Biometrika (2015), **102**, 4, *pp.* 809–827
Printed in Great Britain
doi: 10.1093/biomet/asv051
Advance Access publication 17 November 2015

Bayesian inference for partially observed stochastic differential equations driven by fractional Brownian motion

By A. Beskos

Department of Statistical Science, University College London, 1-19 Torrington Place, London WC1E 7HB, U.K.
a.beskos@ucl.ac.uk

J. Dureau and K. Kalogeropoulos

Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, U.K.
dureau.joseph@gmail.com k.kalogeropoulos@lse.ac.uk

Summary

We consider continuous-time diffusion models driven by fractional Brownian motion. Observations are assumed to possess a nontrivial likelihood given the latent path. Due to the non-Markovian and high-dimensional nature of the latent path, estimating posterior expectations is computationally challenging. We present a reparameterization framework based on the Davies and Harte method for sampling stationary Gaussian processes and use it to construct a Markov chain Monte Carlo algorithm that allows computationally efficient Bayesian inference. The algorithm is based on a version of hybrid Monte Carlo simulation that delivers increased efficiency when used on the high-dimensional latent variables arising in this context. We specify the methodology on a stochastic volatility model, allowing for memory in the volatility increments through a fractional specification. The method is demonstrated on simulated data and on the S&P 500/VIX time series. In the latter case, the posterior distribution favours values of the Hurst parameter smaller than $1/2$, pointing towards medium-range dependence.

Some key words: Bayesian inference; Davies and Harte algorithm; Fractional Brownian motion; Hybrid Monte Carlo algorithm.

1. Introduction

A natural continuous-time modelling framework for processes with memory uses fractional Brownian motion as the driving noise. This is a zero-mean self-similar Gaussian process, say $B^H = \{B_t^H,\ t \geqslant 0\}$, of covariance $E(B_s^H B_t^H) = (|t|^{2H} + |s|^{2H} - |t-s|^{2H})/2$ for $0 \leqslant s \leqslant t$, parameterized by the Hurst index $H \in (0, 1)$. For $H = 1/2$ we get Brownian motion with independent increments. The $H > 1/2$ case gives smoother paths of infinite variation with positively autocorrelated increments that exhibit long-range dependence, in the sense that the autocorrelations are not summable. For $H < 1/2$ we obtain rougher paths with negatively autocorrelated increments that exhibit medium-range dependence; the autocorrelations are summable but decay more slowly than the exponential rate characterizing short-range dependence.

Since the pioneering work of Mandelbrot & Van Ness (1968), various applications have used fractional noise in models to capture self-similarity, non-Markovianity, or subdiffusivity and superdiffusivity; see, for example, Kou (2008). Closer to our context, numerous studies have

explored the well-posedness of stochastic differential equations driven by B^H,

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t^H \tag{1}$$

for given functions b and σ; see Biagini et al. (2008) and references therein. Unlike most inference methods for models based on (1) in nonlinear settings, which have considered direct and high-frequency observations on X_t (Prakasa Rao, 2010), the focus of this paper is on the partial observation setting. We provide a general framework, which is suitable for incorporating information from additional data sources and potentially from different time scales. The aim is to perform full Bayesian inference for all parameters, including H. The Markov chain Monte Carlo algorithm we develop is relevant in contexts where the observations Y have a nontrivial likelihood, say $p(Y \mid B^H)$, conditionally on the driving noise. We assume that $p(Y \mid B^H)$ is known and genuinely a function of the infinite-dimensional latent path B^H, i.e., we cannot marginalize the model onto finite dimensions. While the focus is on a scalar context, the method applies in principle to several latent processes at greater computational cost, for instance with likelihoods $p(Y \mid B_i^{H_i},\ i = 1, \ldots, \kappa)$ for Hurst parameters H_i $(i = 1, \ldots, \kappa)$.

A first problem in this set-up is the intractability of the likelihood function,

$$p(Y \mid \theta) = \int p(Y \mid X, \theta)\, p(\mathrm{d}X \mid \theta),$$

where $\theta \in \mathbb{R}^q$ represents all the unknown parameters. A data-augmentation approach is adopted, to obtain samples from the joint posterior density

$$\Pi(X, \theta \mid Y) \propto p(Y \mid X, \theta)\, p(X \mid \theta)\, p(\theta).$$

In practice, a time-discretized version of the infinite-dimensional path X must be considered, on a time grid of size N. It is essential to construct an algorithm that has stable performance as N gets large, giving accurate approximation of the theoretical posterior $p(\theta \mid Y)$.

For the standard case of $H = 1/2$, data-augmentation algorithms with mixing time independent of N are available (Roberts & Stramer, 2001; Golightly & Wilkinson, 2008; Kalogeropoulos et al., 2010). However, important challenges arise when $H \neq 1/2$. First, some parameters, including H, can be fully identified by a continuous path of X (Prakasa Rao, 2010), as the joint law of $\{X, H\}$ is degenerate, with $p(H \mid X)$ being a Dirac measure. To avoid slow mixing, the algorithm must decouple this dependence. This decoupling can in general be achieved by suitable reparameterization, see the above references for $H = 1/2$, or by a particle algorithm (Andrieu et al., 2010). In the present setting, the latter approach would require a sequential-in-time realization of B^H paths of cost $O(N^2)$ via the Hosking (1984) algorithm or approximate algorithms of lower cost (Norros et al., 1999). Such a method would then face further computational challenges, such as overcoming path degeneracy and producing unbiased likelihood estimators of small variance. The method developed in this paper is tailored to the particular structure of the models of interest, that of a change of measure from a Gaussian law in high dimensions. Second, typical algorithms for $H = 1/2$ make use of the Markovianity of X. They exploit the fact that given Y, the X-path can be split into small blocks of time with updates on each block involving computations only over its associated time period. For $H \neq 1/2$, X is not Markovian, so a similar block update requires calculations over its complete path. Hence, a potentially efficient algorithm should aim to update large blocks.

In this paper, these issues are addressed in order to develop an effective Markov chain Monte Carlo algorithm. The first issue is tackled via a reparameterization provided by the Davies and

Harte construction of B^H. For the second issue, we resort to a version of the hybrid Monte Carlo algorithm (Duane et al., 1987), adopting ideas from Beskos et al. (2011, 2013a). This algorithm has mesh-free mixing time, and so is particularly appropriate for large N.

The method is applied to a class of stochastic volatility models of importance in finance and econometrics. Use of memory in the volatility is motivated by empirical evidence (Ding et al., 1993; Lobato & Savin, 1998). The autocorrelation function of squared returns is often observed to be slowly decaying towards zero, not in an exponential manner that would suggest short-range dependence, nor implying a unit root that would point to integrated processes. In discrete time, such effects can be captured, for example, with the long-memory stochastic volatility model of Breidt et al. (1998), where the log-volatility is a fractional autoregressive integrated moving average process. In continuous time, Comte & Renault (1998) introduced the model

$$\mathrm{d}S_t = \mu S_t\,\mathrm{d}t + \sigma_S(X_t)\,S_t\,\mathrm{d}W_t, \quad S_0 > 0, \tag{2}$$

$$\mathrm{d}X_t = b_X(X_t, \zeta)\,\mathrm{d}t + \sigma_X(X_t, \zeta)\,\mathrm{d}B_t^H, \quad X_0 = x_0 \in \mathbb{R}, \quad 0 \leqslant t \leqslant \ell. \tag{3}$$

Here, S_t and X_t are the asset price and volatility processes, respectively, and W is standard Brownian motion that is independent of B^H. The definition also involves the length $\ell > 0$ of the time period under consideration, as well as functions $\sigma_S : \mathbb{R} \to \mathbb{R}$, $b_X : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}$ and $\sigma_X : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}$, together with unknown parameters $\mu \in \mathbb{R}$ and $\zeta \in \mathbb{R}^p$ ($p \geqslant 1$). In Comte & Renault (1998), the log-volatility is a fractional Ornstein–Uhlenbeck process, with $H > 1/2$, and the paper argues that incorporating long memory in this way captures the empirically observed strong smile effect for long maturity times. In contrast with previous work, we consider the extended model that allows $H < 1/2$, and we show in § 4 that evidence from data points towards medium-range dependence, $H < 1/2$, in the volatility of the S&P 500 index.

In the setting of (2) and (3), partial observations over X correspond to direct observations from the price process S; that is, for times $0 < t_1 < \cdots < t_n = \ell$ for some $n \geqslant 1$, we have

$$Y_k = \log S_{t_k} \quad (k = 1, \ldots, n), \quad Y = \{Y_1, \ldots, Y_n\}. \tag{4}$$

Given Y, we aim to make inference for all parameters $\theta = (\mu, \zeta, H, x_0)$ in our model. Inference methods available in this partial observation setting are limited. Comte & Renault (1998) and Comte et al. (2012) extracted information on the spot volatility from the quadratic variation of the price, and subsequently used it to estimate θ. Rosenbaum (2008) linked the squared increments of the observed price process to the volatility and constructed a wavelet estimator of H. A common feature of these approaches, and of related ones (Gloter & Hoffmann, 2004), is that they require high-frequency observations. In principle, the method of Chronopoulou & Viens (2012a,b) operates on data of any frequency and estimates H in a non-likelihood manner by calibrating estimated option prices over a grid of values of H against observed market prices. In this paper we develop a computational framework for performing full Bayesian inference based on data augmentation. Our approach is applicable even to low-frequency data. Consistency results for high-frequency asymptotics in a stochastic volatility setting point to slow convergence rates of estimators of H (Rosenbaum, 2008). In our case, we rely on the likelihood to retrieve maximal information from the data at hand, so our method could contribute to developing a better empirical understanding of the amount of such information, strong or weak.

