard Gaussians as the high-dimensional Gaussian reference measure.

We begin with the description of the Hamiltonian dynamics upon which HMC is based. The state space is extended via an auxiliary ‘velocity’

$$v = (v_z, v_\theta) \in \mathbb{R}^{2N+q};$$

the original argument

$$x = (z, \theta) \in \mathbb{R}^{2N+q}$$

can be thought of as ‘location’. We consider the ‘total energy’ function:

$$H(x, v; M) = \Phi(x) + \frac{1}{2} \langle z, z \rangle + \frac{1}{2} \langle v, M v \rangle, \quad (13)$$

for a user-specified positive-definite ‘mass’ matrix $M$, consisted by the ‘potential’ $\Phi(x) + \frac{1}{2} \langle z, z \rangle$ and the ‘kinetic energy’ $\frac{1}{2} \langle v, M v \rangle$. We define the distribution on the $(x, v)$-space:

$$Q_N(x, v) \propto \exp\{-H(x, v; M)\} = \exp\{-\Phi(x) - \frac{1}{2} \langle z, z \rangle - \frac{1}{2} \langle v, M v \rangle\}. \quad (14)$$

Note that under $Q_N(x, v)$ we have $v \sim N(0, M^{-1})$. The Hamiltonian dynamics on $\mathbb{R}^{2N+q}$ express preservation of energy and are defined as follows:

$$\frac{dx}{dt} = M^{-1} \frac{\partial H}{\partial v}, \quad M \frac{dv}{dt} = -\frac{\partial H}{\partial x},$$

which in our context, are equivalent to:

$$\frac{dx}{dt} = v, \quad M \frac{dv}{dt} = -(z, 0)\top - \nabla \Phi(x). \quad (15)$$

In a probabilistic context, the solution of the Hamiltonian equations leave the target $Q_N(x, v)$ invariant. The proof under regularity conditions (see e.g. [14]) is straightforward and based on the fact that the solution operator of the differential equations [15] is volume/energy preserving.
In general, a good choice for the mass matrix is one that resembles the inverse of the covariance structure of the target distribution. In the case of a Gaussian target, choosing as mass matrix the inverse of the covariance of the target corresponds to transforming the target to iid standard Gaussians. In our context, we will choose a particular structure for the mass matrix, guided by the prior structure of \((z, \theta)\), that is:

\[
M = \begin{pmatrix} I_{2N} & 0 \\ 0 & A \end{pmatrix}, \quad A = \text{diag}\{a_i : 1 \leq i \leq q\},
\]

(16)

Under this selection, we can rewrite the Hamiltonian equations in (15) as follows

\[
\frac{dx}{dt} = v, \quad \frac{dv}{dt} = -(z, 0)^\top - M^{-1} \nabla \Phi(x).
\]

(17)

**Remark 1.** The choice of the mass matrix \(M\) is a critical aspect for the properties of the final algorithm.

(a) The choice of \(I_{2N}\) for the upper-left block of the mass matrix \(M\) is motivated by the prior distribution for \(Z\). Indeed, had the target distribution been that of \(2N\) iid standard Gaussians, this choice would be clearly optimal. A related idea here is that the individual elements \(Z_k\) of \(Z\) correspond to the coefficients in a Fourier expansion, thus it is expected that the data will be mostly informative about the low frequencies, i.e. a small fraction of \(Z_k\)’s at the top of the vector, with the high frequencies being a-posteriori close to their prior specification. Thus the choice of \(I_{2N}\) for a part of \(M\) in (16) is nearly optimal for a large part of the high-dimensional vector \(Z\).

(b) For the case of \(A\), in the numerical examples we have chosen diagonal values that resemble the inverse of the posterior marginal variances of the parameters as estimated by a preliminary run of the algorithm.

(c) More advanced procedures could involve better adjusting the mass matrix to the different scales and correlation structure of \(\theta\) (maybe also adjoined by the a small part of \(Z\) corresponding to low frequencies in the Fourier transform); in practice this could be based on preliminary runnings of the algorithm or an adaptive MCMC procedure or even recent Riemannian manifold methods ([16]) using the Fisher information (or its observed version). We will not go into such directions in the paper, as even a less contrived (and computationally cheaper) choice of \(M\) will give efficient methods.
3.1 Standard HMC

The standard HMC algorithm developed in [14] discretises the Hamiltonian equations (15) via a leapfrog scheme:

\[ v_{h/2} = v_0 - \frac{h}{2} (z_0, 0)^\top - \frac{h}{2} M^{-1} \nabla \Phi(x_0), \]

\[ x_h = x_0 + hv_{h/2}, \]

\[ v_h = v_{h/2} - \frac{h}{2} (z_h, 0)^\top - \frac{h}{2} M^{-1} \nabla \Phi(x_h), \]

(18)

giving rise to an operator:

\[ (x_0, v_0) \mapsto (x_h, v_h) = \psi_h(x_0, v_0) \]

which is volume-preserving, with the symmetricity property:

\[ \psi_h(x_h, -v_h) = (x_0, -v_0). \]

HMC looks at Hamiltonian dynamics up to some time horizon \( T > 0 \), via the synthesis of:

\[ I = \left\lfloor \frac{T}{h} \right\rfloor \]

(19)

leapfrog steps, so we define \( \psi_I \) to be the synthesis of \( I \) mappings \( \psi_h \). The standard HMC is given in Table 1 (\( \mathcal{P}_x \) denotes projection on the \( x \)-argument). Due to the properties of the leapfrog operator mentioned above, it is easy to verify (14) that under regulatory conditions the employed acceptance probability provides Markov dynamics with invariant distribution \( \Pi(x) \) in (11).

3.2 Advanced HMC

Applying the standard HMC algorithm in Table 1 would provide an algorithm for which the proposal \( x^* \) would become an increasingly inappropriate candidate for a sample from the target for increasing \( N \) (3); thus, the acceptance probability would vanish with increasing \( N \), assuming parameters \( h, T \) were kept fixed. The results in 5 suggest that one must decrease the step-size \( h \) as \( O(N^{-1/4}) \) to control the acceptance probability for increasing \( N \). The advanced HMC algorithm avoids this degeneracy by exploiting the definition of the target as a change of measure from a Gaussian law and employing a splitting technique for the discretisation of the Hamiltonian equations. The development below is related to the approach in 3 where now the driving noise is due to an fBM (thus, equivalently, due to a number of iid standard Gaussians) instead of a BM and the vector of interest \( x \) also involves some important parameters of interest (whereas in 3 it only involved the posterior for the latent Brownian path). The above considerations have important practical implications, as the advanced HMC will turn out to be extremely more
Standard HMC:

(i) Start with an initial value \( x^{(0)} \in \mathbb{R}^{2N+q} \) and set \( k = 0 \).

(ii) Given \( x^{(k)} \) sample \( v^{(k)} \sim N(0, M^{-1}) \) and propose:

\[
x^* = P_x \psi_h(x^{(k)}, v^{(k)})
\]

(iii) Calculate the acceptance probability

\[
a = 1 \wedge \exp\{-\Delta H(x^{(k)}, v^{(k)})\}
\]

for \( \Delta H(x, v) = H(\psi_h(x, v); M) - H(x, v; M) \).

(iv) Set \( x^{(k+1)} = x^* \) with probability \( a \); otherwise set \( x^{(k+1)} = x^{(k)} \).

(v) Set \( k \rightarrow k + 1 \) and go to (ii).

Table 2: Standard HMC, with target \( \Pi_N(x) = \Pi_N(Z, \theta) \) in [11].

Efficient than its standard version in the high-dimensional numerical applications in the next section.

Hamiltonian equations (17) are split into the following two equations:

\[
\begin{align*}
\frac{dx}{dt} &= 0, \quad \frac{dv}{dt} = -M^{-1} \nabla \Phi(x); \\
\frac{dx}{dt} &= v, \quad \frac{dv}{dt} = -(z, 0)^	op.
\end{align*}
\]

(21) (22)

Notice that both equations can be solved analytically. We construct a numerical integrator for (17) by synthesizing steps on (21) and (22). Analytically, we define the solution operators of (21) and (22):

\[
\Xi_t(x, v) = (x, v - t M^{-1} \nabla \Phi(x));
\]

\[
\tilde{\Xi}_t(x, v) = \left((\cos(t) z + \sin(t) v_z, \theta + t v_{\theta}), (-\sin(t) z + \cos(t) v_z, v_{\theta})\right).
\]

(23) (24)

The numerical integrator for (15) is defined as follows:

\[
\Psi_h = \Xi_{h/2} \circ \tilde{\Xi}_h \circ \Xi_{h/2},
\]

(25)

for small \( h > 0 \). We can synthesize steps up to some time horizon \( T \). Defining \( I \) as in [19], \( \Psi_h^I \) will correspond to the synthesis of \( I \) steps \( \Psi_h \) and will provide the
proposal for an MCMC step. Notice that $\Psi_h$ is a volume-preserving operator (as the synthesis of volume-preserving operators); then it is an easy exercise to also illustrate, for $(x_h, v_h) = \Psi_h(x_0, v_0)$, the symmetricity property:

$$\Psi_h(x_h, -v_h) = (x_0, -v_0).$$

As a consequence of the volume-preserving and this symmetricity property, the acceptance probability will have the same expression as with the standard HMC. The complete advanced HMC algorithm is shown in Table 3.

