ESAIM: PS 18 (2014) 332–341
DOI: 10.1051/ps/2013039

ESAIM: Probability and Statistics
www.esaim-ps.org

CONSISTENT NON-PARAMETRIC BAYESIAN ESTIMATION FOR A TIME-INHOMOGENEOUS BROWNIAN MOTION*

Shota Gugushvili[1] and Peter Spreij[2]

Abstract. We establish posterior consistency for non-parametric Bayesian estimation of the dispersion coefficient of a time-inhomogeneous Brownian motion.

Mathematics Subject Classification. 62G20, 62M05.

Received October 8, 2012. Revised April 23, 2013.

1. Introduction

Consider a simple linear stochastic differential equation

$$\mathrm{d}X_t = \sigma(t)\mathrm{d}W_t, \quad X_0 = x, \quad t \in [0,1], \tag{1.1}$$

where W is a Brownian motion on some given probability space and the initial condition x and the square integrable dispersion coefficient σ are deterministic. We interpret equation (1.1) as a short-hand notation for the integral equation

$$X_t = x + \int_0^t \sigma(s)\mathrm{d}W_s, \quad t \in [0,1],$$

where the integral is the Wiener integral of σ with respect to the Brownian motion W. The process X is thus a time-inhomogeneous Brownian motion. The function σ can be viewed as a signal transmitted through a noisy channel, where the noise (modelled by the Brownian motion) is multiplicative. Note that X is a Gaussian process with mean $m(t) = x$ and covariance $\rho(s,t) = \int_0^{s\wedge t} \sigma^2(u)\mathrm{d}u$. By $\mathbb{P}_\sigma$ we will denote the law of the solution X to (1.1).

Assume for simplicity that $x = 0$ and denote $t_{i,n} = i/n, i = 0, \ldots, n$. Suppose that corresponding to the true dispersion coefficient $\sigma = \sigma_0$, one has a sample $X_{t_{i,n}}, i = 1, \ldots, n$, from the process X at his disposal. Assuming that σ_0 belongs to some non-parametric class $\mathcal{X}$ of dispersion coefficients, our goal is to estimate σ_0. This problem for a similar model was treated in [5, 9, 19] using a frequentist approach. However, a non-parametric

Keywords and phrases. Dispersion coefficient, non-parametric Bayesian estimation, posterior consistency, time-inhomogeneous brownian motion.

* *The research of the first author was supported by The Netherlands Organisation for Scientific Research (NWO).*

[1] Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands.
shota.gugushvili@math.leidenuniv.nl

[2] Korteweg-de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands. spreij@uva.nl

Article published by EDP Sciences

© EDP Sciences, SMAI 2014

https://doi.org/10.1051/ps/2013039 Published online by Cambridge University Press

Bayesian approach to estimation of σ_0 is also possible. The likelihood corresponding to the observations $X_{t_{i,n}}$ is given by

$$L_n(\sigma) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi \int_{t_{i-1,n}}^{t_{i,n}} \sigma^2(u)\mathrm{d}u}} \psi \left(\frac{X_{t_{i,n}} - X_{t_{i-1,n}}}{\sqrt{\int_{t_{i-1,n}}^{t_{i,n}} \sigma^2(u)\mathrm{d}u}} \right) \right\}, \tag{1.2}$$

where $\psi(u) = \exp(-u^2/2)$. For a prior Π on $\mathcal{X}$, Bayes' formula yields the posterior measure

$$\mathit{\Pi}(\Sigma | X_{t_{0,n}} \ldots, X_{n,n}) = \frac{\int_{\Sigma} L_n(\sigma) \mathit{\Pi}(\mathrm{d}\sigma)}{\int_{\mathcal{X}} L_n(\sigma) \mathit{\Pi}(\mathrm{d}\sigma)}$$

of any measurable set $\Sigma \subset \mathcal{X}$. In the Bayesian paradigm, the posterior encodes all the information required for inferential purposes. Once the posterior is available, one can proceed to computation of other quantities of interest in Bayesian statistics, such as Bayes point estimates, Bayes factors and so on.

It has been recognised since long that Bayesian procedures should be theoretically grounded through establishing posterior consistency, see *e.g.* [3]. In our context posterior consistency will mean that for every neighbourhood U_{σ_0} of σ_0 (in a suitable topology)

$$\Pi(U_{\sigma_0}^c | X_{t_{0,n}}, \ldots, X_{t_{n,n}}) \xrightarrow{\mathbb{P}_{\sigma_0}} 0 \tag{1.3}$$

as $n \to \infty$ (the notation $\xi_n \xrightarrow{\mathbb{P}_{\sigma_0}} \xi$ in (1.3) and below stands for convergence of a sequence of random variables ξ_n to a random variable ξ in $\mathbb{P}_{\sigma_0}$-probability). In other words, a consistent Bayesian procedure asymptotically puts posterior mass equal to one on every fixed neighbourhood of the true parameter. This is similar to the study of consistency of frequentist estimators. A method that does not appear to work in the idealised setting when an infinite amount of data is available (formalised by assuming that the sample size $n \to \infty$) should also be unattractive in the finite sample setting. Hence the importance of a study of posterior consistency. The situation is typically quite subtle in the infinite-dimensional Bayesian setting: it is known that a careless choice of the prior might render a Bayes procedure inconsistent. For an introduction to consistency issues in Bayesian non-parametric statistics see *e.g.* [6, 25].