Our algorithm has the following characteristics. First, the computational cost per algorithmic step is $O(N \log N)$. Second, the algorithmic mixing time is mesh-free, $O(1)$, with respect to N; that is, reducing the discretization error will not worsen the convergence properties, since the algorithm is well-defined even when considering the complete infinite-dimensional latent

path X. Third, the algorithm decouples the full dependence between X and H. Finally, it is based on a version of hybrid Monte Carlo simulation, employing Hamiltonian dynamics to allow large steps in the state space while treating big blocks of X. In examples the whole of the X-path and parameter θ are updated simultaneously.

Markov chain Monte Carlo methods with mesh-free mixing times for distributions that are changes of measures from Gaussian laws in infinite dimensions have already appeared (Cotter et al., 2013), with the closest references for hybrid Monte Carlo being Beskos et al. (2011, 2013a). A contribution of the present work is to assemble various techniques, including the Davies and Harte reparameterization to re-express the latent-path part of the posterior as a change of measure from an infinite-dimensional Gaussian law, a version of hybrid Monte Carlo which is particularly effective when run on the contrived infinite-dimensional latent-path space, and a careful joint update procedure for the path and parameters, enforcing $O(N \log N)$ costs for the complete algorithm.

2. DAVIES AND HARTE SAMPLING AND REPARAMETERIZATION

2·1. *Fractional Brownian motion sampling*

Our Monte Carlo algorithm considers the driving fractional noise on a grid of discrete times. We use the Davies and Harte method, sometimes called the circulant method, to construct $\{B_t^H,\ 0 \leqslant t \leqslant \ell\}$ on the regular grid $\{\delta, 2\delta, \ldots, N\delta\}$ for some $N \geqslant 1$ and mesh size $\delta = \ell/N$. The algorithm samples the grid points via a linear transform from independent standard Gaussians. This transform will be used in § 2·2 to decouple the latent variables from the Hurst parameter H. The computational cost is $O(N \log N)$ owing to use of the fast Fourier transform. The method is based on the stationarity of the increments of fractional Brownian motion on the regular grid and, in particular, exploits the Toeplitz structure of the covariance matrix of the increments; see Wood & Chan (1994) for a complete description.

We briefly describe the Davies and Harte method, following Wood & Chan (1994). We define the $2N \times 2N$ unitary matrix P with elements $P_{jk} = (2N)^{-1/2} \exp\{-\pi\, \mathrm{i} jk/N\}$ $(0 \leqslant j, k \leqslant 2N - 1)$, where $\mathrm{i}^2 = -1$. Consider also the $2N \times 2N$ matrix

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

with the $N \times N$ submatrices defined as follows: $Q_{11} = \mathrm{diag}(1, 2^{-1/2}, \ldots, 2^{-1/2})$; $Q_{12} = (q_{ij})$ where $q_{i,i-1} = 2^{-1/2}$ for $i = 1, \ldots, N-1$ and otherwise $q_{ij} = 0$; $Q_{21} = (q_{ij})$ where $q_{i,N-i} = 2^{-1/2}$ for $i = 1, \ldots, N-1$ and otherwise $q_{ij} = 0$; $Q_{22} = \mathrm{diag}_{\mathrm{inv}}(1, -\mathrm{i}\, 2^{-1/2}, \ldots, -\mathrm{i}\, 2^{-1/2})$, where $\mathrm{diag}_{\mathrm{inv}}$ denotes a matrix with nonzero entries on the inverse diagonal. We define the diagonal matrix $\Lambda_H = \mathrm{diag}(\lambda_0, \lambda_1, \ldots, \lambda_{2N-1})$ with the values

$$\lambda_k = \sum_{j=0}^{2N-1} c_j \exp(-\pi\, \mathrm{i} jk/N) \quad (k = 0, \ldots, 2N-1).$$

Here $(c_0, c_1, \ldots, c_{2N-1}) = \{g(0), g(1), \ldots, g(N-1), 0, g(N-1), \ldots, g(1)\}$, where $g(k)$ denotes the covariance of increments of B^H of lag $k = 0, 1, \ldots$, i.e.,

$$g(k) = E\{B_1^H (B_{k+1}^H - B_k^H)\} = \tfrac{1}{2}|k+1|^{2H} + \tfrac{1}{2}|k-1|^{2H} - |k|^{2H}.$$

The definition of the c_j $(j = 0, 1, \ldots, 2N-1)$ implies that the λ_k are all real numbers. The Davies and Harte method for generating B^H is shown in Algorithm 1. Finding QZ costs $O(N)$. Finding Λ_H and then calculating $P\Lambda_H^{1/2}QZ$ costs $O(N \log N)$ due to a fast Fourier transform. Separate approaches show that the λ_k are nonnegative for any $H \in (0, 1)$, and hence $\Lambda_H^{1/2}$ is well-posed (Craigmile, 2003). There are several other ways to sample a fractional Brownian motion; see, for instance, the 2004 University of Twente MSc thesis of T. Dieker. However, the Davies and Harte method is, to the best of our knowledge, the fastest exact method on a regular grid and boils down to a simple linear transform that can be easily differentiated, which is needed for our method.

Algorithm 1. Simulation of stationary increments $\left(B_\delta^H, B_{2\delta}^H - B_\delta^H, \ldots, B_{N\delta}^H - B_{(N-1)\delta}^H\right)$.

(i) Sample $Z \sim N(0, I_{2N})$.
(ii) Calculate $Z' = \delta^H P \Lambda_H^{1/2} QZ$.
(iii) Return the first N elements of Z'.

2·2. *Reparameterization*

Algorithm 1 gives rise to a linear mapping $Z \mapsto (B_\delta^H, \ldots, B_{N\delta}^H)$ to generate B^H on a regular grid of size N from $2N$ independent standard Gaussian variables. Thus, the latent variable principle described in § 1 is implemented using the vector Z, a priori independent of H, rather than using the solution X of (1). Indeed, we work with the joint posterior of (Z, θ), which has a density with respect to $\bigotimes_{i=1}^{2N} N(0, 1) \otimes \mathrm{Leb}_q$, namely the product of $2N$ standard Gaussian laws and the q-dimensional Lebesgue measure. Analytically, the posterior distribution Π_N for (Z, θ) is specified as follows:

$$\frac{\mathrm{d}\Pi_N}{\mathrm{d}\{\bigotimes_{i=1}^{2N} N(0, 1) \times \mathrm{Leb}_q\}}(Z, \theta \mid Y) \propto p(\theta)\, p_N(Y \mid Z, \theta). \tag{5}$$

In (5) and the following, the subscript N emphasizes the finite-dimensional approximations due to using an N-dimensional proxy for the infinite-dimensional path X. Some care is needed here, as standard Euler schemes may not converge when used to approximate stochastic integrals driven by fractional Brownian motion. We explain this in § 2·3 and detail the numerical scheme in the Supplementary Material. The target density can be written as

$$\Pi_N(Z, \theta) \propto \exp\left\{-\tfrac{1}{2}\langle Z, Z\rangle - \Phi(Z, \theta)\right\} \tag{6}$$

where, in agreement with (5), we have defined

$$\Phi(Z, \theta) = -\log p(\theta) - \log p_N(Y \mid Z, \theta). \tag{7}$$

In § 3 we describe an efficient Markov chain Monte Carlo sampler tailored to sampling (6).

2·3. *Diffusions driven by fractional Brownian motion*

An extensive literature exists on the stochastic differential equation (1) and its nonscalar extensions, involving various definitions of stochastic integration with respect to B^H and ways of determining a solution; see Biagini et al. (2008). For scalar B^H, the Doss–Sussmann representation (Sussmann, 1978) provides the simplest framework for interpreting (1) for all $H \in (0, 1)$. It involves a pathwise approach, whereby for any $t \mapsto B_t^H(\omega)$ one obtains a solution of the

differential equation for all continuously differentiable paths in a neighbourhood of $B_\cdot^H(\omega)$ and considers the value of this mapping at $B_\cdot^H(\omega)$. Conveniently, the solution found in this way follows the rules of standard calculus and coincides with the Stratonovich representation when $H = 1/2$.

The numerical solution of fractional stochastic differential equations is a topic of intensive research (Mishura, 2008). As shown in a 2013 Harvard University technical report by M. Lysy and N. S. Pillai, care is needed because a standard Euler scheme applied to B^H-driven multiplicative stochastic integrals may diverge to infinity for $H < 1/2$. When allowing $H < 1/2$, we must restrict attention to a particular family of models to get a practical method. For the stochastic volatility class in (2) and (3), we can assume a Sussmann solution for the volatility equation (1). In order to get the corresponding numerical scheme, one can follow the approach in the 2013 technical report by Lysy and Pillai and use the Lamperti transform,

$$F_t = \int^{X_t} \sigma_X^{-1}(u, \zeta)\, \mathrm{d}u,$$

so that F_t has additive noise. A standard Euler scheme for F_t will then converge to the analytical solution in an appropriate mode, under regularity conditions. This approach can in principle be followed for general models with a scalar differential equation and driving noise B^H. The price process differential equation (2) is then interpreted in the usual Itô way. In § 4 we will extend the model in (2) and (3) to allow for a leverage effect. In that case, the likelihood $p(Y \mid B_H)$ will involve a multiplicative stochastic integral over B_H. Due to the particular structure of this class of models, the integral can be replaced with a Riemannian one, allowing the use of a standard finite difference approximation scheme. The Supplementary Material details the numerical method used in the applications. For multi-dimensional models one cannot avoid multiplicative stochastic integrals. For $H > 1/2$ there is a well-defined framework for the numerical approximation of multiplicative stochastic integrals driven by B^H; see an unpublished 2013 University of Kansas manuscript available from Y. Hu. For $1/3 < H < 1/2$ one can use a Milstein-type scheme, and third-order schemes are required for $1/4 < H \leqslant 1/3$ (Deya et al., 2012).