### Advanced HMC:

- **(i)** Start with an initial value $x^{(0)} \sim \otimes_{i=1}^{2N} N(0, 1) \times p(\theta)$ and set $k = 0$.

- **(ii)** Given $x^{(k)}$ sample $v^{(k)} \sim N(0, M^{-1})$ and propose

$$x^* = P_x \Psi_h^T(x^{(k)}, v^{(k)}).$$

- **(iii)** Consider

$$a = 1 \land \exp\{-\Delta H(x^{(k)}, v^{(k)})\}$$

for $\Delta H(x, v) = H(\Psi_h^T(x, v); M) - H(x, v; M)$.

- **(iv)** Set $x^{(k+1)} = x^*$ with probability $a$; otherwise set $x^{(k+1)} = x^{(k)}$.

- **(v)** Set $k \rightarrow k + 1$ and go to (ii).

Table 3: Advanced HMC, with target $\Pi_N(x) = \Pi_N(Z, \theta)$ in (11).

### 3.3 Well-Definition of Advanced HMC when $N = \infty$.

As already noted, an important property of advanced HMC is its mesh-free mixing time: for fixed $h, T > 0$, as $N$ increases the convergence/mixing properties of the Markov chain will not deteriorate. To illustrate this, we will briefly show in this section that there is a well-defined algorithm in the limit $N = \infty$. The construction has many similarities with the proof of Theorem 3.1 in [4], but the difference is that we now work with fBM (instead of BM) and our advanced HMC method updates jointly latent path and model parameters.

In the context of this section, we will assume that in fact $z \in \mathbb{R}^\infty$, and that the distribution of interest corresponds to $\Pi_N$ in [9] or [11] for $N = \infty$, denoted now
by Π. Thus, our target Π can be thought of as defined on the infinite-dimensional space:

\[ \mathcal{H} := \mathbb{R}^\infty \times \mathbb{R}^q, \]

(equipped with the relevant Borel σ-algebra) via the change of measure:

\[
\frac{d\Pi}{d\big(\otimes_{i=1}^\infty N(0, 1) \times \text{Leb}_q\big)}(Z, \theta \mid Y) \propto e^{-\Phi(Z, \theta)},
\]

(27)

for a function \( \Phi : \mathcal{H} \mapsto \mathbb{R} \). Also, We will need the infinite-dimensional vector of partial derivatives \( \nabla \Phi : \mathcal{H} \mapsto \mathcal{H} \). Then, the velocity \( v = (v_z, v_\theta) \) will also lie in the same space, \( v \in \mathcal{H} \), and the mass matrix \( M \), specified in (16) for finite dimension, will now have the infinite-dimensional identity matrix \( I_\infty \) at it’s upper-left block instead of \( I_2 \). This, gives rise to the infinite-dimensional analogue of the bivariate target \( Q_N(dx, dv) \) in (14), now denoted by \( Q(dx, dv) \). Accordingly, we now have that \( \Xi_{h/2}, \tilde{\Xi}_h, \Psi_h : \mathcal{H} \times \mathcal{H} \mapsto \mathcal{H} \times \mathcal{H} \).

The main idea here is that, under a condition on \( \nabla \Phi \), the leapfrog mapping \( \Psi_h \) in (25) projects \( (x_0, v_0) \sim Q \) to \( (x_h, v_h) \) having a distribution which is absolutely continuous with respect to \( Q \), an attribute that implies existence of a non-zero acceptance probability when \( N = \infty \). This is apparent for step \( \tilde{\Xi}_h \) in (24), as it performs a rotation in the \( (z, v_z) \)-space which is invariant for \( \prod_{i=1}^\infty N(0, 1) \otimes \prod_{i=1}^\infty N(0, 1) \), thus the overall step will preserve the absolute continuity properties of \( Q(dx, dv) \). Then, for step \( \Xi_{h/2} \) in (23), the gradient \( \nabla_z \Phi(z, \theta) \) needs to be in the Cameron-Martin space of \( \prod_{i=1}^\infty N(0, 1) \) for the translation \( v \mapsto v - \frac{h}{2} M^{-1} \nabla \Phi(x) \) to preserve the absolute continuity properties of the \( v \)-marginal \( Q(dv) \). This Cameron-Martin space is the \( \ell_2 \) space of squared summable infinite vectors \( \ell_2 \) (see for instance Chapter 2 of [12]). In contrast, for the standard HMC leapfrog mapping in (18), one can even consider the simple case when \( \Phi(x) = \text{const.} \), so that \( \nabla \Phi \equiv 0 \), to see that, immediately from the first step, a input sample from the target \( Q \) gets projected to a variable with a law which is singular with respect to \( Q \) when \( N = \infty \), thus will have zero acceptance probability.

To provide a formal result, we begin by defining the following reference measure on the joint \((x, v)\)-space:

\[
Q_0 = Q_0(dx, dv) = \left( \prod_{i=1}^\infty N(0, 1) \otimes \text{Leb}_q \right) \otimes \left( \prod_{i=1}^\infty N(0, 1) \otimes N_q(0, A^{-1}) \right),
\]

so that the joint target is:

\[
Q(dx, dv) \propto \exp\{-\Phi(x)\} Q_0(dx, dv).
\]

We also consider the sequence of probability measures on \( \mathcal{H} \times \mathcal{H} \):

\[
Q^{(i)} = Q \circ \Psi_h^{-i}, \quad 1 \leq i \leq I,
\]

15
corresponding precisely to the push-forward projection flow of the target measure $Q(dx, dv)$ under application of the leapfrog mappings. Also, for given $(x_0, v_0)$ we will write $(x_i, v_i) = \Psi^i_h(x_0, v_0)$. Note that the difference in the energy $\Delta H(x_0, v_0)$ appearing in the statement of our proposition below is still defined as $\Delta H(x_0, v_0) = H(x_1, v_1; M) - H(x_0, v_0; M)$ for the energy function in (13) with the apparent extension of the involved inner product on $\mathbb{R}^\infty$. Note that even if $H(x_0, v_0; M) = \infty$ on its own with probability 1, the difference $\Delta H(x_0, v_0)$ does not explode as implied by the statement itself of the proposition and the analytic expression for the density in part (i) of the proposition given in the proof in the Appendix.

**Proposition 1.** Assume that for $(z, \theta)$ $Q_0$-a.e., we have $\nabla_z \Phi(z, \theta) \in \ell_2$. Then:

i) $Q(I)$ is absolutely continuous w.r.t. $Q_0$ with probability density:

$$\frac{dQ(I)}{dQ_0}(x_I, v_I) = \exp\{\Delta H(x_0, v_0) - \Phi(x_I)\} .$$

ii) The Markov chain with transition dynamics (for current position $x_0 \in \mathcal{H}$):

$$x' = I [U \leq a(x_0, v_0)] x_I + I [U > a(x_0, v_0)] x_0 ,$$

for a uniform $U \sim \text{Un}[0, 1]$ and noise $v_0 \sim \prod_{i=1}^\infty N(0, 1) \otimes N_q(0, A^{-1})$, has invariant distribution $\Pi(dx)$ in (27).