Our task in this work is to establish (1.3) under suitable assumptions on the class of dispersion coefficients σ and the prior Π. Asymptotic properties of Bayesian procedures in estimation problems for stochastic differential equations have been already considered under various setups in [8, 12–14, 17], primarily in the context of non-parametric Bayesian estimation of the drift coefficient of a stochastic differential equation. Computational approaches to non-parametric Bayesian inference for stochastic differential equations were studied in [11, 15]. Convenient overviews of the available results are given in [16, 26]. However, in the above works dealing with Bayesian asymptotics it is assumed that either a continuous record of observations is available on the solution to a stochastic differential equation, or that the solution is observed at equispaced time points $\Delta, 2\Delta, \ldots, n\Delta$, with asymptotics treated in the latter case under the assumption that Δ is independent of n and $n \to \infty$. Our problem, on the other hand, requires a different approach due to a different sampling scheme and the fact that ergodicity of the solution to a stochastic differential equation, that played a prominent role in most of the previous works on non-parametric Bayesian approach to statistical inference for stochastic differential equations, is irrelevant in our case. Although the setup we consider looks simple, to the best of our knowledge our work is the first one to treat an inference problem for a stochastic differential equation in the so called high-frequency data case when $\Delta = \Delta_n \to 0$ as $n \to \infty$ using a non-parametric Bayesian approach. The high-frequency data setting is particularly relevant in financial mathematics, where asset prices are often modelled through stochastic differential equations and where huge amounts of observations on them separated by very short time instances are available. Perhaps the most interesting feature of the present work is the method of proof of posterior consistency, which differs in certain respects from the currently used techniques. See Section 4 for a discussion. Also the simplicity of our model should not necessarily be considered a disadvantage: indeed,

https://doi.org/10.1051/ps/2013039 Published online by Cambridge University Press

the model is somewhat similar to the Gaussian white noise model (see *e.g.* Chap. 7, Sect. 4 in [10]), which, as is known, has triggered some important developments in mathematical statistics.

The paper is organised as follows: in the next section we formulate our main result dealing with posterior consistency for non-parametric estimation of the dispersion coefficient. Since posterior consistency is closely linked with properties of a prior Π, in Section 3 we provide an example of a reasonable prior satisfying the assumptions made in Section 2. Section 4 contains a brief discussion on the obtained result. The proof of our main theorem is deferred until Section 5, while the Appendix contains two technical lemmas used in Section 5.

2. Results

The non-parametric class of dispersion coefficients we will be looking at is given in the following definition.

Definition 2.1. Let $\mathcal{X}$ be the collection of dispersion coefficients $\sigma : [0,1] \to [\kappa, K]$, such that $\sigma \in \mathcal{X}$ is Lipschitz with Lipschitz constant M. Here $0 < \kappa < K < \infty$ and $0 < M < \infty$ are three constants independent of the particular $\sigma \in \mathcal{X}$.

Remark 2.2. Note that for a constant σ we have $\mathbb{P}_\sigma = \mathbb{P}_{-\sigma}$. A positivity assumption on $\sigma \in \mathcal{X}$ in Definition 2.1 can hence be viewed as a simple and natural identifiability requirement. Strict positivity assumption $\sigma \geq \kappa > 0$ allows one to escape technical complications when manipulating the likelihood (1.2) (this condition has already appeared *e.g.* in [9]), while the upper bound $\sigma(t) \leq K, t \in [0,1]$, restricts the size of the non-parametric class $\mathcal{X}$ and is reasonable in light of Definition 2.3 given below. Finally, Lipschitz continuity of σ comes in handy at various stages of the proof of posterior consistency.

The notion of posterior consistency depends on a topology on $\mathcal{X}$.

Definition 2.3. The topology $\mathcal{T}$ on $\mathcal{X}$ is the topology induced by the L_2-norm $\|\cdot\|_2$.

We now formalise the concept of posterior consistency.

Definition 2.4. Let the prior Π be defined on $\mathcal{X}$. We say that posterior consistency holds, if for any fixed $\sigma_0 \in \mathcal{X}$ and every neighbourhood U_{σ_0} of σ_0 in the topology $\mathcal{T}$ from Definition 2.3 we have

$$\Pi(U^c_{\sigma_0} | X_{t_{0,n}} \ldots, X_{n,n}) \xrightarrow{\mathbb{P}_{\sigma_0}} 0$$

as $n \to \infty$.

We summarise our assumptions.

Assumption 2.5. Assume that

(a) the model (1.1) is given with $x = 0$ and $\sigma \in \mathcal{X}$, where $\mathcal{X}$ is defined in Definition 2.1,
(b) $\sigma_0 \in \mathcal{X}$ denotes the true dispersion coefficient,
(c) a discrete-time sample $\{X_{t_{i,n}}\}$ from the solution to (1.1) corresponding to σ_0 is available, where $t_{i,n} = i/n, i = 0, \ldots, n$.

Let

$$V_{\sigma_0,\varepsilon} = \{\sigma \in \mathcal{X} : \|\sigma - \sigma_0\|_\infty < \varepsilon\},$$

where $\|\cdot\|_\infty$ denotes the L_∞-norm. The following is the main result of the paper.