3. AN EFFICIENT MARKOV CHAIN MONTE CARLO SAMPLER

3·1. *Standard hybrid Monte Carlo algorithm*

We use a hybrid Monte Carlo algorithm to explore the posterior of Z and θ in (6). The standard method was introduced by Duane et al. (1987), but we employ an advanced version, tailored to the structure of the distributions of interest and closely related to algorithms developed in Beskos et al. (2011, 2013a) for effective sampling of changes of measures from Gaussian laws in infinite dimensions. First we briefly describe the standard algorithm.

The state space is extended via the velocity $v = (v_z, v_\theta) \in \mathbb{R}^{2N+q}$. The original arguments $x = (z, \theta) \in \mathbb{R}^{2N+q}$ can be thought of as location. The total energy function is, for Φ in (7),

$$H(x, v; M) = \Phi(x) + \tfrac{1}{2}\langle z, z\rangle + \tfrac{1}{2}\langle v, Mv\rangle, \tag{8}$$

with a user-specified positive-definite mass matrix M, involving the potential $\Phi(x) + \langle z, z\rangle/2$ and kinetic energy $\langle v, Mv\rangle/2$. Hamiltonian dynamics on $\mathbb{R}^{2N+q}$ express conservation of energy and are defined via the system of differential equations $\mathrm{d}x/\mathrm{d}t = M^{-1}(\partial H/\partial v)$, $M(\mathrm{d}v/\mathrm{d}t) = -\partial H/\partial x$, which, in the context of (8), become $\mathrm{d}x/\mathrm{d}t = v$, $M(\mathrm{d}v/\mathrm{d}t) = -(z, 0)^{\mathrm{T}} - \nabla\Phi(x)$. In general, a good choice of M would resemble the inverse covariance of the target $\Pi_N(x)$. In our

context, guided by the prior structure of (z, θ), we set

$$M = \begin{pmatrix} I_{2N} & 0 \\ 0 & A \end{pmatrix}, \quad A = \mathrm{diag}(a_i, \ldots, a_q) \tag{9}$$

and rewrite the Hamiltonian equations as

$$\frac{\mathrm{d}x}{\mathrm{d}t} = v, \quad \frac{\mathrm{d}v}{\mathrm{d}t} = -\begin{pmatrix} z \\ 0 \end{pmatrix} - M^{-1}\nabla\Phi(x). \tag{10}$$

The standard hybrid Monte Carlo algorithm discretizes (10) via a leapfrog scheme, i.e.,

$$\begin{aligned} v_{h/2} &= v_0 - \frac{h}{2}(z_0, 0)^{\mathrm{T}} - \frac{h}{2} M^{-1}\nabla\Phi(x_0), \\ x_h &= x_0 + h v_{h/2}, \\ v_h &= v_{h/2} - \frac{h}{2}(z_h, 0)^{\mathrm{T}} - \frac{h}{2} M^{-1}\nabla\Phi(x_h), \end{aligned} \tag{11}$$

where $h > 0$. Scheme (11) gives rise to the operator $(x_0, v_0) \mapsto \psi_h(x_0, v_0) = (x_h, v_h)$. The sampler runs up to a time horizon $T > 0$ via the synthesis of $I = \lfloor T/h \rfloor$ leapfrog steps, so we define ψ_h^I to be the synthesis of I mappings ψ_h. The dynamics in (10) preserve the total energy and are invariant for the density $\exp\{-H(x, v; M)\}$, but their discretized version requires an accept/reject correction. The full method is shown in Algorithm 2, with $\mathcal{P}_x$ denoting the projection on x. The proof that Algorithm 2 gives a Markov chain which preserves $\Pi_N(x)$ is based on ψ_h^I being volume-preserving and having the symmetricity property $\psi_h^I(x_I, -v_I) = (x_0, -v_0)$, as with the exact solver of the Hamiltonian equations; see, for example, Duane et al. (1987). For $\alpha, \beta > 0$ we denote by $\alpha \wedge \beta$ their minimum.

Algorithm 2. Standard hybrid Monte Carlo algorithm, with target $\Pi_N(x) = \Pi_N(Z, \theta)$ in (6).

(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 $x^\star = \mathcal{P}_x\, \psi_h^I(x^{(k)}, v^{(k)})$.
(iii) Calculate $a = 1 \wedge \exp[H(x^{(k)}, v^{(k)}; M) - H\{\psi_h^I(x^{(k)}, v^{(k)}); M\}]$.
(iv) Set $x^{(k+1)} = x^\star$ with probability a; otherwise set $x^{(k+1)} = x^{(k)}$.
(v) Set $k \to k + 1$ and go to (ii).

Remark 1. The index t of the Hamiltonian equations must not be confused with the index t of the diffusion processes in the models of interest. When applied here, each hybrid Monte Carlo step updates a complete sample path, so the t-index for paths can be regarded as a space direction.

3·2. *Advanced hybrid Monte Carlo algorithm*

Algorithm 2 provides an inappropriate proposal $x^\star$ for increasing N (Beskos et al., 2011), with the acceptance probability approaching 0, when h and T are fixed. Beskos et al. (2013b) suggested that controlling the acceptance probability requires a step size of $h = O(N^{-1/4})$. Advanced hybrid Monte Carlo simulation avoids this degeneracy by employing a modified leapfrog scheme that yields better performance in high dimensions.

Remark 2. Choice of the mass matrix M as in (9) is critical for the final algorithm. Our choosing I_{2N} for the upper-left block of M is motivated by the prior for Z. We will see in § 3·3

that this choice also ensures the well-posedness of the algorithm as $N \to \infty$. A posteriori, we have found that the information in the data spreads fairly uniformly over $Z_1, \ldots, Z_{2N}$, so I_{2N} seems a sensible choice also under the posterior distribution. For the choice of the diagonal A, in the numerical work we used the inverse of the marginal posterior variances of θ as estimated by preliminary runs. More automated choices could involve adaptive Markov chain Monte Carlo or Riemannian manifold approaches (Girolami & Calderhead, 2011) using the Fisher information.

Remark 3. The development below is closely related to the approach taken by Beskos et al. (2011), who illustrated the mesh-free mixing property of the algorithm in the context of distributions of diffusion paths driven by Brownian motion. In this paper, the algorithm is extended to treat also the model parameters and the different set-up with a product of standard Gaussians as the high-dimensional Gaussian reference measure.

We develop the method as follows. The Hamiltonian equations (10) are now split into two parts:

$$\mathrm{d}x/\mathrm{d}t = 0, \quad \mathrm{d}v/\mathrm{d}t = -M^{-1}\nabla\Phi(x), \tag{12}$$

$$\mathrm{d}x/\mathrm{d}t = v, \quad \mathrm{d}v/\mathrm{d}t = -(z, 0)^{\mathrm{T}}, \tag{13}$$

where the ordinary differential equations (12) and (13) can both be solved analytically. We obtain a numerical integrator for (10) by synthesizing the steps of (12) and (13). We define the solution operators of (12) and (13) to be

$$\Xi_t(x, v) = \left\{x,\ v - tM^{-1}\nabla\Phi(x)\right\}, \tag{14}$$

$$\tilde{\Xi}_t(x, v) = \left[\{\cos(t)\, z + \sin(t)\, v_z,\ \theta + tv_\theta\},\ \{-\sin(t)\, z + \cos(t)\, v_z,\ v_\theta\}\right]. \tag{15}$$

The numerical integrator for (10) is

$$\Psi_h = \Xi_{h/2} \circ \tilde{\Xi}_h \circ \Xi_{h/2} \tag{16}$$

for small $h > 0$. As with the standard hybrid Monte Carlo algorithm, we synthesize $I = \lfloor T/h \rfloor$ leapfrog steps Ψ_h and denote the complete mapping by Ψ_h^I. Notice that Ψ_h is volume-preserving and that, for $(x_h, v_h) = \Psi_h(x_0, v_0)$, the symmetricity property $\Psi_h(x_h, -v_h) = (x_0, -v_0)$ holds. Owing to these properties, the acceptance probability has the same expression as for the standard hybrid Monte Carlo algorithm. The full method is shown in Algorithm 3.

Algorithm 3. Advanced hybrid Monte Carlo algorithm, with target $\Pi_N(x) = \Pi_N(Z, \theta)$ in (6).

(i) Start with an initial value $x^{(0)} \sim \bigotimes_{i=1}^{2N} N(0, 1) \times p(\theta)$ and set $k = 0$.
(ii) Given $x^{(k)}$, sample $v^{(k)} \sim N(0, M^{-1})$ and propose $x^\star = \mathcal{P}_x\, \Psi_h^I(x^{(k)}, v^{(k)})$.
(iii) Calculate $a = 1 \wedge \exp[H(x^{(k)}, v^{(k)}; M) - H\{\Psi_h^I(x^{(k)}, v^{(k)}); M\}]$.
(iv) Set $x^{(k+1)} = x^\star$ with probability a; otherwise set $x^{(k+1)} = x^{(k)}$.
(v) Set $k \to k + 1$ and go to (ii).