**Proof.** It follows the approach of Theorem 3.1 in [4] and is given in the Appendix.

**Remark 2.** With regards to the assumption $\nabla_z \Phi(z, \theta) \in \ell_2$, $Q_0$-a.e., to connect with the finite-dimensional algorithm applied in practice, one could try to show that the practical algorithm corresponds to a finite-dimensional projection (e.g. after applying finite-difference approximation) of an algorithm in theory defined on $\mathcal{H}$ for which the regularity condition is indeed satisfied. This is the approach we followed for BM-driven diffusion bridges in [4]. We will not pursue that in this work, as it would require a quite substantial number of tedious but otherwise not-interesting algebraic calculations (one looks at the practical algorithm for increasing dimension to form the infinite-dimensional one). We believe that Proposition 1 already highlights the structurally important mesh-free property of the advanced HMC method. We will only mention that the intuition for why such a summability condition might hold (maybe under some regularity conditions on the functions involved in the specification of the algorithm) is that the data possess a finite amount of information for the Z-vector, thus after a large enough number of co-ordinates the partial derivatives will be diminishing towards zero.
4 Fractional Stochastic Volatility models

4.1 Data and Model

In order to illustrate the algorithm and assess its performance, we apply in this section the algorithm to the fractional SV model of [11, 8, 9]:

\[
dU_t = \left( \mu - \sigma S(X_t)^2 / 2 \right) dt + \exp(X_t/2) \left( \sqrt{1 - \rho^2} dW_t + \rho dB_t^H \right),
\]
\[
dX_t = \kappa(\mu_X - X_t)dt + \sigma_X dB_t^H. \tag{28}
\]

Thus, the volatility is modeled via the well-studied fractional Ornstein-Uhlenbeck process. As mentioned earlier we incorporate the ‘leverage’ parameter \( \rho \) to our parameter vector \( \theta \) as well as \( \mu_X \) that appears in [9]. Moreover, we extend the range of \( H \in (0, 1) \) thus allowing for long, medium and short range dependence as opposed to previous literature who restricted attention to \( H \in (1/2, 1) \). In the case where the available data consist of direct observations on \( S_t \), the model is specified by (11) and (10) where the algorithm of Section 3 can be applied.

We also consider an alternative type of dataset consisting of volatility proxies constructed from option prices, in addition to the observations of \( S_t \). In [1], the VIX index was used to construct proxies for the volatility. They initially considered a simple approach (unadjusted proxy) by using VIX directly to obtain \( \sigma_S(X_t) \) and therefore \( X_t \). In addition to the unadjusted proxy, an adjusted version (integrated volatility proxy) was also considered by making the assumption that the pricing measure has a linear drift and using Fubini’s theorem; see also [25]. In the Bayesian setting, the integrated volatility proxy was also used by [21] and [29] to provide observations of \( \sigma_S(X_t) \) where the authors incorporated additional measurement error. In the following applications we take a rather simplified approach to illustrate the proposed methodology, and use the unadjusted volatility proxy as a noisy measurement device for \( \sigma_S(X_t) \). This is done for two reasons. First, our focus lies mainly on exploring the behaviour of our algorithm under a different observation regime (noisy observations) and we don’t want to introduce additional subject-specific considerations such as making assumptions on the pricing measure. Second, the difference between the two approaches is often negligible, see for example the relevant simulation experiments in [1] for the Heston model, and can possibly be omitted (or left to the error term) in order to avoid introducing additional complexity. We do note however that the approaches of [21, 29] can be implemented under our framework and, more generally, the problem of combining option and asset prices needs to be investigated further even in the context of diffusion driven models. We are currently working towards this direction.

The additional noisy observations corresponding to the VIX data, are denoted
by \( Y^x_k \) and are assumed to provide information on \( X_{t,k} \) via the model below:

\[
Y^x_k = X_{t,k} + \epsilon_k, \quad \epsilon_k \overset{i.i.d.}{\sim} N(0, \tau^2), \quad 1 \leq k \leq n.
\]  

(29)

Note that in the Appendix we give the details for the calculation of \( p_N(Y|Z, \theta) \) and the derivatives \( \nabla_Z p_N(Y|Z, \theta) \), \( \nabla_{\theta} p_N(Y|Z, \theta) \) only for the case when data \( Y \) is as in (4), i.e. direct observations of the price process \( U_t \), as including also the terms due to the extra data in (29) is fairly trivial.

In order to carry out Bayesian inference, the model is completed with priors that were chosen according to related literature on diffusion driven SV models, such as [7]. Note that in all applications in the sequel we considered daily data over a year, thus corresponding to \( n = 252 \) time instances when observations where collected. A vague uniform prior distribution between 1 and 100 days\(^{-1} \) was chosen for \( \kappa \). The prior distribution for \( \mu_X \) has been defined as a normal density with 95\% credible interval spanning between the minimum and maximum simulated volatility values or the real VIX observations when these are used, over the entire period under consideration. A standard inverse gamma distribution was used to define the prior of \( \sigma^2_X \), with shape and scale parameters \( \alpha = 2 \) and \( \beta = \alpha \times 0.03 \times \sqrt{252} \) respectively. Vague priors were chosen for the remaining parameters, namely uniform distribution between \((0, 1)\) and \((-1, 1)\) for \( H \) and \( \rho \) respectively and \( N(0, 10^6) \) for \( \mu \).

We also note that parameter \( \tau \) was assumed to be known. This parameter controls the weight placed to the volatility proxies in order to create a weighted averaged volatility measurement combining information from asset and option prices. We have therefore treated this as a user-specified parameter that can be chosen on the basis of additional considerations relevant to the application at hand. If needed it is possible to estimate this parameter using the algorithm of this paper, although according to our experience this has a small effect to the performance of the MCMC sampler. We set \( \tau = 0.05 \) based on an estimate obtained by a preliminary run of the full model to S&P500 / VIX data. Also, in all cases we used discretisation increment \( \delta = 0.1 \) for the Euler approximation of the continuous-time SDEs, thus the dimension of the latent variable \( Z \) was \( 2N = 2 \times 252 \times 10 = 5,040 \).

The analysis based on the data including S&P500 prices only, resulted in a posterior that is centered around values \( H < 1/2 \). Hence, despite the somewhat exploratory nature of the unadjusted volatility proxies approach, the estimation based on the joint S&P500/VIX series allowed us to get a rough idea on whether the evidence in option prices suggests otherwise; in fact it doesn’t as both sources of data suggest medium range dependence. It is certainly worth further investigating this issue, e.g. by clarifying the effect of the selected model, but we should note for now that some empirical evidence, independent of the model selection, from the VIX data we have used in our simulations, see Fig.\( \text{[Fig]} \) below, do also
seem to give some preliminary evidence of negative correlation. Of course, there are many cases when long memory has been empirically verified to characterize fBM-driven models (see e.g. Chapter 1 of [26] and the references therein) and our methods certainly are relevant also to these cases.

4.2 Illustration to Simulated Data

We first apply the advanced HMC algorithm in Table 3 jointly updating \((Z, \theta)\), to simulated data. We generated 250 observations from the model (28), corresponding roughly to a year of data. We then considered two datasets:

i) Dataset Sim-A: with 250 daily observations on \(S_t\) only, as in (4).

ii) Dataset Sim-B: with additional daily observations on \(X_t\) for the same time period contaminated with measurement errors as in (29).

Tables 4 and 5 show posterior estimates (such 95% credible intervals, mean and median) obtained from running advanced HMC for Datasets Sim-A and Sim-B respectively, together with the true parameter values. The true parameters were chosen to be similar with those in previous analyses on the S&P500/VIX indices based on diffusion-driven SV models ([1, 7]) and the ones we found ourselves from the real data analysis in the sequel. For both Datasets, we repeated the procedure for three different values of \(H\) (i.e. 0.3, 0.5 and 0.7) corresponding to negative autocorrelation, independence and positive autocorrelation at the fBM increments. The advanced HMC algorithm updated paths and parameters jointly in a single block, and consisted of respectively 30 and 50 leapfrog steps for Datasets Sim-A and Sim-B. The value of \(h\) was tuned to achieve an average acceptance rate of around 70%: it was set to 0.03 and 0.01 for Datasets Sim-A and Sim-B respectively. MCMC traceplots for the case \(H = 0.3\) for Datasets Sim-A and Sim-B are shown in Figures 1 and 2. We did not notice substantial difference for \(H = 0.5\) and \(H = 0.7\), so we do not provide the corresponding traceplots for space considerations. Overall, the mixing of the chain appears to be quite good considering the complexity of the model. More details on this are provided in Section 4.4.