Theorem 2.6. *Under Assumption 2.5 posterior consistency as in Definition 2.4 holds, provided the prior* Π *on* $\mathcal{X}$ *satisfies*

$$\Pi(V_{\sigma,\varepsilon}) > 0 \tag{2.1}$$

for any $\varepsilon > 0$ *and at any* $\sigma \in \mathcal{X}$.

https://doi.org/10.1051/ps/2013039 Published online by Cambridge University Press

3. Example of a prior

In this section we provide an example of a prior satisfying condition (2.1). Fix $0 < \kappa < K < \infty$ and take a fixed Lipschitz continuous function $f : \mathbb{R} \to [0, K-\kappa]$ with Lipschitz constant $N > 0$ and set $\sigma(t) = \kappa + \int_0^t f(h(s))\mathrm{d}s$, where $h : [0,1] \to \mathbb{R}$ ranges over the set of Hölder continuous functions of order $\beta \in (0, 1/2)$ on $[0,1]$ for some fixed β. Then each σ maps the interval $[0,1]$ into the interval $[\kappa, K]$ and σ is also Lipschitz with Lipschitz constant K, because

$$|\sigma(t) - \sigma(s)| = \left| \int_s^t f(h(u))\mathrm{d}u \right| \leq K|t-s|.$$

We take the collection of these functions σ as the collection $\mathcal{X}$ from Assumption 2.5(a). We will now construct a prior Π on $\mathcal{X}$. Let $\widetilde{W} = (\widetilde{W}_t)_{0\leq t\leq 1}$ be a standard Brownian motion over the time interval $[0,1]$ and let Z be a standard normal random variable independent of $\widetilde{W}$. Define the Brownian motion $\overline{W} = (\overline{W}_t)_{0\leq t\leq 1}$ initialised at Z by $\overline{W}_t = Z + \widetilde{W}_t$ and introduce the process $Y = (Y_t)_{0\leq t\leq 1}$, where

$$Y_t = \kappa + \int_0^t f(\overline{W}_s)\mathrm{d}s.$$

Our prior Π on $\mathcal{X}$ will be the law of the process Y.

We have to check that the prior Π satisfies (2.1). To that end take a fixed $\sigma_0(t) = \kappa + \int_0^t f(h_0(s))\mathrm{d}s$ and let $\overline{w}$ be a generic realisation of the process $\overline{W}$, so that $\overline{y}_t = \kappa + \int_0^t f(\overline{w}_s)\mathrm{d}s$ is the corresponding generic realisation of the process Y. We have

$$\Pi(V_{\sigma_0,\varepsilon}) = \Pi(\overline{y} : \|\overline{y} - \sigma_0\|_\infty < \varepsilon) \geq \Pi\left(\overline{w} : \|\overline{w} - h_0\|_\infty < \frac{\varepsilon}{N}\right),$$

because

$$\sup_{t\in[0,1]} \left| \int_0^t [f(\overline{w}_s) - f(h_0(s))]\mathrm{d}s \right| \leq \|f(\overline{w}) - f(h_0)\|_\infty \leq N\|\overline{w} - h_0\|_\infty.$$

By Lemma 5.3 in [22],

$$\Pi\left(\overline{w} : \|\overline{w} - h_0\|_\infty < \frac{\varepsilon}{N}\right) \geq \exp\left(-\inf_{g:\|g-h_0\|_\infty<\varepsilon/(2N)} \frac{1}{2}\|g\|_H^2\right) \Pi\left(\|\overline{W}\|_\infty < \frac{\varepsilon}{2N}\right).$$

Here H denotes the Reproducing Kernel Hilbert Space (RKHS) of the process $\overline{W}$, $g \in H$, while $\|\cdot\|_H$ is the RKHS norm (see [22] for a detailed treatment of these concepts with a view towards non-parametric Bayesian statistics). In our case H consists of absolutely continuous functions $g : [0,1] \to \mathbb{R}$, such that $\|g'\|_2 < \infty$, while the RKHS norm is given by $\|g\|_H = \sqrt{g^2(0) + \|g'\|_2^2}$, see page 1446 in [21]. Note that the set $\{g \in H : \|g - h_0\|_\infty < \varepsilon/(2N)\}$ is not empty, as h_0 can be approximated arbitrarily closely in the L_∞-norm by the convolution $h_0 * k_b$ of h_0 with a smooth kernel $k_b(\cdot) = (1/b)k(\cdot/b)$ (*cf.* p. 1446 in [21]; we assume that $\|k'\|_2 < \infty$ and $b \to 0$). Furthermore, $\Pi(\|\overline{W}\|_\infty < \varepsilon/(2N)) > 0$, because $\|\overline{W}\|_\infty$ has a strictly positive density. Condition (2.1) easily follows. In case one is interested in a smoother class of dispersion coefficients σ than what we have just constructed, one can simply take in the above construction of the prior Π a smoother, say β times differentiable function f, and replace the Brownian motion $\overline{W}$ with a Riemann–Liouville process $R = (R_t)_{0\leq t\leq 1}$ with Hurst parameter β,

$$R_t = \sum_{k=0}^{\beta} Z_k t^k + \int_0^t (t-s)^{\beta-1/2}\mathrm{d}\widetilde{W}_t,$$

where Z_k's are standard normal random variables, $\widetilde{W}$ is a standard Brownian motion and $Z_0, Z_1, \ldots, Z_\beta, \widetilde{W}$ are independent. See Section 4.2 in [21] for more information on the Riemann–Liouville processes. Arguments similar to the ones given above yield that in this case as well (2.1) is satisfied.

https://doi.org/10.1051/ps/2013039 Published online by Cambridge University Press