3·3. *Advanced hybrid Monte Carlo algorithm with $N \to \infty$.*

An important property of the advanced method is its mesh-free mixing time. As N increases while h and T are held fixed, the convergence/mixing properties of the Markov chain do not deteriorate. To illustrate this, we show that there is a well-defined algorithm in the limit as $N \to \infty$.

Remark 4. We follow closely the arguments in Beskos et al. (2013a), with the differences discussed in Remark 3. We include here a proof of the well-posedness of the advanced hybrid Monte Carlo algorithm in the case where the state space is infinite-dimensional, as it cannot be deduced directly from Beskos et al. (2013a). The proof provides insight into the algorithm, for instance highlighting those aspects that lead to mesh-free mixing.

Let ∇_z denote the vector of partial derivatives over the z-component, so that $\nabla_x = (\nabla_z, \nabla_\theta)^{\mathrm{T}}$. Here $z \in \mathbb{R}^\infty$, and the distribution of interest corresponds to Π_N in (6) as $N \to \infty$, denoted by Π and defined on the infinite-dimensional space $\mathcal{H} = \mathbb{R}^\infty \times \mathbb{R}^q$ via the change of measure

$$\frac{\mathrm{d}\Pi}{\mathrm{d}\{\bigotimes_{i=1}^{\infty} N(0,1) \times \mathrm{Leb}_q\}}(Z, \theta \mid Y) \propto \exp\{-\Phi(Z, \theta)\} \tag{17}$$

for a function $\Phi : \mathcal{H} \to \mathbb{R}$. We also need the vector of partial derivatives $\nabla\Phi : \mathcal{H} \to \mathcal{H}$. We have the velocity $v = (v_z, v_\theta) \in \mathcal{H}$, and the matrix M, specified in (9) for finite dimensions, has the infinite-dimensional identity matrix I_∞ in its upper-left block instead of I_{2N}; that is, $M : \mathcal{H} \to \mathcal{H}$ is the linear operator $(z, \theta)^{\mathrm{T}} \mapsto M(z, \theta)^{\mathrm{T}} = (z, A\theta)^{\mathrm{T}}$. Accordingly, $\Xi_{h/2}$, $\tilde{\Xi}_h$, $\Psi_h : \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$ are defined as in (14)–(16) with domain and range of values over the infinite-dimensional vector space $\mathcal{H}$.

We consider the joint location-velocity law on (x, v), $Q(\mathrm{d}x, \mathrm{d}v) = \Pi(\mathrm{d}x) \otimes N(0, M^{-1})(\mathrm{d}v)$. The main idea is that Ψ_h in (16) projects $(x_0, v_0) \sim Q$ to (x_h, v_h) having a distribution absolutely continuous with respect to Q, an attribute that implies existence of a nonzero acceptance probability when $N = \infty$, under conditions on $\nabla\Phi$. This is clear for $\tilde{\Xi}_h$ in (15), as it applies 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 preserves absolute continuity of $Q(\mathrm{d}x, \mathrm{d}v)$. Then, for step $\Xi_{h/2}$ in (14), the gradient $\nabla_z\Phi(z, \theta)$ must lie in the so-called Cameron–Martin space of $\prod_{i=1}^{\infty} N(0,1)$ for the translation $v \mapsto v - (h/2)M^{-1}\nabla\Phi(x)$ to preserve absolute continuity of the v-marginal $Q(\mathrm{d}v)$. This Cameron–Martin space is the space of squared summable infinite vectors, which we denote by ℓ_2 (Da Prato & Zabczyk, 1992, ch. 2). In contrast, for the standard hybrid Monte Carlo algorithm one can consider even the case of $\Phi(x)$ being a constant, so that $\nabla\Phi \equiv 0$, to see that, immediately from the first step in the leapfrog update in (11), an input sample from the target Q gets projected to a variable that has singular law with respect to Q when $N = \infty$, and therefore has zero acceptance probability.

For a rigorous result, we first define a reference measure on the (x, v) space,

$$Q_0 = Q_0(\mathrm{d}x, \mathrm{d}v) = \left\{ \prod_{i=1}^{\infty} N(0,1) \otimes \mathrm{Leb}_q \right\}(\mathrm{d}x) \otimes N(0, M^{-1})(\mathrm{d}v),$$

so that the joint target is $Q(\mathrm{d}x, \mathrm{d}v) \propto \exp\{-\Phi(x)\}\, Q_0(\mathrm{d}x, \mathrm{d}v)$. We also consider the sequence of probability measures on $\mathcal{H} \times \mathcal{H}$ defined by $Q^{(i)} = Q \circ \Psi_h^{-i}$ $(i = 1, \ldots, I)$, corresponding to the push-forward projection of Q via the leapfrog steps. For given (x_0, v_0), we write $(x_i, v_i) = \Psi_h^i(x_0, v_0)$. The difference in energy $\Delta H(x_0, v_0)$ appearing in the statement of Proposition 1 below is still defined as $\Delta H(x_0, v_0) = H(x_I, v_I; M) - H(x_0, v_0; M)$ for the energy function in (8), with the obvious extension to $\mathbb{R}^\infty$ of the inner product involved. Even if $H(x_0, v_0; M) = \infty$ with probability 1, the difference $\Delta H(x_0, v_0)$ does not explode, as implied by the analytic expression for $\Delta H(x_0, v_0)$ given in the proof of Proposition 1 in the Appendix. We denote the indicator function by $\mathbb{I}$, so that $\mathbb{I}_E = 1$ if a given statement E is true and 0 otherwise.

PROPOSITION 1. *Assume that* $\nabla_z \Phi(z, \theta) \in \ell_2$ *almost surely under* $\prod_{i=1}^{\infty} N(0, 1) \otimes p(\mathrm{d}\theta)$. *Then:*

(i) $Q^{(I)}$ *is absolutely continuous with respect to* Q_0, *with probability density*

$$\frac{\mathrm{d}Q^{(I)}}{\mathrm{d}Q_0}(x_I, v_I) = \exp\{\Delta H(x_0, v_0) - \Phi(x_I)\};$$

(ii) *the Markov chain with transition dynamics for current position* $x_0 \in \mathcal{H}$,

$$x' = \mathbb{I}_{U \leqslant a(x_0, v_0)}\, x_I + \mathbb{I}_{U > a(x_0, v_0)} x_0,$$

where $U \sim \mathrm{Un}[0, 1]$ *and the noise is* $v_0 \sim \prod_{i=1}^{\infty} N(0, 1) \otimes N_q(0, A^{-1})$, *has invariant distribution* $\Pi(\mathrm{d}x)$ *as in* (17).

The proof is given in the Appendix.

Remark 5. The condition $\nabla_z \Phi(z, \theta) \in \ell_2$ relates to the fact that the data have a finite amount of information about Z, so the sensitivity of the likelihood for each individual Z_i can be small for large N. We have not pursued this further analytically, as Proposition 1 already highlights the structurally important mesh-free property of the method.

4. FRACTIONAL STOCHASTIC VOLATILITY MODELS

4·1. *Data and model*

To illustrate the application of Algorithm 3, we return to the fractional stochastic volatility models. Starting from (2) and (3), we henceforth work with $U_t = \log(S_t)$ and use Itô's formula to rewrite the equations in terms of U_t and X_t. We also extend the model to allow correlation between $\mathrm{d}U_t$ and $\mathrm{d}X_t$:

$$\begin{aligned} \mathrm{d}U_t &= \left\{\mu - \sigma_S(X_t)^2/2\right\} \mathrm{d}t + \sigma_S(X_t) \left\{(1 - \rho^2)^{1/2}\, \mathrm{d}W_t + \rho\, \mathrm{d}B_t^H\right\}, \\ \mathrm{d}X_t &= b_X(X_t, \zeta)\, \mathrm{d}t + \sigma_X(X_t, \zeta)\, \mathrm{d}B_t^H, \quad 0 \leqslant t \leqslant \ell, \end{aligned} \tag{18}$$

for a parameter $\rho \in (-1, 1)$; so henceforth $\theta = (\mu, \zeta, H, \rho, x_0) \in \mathbb{R}^q$ with $q = p + 4$. We set $H \in (0, 1)$, thus allowing for medium-range dependence, in contrast to previous works, which typically restricted attention to $H \in (1/2, 1)$. Given the observations Y from the log-price process in (4), there is a well-defined likelihood $p(Y \mid B^H, \theta)$. Conditionally on the latent driving noise B^H, the log-price process U is Markovian. From the specification of the model, we have that

$$Y_k \mid Y_{k-1}, B^H, \theta \sim N\left\{m_k(B^H, \theta),\ \Sigma_k(B^H, \theta)\right\} \quad (k = 1, \ldots, n), \tag{19}$$

where $Y_0 \equiv U_0$ is assumed fixed, with mean and variance parameters

$$\begin{aligned} m_k(B^H, \theta) &= Y_{k-1} + \int_{t_{k-1}}^{t_k} \left\{\mu - \sigma_S(X_t)^2/2\right\} \mathrm{d}t + \rho \int_{t_{k-1}}^{t_k} \sigma_S(X_t)\, \mathrm{d}B_t^H, \\ \Sigma_k(B^H, \theta) &= (1 - \rho^2) \int_{t_{k-1}}^{t_k} \sigma_S(X_t)^2\, \mathrm{d}t. \end{aligned}$$

From (19), it is trivial to write down a complete expression for the likelihood $p(Y \mid B^H, \theta)$.

