Biometrika (2019), **106**, 4, *pp.* 941–956
Printed in Great Britain
doi: 10.1093/biomet/asz042
Advance Access publication 16 September 2019

Semiparametric segment M-estimation for locally stationary diffusions

By P.-Y. DELÉAMONT and D. LA VECCHIA

Research Center for Statistics, University of Geneva, Blv. Pont d'Arve 40, 1211 Geneva, Switzerland

Pierre-Yves.Deleamont@unige.ch   davide.lavecchia@unige.ch

Summary

We develop and implement a novel M-estimation method for locally stationary diffusions observed at discrete time-points. We give sufficient conditions for the local stationarity of general time-inhomogeneous diffusions. Then we focus on locally stationary diffusions with time-varying parameters, for which we define our M-estimators and derive their limit theory.

Some key words: Biological signal; Kernel estimation; Martingale estimating function; Polynomial diffusion.

1. Introduction

Many statistical procedures have been developed for conducting inference on diffusions observed at discrete time-points; see, for example, Kessler & Sørensen (1999), Aït-Sahalia (2002, 2008), Aït-Sahalia & Mykland (2004), Bibby et al. (2010), La Vecchia & Trojani (2010), Kessler et al. (2012) and Li (2013). All of these methods rely on the assumption of stationarity, but this rarely holds in applications. For instance, in medicine, multivariate biological signals measured through electroencephalograms of epileptic patients display a time-varying scale. In some cases, a simple transformation can be used to recover a stationary series for which standard methods apply, but in the vast majority of situations no such transformation is available. Much emphasis has therefore been placed on developing inferential procedures for nonstationary processes.

One possible approach is related to the idea of recurrence (Bandi & Phillips, 2010; Park, 2014). Another possibility is to consider local stationarity (Dahlhaus, 1997, 2012). In that framework, the object of interest is a stochastic process whose characteristics change smoothly over time, in such a way that the nonstationary process can be approximated locally by a stationary one. The concept of local stationarity has been applied to various types of processes, including linear autoregressive (Dahlhaus, 1997), autoregressive with conditional heteroskedasticity (Dahlhaus & Subba Rao, 2006; Fryzlewicz et al., 2008), generalized autoregressive with conditional heteroskedasticity (Hafner & Linton, 2010) and nonlinear autoregressive (Vogt, 2012). Estimators have been successfully defined for these processes and their asymptotic theory is well established. Most results are in the univariate, discrete-time setting. Partial exceptions include the multivariate regression functions in Vogt (2012) and the multivariate generalized autoregressive processes with conditional heteroskedasticity in Hafner & Linton (2010) in a discrete-time setting, as well as the scalar diffusions in Koo & Linton (2012).

Extending these results to a continuous-time and possibly multivariate setting is not straightforward for essentially two reasons. First, establishing sufficient conditions for the local stationarity of general multivariate diffusions with possibly nonlinear drift and diffusion functions requires extension of the results available for discrete-time nonlinear processes and for scalar diffusions. Second, the continuous-time nature of diffusions poses statistical challenges for semiparametric estimation. The method of Koo & Linton (2012) does not lend itself to a natural extension to the multivariate setting and relies on kernels in time and space; it is unlikely to perform well in moderate samples.

In this paper we develop a novel theory for semiparametric estimation of discretely observed, multivariate locally stationary diffusions. We introduce a new class of M-estimators for time-inhomogeneous diffusions characterized by time-varying parameters. The estimators have a natural interpretation: they are defined by equations obtained from a combination of martingale estimating functions and a kernel in time.

2. Locally stationary diffusions

On $\mathcal{S} \subseteq \mathbb{R}^d$ ($d \geqslant 1$), consider a time-inhomogeneous diffusion $\{X_t,\ 0 \leqslant t \leqslant T\}$ that solves the stochastic differential equation

$$\mathrm{d}X_t = \mu(t/T, X_t)\,\mathrm{d}t + \sigma(t/T, X_t)\,\mathrm{d}W_t, \quad X_0 = Z_0, \tag{1}$$

where $\{W_t,\ t \geqslant 0\}$ is a d-dimensional Brownian motion and Z_0 is a random vector, independent of the Brownian motion, satisfying $E(\|Z_0\|^2) < \infty$. The drift vector μ and the dispersion matrix σ are functions of the rescaled time t/T, taking values in $[0, 1]$. We will occasionally write $X_{t,T}$ instead of X_t to emphasize the dependence on T, and to indicate that we are dealing with a triangular array of stochastic differential equations as in (1).

We define a family, indexed by u, of strictly stationary diffusions $\{\tilde{X}_t(u),\ t \geqslant 0\}$ on $\mathcal{S}$ solving

$$\mathrm{d}\tilde{X}_t(u) = \mu\{u, \tilde{X}_t(u)\}\,\mathrm{d}t + \sigma\{u, \tilde{X}_t(u)\}\,\mathrm{d}W_t, \quad \tilde{X}_0(u) = \tilde{Z}_0(u) \tag{2}$$

for every $u \in [0, 1]$, where $\tilde{Z}_0(u)$ is a random vector with stationary distribution, indexed by u, satisfying $E(\|\tilde{Z}_0(u)\|^2) < \infty$. The drift and dispersion in (2) have the same structure as in (1), but for a fixed time argument. The existence of a unique strong solution to equations (1) and (2) is guaranteed under general conditions; see, e.g., Stroock & Varadhan (2006). The requirements needed to ensure stationarity of the solution to (2) are best given case by case. From now on, we assume the stationarity of $\{\tilde{X}_t(u),\ t \geqslant 0\}$.