4. Discussion

In the present work we established posterior consistency for a statistical model obtained from a simple linear stochastic differential equation. General techniques for proving posterior consistency for a wide range of statistical models are by now well-developed. In the i.i.d. setting, broadly speaking, two main approaches exist in the literature: a 'classical' approach as epitomised *e.g.* by [1,18] (we combine these two papers into one category, because they in some sense make use of assumptions of similar type, although their actual assertions are different), and a martingale approach developed more recently in [23,24]. The first approach was extended to the setting of independent non-identically distributed observations in [2], see in particular Theorem A.1 there. The second approach was extended to the case of discretely observed Markov processes in [7]. A general theorem for posterior consistency in [2] makes two requirements: firstly, the prior must put sufficient mass in arbitrarily small neighbourhoods of the true parameter (in an appropriate topology), and secondly, a sequence of sieves (increasing sequence of subsets of the parameter set) guaranteeing existence of certain exponentially consistent tests has to be exhibited; see page 1056 in [2] for additional details. Although in our setting the observations $X_{t_{i,n}}, i = 1, \dots, n$, are not independent, the increments $X_{t_{i,n}} - X_{t_{i-1,n}}$ are, and it appears conceivable that Theorem A.1 in [2] could be used to establish posterior consistency in our model as well. However, we opted for a different approach, see the proof of our posterior consistency result, Theorem 2.6. A similarity shared by Theorem A.1 in [2] and Theorem 2.6 is that both theorems require that the prior puts sufficient mass in arbitrarily small neighbourhoods of the true parameter (in appropriate topologies). A difference is that due to the special structure of our model we do not need to make any reference to tests and sieves, but can establish posterior consistency by directly manipulating the posterior; see the Proof of Theorem 2.6 for details. In this sense our approach to proving posterior consistency appears to be more direct and more elementary than the one that would employ Theorem A.1 in [2]. Neither do we make reference to entropy arguments as done *e.g.* in [1]. As far as the martingale approach to posterior consistency for ergodic Markov processes is concerned, we can be brief here: ergodicity is irrelevant in our setting and in fact our special sampling scheme seems to make generalisation or modification of the arguments from [7], [23,24] impossible.

Next a brief remark on condition (2.1) on the prior Π is in order. Although it is formulated in terms of neighbourhoods in the L_∞-norm, the assertion returned by Theorem 2.6 employs the topology induced by the L_2-norm. A 'discrepancy' between norms used is however not uncommon in posterior consistency results. See for instance [1].

5. Proofs

Proof of Theorem 2.6. Let U_{σ_0} be an arbitrary, but fixed neighbourhood of σ_0 in the topology $\mathcal{T}$ and let $\widetilde{U}_{\sigma_0,\varepsilon} = \{\sigma \in \mathcal{X} : \|\sigma - \sigma_0\|_2 < \varepsilon\}$. There exists $\varepsilon > 0$, such that $\widetilde{U}_{\sigma_0,\varepsilon} \subset U_{\sigma_0}$, and hence $U^c_{\sigma_0} \subset \widetilde{U}^c_{\sigma_0,\varepsilon}$. Fix such an ε. In order to prove the theorem, it thus suffices to show that

$$\Pi(\widetilde{U}^c_{\sigma_0,\varepsilon} | X_{t_{0,n}} \dots, X_{t_{n,n}}) \xrightarrow{\mathbb{P}_{\sigma_0}} 0 \tag{5.1}$$

as $n \to \infty$. Write

$$\Pi(\widetilde{U}^c_{\sigma_0,\varepsilon} | X_{t_{0,n}} \dots, X_{t_{n,n}}) = \frac{\int_{\widetilde{U}^c_{\sigma_0,\varepsilon}} L_n(\sigma) \Pi(\mathrm{d}\sigma)}{\int_{\mathcal{X}} L_n(\sigma) \Pi(\mathrm{d}\sigma)} = \frac{\int_{\widetilde{U}^c_{\sigma_0,\varepsilon}} R_n(\sigma) \Pi(\mathrm{d}\sigma)}{\int_{\mathcal{X}} R_n(\sigma) \Pi(\mathrm{d}\sigma)}, \tag{5.2}$$

where $R_n(\sigma) = L_n(\sigma)/L_n(\sigma_0)$ denotes the likelihood ratio. We will separately bound the numerator and denominator on the right-hand side of the last equality in (5.2) (we will use the notation D_n for the denominator and N_n for the numerator) and then combine the bounds to establish (5.1). As we will see, the left-hand side of (5.1) in fact decays exponentially fast to zero. Note that when establishing posterior consistency, [1] and [24] also treat the numerator and denominator in the expression for the posterior separately, but similarity of our approach to the one in those papers largely ends here.

https://doi.org/10.1051/ps/2013039 Published online by Cambridge University Press

Let $S_n(\sigma) = n^{-1}\log R_n(\sigma)$. Then $D_n = \int_{\mathcal{X}} \exp(nS_n(\sigma))\Pi(\mathrm{d}\sigma)$. Now

$$\begin{aligned} S_n(\sigma) &= \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^2(u)\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma^2(u)\mathrm{d}u}\right) - \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\left[\frac{(X_{t_{i,n}}-X_{t_{i-1,n}})^2}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma^2(u)\mathrm{d}u} - \frac{(X_{t_{i,n}}-X_{t_{i-1,n}})^2}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^2(u)\mathrm{d}u}\right] \\ &= T_{1,n}(\sigma) + T_{2,n}(\sigma), \end{aligned}$$

with obvious definitions of $T_{1,n}(\sigma)$ and $T_{2,n}(\sigma)$. Let $\widetilde{\varepsilon} > 0$ be a constant with its value to be chosen appropriately later on. Since $D_n \geq \int_{V_{\sigma_0,\widetilde{\varepsilon}}} R_n(\sigma)\Pi(\mathrm{d}\sigma)$, Lemmas A.1 and A.2 from the Appendix and formula (A.1) give that with probability tending to one,

$$D_n \geq \int_{V_{\sigma_0,\widetilde{\varepsilon}}} \exp\left(-\frac{4K}{\kappa^2}\widetilde{\varepsilon}n\right)\Pi(\mathrm{d}\sigma) = \exp\left(-\frac{4K}{\kappa^2}\widetilde{\varepsilon}n\right)\Pi(V_{\sigma_0,\widetilde{\varepsilon}}).$$