Recalling the mapping $Z \mapsto (B^H_\delta, \ldots, B^H_{N\delta})$ from the Davies and Harte method described in § 2, for $N \geqslant 1$ and discretization step $\delta = \ell/N$, the expression for $p(Y \mid B^H, \theta)$ in continuous time will provide an expression for $p_N(Y \mid Z, \theta)$ in discrete time upon consideration of a numerical scheme. In § 2·3 we described the Doss–Sussmann interpretation of the stochastic volatility model; in the Supplementary Material we present in detail the corresponding numerical scheme. Expressions for $p_N(Y \mid Z, \theta)$ and the derivatives $\nabla_Z \log p_N(Y \mid Z, \theta)$ and $\nabla_\theta \log p_N(Y \mid Z, \theta)$ required by the Hamiltonian methods are in the Supplementary Material.

The methodology provided in this paper allows us to handle data from different sources and on different scales, with little additional effort. To illustrate this, we analyse two extended sets of data that contain additional information over just using daily observations of U_t. The first extension treats volatility proxies, constructed from option prices, as direct observations on X_t, as in Aït-Sahalia & Kimmel (2007), Jones (2003) and Stramer & Bognar (2011). Aït-Sahalia & Kimmel (2007) used two proxies from the VIX index. First, they considered a simple unadjusted proxy that uses VIX data to directly obtain $\sigma_S(X_t)$ and therefore X_t. Second, an adjusted integrated volatility proxy is considered, assuming that the pricing measure has a linear drift; see Aït-Sahalia & Kimmel (2007, § 5.1). The integrated volatility proxy was also used by Jones (2003) and Stramer & Bognar (2011) to provide observations of $\sigma_S(X_t)$, where additional measurement error is incorporated into the model. We take the simpler approach and use the unadjusted volatility proxy as a noisy measurement device for $\sigma_S(X_t)$, for two reasons. First, our focus is mainly on exploring the behaviour of our algorithm in a different observation regime, so we want to avoid additional subject-specific considerations, such as assumptions on the pricing measure. Second, the difference between the two approaches is often negligible; see, for example, the simulation experiments in Aït-Sahalia & Kimmel (2007) for the Heston model. The approaches of Jones (2003) and Stramer & Bognar (2011) can still be incorporated into our framework. More generally, the combination of option and asset prices must be investigated further even in the context of standard Brownian motion.

Following the above discussion, we denote the additional noisy observations from VIX proxies by Y^x_k and assume that they provide information on X_{t_k} via

$$Y^x_k = X_{t_k} + \epsilon_k \quad (k = 1, \ldots, n), \tag{20}$$

where the ϵ_k are independent $N(0, \tau^2)$ variates. We refer to the dataset consisting of observations Y as type A and the dataset consisting of Y and Y^x as type B. The second extension builds on the type B dataset and incorporates intraday observations on Y, thus encompassing two observation frequency regimes; this is referred to as type C.

The parameter τ controls the weight placed on the volatility proxies in order to form a weighted averaged volatility measurement that combines information from asset and option prices. Hence we treat τ as a user-specified parameter. In the following numerical examples, we set $\tau = 0{\cdot}05$ based on estimates from a preliminary run of the full model applied to the S&P 500/VIX time series. In the Supplementary Material we give $p_N(Y \mid Z, \theta)$, $\nabla_Z \log p_N(Y \mid Z, \theta)$ and $\nabla_\theta \log p_N(Y \mid Z, \theta)$ only for the type A case, but it is straightforward to include terms due to the extra data in (20).

4·2. *Illustration on simulated data*

We apply our method to the model of Comte & Renault (1998), also considered in Chronopoulou & Viens (2012a,b), but we further include an extension for correlated noise as

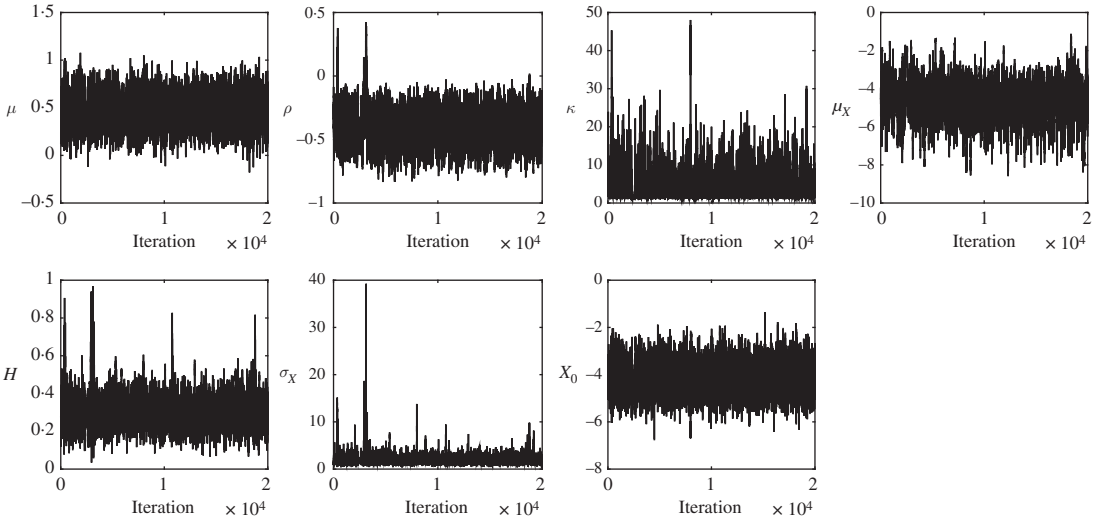

Fig. 1. Traceplots from 2×10^4 iterations of the advanced hybrid Monte Carlo algorithm for dataset Sim-A. True parameter values are as in Table 1 with $H = 0{\cdot}3$. The execution time was about 5 h using Matlab code.

in (18); that is, we have

$$\begin{aligned} \mathrm{d}U_t &= \{\mu - \exp(X_t)/2\}\,\mathrm{d}t + \exp(X_t/2)\left\{(1-\rho^2)^{1/2}\,\mathrm{d}W_t + \rho\,\mathrm{d}B_t^H\right\}, \\ \mathrm{d}X_t &= \kappa(\mu_X - X_t)\,\mathrm{d}t + \sigma_X\,\mathrm{d}B_t^H. \end{aligned} \tag{21}$$

Similar to related work, the model is completed with priors; see, for example, an unpublished 2010 Washington University technical report by S. Chib. The prior for μ_X is normal, with 95% credible interval spanning the range from the minimum to the maximum volatility values over the entire period under consideration. The prior for σ_X^2 is an inverse gamma distribution with shape and scale parameters $\alpha = 2$ and $\beta = \alpha \times 0{\cdot}03 \times 252^{1/2}$. Vague priors are chosen for the remaining parameters: Un(0, 1) and Un(-1, 1) for H and ρ, and $N(0, 10^6)$ for μ.

We first apply Algorithm 3 to simulated data. We generated 250 observations from model (21), corresponding roughly to a year of data. We considered two datasets: Sim-A, with 250 daily observations on S_t only, as in (4); and Sim-B, with additional daily observations on X_t for the same time period, contaminated with measurement error as in (20). We consider $H = 0{\cdot}3$, $0{\cdot}5$ and $0{\cdot}7$, and use a discretization step $\delta = 0{\cdot}1$ for the Euler approximation of the path of Z, resulting in $2N = 2 \times 250 \times 10 = 5000$. The true values of the parameters were chosen to be similar to those in previous analyses of the S&P 500/VIX indices based on standard Markov models (Aït-Sahalia & Kimmel, 2007) and to those we found from the data analysis in § 4·3. The Hamiltonian integration horizon was set to $T = 0{\cdot}9$ for Sim-A and $T = 1{\cdot}5$ for Sim-B. The number of leapfrog steps was tuned, ranging from 10 to 50, to achieve an average acceptance rate between 70% and 80% across the different simulated datasets.

Figures 1 and 2 show traceplots for $H = 0{\cdot}3$; the plots for $H = 0{\cdot}5$ and $H = 0{\cdot}7$ are similar. The mixing of the chain appears to be quite good, considering the complexity of the model. Table 1 shows posterior estimates obtained from running the advanced hybrid Monte Carlo algorithm on datasets Sim-A and Sim-B.