In the case of Dataset Sim-A, the results in Table 4 show reasonable agreement between posterior distribution and true parameter values, with the true values being within the 95% credible intervals in all but two cases. More interestingly, though, most of the credible intervals are quite wide reflecting the limited number of data or the small amount of information in Dataset Sim-A for particular parameters. For the Hurst parameter in particular, in the case of negatively autocorrelated increments, the output clearly suggests a value of \(H\) below 0.5, in contrast to the case of \(H = 0.5\) or \(H = 0.7\) when observations from the price process seem to contain very small amount of information for \(H\).
In the case of Dataset Sim-B, the posterior is now much more informative for all parameters. The 95% credible intervals contain the true values in all cases and are much narrower making it possible to produce accurate estimates of $H$ when the true value is below, equal or above 0.5.

| Model         | Parameter | True value | 2.5% | 97.5% | Mean | Median |
|---------------|-----------|------------|------|-------|------|--------|
| Data from H=0.3 | $\mu$     | 0.25       | -0.14| 0.67  | 0.28 | 0.29   |
|               | $\rho$    | -0.75      | -0.91| -0.31 | -0.63| -0.63  |
|               | $\kappa$  | 4.00       | 1.60 | 10.80 | 4.90 | 4.40   |
|               | $\mu_X$   | -5.00      | -4.97| -2.49 | -3.79| -3.82  |
|               | $H$       | 0.30       | 0.24 | 0.38  | 0.31 | 0.31   |
|               | $\sigma_X$| 2.00       | 1.08 | 3.11  | 1.90 | 1.83   |
|               | $X_0$     | -5.00      | -5.24| -4.19 | -4.72| -4.72  |
| Data from H=0.5 | $\mu$     | 0.25       | 0.17 | 0.49  | 0.33 | 0.33   |
|               | $\rho$    | -0.75      | -0.73| 0.19  | -0.29| -0.31  |
|               | $\kappa$  | 4.00       | 1.71 | 13.1  | 5.52 | 4.85   |
|               | $\mu_X$   | -5.00      | -6.21| -3.66 | -4.97| -4.98  |
|               | $H$       | 0.50       | 0.30 | 0.79  | 0.55 | 0.55   |
|               | $\sigma_X$| 2.00       | 1.32 | 6.55  | 2.91 | 2.57   |
|               | $X_0$     | -5.00      | -5.17| -4.80 | -4.99| -4.99  |
| Data from H=0.7 | $\mu$     | 0.25       | 0.11 | 0.38  | 0.25 | 0.25   |
|               | $\rho$    | -0.76      | -0.78| 0.47  | -0.25| -0.25  |
|               | $\kappa$  | 4.00       | 1.90 | 23.60 | 8.40 | 6.75   |
|               | $\mu_X$   | -5.00      | -6.20| -4.53 | -5.45| -5.48  |
|               | $H$       | 0.70       | 0.25 | 0.80  | 0.55 | 0.56   |
|               | $\sigma_X$| 2.00       | 0.83 | 5.72  | 2.37 | 2.09   |
|               | $X_0$     | -5.00      | -5.20| -4.82 | -5.01| -5.01  |

Table 4: Posterior summaries for the model parameters as obtained from applying advanced HMC on the posterior distribution from Dataset Sim-A.

### 4.3 Real-Data from S&P500/VIX

In this section we apply the advanced HMC algorithm to data from the S&P500 index. Notice that Figure 3 examines autocorrelations for the values of the VIX index, transformed to the scale of $X$ as a rough proxy for this process. Of course, these data utilise information from a different source, i.e. option prices on the S&P500, so there is no guarantee that there will be strict agreement with the information on the volatility contained in the S&P500 values. Nevertheless it is
| Model          | Parameter | True value | 2.5%  | 97.5% | Mean  | Median |
|----------------|-----------|------------|-------|-------|-------|--------|
| Data from H=0.3| μ         | 0.25       | -0.07 | 0.25  | 0.09  | 0.09   |
|                | ρ         | -0.75      | -0.80 | -0.55 | -0.68 | -0.68  |
|                | κ         | 4.00       | 3.20  | 13.40 | 8.00  | 7.90   |
|                | μ_X       | -5.00      | -6.00 | -4.86 | -5.35 | -5.33  |
|                | H         | 0.30       | 0.25  | 0.30  | 0.28  | 0.28   |
|                | σ_X       | 2.00       | 1.46  | 2.12  | 1.92  | 1.81   |
|                | X_0       | -5.00      | -5.16 | -4.83 | -5.00 | -5.00  |
| Data from H=0.5| μ         | 0.25       | 0.090 | 0.39  | 0.24  | 0.25   |
|                | ρ         | -0.75      | -0.84 | -0.71 | -0.78 | -0.78  |
|                | κ         | 4.00       | 2.57  | 7.50  | 4.53  | 4.35   |
|                | μ_X       | -5.00      | -5.79 | -4.19 | -4.96 | -4.94  |
|                | H         | 0.50       | 0.47  | 0.52  | 0.50  | 0.50   |
|                | σ_X       | 2.00       | 1.66  | 2.09  | 1.86  | 1.86   |
|                | X_0       | -5.00      | -5.10 | -4.85 | -4.97 | -4.97  |
| Data from H=0.7| μ         | 0.25       | 0.14  | 0.48  | 0.28  | 0.28   |
|                | ρ         | -0.75      | -0.79 | -0.68 | -0.75 | -0.75  |
|                | κ         | 4.00       | 3.54  | 21.90 | 11.20 | 10.70  |
|                | μ_X       | -5.00      | -5.43 | -4.70 | -5.05 | -5.05  |
|                | H         | 0.50       | 0.68  | 0.86  | 0.76  | 0.76   |
|                | σ_X       | 2.00       | 1.36  | 3.91  | 2.27  | 2.11   |
|                | X_0       | -5.00      | -5.09 | -4.92 | -5.00 | -5.00  |

Table 5: Posterior summaries for the model parameters as obtained from applying advanced HMC on the posterior distribution from Dataset Sim-B.

useful to contrast these two sources of information, regarding the memory properties of the volatility in particular. We consider VIX values since 1990 in an attempt to obtain a bigger picture regarding the behaviour of the series. A very interesting empirical characteristic is obtained when we examine the autocorrelation of the increments of the log-VIX values as they appear to exhibit negative autocorrelation at the first lags. The fractional SV model considered in this paper provides a natural parametric model for such observed negative autocorrelations. In this case one would expect a value \( H < 0.5 \) that would deviate from a BM-driven SV model. In fact it is not possible to reproduce such autocorrelation plots, at this range of parameter values and at that sampling frequency, when drawing from BM-driven SV models, suggesting that a fBM-driven approach would be interesting.

We fitted model \(^{[28]}\) to daily data using advanced HMC algorithm in Table 3.
Figure 1: Traceplots of all parameters from $2 \times 10^4$ iterations of advanced HMC, for the posterior distribution corresponding to Dataset Sim-A. True parameter values are as in Table 4 with true Hurst parameter value $H = 0.3$. Execution time was about 5 hours, on a standard PC, with code in Matlab.

Figure 2: Traceplots as in Figure 1, corresponding now to Dataset Sim-B. True parameter values are now as in Table 5 with true Hurst parameter value $H = 0.3$. Execution time was this time about 7 hours (due using more leapfrog steps (i.e. 50) compared to Dataset Sim-A (i.e. 30)).

to sample from the related posterior. As with the simulations in Section 4.2 we considered two types of datasets:
i) Dataset A: with S&P500 values only (discrete-time observations of the price process). We considered daily S&P500 values from two different time periods close to the recession. The first period was from 5 March 2007 to 5 March 2008, i.e. a year of data before the Bear Stearns closure (before crisis data). The second period was from 15 September 2008 to 15 September 2009, i.e. a year of data after the Lehman Brothers bankruptcy (after crisis data).

ii) Dataset B: the same as with Dataset A, but with the addition of daily VIX values for the same time period as with the S&P500 values.

Tables 6 and 7 show posterior estimates obtained after applying our advanced HMC algorithm to the posterior distribution corresponding to Dataset A and B respectively. In all cases the posterior distribution of $H$ is concentrated well below 0.5 thus confirming the empirical evidence of Figure 3. Looking closer at Table 7 we note that the period after the crisis seems to be associated with higher negative autocorrelation (lower values of $H$) as well as a higher level of volatility. It is also worth mentioning that the value of $\rho$ is consistently negative suggesting the presence of a leverage effect.
| Model                     | Parameter | 2.5%   | 97.5%  | Mean   | Median |
|--------------------------|-----------|--------|--------|--------|--------|
| 5 Mar 2007 - 5 Mar 2008  | \( \mu \) | -0.25  | 0.21   | -0.03  | -0.04  |
| before Bear Stearns closure | \( \rho \) | -0.97  | -0.44  | -0.80  | -0.83  |
|                          | \( \kappa \) | 4.23   | 21.73  | 11.34  | 10.79  |
|                          | \( \mu_X \) | -5.05  | -3.56  | -4.37  | -4.39  |
|                          | \( H \)   | 0.34   | 0.47   | 0.41   | 0.41   |
|                          | \( \sigma_X \) | 0.81   | 3.31   | 1.86   | 1.78   |
|                          | \( X_0 \) | -5.71  | -4.75  | -5.23  | -5.23  |
| 15 Sep 2008 - 15 Sep 2009| \( \mu \) | -0.56  | 0.24   | -0.16  | -0.15  |
| after Lehman Brothers closure | \( \rho \) | -0.97  | -0.53  | -0.79  | -0.80  |
|                          | \( \kappa \) | 2.72   | 13.90  | 7.18   | 6.84   |
|                          | \( \mu_X \) | -5.08  | -2.80  | -3.78  | -3.72  |
|                          | \( H \)   | 0.33   | 0.44   | 0.38   | 0.38   |
|                          | \( \sigma_X \) | 1.21   | 3.70   | 2.20   | 2.10   |
|                          | \( X_0 \) | -4.04  | -3.06  | -3.51  | -3.50  |