By assumption (2.1), $\Pi(V_{\sigma_0,\widetilde{\varepsilon}}) > 0$. Fix a constant $\beta = 5K\widetilde{\varepsilon}/\kappa^2$. Then for all n large enough,

$$\exp\left(-\frac{4K}{\kappa^2}\widetilde{\varepsilon}n\right)\Pi(V_{\sigma_0,\widetilde{\varepsilon}}) \geq \mathrm{e}^{-\beta n}.$$

As a consequence, with probability tending to one,

$$D_n \geq \mathrm{e}^{-\beta n} \tag{5.3}$$

as $n \to \infty$. This is our required lower bound for the denominator D_n.

Using similar techniques, we will next treat the numerator N_n. Firstly, note that by elementary arguments one can show that for an arbitrary fixed constant $C > 0$ there exists another constant $c > 0$, such that the inequality

$$\log(1+y) \leq y - cy^2, \quad -1 < y \leq C$$

holds (one can take $c = [2(C+1)]^{-1}$). Hence

$$\log\left(\frac{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^2(u)\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma^2(u)\mathrm{d}u}\right) \leq \frac{\int_{t_{i-1,n}}^{t_{i,n}}[\sigma_0^2(u)-\sigma^2(u)]\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma^2(u)\mathrm{d}u} - c\left(\frac{\int_{t_{i-1,n}}^{t_{i,n}}[\sigma_0^2(u)-\sigma^2(u)]\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma^2(u)\mathrm{d}u}\right)^2$$

for some constant c independent of $\sigma \in \mathcal{X}, i$ and n. Therefore, after a simple, but lengthy computation employing Assumption 2.5(a), *cf.* the proof of Lemma A.1,

$$\begin{aligned} T_{1,n}(\sigma) &\leq \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\frac{\int_{t_{i-1,n}}^{t_{i,n}}[\sigma_0^2(u)-\sigma^2(u)]\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma^2(u)\mathrm{d}u} - \frac{c}{2}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\int_{t_{i-1,n}}^{t_{i,n}}[\sigma_0^2(u)-\sigma^2(u)]\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma^2(u)\mathrm{d}u}\right)^2 \\ &= \frac{1}{2}\int_0^1\frac{\sigma_0^2(u)-\sigma^2(u)}{\sigma^2(u)}\mathrm{d}u - \frac{c}{2}\int_0^1\frac{(\sigma_0^2(u)-\sigma^2(u))^2}{\sigma^4(u)}\mathrm{d}u + O\left(\frac{1}{n}\right), \end{aligned}$$

where the remainder term is of order n^{-1} uniformly in $\sigma \in \mathcal{X}$. Hence

$$S_n(\sigma) \leq -\frac{c}{2}\int_0^1\frac{(\sigma_0^2(u)-\sigma^2(u))^2}{\sigma^4(u)}\mathrm{d}u \tag{5.4}$$

$$+ T_{2,n}(\sigma) + \frac{1}{2}\int_0^1\frac{\sigma_0^2(u)-\sigma^2(u)}{\sigma^2(u)}\mathrm{d}u \tag{5.5}$$

$$+ O\left(\frac{1}{n}\right). \tag{5.6}$$

https://doi.org/10.1051/ps/2013039 Published online by Cambridge University Press

To bound from above the term on the right-hand side of inequality (5.4), use the fact that

$$\frac{c}{2}\int_0^1 \frac{(\sigma_0^2(u)-\sigma^2(u))^2}{\sigma^4(u)}\mathrm{d}u \geq \frac{2\kappa^2 c}{K^4}\varepsilon^2$$

for $\sigma \in \widetilde{U}^c_{\sigma_0,\varepsilon}$. Furthermore, by Lemma A.2, uniformly in $\sigma \in \widetilde{U}^c_{\sigma_0,\varepsilon}$ and with probability tending to one as $n \to \infty$, the term (5.5) is smaller than any positive number fixed beforehand. So is the term (5.6). Therefore, uniformly in $\sigma \in \widetilde{U}^c_{\sigma_0,\varepsilon}$ and with probability tending to one as $n \to \infty$,

$$S_n(\sigma) \leq -\frac{\kappa^2 c}{K^4}\varepsilon^2,$$

say. Thus with probability tending to one as $n \to \infty$,

$$N_n = \int_{\widetilde{U}^c_{\sigma_0,\varepsilon}} \exp(nS_n(\sigma))\Pi(\mathrm{d}\sigma) \leq \exp\left(-\frac{\kappa^2 c}{K^4}\varepsilon^2 n\right). \tag{5.7}$$

This finishes bounding from above the numerator N_n.