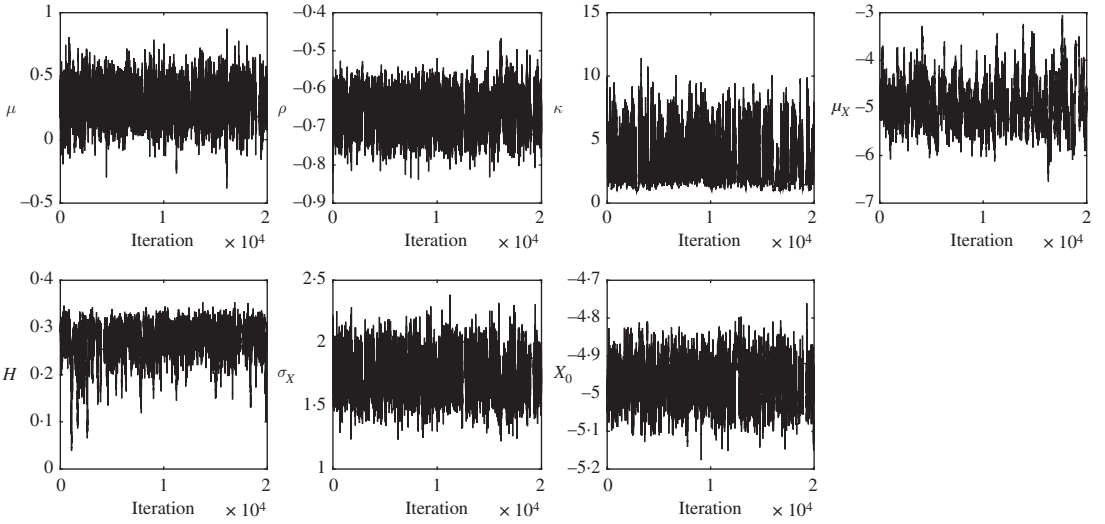

Fig. 2. Traceplots as in Fig. 1 but for dataset Sim-B, with true parameter values as in Table 1 and $H = 0{\cdot}3$. The execution time was about 7 h, due to using 50 leapfrog steps, whereas the algorithm for Sim-A used 30 leapfrog steps.

Table 1. *Posterior summary statistics for model* (21) *for datasets Sim-A and Sim-B; the statistics shown are the estimates of the* 2·5 *and* 97·5 *percentiles, the mean and the median as obtained from the advanced hybrid Monte Carlo algorithm* (*Algorithm* 3)

			Dataset Sim-A				Dataset Sim-B			
Dataset	Parameter	True value	2·5%	97·5%	Mean	Median	2·5%	97·5%	Mean	Median
$H = 0{\cdot}3$	μ	0·25	0·18	0·76	0·46	0·46	0·01	0·55	0·28	0·28
	ρ	−0·75	−0·69	−0·12	−0·40	−0·40	−0·75	−0·57	−0·67	−0·67
	κ	4·00	1·13	12·15	3·79	2·79	1·01	7·40	3·22	2·74
	μ_X	−5·00	−5·62	−3·44	−4·46	−4·42	−5·85	−3·74	−4·95	−4·98
	H	0·30	0·20	0·44	0·30	0·30	0·18	0·32	0·27	0·28
	σ_X	2·00	0·90	3·90	1·95	1·78	1·45	2·07	1·75	1·75
	X_0	−5·00	−5·05	−4·07	−4·59	−4·60	−5·08	−4·87	−4·97	−4·97
$H = 0{\cdot}5$	μ	0·25	0·01	0·99	0·48	0·470	−0·14	0·39	0·14	0·15
	ρ	−0·75	−0·91	−0·13	−0·60	−0·62	−0·88	−0·75	−0·82	−0·82
	κ	4·00	1·33	19·94	7·38	6·24	2·49	6·53	3·96	3·75
	μ_X	−5·00	−5·41	−3·94	−4·83	−4·90	−5·90	−3·87	−4·71	−4·61
	H	0·50	0·29	0·74	0·50	0·49	0·48	0·55	0·52	0·52
	σ_X	2·00	0·83	4·60	2·29	2·14	1·74	2·53	2·10	2·09
	X_0	−5·00	−5·75	−4·56	−5·15	−5·13	−5·04	−4·87	−4·96	−4·96
$H = 0{\cdot}7$	μ	0·25	0·19	0·38	0·28	0·28	−0·09	0·39	0·15	0·14
	ρ	−0·75	−0·78	−0·25	−0·60	−0·62	−0·79	−0·68	−0·72	−0·73
	κ	4·00	1·13	12·12	4·89	4·31	2·18	15·57	6·82	7·97
	μ_X	−5·00	−5·65	−4·93	−5·38	−5·42	−5·52	−4·38	−5·02	−5·00
	H	0·70	0·47	0·80	0·61	0·59	0·62	0·83	0·74	0·73
	σ_X	2·00	0·90	3·15	1·72	1·61	1·22	5·33	2·92	3·04
	X_0	−5·00	−5·47	−4·88	−5·07	−5·03	−5·15	−4·97	−5·06	−5·06

The results for dataset Sim-A in Table 1 show reasonable agreement between the posterior distribution and the true parameter values. Several of the credible intervals are wide, reflecting the small amount of information in Sim-A for particular parameters. In the case of medium-range

memory with $H < 0{\cdot}3$, the 95% credible interval is [0·20, 0·44]. When $H = 0{\cdot}5$ or 0·7, the credible intervals are wider. In particular, for $H = 0{\cdot}7$, this may suggest that the data do not provide substantial evidence of long-range memory. In such cases, one option is to consider richer datasets such as Sim-B where, as can be seen from Table 1, the credible interval is tighter and does not contain 0·5. Another option, which does not involve using volatility proxies, is to consider a longer or more frequently observed time series using intraday data. For example, rerunning the algorithm with a denser version of the Sim-A dataset containing two equispaced observations per day yields a 95% credible interval of [0·58, 0·74] for H. The posterior distribution for Sim-B is more informative for all parameters and provides accurate estimates of H. The 95% credible interval for H is below 0·5 when $H = 0{\cdot}3$ and above 0·5 when $H = 0{\cdot}7$.

4·3. *Real data from S&P* 500 *and VIX time series*

Dataset A consists of S&P 500 values only, i.e., discrete-time observations of U. We considered daily S&P 500 values from 5 March 2007 to 5 March 2008, before the Bear Stearns closure, and from 15 September 2008 to 15 September 2009, after the Lehman Brothers closure.

Dataset B is as dataset A but with daily VIX values for the same periods added.

Dataset C is as dataset B but with intraday observations of S&P 500 added; for each day we extracted three equispaced observations from 8:30 to 15:00.

Table 2 shows posterior estimates obtained from our algorithm for datasets A, B and C. The integration horizon T was set to 0·9, 1·5 and 1·5 for datasets A, B and C, respectively, and the numbers of leapfrog steps were chosen to achieve acceptance probabilities between 0·7 and 0·8.

The primary purpose of this analysis was to illustrate application of the algorithm in various observation regimes, so we do not attempt to draw strong conclusions from the results. Both extensions of the fractional stochastic volatility model considered in this paper, allowing $H < 0{\cdot}5$ and $\rho \neq 0$, seem to provide useful additions. In all cases, the concentration of the posterior distribution of H below 0·5 suggests medium-range dependence, in agreement with the results of an unpublished 2014 City University of New York manuscript by J. Gatheral. Moreover, the value of ρ is negative in all cases, suggesting the presence of a leverage effect. Our modelling and inferential framework provides a useful tool for further investigation of the S&P 500 index in other time periods, with different types of datasets and over various time scales.

4·4. *Comparison of different hybrid Monte Carlo schemes*

The results in §§ 4·2 and 4·3 were obtained by updating jointly the latent path and parameters with the advanced hybrid Monte Carlo method in Algorithm 3, labelled Scheme 1 in Table 3. In this subsection we compare this Markov chain Monte Carlo scheme with four variants. Scheme 2 is the Gibbs counterpart of Scheme 1. Scheme 3 performs joint updates of paths and parameters, like Scheme 1, but according to the standard hybrid Monte Carlo method, Algorithm 2. Schemes 4 and 5 are the same as Schemes 1 and 3, respectively, but with a smaller time discretization step. In each case, the same mass matrix was used, of the form (9). The integration horizon was fixed at $T = 0{\cdot}9$ and $T = 1{\cdot}5$ for datasets Sim-A and Sim-B, respectively, and the acceptance probability was between 0·7 and 0·8, based on previous experience. The time discretization step of the differential equations was set to $\delta = 0{\cdot}1$ for Schemes 1, 2 and 3, whereas for Schemes 4 and 5 it was set to $\delta = 0{\cdot}01$ to illustrate the behaviour of the standard and advanced hybrid Monte Carlo algorithms at finer resolution. Comparisons of the sampling efficiency were made by looking at the minimum effective sample sizes (Geyer, 1992) over θ and z, denoted by $\min_\theta(\text{ESS})$, $\min_z(\text{ESS})$ and $\min_{\theta,z}(\text{ESS})$; these quantities were computed from the lagged autocorrelations of the traceplots, and link to the percentage of the total number of Monte Carlo draws that can be