Table 6: Posterior summaries for the model parameters as obtained from applying advanced HMC on the posterior distribution from Dataset A.

| Model                     | Parameter | 2.5%   | 97.5%  | Mean   | Median |
|--------------------------|-----------|--------|--------|--------|--------|
| 5 Mar 2007 - 5 Mar 2008  | \( \mu \) | -0.22  | 0.07   | -0.07  | -0.07  |
| before Bear Stearns closure | \( \rho \) | -0.82  | -0.67  | -0.75  | -0.76  |
|                          | \( \kappa \) | 3.82   | 8.77   | 6.27   | 6.27   |
|                          | \( \mu_X \) | -5.27  | -4.71  | -5.00  | -5.00  |
|                          | \( H \)   | 0.33   | 0.37   | 0.35   | 0.35   |
|                          | \( \sigma_X \) | 0.77   | 0.96   | 0.86   | 0.87   |
|                          | \( X_0 \) | -5.59  | -5.38  | -5.48  | -5.49  |
| 15 Sep 2008 - 15 Sep 2009| \( \mu \) | -0.48  | 0.09   | -0.19  | -0.19  |
| after Lehman Brothers closure | \( \rho \) | -0.71  | -0.56  | -0.64  | -0.64  |
|                          | \( \kappa \) | 1.73   | 5.37   | 3.39   | 3.33   |
|                          | \( \mu_X \) | -4.64  | -3.65  | -4.07  | -4.06  |
|                          | \( H \)   | 0.28   | 0.33   | 0.31   | 0.31   |
|                          | \( \sigma_X \) | 0.63   | 0.80   | 0.72   | 0.72   |
|                          | \( X_0 \) | -4.44  | -4.19  | -4.31  | -4.31  |

Table 7: Posterior summaries for the model parameters as obtained from applying advanced HMC on the posterior distribution from Dataset B.
4.4 Comparison of Different HMC Implementations

The previous results have been obtained by updating jointly the paths and parameters of the model with the advanced version of the HMC (referred to as aHMC in the tables that follow). We now compare the performance of this algorithm to the Gibbs version of the advanced HMC, in which paths and parameters are updated alternatively. In addition, we compare the performance of the advanced HMC with standard HMC [14]. In each case, the same mass matrix is utilised, determined from the empirical covariance of previous explorations of the posterior density. The number of leapfrogs is set to 30 and 50 on Datasets Sim-A and Sim-B respectively. The integration time step for the Hamiltonian equations is set to 0.03 and 0.01 in the Joint and Gibbs versions of the advanced HMC. It had to be decreased to 0.0003 and 0.0002 for the standard HMC to achieve an acceptance rate over 70%, fact that provides an initial evidence that the performance of standard HMC is not as good as in the advanced HMC versions.

The computation time of each iteration, in each of the three implementations, is similar. One iteration takes between 1.84 and 1.86 seconds on Dataset Sim-A, using 30 leapfrog steps. On Dataset Sim-B, iterations take between 2.95 and 2.98 seconds, due to an increased number of leapfrog steps. One way to assess the different performances of each algorithm is the Effective Sample Size (ESS), computed as in [15] from the lagged autocorrelations of the traceplots. We focus on the minimum ESS over the different components of \( \theta \) and \( Z \) as an overall measure for the sampling efficiency of the algorithm, linking to the the percentage out of the total MCMC draws that can be considered as independent draws from the

| Sampler   | \( \min_\theta(ESS) \) | \( \min_\theta(ESS) \) | time | \( 100 \times \frac{\min_\theta(ESS)}{\text{time}} \) | rel. \( \frac{\min_\theta(ESS)}{\text{time}} \) |
|-----------|-------------------------|------------------------|------|-----------------------------------------------|-----------------------------------------------|
| aHMC\((\text{Joint})\) | 0.90%                   | 1.40%                  | 1.84 | 0.49                                          | 22.5                                          |
| aHMC\((\text{Gibbs})\)  | 0.42%                   | 10.8%                  | 1.86 | 0.23                                          | 10.4                                          |
| HMC\((\text{Joint})\)   | 0.04%                   | 0.49%                  | 1.84 | 0.02                                          | 1.00                                          |

Table 8: Relative efficiency of the different versions of the HMC applied on Dataset Sim-A (H=0.3). Comparison is made via the minimum ESS (%) and CPU times (seconds).

| Sampler   | \( \min_\theta(ESS) \) | \( \min_\theta(ESS) \) | time | \( 100 \times \frac{\min_\theta(ESS)}{\text{time}} \) | rel. \( \frac{\min_\theta(ESS)}{\text{time}} \) |
|-----------|-------------------------|------------------------|------|-----------------------------------------------|-----------------------------------------------|
| aHMC\((\text{Joint})\) | 1.80%                   | 2.70%                  | 2.95 | 0.61                                          | 45.0                                          |
| aHMC\((\text{Gibbs})\)  | 0.06%                   | 3.98%                  | 2.98 | 0.02                                          | 1.48                                          |
| HMC\((\text{Joint})\)   | 0.04%                   | 0.49%                  | 2.95 | 0.01                                          | 1.00                                          |

Table 9: Relative efficiency of the different versions of the HMC applied on Dataset Sim-B (H=0.3). Comparison is made via the minimum ESS (%) and CPU times (seconds).
In each case, the sampling efficiency over parameters is lower than the one over path components. The comparison of the Joint and Gibbs versions of the advanced HMC provides further insight. The Joint version is respectively 2.16 and 30.4 times more efficient than its Gibbs counterpart, illustrating the effect of the strong posterior correlation between $Z$ and $\theta$. This correlation increases significantly when noisy observations of the volatility path are introduced. This correlation is due to the data since, given our construction, these two components are a-priori independent.

These simulations illustrate the gain provided by the advanced implementation of the HMC algorithm over its standard counterpart. A factor 22.5 and 45.0 is respectively observed on Datasets Sim-A and Sim-B, due to the necessary decrease in the integration horizon of standard HMC to maintain a 70% acceptance rate.

5 Summary and Further Directions

In this paper we have developed a Bayesian inference methodology for SDE models driven by fBM and applied our advanced version of the HMC algorithm to tackle the computationally challenging task of their estimation. Such a modeling structure allows for the presence of memory that can be observed in various applications including the volatility of asset prices and indices such as the S&P500. As a prototypical example, we therefore considered the class of fractional SV models and applied the advanced version of the HMC algorithm to carry out Bayesian inference for real and simulated data. The developed HMC algorithm performed very well and provides, to our knowledge, one of the few options for routine Bayesian likelihood-based estimation in the context of discretely observed fBM-driven diffusion processes. A conclusion here is that current computational capabilities together with algorithmic innovation have provided tools that can now allow practitioners to experiment with non-Markovian model structures in the form of fBM-driven SDEs, in generic non-linear contexts.

Thinking of the generality and scope of our research, we can make the following remarks:

Direct Observations: We have assumed existence of a non-trivial (Lebesgue) density for the observations given the latent fBM path and the parameters. This will not be the case when the data correspond to direct observations of the fBM-driven SDE (in our numerical applications, the observations on the volatility included some error). In such a scenario, to extend our approach (and similarly to the case for algorithms for standard BM-driven SDEs) we will need to develop a framework for bridges (i.e. paths conditioned on hitting the data) for the fBM-driven SDEs, and work with their mathematical properties.
Further Advancing the HMC dynamics: A very interesting direction of investigation involves combining the advanced HMC algorithm in this paper (giving emphasis to computational robustness in high dimensions) with recent Riemannian manifold HMC methods [10] allowing for automation at the specification of the mass matrix and efficient HMC transitions on parametric distributions with highly irregular contour structure.

Gaussian Processes Beyond fBM: Our algorithmic framework can also be applied for models when the latent variables correspond to general stationary Gaussian processes, as the initial Davies and Harte transform and all other steps in the development of our method can be carried forward in this context.