We now combine bounds (5.3) and (5.7) to conclude that with probability tending to one as $n \to \infty$,

$$\Pi(\widetilde{U}^c_{\sigma_0,\varepsilon}|X_{t_{0,n}}\ldots,X_{t_{n,n}}) \leq \exp\left(-\left(\frac{\kappa^2 c}{K^4}\varepsilon^2 - \frac{5K}{\kappa^2}\widetilde{\varepsilon}\right)n\right).$$

Picking $\widetilde{\varepsilon}$ small enough, so that

$$\frac{\kappa^2 c}{K^4}\varepsilon^2 - \frac{5K}{\kappa^2}\widetilde{\varepsilon} > 0,$$

implies (5.1) and completes the proof of the theorem. □

Appendix A

Lemma A.1. *Under the same assumptions as in Theorem 2.6, for $T_{1,n}(\sigma)$ as in the Proof of Theorem 2.6, all $\sigma \in V_{\sigma_0,\widetilde{\varepsilon}}$ simulatenously and for n large enough, $T_{1,n}(\sigma) \geq -2K\widetilde{\varepsilon}/\kappa^2$.*

Proof. The elementary inequality

$$\frac{y}{y+1} \leq \log(1+y), \quad y > -1$$

gives that

$$\log\left(\frac{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^2(u)\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma^2(u)\mathrm{d}u}\right) \geq \frac{\int_{t_{i-1,n}}^{t_{i,n}}[\sigma_0^2(u)-\sigma^2(u)]\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^2(u)\mathrm{d}u}.$$

Next, employing Assumption 2.5(a) and (c), by a simple computation one can show that

$$\frac{\int_{t_{i-1,n}}^{t_{i,n}}[\sigma_0^2(u)-\sigma^2(u)]\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^2(u)\mathrm{d}u} = \frac{\sigma_0^2(t_{i-1,n})-\sigma^2(t_{i-1,n})}{\sigma_0^2(t_{i-1,n})} + O\left(\frac{1}{n}\right),$$

where the remainder term is of order n^{-1} uniformly in $\sigma \in \mathcal{X}$. Therefore,

$$T_{1,n}(\sigma) \geq \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\frac{\sigma_0^2(t_{i-1,n})-\sigma^2(t_{i-1,n})}{\sigma_0^2(t_{i-1,n})} + O\left(\frac{1}{n}\right).$$

https://doi.org/10.1051/ps/2013039 Published online by Cambridge University Press

By another simple computation,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\sigma_0^2(t_{i-1,n})-\sigma^2(t_{i-1,n})}{\sigma_0^2(t_{i-1,n})}=\int_0^1\frac{\sigma_0^2(u)-\sigma^2(u)}{\sigma_0^2(u)}\mathrm{d}u+O\left(\frac{1}{n}\right),$$

where the remainder term is of order n^{-1} uniformly in $\sigma\in\mathcal{X}$. For $\sigma\in V_{\sigma_0,\widetilde{\varepsilon}}$ we have under Assumption 2.5(a) that

$$\left|\int_0^1\frac{\sigma_0^2(u)-\sigma^2(u)}{\sigma_0^2(u)}\mathrm{d}u\right|\le\frac{2K}{\kappa^2}\widetilde{\varepsilon}. \tag{A.1}$$

This implies the statement of the lemma. □

Lemma A.2. *Denote*

$$Q_n(\sigma)=\left|T_{2,n}(\sigma)+\frac{1}{2}\int_0^1\frac{\sigma_0^2(u)-\sigma^2(u)}{\sigma^2(u)}\mathrm{d}u\right|,$$

where $T_{2,n}(\sigma)$ is defined in the Proof of Theorem 2.6. Then under the same assumptions as in Theorem 2.6, $\sup_{\sigma\in\mathcal{X}}Q_n(\sigma)\xrightarrow{\mathbb{P}_{\sigma_0}}0$ as $n\to\infty$. Furthermore, for any fixed $\widetilde{\varepsilon}>0$, for all $\sigma\in V_{\sigma_0,\widetilde{\varepsilon}}$ simultaneously, with probability tending to one as $n\to\infty$, $T_{2,n}(\sigma)\ge-2K\widetilde{\varepsilon}/\kappa^2$.

Proof. The first statement of the lemma will be derived from an application of Theorem 18.14 from [20]. In particular, viewing Q_n as a process on $\mathcal{X}$ with bounded sample paths, we will show that it converges in distribution to a zero process on $\mathcal{X}$. The first statement of the lemma will then be a consequence of equivalence of convergence in distribution and in probability for constant limits. Note that in order to circumvent possible (non)-measurability issues, outer probability is employed in the formulation of Theorem 18.14 from [20] (see Sect. 18.2 in [20] for more information on outer probability). Since no such problems will arise in our setting, we can instead directly work under probability $\mathbb{P}_{\sigma_0}$. Indeed, the summands in $T_{2,n}(\sigma)$ are of the form $F_{i,n}(\sigma)(X_{t_{i,n}}-X_{t_{i-1,n}})^2$, with $F_{i,n}$ the obvious functional of σ. Hence taking the supremum over σ does not affect the measurability property of $T_{2,n}(\sigma)$.