Table 2. *Posterior summary statistics for model* (21) *for datasets A, B and C*

Dataset			Parameters μ	ρ	κ	μ_X	H	σ_X	X_0
A	05/03/07–05/03/08	2·5%	−0·28	−0·77	3·33	−5·31	0·13	0·39	−6·07
	(before Bear	97·5%	0·19	−0·13	60·37	−4·27	0·40	1·42	−5·37
	Stearns closure)	Mean	−0·02	−0·47	26·03	−4·79	0·30	0·75	−5·71
		Median	−0·01	−0·48	24·67	−4·78	0·31	0·70	−5·71
	15/09/08–15/09/09	2·5%	−0·26	−0·73	1·06	−4·34	0·17	0·72	−4·84
	(after Lehman	97·5%	0·47	−0·19	27·26	−2·93	0·46	3·56	−3·94
	Brothers closure)	Mean	0·10	−0·49	8·07	−3·63	0·36	1·61	−4·39
		Median	0·10	−0·49	5·83	−3·61	0·38	1·42	−4·39
B	05/03/07–05/03/08	2·5%	−0·12	−0·75	1·81	−5·28	0·25	0·59	−5·84
	(before Bear	97·5%	0·27	−0·50	7·46	−4·44	0·33	0·90	−5·65
	Stearns closure)	Mean	0·07	−0·62	4·47	−4·93	0·29	0·72	−5·74
		Median	0·03	−0·62	4·47	−4·95	0·29	0·72	−5·74
	15/09/08–15/09/09	2·5%	0·08	−0·48	1·01	−4·50	0·34	0·60	−4·26
	(after Lehman	97·5%	0·23	−0·19	2·13	−2·93	0·42	0·84	−4·08
	Brothers closure)	Mean	0·08	−0·49	1·35	−3·63	0·38	0·71	−4·17
		Median	0·08	−0·49	1·27	−3·61	0·38	0·71	−4·17
C	05/03/07–05/03/08	2·5%	−0·14	−0·56	1·10	−5·54	0·26	0·68	−5·80
	(before Bear	97·5%	0·32	−0·27	4·14	−4·61	0·35	0·93	−5·61
	Stearns closure)	Mean	0·10	−0·42	2·07	−5·07	0·31	0·80	−5·71
		Median	0·10	−0·43	1·86	−5·10	0·32	0·81	−5·71
	15/09/08–15/09/09	2·5%	−0·59	−0·48	1·21	−4·11	0·28	0·45	−4·32
	(after Lehman	97·5%	−0·33	−0·30	2·31	−3·37	0·37	0·74	−4·08
	Brothers closure)	Mean	−0·47	−0·39	1·54	−3·75	0·33	0·59	−4·20
		Median	−0·47	−0·39	1·46	−3·74	0·33	0·58	−4·21

Table 3. *Relative efficiency of five versions of hybrid Monte Carlo schemes on datasets Sim-A and Sim-B*

Dataset	Sampler	$\min_\theta(\text{ESS})$ (%)	$\min_z(\text{ESS})$ (%)	Leapfrogs	Time (s)	$\frac{\min_{\theta,z}(\text{ESS})}{\text{time}}$	Rel. $\frac{\min_{\theta,z}(\text{ESS})}{\text{time}}$
Sim-A	Scheme 1	1·47	3·95	10	0·87	1·70	9·98
	Scheme 2	0·15	4·05	10	0·88	0·17	1·00
	Scheme 3	1·15	1·20	10	0·88	1·33	7·81
	Scheme 4	1·48	4·35	10	1·27	1·17	4·39
	Scheme 5	1·35	3·50	40	5·06	0·27	1·00
Sim-B	Scheme 1	3·19	8·81	50	3·35	0·95	5·32
	Scheme 2	0·60	5·00	50	3·41	0·18	1·00
	Scheme 3	1·20	3·40	50	3·35	0·36	2·00
	Scheme 4	1·94	8·40	50	6·13	0·32	3·76
	Scheme 5	1·03	6·95	100	12·26	0·08	1·00

considered as independent samples from the posterior. The computing time per iteration was also recorded.

The schemes were run on the datasets Sim-A and Sim-B with $H = 0{\cdot}3$; see Table 3. We first compare Schemes 1 and 2; on Sim-A and Sim-B, Scheme 1 was 9·98 and 5·32 times more efficient, respectively, illustrating the effect of strong posterior dependence between Z and θ. This dependence is introduced by the data, since Z and θ are independent a priori. The comparison

also illustrates the gain provided by the advanced hybrid Monte Carlo algorithm. In line with the associated theory, this gain increases as the discretization step δ becomes smaller, as Scheme 4 is 4·39 and 3·76 times more efficient than Scheme 5 on the Sim-A and Sim-B datasets, respectively.

5. DISCUSSION

Our method performs reasonably well and provides one of the few options, as far as we know, for routine Bayesian likelihood-based estimation for partially observed diffusions driven by fractional noise. Current computational capabilities and algorithmic improvements allow practitioners to experiment with non-Markovian model structures of the class considered in this paper in generic nonlinear contexts.

It is of interest to investigate the implications of the fractional model in option pricing for $H < 0{\cdot}5$. The joint estimation of physical and pricing measures based on asset and option prices can be studied in more depth, for both white and fractional noise. Moreover, the samples from the joint posterior of H and the other model parameters can be used to incorporate parameter uncertainty into the option pricing procedure. The posterior samples can also be used for Bayesian hypothesis testing, although this may require the marginal likelihood. Also, models with time-varying H are worth investigating when considering long time series. The Davies and Harte method, applied to blocks of periods of constant H given a stream of standard normal variates, would typically create discontinuities in conditional likelihoods, so a different and sequential method could turn out to be more appropriate in this context.

Another direction for investigation involves combining the algorithm in this paper, which focuses on computational robustness in high dimensions, with recent Riemannian manifold methods (Girolami & Calderhead, 2011), which automate the specification of the mass matrix and perform efficient Hamiltonian transitions on distributions with highly irregular contours.

Looking to general Gaussian processes beyond fractional Brownian motion, our method can also be applied to models in which 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 out in this context. For instance, a potential area of application is to Gaussian prior models for infinite-dimensional spatial processes.

We have assumed existence of a nontrivial Lebesgue density for observations given the latent diffusion path and parameters; however, this is not the case when data correspond to direct observations of the process, where one needs to work with Girsanov densities for diffusion bridges. The 2013 Harvard University technical report by M. Lysy and H. S. Pillai looks at this set-up.

Finally, another application would be to parametric inference for generalized Langevin equations with fractional noise, which arise as models in physics and biology (Kou & Xie, 2004).

ACKNOWLEDGEMENT

The first author was supported by the Leverhulme Trust, and the second and third authors were supported by the U.K. Engineering and Physical Sciences Research Council. We thank the reviewers for suggestions that have greatly improved the paper.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online presents the likelihood $p_N(Y \mid Z, \theta)$ and the derivatives $\nabla_Z p_N(Y \mid Z, \theta)$ and $\nabla_\theta \log p_N(Y \mid Z, \theta)$ required by the Hamiltonian methods, for the stochastic volatility class of models in (18) under observation regime (4).

Appendix

Proof of Proposition 1

The proof that standard hybrid Monte Carlo preserves $Q_N(x, v) = \exp\{-H(x, v; M)\}$, with H as in (8), is based on the volume preservation of ψ_h^I. That is, for the reference measure $Q_{N,0} \equiv \mathrm{Leb}_{4N+2q}$ we have $Q_{N,0} \circ \psi_h^{-I} \equiv Q_{N,0}$, enabling a simple change of variables when integrating (Duane et al., 1987). In infinite dimensions, a similar equality for Q_0 does not hold, so instead we adopt a probabilistic approach. To prove (i), we obtain a recursive formula for the densities $\mathrm{d}Q^{(i)}/\mathrm{d}Q_0$ $(i = 1, \ldots, I)$. We set

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

with $A = \mathrm{diag}(a_1, \ldots, a_q)$. We also set $g(x) = -\mathcal{C}^{1/2}\nabla\Phi(x)$ for $x \in \mathcal{H}$. From the definition of Ψ_h in (16), we have $Q^{(i)} = Q^{(i-1)} \circ \Xi_{h/2}^{-1} \circ \tilde{\Xi}_h^{-1} \circ \Xi_{h/2}^{-1}$. The map $\Xi_{h/2}(x, v) = \{x, v - (h/2)\mathcal{C}\nabla\Phi(x)\}$ keeps x fixed and translates v. The assumption $\nabla_z\Phi(z, \theta) \in \ell_2$ implies that $-(h/2)\mathcal{C}\nabla\Phi(x)$ is an element in the Cameron–Martin space of the v-marginal under Q_0, this marginal being $\prod_{i=1}^{\infty} N(0, 1) \otimes N(0, A^{-1})$. So, from standard theory for Gaussian laws on general spaces (Da Prato & Zabczyk, 1992, Proposition 2.20), we have that $Q_0 \circ \Xi_{h/2}^{-1}$ and Q_0 are absolutely continuous with respect to each other, with density

$$G(x, v) = \exp\left\{\left\langle \frac{h}{2}g(x),\ \mathcal{C}^{-1/2}v\right\rangle - \frac{1}{2}\left|\frac{h}{2}g(x)\right|^2\right\}. \tag{A1}$$

The assumption $\nabla_z\Phi(z, \theta) \in \ell_2$ guarantees that all inner products appearing in (A1) are finite. Recall that ℓ_2 denotes the space of squared summable infinite vectors. Hence,

$$\begin{aligned}\frac{\mathrm{d}Q^{(i)}}{\mathrm{d}Q_0}(x_i, v_i) &= \frac{\mathrm{d}\{Q^{(i-1)} \circ \Xi_{h/2}^{-1} \circ \tilde{\Xi}_h^{-1} \circ \Xi_{h/2}^{-1}\}}{\mathrm{d}Q_0}(x_i, v_i) \\ &= \frac{\mathrm{d}\{Q^{(i-1)} \circ \Xi_{h/2}^{-1} \circ \tilde{\Xi}_h^{-1} \circ \Xi_{h/2}^{-1}\}}{\mathrm{d}\{Q_0 \circ \Xi_{h/2}^{-1}\}}(x_i, v_i) \times \frac{\mathrm{d}\{Q_0 \circ \Xi_{h/2}^{-1}\}}{\mathrm{d}Q_0}(x_i, v_i) \\ &= \frac{\mathrm{d}\{Q^{(i-1)} \circ \Xi_{h/2}^{-1} \circ \tilde{\Xi}_h^{-1}\}}{\mathrm{d}Q_0}\left\{\Xi_{h/2}^{-1}(x_i, v_i)\right\} \times G(x_i, v_i).\end{aligned} \tag{A2}$$