Applications of the fractional SV models: It is of great interest to see the implication of the fractional SV model in pricing, especially for values of $H < 1/2$. Moreover, the problem of joint estimation of the physical and pricing measure based on asset and option prices can be studied in more depth, even in the case of diffusion SV models, and certainly to the case of fractional BM.

The authors of this paper are currently pursuing research in all above directions. Finally, another interesting direction can involve Bayesian parametric inference for generalised Langevin Equations with fractional noise, with such models arising in applications in physics and biology (see e.g. [23]).

**Acknowledgements**

This work was supported by the EPSRC grant EP/K001264/1.

**Appendix**

A Calculation of $p_N(Y|Z, \theta)$

First, for given $Z = (Z_k)_{k=1}^{2N}$, we have the following transform from the Davies and Harte method:

$$G_N = (G(i))_{i=1}^{2N} = P_{1:N} \{ P \sqrt{A_H} Q Z \} ,$$

where $P_{1:N} \{ \cdot \}$ denotes projection of the $2N$-dimensional input vector to it’s first $N$ co-ordinates. Once we have obtained fGN, we can obtain the increments of $B^H$ over all intervals $[(i-1)\delta, i\delta]$, for $1 \leq i \leq N$, and use them within an Euler
discretisation scheme for the SV-process X. Analytically, due to the self-similarity property of the fBm:
\[
\{B_t^H\}_{t \geq 0} = \{c^H B_{t/c}^H\}_{t \geq 0},
\]
for any \(c > 0\), thus we have \(B_t^H - B_{(i-1)\delta}^H = \delta^H G(i)\), and the Euler approximation of \(X\) writes as follows:
\[
X_i - X_{i-1} = \mu_X (X_{i-1}, \zeta) \delta + \sigma_X (X_{i-1}, \zeta) \delta^H G(i),
\]
for \(1 \leq i \leq N\). It remains to also discretise the likelihood factors \(p(Y_k|Y_{k-1}, B^H, \theta)\) in (10). For \(1 \leq k \leq n\), and \(j_k = \lfloor t_k/\delta \rfloor\), we find:
\[
m_{k,N} = Y_{k-1} + \sum_{i=j_{k-1}+1}^{j_k} (\mu - \sigma S(X_{i-1})^2/2) \delta + \rho \sum_{i=j_{k-1}+1}^{j_k} \sigma S(X_{i-1}) \delta^H G(i);
\]
\[
V_{k,N} = (1 - \rho^2) \sum_{i=j_{k-1}+1}^{j_k} \sigma S(X_{i-1})^2 \delta.
\]
Thus, denoting by \(\Phi(\cdot; \mu, v)\) the probability density function of a normal distribution with mean \(\mu \in \mathbb{R}\) and variance \(v > 0\), the required likelihood term is:
\[
p_N(Y|Z, \theta) = \prod_{k=1}^{N} \Phi(Y_k; m_{k,N}, V_{k,N}).
\]

### B Calculation of \(\nabla_Z p_N(Y|Z, \theta)\) and \(\nabla_\theta p_N(Y|Z, \theta)\)

One needs to apply the chain rule over the composition of deterministic mappings:
\[
Z \mapsto G_N \mapsto (X_1, X_2, \ldots, X_N),
\]
taking care so that unnecessary costs are not added to the calculations. It will be useful to separate the parameters here:
\[
\theta_y = (\mu, \rho); \quad \theta_x = (\zeta, H).
\]
From the chain rule we get immediately that:
\[
\nabla_Z \log p_N(Y|Z, \theta) = (\frac{dG_N}{dZ})^\top (\frac{dX}{dG_N})^\top \nabla_X \log p_N(Y|X, \theta_y);
\]
\[
\nabla_\zeta \log p_N(Y|Z, \theta) = (\frac{dX}{d\zeta})^\top \nabla_X \log p_N(Y|X, \theta_y);
\]
\[
\partial_H \log p_N(Y|Z, \theta) = (\frac{dG_N}{dH})^\top (\frac{dX}{dG_N})^\top \nabla_X \log p_N(Y|X, \theta_y),
\]
where we have set:

\[
\frac{dG}{dX} = (\partial G(i)/\partial X)_{ij} \in \mathbb{R}^{N \times N}; \quad \frac{dG}{dZ} = (\partial G(i)/\partial Z)_{ij} \in \mathbb{R}^{N \times (2N)}; \quad \frac{dX}{d\zeta} = (\partial X_i/\partial \zeta)_{ij} \in \mathbb{R}^{N \times p}; \quad \frac{dG}{dH} = (dG(i)/dH)_{i} \in \mathbb{R}^{N}.
\]

The exact calculation of the derivatives for the last step of the hierarchy in the model:

\[
\nabla_{X} \log p_{N}(Y|X, \theta) , \quad \nabla_{\theta} \log p_{N}(Y|X, \theta)
\]

is particular for the likelihood expression (conditionally on the latent process X) of the SV model and is given later in Appendix B.1. We give now the derivatives over the other steps in the chain, that could also be relevant for fBM-driven diffusion models beyond the SV setup.

We start from \(dG_N/dZ\). Recall the Euler approximation of \(X\) in (30). Setting:

\[
\sigma_{j-1} = \sigma_{X}(X_{j-1}, \zeta) ; \quad \mu_{j-1} = b_{X}(X_{j-1}, \zeta) ; \quad \partial x \sigma_{j-1} = \partial x \sigma_{X}(X_{j-1}, \zeta), \quad f_{j-1} = (1 - \mu_{j-1} \delta)x_{j-1} - (X_{j} - X_{j-1} - \mu_{j-1} \delta) \partial x \sigma_{j-1} \sigma^{2}_{j-1},
\]

for \(1 \leq j \leq N\), one can immediately obtain the following:

\[
\frac{dG}{dX} = \delta^{-2H} \begin{pmatrix}
\sigma_{0}^{-1} & 0 & \cdots & 0 & 0 \\
0 & \sigma_{1}^{-1} & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & f_{N-1} & \sigma_{N-1}^{-1}
\end{pmatrix}.
\]

Then, following Davies and Harte algorithm in Section 2, we have:

\[
\frac{dG}{dZ} = \mathcal{P}_{1:N,1:(2N)}\{P \sqrt{\Lambda_{H}} Q\},
\]

where \(\mathcal{P}_{1:N,1:(2N)}\) denotes projection of the \((2N) \times (2N)\)-dimensional input matrix to its first \(N\) rows. The \(N\) rows of \(dX/d\zeta\) are obtained recursively via the Euler scheme (30) starting from:

\[
\nabla_{\zeta}X_1 = \nabla_{\zeta}b_{X}(X_0, \zeta) \delta + \nabla_{\zeta} \sigma_{X}(X_0, \zeta) \delta^H G(1)
\]

and proceeding as follows for \(i = 2, \ldots, N\):

\[
\nabla_{\zeta}X_i = \nabla_{\zeta}X_{i-1} + \nabla_{\zeta}b_{X}(X_{i-1}, \zeta) \delta + \nabla_{\zeta} \sigma_{X}(X_{i-1}, \zeta) \delta^H G(i).
\]

29
Finally, following the Davies and Harte method, we have that:

\[ \frac{dG_N}{dH} = \mathcal{P}_{1:N,1:(2N)} \left\{ P \frac{dN_H^{1/2}}{dH} QZ \right\} . \]

Thus, upon recalling the expression in (8) for the \(2N\) elements \(\{\lambda_k\}_{k=0}^{2N-1}\) of the diagonal matrix \(\Lambda_H\), we calculate:

\[ \frac{d\lambda_k}{dH} = \frac{1}{2\lambda_k^{1/2}} \sum_{j=0}^{2N-1} \frac{dc_{0,j}}{dH} \exp \left( -2\pi i \frac{jk}{2N} \right) . \] (32)

The remaining derivatives \(dc_{0,j}/dH\) are easy to obtain via \(g(k)/dH\) where \(g(\cdot)\) are the lagged autocovariances of \(f_Gn\) in (7). From there, we have that:

\[ \frac{dg(k)}{dH} \delta^{-2H} = \begin{cases} 
2 \log(\delta), & k = 0; \\
2 \log(\delta)(2^{2H-1} - 1) + \log(2)2^{2H}, & k = 1; \\
2 \log(\delta)(\frac{1}{2}(k+1)^{2H} + \frac{1}{2}(k-1)^{2H} - k^{2H}) + \log(k+1)(k+1)^{2H} + \log(k-1)(k-1)^{2H} - 2 \log(k)k^{2H}, & k \geq 2 .
\end{cases} \]

Notice that calculation of \(dG_N/dH\) requires \(O(N \log N)\) operations using FFT.