In order to apply Theorem 18.14 from [20], we need to verify its conditions. In our setting they reduce to the following ones: firstly, marginal vectors of Q_n must converge in distribution to zero vectors, *i.e.*

$$(Q_n(\sigma_1),\ldots,Q_n(\sigma_\ell))\xrightarrow{\mathcal{D}_{\sigma_0}}(\underbrace{0,\ldots,0}_{\ell}),\quad \ell\in\mathbb{N}. \tag{A.2}$$

Secondly, the tightness condition must be satisfied: for arbitrary constants $\eta>0$ and $\xi>0$, one must be able to find a partition of $\mathcal{X}$ into finitely many $\mathcal{X}_1,\ldots,\mathcal{X}_\ell$, such that

$$\limsup_{n\to\infty}\mathbb{P}_{\sigma_0}\left(\sup_{1\le k\le\ell}\sup_{\sigma_1,\sigma_2\in\mathcal{X}_k}|Q_n(\sigma_1)-Q_n(\sigma_2)|\ge\xi\right)\le\eta. \tag{A.3}$$

Denote

$$\mathcal{F}_{i,n}=\sigma(X_{t_j,n},j=1,\ldots,i)$$

and

$$\chi_{i,n}(\sigma)=-\frac{1}{2}\frac{1}{n}(X_{t_i,n}-X_{t_{i-1,n}})^2\frac{\int_{t_{i-1,n}}^{t_{i,n}}[\sigma_0^2(u)-\sigma^2(u)]\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma_0^2(u)\mathrm{d}u\int_{t_{i-1,n}}^{t_{i,n}}\sigma^2(u)\mathrm{d}u}.$$

Also let $\mathbb{E}_{\sigma_0}$ be the expectation operator with respect to measure $\mathbb{P}_{\sigma_0}$. Note that

$$\sum_{i=1}^{n}\mathbb{E}_{\sigma_0}\left(\chi_{i,n}(\sigma)|\mathcal{F}_{i-1,n}\right)=-\frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}\frac{\int_{t_{i-1,n}}^{t_{i,n}}[\sigma_0^2(u)-\sigma^2(u)]\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma^2(u)\mathrm{d}u}=-\frac{1}{2}\int_0^1\frac{\sigma_0^2(u)-\sigma^2(u)}{\sigma^2(u)}\mathrm{d}u+O\left(\frac{1}{n}\right),$$

https://doi.org/10.1051/ps/2013039 Published online by Cambridge University Press

where the remainder term is of order n^{-1} uniformly in $\sigma \in \mathcal{X}$. Furthermore, Assumption 2.5(a) yields that

$$\mathbb{E}_{\sigma_0}\left(\chi^2_{i,n}(\sigma)|\mathcal{F}_{i-1,n}\right) = \frac{3}{4n^2}\left(\frac{\int_{t_{i-1,n}}^{t_{i,n}}[\sigma_0^2(u)-\sigma^2(u)]\mathrm{d}u}{\int_{t_{i-1,n}}^{t_{i,n}}\sigma^2(u)\mathrm{d}u}\right)^2 = O\left(\frac{1}{n^2}\right),$$

where the order bound is uniform in $\sigma \in \mathcal{X}$. It follows that

$$\sum_{i=1}^{n}\mathbb{E}_{\sigma_0}\left(\chi^2_{i,n}(\sigma)|\mathcal{F}_{i-1,n}\right) \to 0.$$

Lemma 9 in [4] then implies that $Q_n(\sigma) \xrightarrow{\mathbb{P}_{\sigma_0}} 0$. This verifies (A.2).

We will now check (A.3). Fix ξ and η in (A.3). By a lengthy, but simple computation employing Assumption 2.5(a) and the triangle inequality,

$$|Q_n(\sigma_1)-Q_n(\sigma_2)| \le \frac{K}{\kappa^4}\|\sigma_1-\sigma_2\|_\infty\sum_{i=1}^{n}(X_{t_{i,n}}-X_{t_{i-1,n}})^2 + \frac{K^3}{\kappa^4}\|\sigma_1-\sigma_2\|_\infty. \tag{A.4}$$

By the Arzelà–Ascoli theorem, under Assumption 2.5(a) the family $\mathcal{X}$ is totally bounded for the supremum metric. By definition this means that for every $\zeta > 0$ there exists a finite set $\widetilde{\mathcal{X}} \subset \mathcal{X}$, such that for any $\sigma \in \mathcal{X}$ there is some $\widetilde{\sigma} \in \widetilde{\mathcal{X}}$ with $\|\sigma-\widetilde{\sigma}\|_\infty < \zeta/2$. This and the triangle inequality imply existence of a finite partition $\mathcal{X}_1,\dots,\mathcal{X}_\ell$ of $\mathcal{X}$, such that

$$\sup_{1\le k\le \ell}\sup_{\sigma_1,\sigma_2\in\mathcal{X}_k}\|\sigma_1-\sigma_2\|_\infty < \zeta. \tag{A.5}$$

Furthermore, by the definition of the quadratic variation of the process X,

$$\sum_{i=1}^{n}(X_{t_{i,n}}-X_{t_{i-1,n}})^2 \xrightarrow{\mathbb{P}_{\sigma_0}} \int_0^1\sigma_0^2(u)\mathrm{d}u. \tag{A.6}$$

Combination of (A.4)–(A.6) yields (A.3) for ζ small enough, and consequently the first statement of the lemma too. The second statement of the lemma is a consequence of the first one, the fact that $\sigma \in V_{\sigma_0,\widetilde{\varepsilon}}$, an analogue of inequality (A.1),

$$\left|\frac{1}{2}\int_0^1\frac{\sigma_0^2(u)-\sigma^2(u)}{\sigma^2(u)}\mathrm{d}u\right| \le \frac{K}{\kappa^2}\widetilde{\varepsilon},$$

and of a simple rearrangement

$$T_{2,n}(\sigma) = \left\{T_{2,n}(\sigma)+\frac{1}{2}\int_0^1\frac{\sigma_0^2(u)-\sigma^2(u)}{\sigma^2(u)}\mathrm{d}u\right\} - \frac{1}{2}\int_0^1\frac{\sigma_0^2(u)-\sigma^2(u)}{\sigma^2(u)}\mathrm{d}u.$$

This completes the proof of the lemma. □

Acknowledgements. The authors would like to thank the referees for a careful reading of the manuscript and useful suggestions.