We have $Q_0 \circ \tilde{\Xi}_h^{-1} \equiv Q_0$, as $\tilde{\Xi}_h$ rotates the infinite-dimensional products of independent standard Gaussians for the z- and v_z-components of Q_0 and translates the Lebesgue measure for the θ-component; thus overall $\tilde{\Xi}_h$ preserves Q_0. We also have $(\tilde{\Xi}_h^{-1} \circ \Xi_{h/2}^{-1})(x_i, v_i) \equiv \Xi_{h/2}(x_{i-1}, v_{i-1})$, so

$$\begin{aligned}\frac{\mathrm{d}\{Q^{(i-1)} \circ \Xi_{h/2}^{-1} \circ \tilde{\Xi}_h^{-1}\}}{\mathrm{d}Q_0}\left\{\Xi_{h/2}^{-1}(x_i, v_i)\right\} &= \frac{\mathrm{d}\{Q^{(i-1)} \circ \Xi_{h/2}^{-1} \circ \tilde{\Xi}_h^{-1}\}}{\mathrm{d}\{Q_0 \circ \tilde{\Xi}_h^{-1}\}}\left\{\Xi_{h/2}^{-1}(x_i, v_i)\right\} \\ &= \frac{\mathrm{d}\{Q^{(i-1)} \circ \Xi_{h/2}^{-1}\}}{\mathrm{d}Q_0}\left\{\Xi_{h/2}(x_{i-1}, v_{i-1})\right\} \\ &= \frac{\mathrm{d}Q^{(i-1)}}{\mathrm{d}Q_0}(x_{i-1}, v_{i-1}) \times G\{\Xi_{h/2}(x_{i-1}, v_{i-1})\},\end{aligned}$$

where for the last equation we divided and multiplied by $Q_0 \circ \Xi_{h/2}^{-1}$, as in the calculations in (A2), and made use of (A1). Hence, recalling the explicit expression for $\Xi_{h/2}$, overall we have that

$$\frac{\mathrm{d}Q^{(i)}}{\mathrm{d}Q_0}(x_i, v_i) = \frac{\mathrm{d}Q^{(i-1)}}{\mathrm{d}Q_0}(x_{i-1}, v_{i-1}) \times G(x_i, v_i) \times G\left\{x_{i-1},\ v_{i-1} + \frac{h}{2}\mathcal{C}^{1/2}g(x_{i-1})\right\}.$$

From here one can follow precisely the steps in Beskos et al. (2013a, § 3.4) to obtain, for $L = \mathcal{C}^{-1}$,

$$\log\left[G(x_i, v_i)\, G\left\{x_{i-1},\ v_{i-1} + \frac{h}{2}\mathcal{C}^{1/2} g(x_{i-1})\right\}\right]$$
$$= \frac{1}{2}\langle x_i, Lx_i\rangle + \frac{1}{2}\langle v_i, Lv_i\rangle - \frac{1}{2}\langle x_{i-1}, Lx_{i-1}\rangle - \frac{1}{2}\langle v_{i-1}, Lv_{i-1}\rangle.$$

Thus, due to the cancellations upon summing, we have obtained the expression for $(\mathrm{d}Q^{(I)}/\mathrm{d}Q_0)(x_I, v_I)$ given in statement (i) of Proposition 1.

Given (i), the proof of (ii) follows precisely as in the proof of Theorem 3.1 in Beskos et al. (2013a).

REFERENCES

AÏT-SAHALIA, Y. & KIMMEL, R. (2007). Maximum likelihood estimation of stochastic volatility models. *J. Finan. Econ.* **83**, 413–52.

ANDRIEU, C., DOUCET, A. & HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Statist. Soc.* B **72**, 269–342.

BESKOS, A., PINSKI, F. J., SANZ-SERNA, J. M. & STUART, A. M. (2011). Hybrid Monte Carlo on Hilbert spaces. *Stoch. Proces. Appl.* **121**, 2201–30.

BESKOS, A., KALOGEROPOULOS, K. & PAZOS, E. (2013a). Advanced MCMC methods for sampling on diffusion pathspace. *Stoch. Proces. Appl.* **123**, 1415–53.

BESKOS, A., PILLAI, N., ROBERTS, G., SANZ-SERNA, J.-M. & STUART, A. (2013b). Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli* **19**, 1501–34.

BIAGINI, F., HU, Y., ØKSENDAL, B. & ZHANG, T. (2008). *Stochastic Calculus for Fractional Brownian Motion and Applications*. London: Springer.

BREIDT, F., CRATO, N. & DE LIMA, P. (1998). The detection and estimation of long memory in stochastic volatility. *J. Economet.* **83**, 325–48.

CHRONOPOULOU, A. & VIENS, F. (2012a). Estimation and pricing under long-memory stochastic volatility. *Ann. Finan.* **8**, 379–403.

CHRONOPOULOU, A. & VIENS, F. (2012b). Stochastic volatility and option pricing with long-memory in discrete and continuous time. *Quant. Finan.* **12**, 635–49.

COMTE, F. & RENAULT, E. (1998). Long memory in continuous-time stochastic volatility models. *Math. Finan.* **8**, 291–323.

COMTE, F., COUTIN, L. & RENAULT, E. (2012). Affine fractional stochastic volatility models. *Ann. Finan.* **8**, 337–78.

COTTER, S. L., ROBERTS, G. O., STUART, A. M. & WHITE, D. (2013). MCMC methods for functions: Modifying old algorithms to make them faster. *Statist. Sci.* **28**, 424–46.

CRAIGMILE, P. F. (2003). Simulating a class of stationary Gaussian processes using the Davies–Harte algorithm, with application to long memory processes. *J. Time Ser. Anal.* **24**, 505–11.

DA PRATO, G. & ZABCZYK, J. (1992). *Stochastic Equations in Infinite Dimensions,* vol. 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge: Cambridge University Press.

DEYA, A., NEUENKIRCH, A. & TINDEL, S. (2012). A Milstein-type scheme without Lévy area terms for SDEs driven by fractional Brownian motion. *Ann. Inst. Henri Poincaré Prob. Statist.* **48**, 518–50.

DING, Z., GRANGER, C. & ENGLE, R. (1993). A long memory property of stock market returns and a new model. *J. Empirical Finan.* **1**, 83–106.

DUANE, S., KENNEDY, A., PENDLETON, B. & ROWETH, D. (1987). Hybrid Monte Carlo. *Phys. Lett.* B **195**, 216–22.

GEYER, C. J. (1992). Practical Markov chain Monte Carlo (with Discussion). *Statist. Sci.* **7**, 473–83.

GIROLAMI, M. & CALDERHEAD, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods (with discussion and a reply by the authors). *J. R. Statist. Soc.* B **73**, 123–214.

GLOTER, A. & HOFFMANN, M. (2004). Stochastic volatility and fractional Brownian motion. *Stoch. Proces. Appl.* **113**, 143–72.

GOLIGHTLY, A. & WILKINSON, D. J. (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Comp. Statist. Data Anal.* **52**, 1674–93.

HOSKING, J. R. M. (1984). Fractional differencing. *Water Resources Res.* **20**, 1898–908.

JONES, C. (2003). The dynamics of stochastic volatility: Evidence from underlying and options markets. *J. Economet.* **116**, 181–224.

KALOGEROPOULOS, K., ROBERTS, G. O. & DELLAPORTAS, P. (2010). Inference for stochastic volatility models using time change transformations. *Ann. Statist.* **38**, 784–807.

KOU, S. C. (2008). Stochastic modeling in nanoscale biophysics: Subdiffusion within proteins. *Ann. Appl. Statist.* **2**, 501–35.

KOU, S. C. & XIE, X. S. (2004). Generalized Langevin equation with fractional Gaussian noise: Subdiffusion within a single protein molecule. *Phys. Rev. Lett.* **93**, article no. 180603.

LOBATO, I. N. & SAVIN, N. E. (1998). Real and spurious long memory properties of stock market data. *J. Bus. Econ. Statist.* **16**, 261–8.

MANDELBROT, B. B. & VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10**, 422–37.

MISHURA, Y. S. (2008). *Stochastic Calculus for Fractional Brownian Motion and Related Processes,* vol. 1929 of *Lecture Notes in Mathematics*. Berlin: Springer.

NORROS, I., VALKEILA, E. & VIRTAMO, J. (1999). An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions. *Bernoulli* **5**, 571–87.

PRAKASA, Rao, B. L. S. (2010). *Statistical Inference for Fractional Diffusion Processes*. Chichester: John Wiley & Sons.

ROBERTS, G. O. & STRAMER, O. (2001). On inference for partial observed nonlinear diffusion models using the Metropolis–Hastings algorithm. *Biometrika* **88**, 603–21.

ROSENBAUM, M. (2008). Estimation of the volatility persistence in a discretely observed diffusion model. *Stoch. Proces. Appl.* **118**, 1434–62.

STRAMER, O. & BOGNAR, M. (2011). Bayesian inference for irreducible diffusion processes using the pseudo-marginal approach. *Bayesian Anal.* **6**, 231–58.

SUSSMANN, H. J. (1978). On the gap between deterministic and stochastic ordinary differential equations. *Ann. Prob.* **6**, 19–41.

WOOD, A. & CHAN, G. (1994). Simulation of stationary Gaussian processes in $[0, 1]^d$. *J. Comp. Graph. Statist.* **3**, 409–32.

[*Received July* 2013. *Revised May* 2015]