B.1 Calculation of \(\nabla_X \log p_N(Y|X,\theta_y)\) and \(\nabla_{\theta_y} \log p_N(Y|X,\theta_y)\)

We now give the derivatives for the log-likelihood (given latent variables) of the adopted fractional SV model. The expressions correspond to the case of observing the price only, but can be easily adapted to the case of additional noisy volatility observations. Working with expression (31), with \(p_N(Y|Z,\theta)\) now viewed as a function of \((X,\theta_y)\), thus expressed now as \(p_N(Y|X,\theta_y)\), we obtain the following:

\[ \nabla_X \log p(Y|X,\theta_y) = -\frac{1}{2} \sum_{k=1}^{n} \left\{ \frac{\nabla_X V_{k,N}}{V_{k,N}} - \frac{(Y_{k-m_{k,N}})(2V_{k,N} \nabla_X m_{k,N} + (Y_{k-m_{k,N}}) \nabla_X V_{k,N})}{V_{k,N}^2} \right\} \] (33)

with the gradient terms, for \(1 \leq i \leq n:\)

\[ (\nabla_X m_{k,N})_i = \mathbb{I} \{ j_{k-1} \leq i \leq j_k - 1 \} \cdot \left( -\sigma_S(X_i)\sigma'_S(X_i) \delta + \rho \sigma'_S(X_i) \delta H G(i) \right) ; \]

\[ (\nabla_X V_{k,N})_i = \mathbb{I} \{ j_{k-1} \leq i \leq j_k - 1 \} \cdot \left( 2(1-\rho^2)\sigma_S(X_i)\sigma'_S(X_i) \delta \right) , \]
Finally, for $\nabla_{\theta_y} p(Y|X, \theta_y)$ one can use again expression (33), but with the gradient terms now being w.r.t. $\theta_y$, so that in this case:

$$\nabla_{\theta_y} m_{k,N} = \left( (j_k - j_{k-1}) \delta, \sum_{i=j_{k-1}+1}^{j_k} \sigma_S(X_{i-1}) \delta^H G(i) \right)^\top;$$

$$\nabla_{\theta_y} V_{k,N} = \left( 0, -2 \rho \sum_{i=j_{k-1}+1}^{j_k} \sigma_S(X_{i-1})^2 \delta \right)^\top.$$

### C Proof of Proposition 1

The proof of (i) is based on obtaining a recursive formula for the Radon-Nikodym derivatives $dQ_i/dQ_0$ for $i = 1, 2, \ldots, I$. We set:

$$C = M^{-1} = \begin{pmatrix} I_{\infty} & 0 \\ 0 & A^{-1} \end{pmatrix},$$

where recall that $A = \text{diag}\{a_i : 1 \leq i \leq q\}$. We also set:

$$g(x) := -C^{1/2} \nabla \Phi(x), \quad x \in \mathcal{H}.$$  

From the definition of $\Psi_h$ from (25), we have the following recursion for probability measures:

$$Q^{(i)} = Q^{(i-1)} \circ \Xi_{h/2}^{-1} \circ \Xi_{h}^{-1} \circ \Xi_{h/2}^{-1}.$$  

We will need the following result. Given the assumption $\nabla_z \Phi(z, \theta) \in \ell_2$, $Q_0$-a.e, we have that $Q_0 \circ \Xi_{h/2}^{-1}$ and $Q_0$ are absolutely continuous w.r.t. each other, with Radon-Nikodym derivative:

$$\frac{dQ_0 \circ \Xi_{h/2}^{-1}}{dQ_0}(x, v) = \exp \left\{ \left( \frac{h}{2} g(x), C^{-1/2} v \right) - \frac{1}{2} \frac{h}{2} |g(x)|^2 \right\} =: G(x, v). \quad (34)$$

Note that the inner products (and the related norm) are defined on $\mathcal{H} = \mathbb{R}^{\infty} \times \mathbb{R}$ in the standard way, i.e. for $a = (a_i)_{i=1}^{\infty}$, $b = (b_i)_{i=1}^{\infty}$, we have that $\langle a, b \rangle = \sum_{i=1}^{\infty} a_i b_i$. The stated assumption in $\nabla_z \Phi(z, \theta)$ guarantees that the inner products appearing in the density will converge. Otherwise, since $x$ is fixed under $\Xi_{h/2}$, the result is simply the limit of the finite-dimensional densities of two Gaussian laws with identity covariance matrices and different means, see Proposition 2.20 of [12].
Thus, we have that:

\[
\frac{dQ^{(i)}_0}{dQ_0}(x_i, v_i) = \frac{d\{Q^{(i-1)} \circ \Xi^{-1}_{h/2} \circ \tilde{\Xi}^{-1} \circ \Xi^{-1}_{h/2}\}(x_i, v_i)}{dQ_0} = \frac{d\{Q^{(i-1)} \circ \Xi^{-1}_{h/2} \circ \tilde{\Xi}^{-1}\}(x_i, v_i)}{dQ_0} \times \frac{d\{Q_0 \circ \Xi^{-1}_{h/2}\}(x_i, v_i)}{dQ_0} = \frac{d\{Q^{(i-1)} \circ \Xi^{-1}_{h/2}\}(\Xi^{-1}_{h/2}(x_i, v_i)) \times G(x_i, v_i)}{dQ_0} \ .
\]  

(35)

Notice now that \(Q_0 \circ \Xi^{-1}_{h/2} \equiv Q_0\), as mapping \(\Xi\) will rotate the infinite-dimensional products of iid standard Gaussian for the \(z, v\)-components of \(Q_0\) (thus will preserve them) and will translate the Lebesque measure for the \(\theta\)-component, thus will also preserve it. Upon noticing also that \((\Xi^{-1}_{h/2} \circ \Xi^{-1}) (x_i, v_i) \equiv \Xi^{-1}_{h/2}(x_{i-1}, v_{i-1})\) we have that:

\[
\frac{d\{Q^{(i-1)} \circ \Xi^{-1}_{h/2} \circ \tilde{\Xi}^{-1}\}(\Xi^{-1}_{h/2}(x_i, v_i))}{dQ_0} = \frac{d\{Q^{(i-1)} \circ \Xi^{-1}_{h/2}\}(\Xi^{-1}_{h/2}(x_{i-1}, v_{i-1}))}{dQ_0} \times \frac{d\{Q_0 \circ \Xi^{-1}_{h/2}\}(\Xi^{-1}_{h/2}(x_{i-1}, v_{i-1}))}{dQ_0} \ .
\]

where for the last equation we have divided and multiplied with \(Q_0 \circ \Xi^{-1}_{h/2}\) as in the calculations in (35) and used again the density in (34). Thus, recalling also the explicit expression for \(\Xi^{-1}_{h/2}\), overall we have shown that:

\[
\frac{dQ^{(i)}_0}{dQ_0}(x_i, v_i) = \frac{dQ^{(i-1)}_0}{dQ_0}(x_{i-1}, v_{i-1}) \cdot G(x_i, v_i) \cdot G(x_{i-1}, v_{i-1} + \frac{h}{2}C^{1/2}g(x_{i-1})) \ .
\]

From this point on one can follow precisely the steps in Section 3.4 of \[4\] to obtain the following identity (for \(L = C^{-1}\)):

\[
\log\{G(x_i, v_i) G(x_{i-1}, v_{i-1} + \frac{h}{2}C^{1/2}g(x_{i-1}))\} = \frac{1}{2} \langle x_i, LX_i \rangle + \frac{1}{2} \langle v_i, LV_i \rangle - \frac{1}{2} \langle x_{i-1}, LX_{i-1} \rangle - \frac{1}{2} \langle v_{i-1}, LV_{i-1} \rangle \ .
\]

Thus, due to the cancellations upon summing up, we have obtained that:

\[
\frac{dQ^{(i)}_0}{dQ_0}(x_I, v_I) = \exp\{\Delta H(x_0, v_0) - \Phi(x_I)\} \ .
\]  

(36)

This completes the proof of part (i) of the Proposition. Given this, the proof of part (ii) follows precisely as in the proof of Theorem 3.1 in \[4\].

32
References

[1] Y. Aït-Sahalia and R. Kimmel. Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics*, 83(2):413 – 452, 2007.

[2] C. Andrieu, A. Doucet, and R. Holenstein. Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342, 2010.

[3] A. Beskos, Stuart A., and Sanz-Serna J.M. Hybrid Monte-Carlo on Hilbert spaces. *Stochastic Process. Appl.*, 121:2201–2230, 2011.