References

[1] A. Barron, M.J. Schervish and L. Wasserman, The consistency of posterior distributions in nonparametric problems. *Ann. Statist.* **27** (1999) 536–561.

[2] N. Choudhuri, S. Ghosal and A. Roy, Bayesian estimation of the spectral density of a time series. *J. Amer. Statist. Assoc.* **99** (2004) 1050–1059.

https://doi.org/10.1051/ps/2013039 Published online by Cambridge University Press

[3] P. Diaconis and D. Freedman, On the consistency of Bayes estimates. With a discussion and a rejoinder by the authors. *Ann. Statist.* **14** (1986) 1–67.
[4] V. Genon-Catalot and J. Jacod, On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. Henri Poincaré Probab. Statist.* **29** (1993) 119–151.
[5] V. Genon-Catalot, C. Laredo and D. Picard, Nonparametric estimation of the diffusion coefficient by wavelets methods. *Scand. J. Statist.* **19** (1992) 317–335.
[6] S. Ghosal, J.K. Ghosh and R.V. Ramamoorthi, Consistency issues in Bayesian nonparametrics. Asymptotics, Nonparametrics, and Time Series. Vol. 158 of *Textbooks Monogr.* Dekker, New York (1999) 639–667.
[7] S. Ghosal and Y. Tang, Bayesian consistency for Markov processes. *Sankhyā* **68** (2006) 227–239.
[8] S. Gugushvili and P. Spreij, Non-parametric Bayesian drift estimation for stochastic differential equations (2012). Preprint `arXiv:1206.4981` [math.ST].
[9] M. Hoffmann, Minimax estimation of the diffusion coefficient through irregular samplings. *Statist. Probab. Lett.* **32** (1997) 11–24.
[10] I.A. Ibragimov and R.Z. Has′minskiĭ, Asimptoticheskaya teoriya otsenivaniya [Asymptotic Theory of Estimation] (Russian). Nauka, Moscow (1979).
[11] F. van der Meulen, M. Schauer and H. van Zanten, Reversible jump MCMC for nonparametric drift estimation for diffusion processes. *Comput. Statist. Data Anal.* **71** (2014) 615–632. Available on http://dx.doi.org/10.1016/j.csda.2013.03.002.
[12] F.H. van der Meulen, A.W. van der Vaart and J.H. van Zanten, Convergence rates of posterior distributions for Brownian semimartingale models. *Bernoulli* **12** (2006) 863–888.
[13] F. van der Meulen and H. van Zanten, Consistent nonparametric Bayesian estimation for discretely observed scalar diffusions. *Bernoulli* **19** (2013) 44–63.
[14] L. Panzar and H. van Zanten, Nonparametric Bayesian inference for ergodic diffusions. *J. Statist. Plann. Inference* **139** (2009) 4193–4199.
[15] O. Papaspiliopoulos, Y. Pokern, G.O. Roberts and A.M. Stuart, Nonparametric estimation of diffusions: a differential equations approach. *Biometrika* **99** (2012) 511–531.
[16] G.A. Pavliotis, Y. Pokern and A.M. Stuart, Parameter estimation for multiscale diffusions: an overview. Statistical Methods for Stochastic Differential Equations. Vol. 124 of *Monogr. Statist. Appl. Probab.* CRC Press, Boca Raton, FL (2012) 429–472.
[17] Y. Pokern, A.M. Stuart and J.H. van Zanten. Posterior consistency *via* precision operators for nonparametric drift estimation in SDEs. *Stoch. Process. Appl.* **123** (2013) 603–628.
[18] L. Schwartz, On Bayes procedures. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **4** (1965) 10–26.
[19] P. Soulier, Nonparametric estimation of the diffusion coefficient of a diffusion process. *Stochastic Anal. Appl.* **16** (1998) 185–200.
[20] A.W. van der Vaart, Asymptotic Statistics. Vol. 3 of *Cambr. Ser. Stat. Probab. Math.* Cambridge University Press, Cambridge (1998).
[21] A.W. van der Vaart and J.H. van Zanten, Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.* **36** (2008a) 1435–1463.
[22] A.W. van der Vaart and J.H. van Zanten, Reproducing kernel Hilbert spaces of Gaussian priors. Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh. Vol. 3 of *Inst. Math. Stat. Collect.* Inst. Math. Statist., Beachwood, OH (2008) 200–222.
[23] S. Walker, On sufficient conditions for Bayesian consistency. *Biometrika* **90** (2003) 482–488.
[24] S. Walker, New approaches to Bayesian consistency. *Ann. Statist.* **32** (2004) 2028–2043.
[25] L. Wasserman, Asymptotic properties of nonparametric Bayesian procedures. Practical Nonparametric and Semiparametric Bayesian Statistics. Vol. 133 of *Lect. Notes Statist.* Springer, New York (1998) 293–304.
[26] H. van Zanten, Nonparametric Bayesian methods for one-dimensional diffusion models. *Math. Biosci.* (2013). Available on http://dx.doi.org/10.1016/j.mbs.2013.03.008.

https://doi.org/10.1051/ps/2013039 Published online by Cambridge University Press