[4] A. Beskos, K. Kalogeropoulos, and E. Pazos. Advanced MCMC methods for sampling on diffusion pathspace. *Stochastic Process. Appl.*, 123:1415–1453, 2013.

[5] A. Beskos, N. Pillai, G. Roberts, A. Stuart, and J.M. Sanz-Serna. Optimal tuning of the Hybrid Monte-Carlo algorithm. Technical report, University of Warwick, 2012. To appear in Bernoulli.

[6] P. Cheridito. Arbitrage in fractional brownian motion models. *Finance and Stochastics*, 7(4):533–553, 2003.

[7] S. Chib, M.K. Pitt, and N. Shephard. Likelihood based inference for diffusion driven state space models. Working paper, 2006.

[8] A. Chronopoulou and F.G. Viens. Estimation and pricing under long-memory stochastic volatility. *Annals of Finance*, 8(2-3):379–403, 2012.

[9] A. Chronopoulou and F.G. Viens. Stochastic volatility and option pricing with long-memory in discrete and continuous time. *Quantitative Finance*, 12(4):635–649, 2012.

[10] F. Comte, L. Coutin, and E. Renault. Affine fractional stochastic volatility models. *Annals of Finance*, 8(2-3):337–378, 2012.

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

[12] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.

[13] Z. Ding, C.W.J. Granger, and R.F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83 – 106, 1993.
[14] S. Duane, A.D. Kennedy, B. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Phys. Lett. B*, 195(2):216–222, 1987.

[15] C.J. Geyer. Practical markov chain monte carlo. *Statist. Sci.*, 7:473–483, 1992.

[16] M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(2):123–214, 2011. With discussion and a reply by the authors.

[17] A. Gloter and M. Hoffmann. Stochastic volatility and fractional brownian motion. *Stochastic Processes and their Applications*, 113(1):143 – 172, 2004.

[18] A. Golightly and D. J. Wilkinson. Bayesian inference for nonlinear multivariate diffusion models observed with error. *Comput. Statist. Data Anal.*, 52(3):1674–1693, 2008.

[19] A.C. Harvey. Long memory in stochastic volatility. In *Forecasting Volatility in Financial Markets*, pages 307–320. J.Knight and S. Satchell (eds.) , Butterworth-Haineman, Oxford, 1998.

[20] F. Jay Breidt, N. Crato, and P. de Lima. The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83(12):325 – 348, 1998.

[21] C.S. Jones. The dynamics of stochastic volatility: evidence from underlying and options markets. *J. Econometrics*, 116(1-2):181–224, 2003. Frontiers of financial econometrics and financial engineering.

[22] K. Kalogeropoulos, G.O. Roberts, and P. Dellaportas. Inference for stochastic volatility models using time change transformations. *Annals of Statistics*, 38(2):784–807, 2010.

[23] S. C. Kou and X. Sunney Xie. Generalized langevin equation with fractional gaussian noise: Subdiffusion within a single protein molecule. *Physical Review Letters*, 93(18):180603, 2004.

[24] I. N. Lobato and N. E. Savin. Real and spurious long memory properties of stock market data. *Journal of Business and Economic Statistics*, 16:261–268, 1998.

[25] A. Mijatović and P. Schneider. Empirical asset pricing with nonlinear risk premia. *Journal of Financial Econometrics*, 2010. To appear.

[26] B. L. S. Prakasa Rao. *Statistical inference for fractional diffusion processes*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2010.
[27] G.O. Roberts and O. Stramer. On inference for partial observed nonlinear diffusion models using the metropolis-hastings algorithm. *Biometrika*, 88(3):603–621, 2001.

[28] M. Rosenbaum. Estimation of the volatility persistence in a discretely observed diffusion model. *Stochastic Processes and their Applications*, 118(8):1434 – 1462, 2008.

[29] O. Stramer and M. Bognar. Bayesian inference for irreducible diffusion processes using the pseudo-marginal approach. *Bayesian Analysis*, 6(2):231–258, 2011.

[30] A. Wood and G. Chan. Simulation of stationary Gaussian processes in $[0, 1]^d$. *J. Comput. Graph. Statist.*, 3(4):409–432, 1994.

References

[1] Y. Aït-Sahalia and R. Kimmel. Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics*, 83(2):413 – 452, 2007.

[2] C. Andrieu, A. Doucet, and R. Holenstein. Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342, 2010.

[3] A. Beskos, Stuart A., and Sanz-Serna J.M. Hybrid Monte-Carlo on Hilbert spaces. *Stochastic Process. Appl.*, 121:2201–2230, 2011.

[4] A. Beskos, K. Kalogeropoulos, and E. Pazos. Advanced MCMC methods for sampling on diffusion pathspace. *Stochastic Process. Appl.*, 123:1415–1453, 2013.

[5] A. Beskos, N. Pillai, G. Roberts, A. Stuart, and J.M. Sanz-Serna. Optimal tuning of the Hybrid Monte-Carlo algorithm. Technical report, University of Warwick, 2012. To appear in Bernoulli.

[6] P. Cheridito. Arbitrage in fractional brownian motion models. *Finance and Stochastics*, 7(4):533–553, 2003.

[7] S. Chib, M.K. Pitt, and N. Shephard. Likelihood based inference for diffusion driven state space models. Working paper, 2006.

[8] A. Chronopoulou and F.G. Viens. Estimation and pricing under long-memory stochastic volatility. *Annals of Finance*, 8(2-3):379–403, 2012.
[9] A. Chronopoulou and F.G. Viens. Stochastic volatility and option pricing with long-memory in discrete and continuous time. *Quantitative Finance*, 12(4):635–649, 2012.

[10] F. Comte, L. Coutin, and E. Renault. Affine fractional stochastic volatility models. *Annals of Finance*, 8(2-3):337–378, 2012.

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

[12] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.

[13] Z. Ding, C.W.J. Granger, and R.F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83 – 106, 1993.

[14] S. Duane, A.D. Kennedy, B. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Phys. Lett. B*, 195(2):216–222, 1987.

[15] C.J. Geyer. Practical markov chain monte carlo. *Statist. Sci.*, 7:473–483, 1992.

[16] M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(2):123–214, 2011. With discussion and a reply by the authors.

[17] A. Gloter and M. Hoffmann. Stochastic volatility and fractional brownian motion. *Stochastic Processes and their Applications*, 113(1):143 – 172, 2004.

[18] A. Golightly and D. J. Wilkinson. Bayesian inference for nonlinear multivariate diffusion models observed with error. *Comput. Statist. Data Anal.*, 52(3):1674–1693, 2008.

[19] A.C. Harvey. Long memory in stochastic volatility. In *Forecasting Volatility in Financial Markets*, pages 307–320. J.Knight and S. Satchell (eds.), Butterworth-Haineman, Oxford, 1998.

[20] F. Jay Breidt, N. Crato, and P. de Lima. The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83(12):325 – 348, 1998.

[21] C.S. Jones. The dynamics of stochastic volatility: evidence from underlying and options markets. *J. Econometrics*, 116(1-2):181–224, 2003. Frontiers of financial econometrics and financial engineering.
[22] K. Kalogeropoulos, G.O. Roberts, and P. Dellaportas. Inference for stochastic volatility models using time change transformations. *Annals of Statistics*, 38(2):784–807, 2010.

[23] S. C. Kou and X. Sunney Xie. Generalized langevin equation with fractional gaussian noise: Subdiffusion within a single protein molecule. *Physical Review Letters*, 93(18):180603, 2004.

[24] I. N. Lobato and N. E. Savin. Real and spurious long memory properties of stock market data. *Journal of Business and Economic Statistics*, 16:261–268, 1998.

[25] A. Mijatović and P. Schneider. Empirical asset pricing with nonlinear risk premia. *Journal of Financial Econometrics*, 2010. To appear.

[26] B. L. S. Prakasa Rao. *Statistical inference for fractional diffusion processes*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2010.

[27] G.O. Roberts and O. Stramer. On inference for partial observed nonlinear diffusion models using the metropolis-hastings algorithm. *Biometrika*, 88(3):603–621, 2001.

[28] M. Rosenbaum. Estimation of the volatility persistence in a discretely observed diffusion model. *Stochastic Processes and their Applications*, 118(8):1434 – 1462, 2008.

[29] O. Stramer and M. Bognar. Bayesian inference for irreducible diffusion processes using the pseudo-marginal approach. *Bayesian Analysis*, 6(2):231–258, 2011.

[30] A. Wood and G. Chan. Simulation of stationary Gaussian processes in [0, 1]^d. *J. Comput. Graph. Statist.*, 3(4):409–432, 1994.