A comparison of (1) and (2) suggests that if $t/T \approx u$, then $X_{t,T}$ and $\tilde{X}_t(u)$ should be close with high probability. Thus, the intuition for local stationarity is that around each rescaled time-point t/T the process $\{X_{t,T}\}$ can be approximated, in a suitable stochastic sense, by a stationary process $\{\tilde{X}_t(u)\}$. Since the drift and the dispersion of $\{X_{t,T}\}$ change smoothly over $[0, T]$, the approximation accuracy depends both on T and on $|t/T - u|$. We formalize these arguments in the following definition.

Definition 1. *Let $T \in\ ]0, \infty[$. The stochastic process $\{X_{t,T},\ 0 \leqslant t \leqslant T\}$ defined as the solution to (1) is said to be locally stationary if for any $u \in [0, 1]$ there exists a stationary diffusion process $\{\tilde{X}_t(u),\ t \geqslant 0\}$ solving (2) such that*

$$\|X_{t,T} - \tilde{X}_t(u)\| \leqslant \big(|t/T - u| + 1/T\big)\, U_{t,T}(u) \tag{3}$$

almost surely, where $\{U_{t,T}(u),\ 0 \leqslant t \leqslant T\}$ is an $\mathbb{R}_+$-valued process with $E\{U_{t,T}(u)^\rho\} < C$ for some C independent of t, T and u and for some $\rho > 0$.

As noted by Vogt (2012), the fact that the first moment of $U_{t,T}(u)$ is uniformly bounded implies that $U_{t,T}(u) = O_{\rm p}(1)$ and, therefore, $\|X_{t,T} - \tilde{X}_t(u)\| = O_{\rm p}(|t/T - u| + 1/T)$. The bound in (3) is the key concept of local stationarity and lies at the core of our theory.

The Appendix gives sufficient conditions for local stationarity, striking a balance between the generality of the model in (1), where μ and σ remain totally unspecified, and the feasibility of proving the following proposition.

PROPOSITION 1 (Local stationarity). *Let $\{X_{t,T},\ 0 \leqslant t \leqslant T\}$ be the solution to (1). Under Assumption A1 in the Appendix, the process is locally stationary in the sense of Definition 1.*

Some of the proposed conditions in Assumption A1, such as condition (a), are standard in the literature on diffusions (see, e.g., Stroock & Varadhan, 2006), while others, such as condition (b), are common in the literature on local stationarity (see, e.g., Vogt, 2012). These conditions can be relaxed by imposing more structure on μ and σ. Below, we suppose that the process $\{X_t,\ 0 \leqslant t \leqslant T\}$ solving (1) is represented by the semiparametric time-inhomogeneous model

$$\mathrm{d}X_t = \mu\{\theta\,(t/T), X_t\}\,\mathrm{d}t + \sigma\{\theta(t/T), X_t\}\,\mathrm{d}W_t, \quad X_0 = Z_0, \tag{4}$$

where the drift and dispersion depend on the time-varying $\theta(\cdot)$, an unspecified function of time. For (4), local stationarity can be proved assuming smoothness of $\theta(\cdot)$, which is the object of inference; see § 5.1 for a detailed illustration with a specific time-inhomogeneous diffusion.

Typically, (4) results from the generalization of parametric models, with fixed parameters replaced by time-varying ones (Fan et al., 2003; Koo & Linton, 2012). For instance, consider

$$\mathrm{d}X_t = \kappa(t/T)\{\alpha(t/T) - X_t\}\,\mathrm{d}t + \sigma(t/T)\,\mathrm{d}W_t,$$

where the parameters change slowly over time. This process can be rewritten as in (4), with $\theta(t/T) = \{\kappa(t/T), \alpha(t/T), \sigma^2(t/T)\}^{\rm T}$. It is related, up to the rescaling of time to the unit interval, to the Hull & White (1990) model in finance, which extends the Vasicek (1977) model with constant parameters.

3. SEMIPARAMETRIC SEGMENT M-ESTIMATION

3.1. *Definition*

Assume that we have a sample $X_1, \ldots, X_T$ of discrete-time observations from the locally stationary process (4), with T fixed. We consider the time between two consecutive observations to be $\Delta \equiv 1$. Our theory can easily be generalized to the case where the observations are unequally spaced, by defining the time interval between observations at t_{i-1} and t_i as $\Delta_i = t_i - t_{i-1}$. Some types of randomly sampled data could also be considered, as in Aït-Sahalia & Mykland (2004).

We are interested in estimating the p-dimensional curve $\theta(\cdot)$ such that $\theta(u) \in \Theta \subset \mathbb{R}^p$. For $\theta = \theta(t/T)$ in (4), suppose that the transition density of the discretely observed locally stationary diffusion is known in closed form. Let us denote the transition density from state x at time s to state y at time $t > s$ by $p_{t,s}(y; x, \theta)$. Building on the Markovian nature of the process, we define the log transition density $\ell_t(\theta) = \log p_{t,t-1}(X_t; X_{t-1}, \theta)$. Then, following Dahlhaus & Subba Rao (2006), we define the weighted loglikelihood at some fixed time-point t_0 by

$$\mathcal{L}^{(b)}_{t_0,T}(\theta) = (bT)^{-1} \sum_{t=1}^{T} K\{(t_0 - t)/(bT)\}\,\ell_t(\theta),$$

where $K : [-1/2, 1/2] \to \mathbb{R}$ is a kernel of bounded variation satisfying $\int_{-1/2}^{1/2} K(u)\,\mathrm{d}u = 1$ and $\int_{-1/2}^{1/2} uK(u)\,\mathrm{d}u = 0$, and b is the bandwidth parameter. These conditions on the kernel are standard in the literature on local stationarity; see Dahlhaus & Subba Rao (2006) and Dahlhaus (2012). The segment maximum likelihood estimator is then $\hat{\theta}_{t_0,T} = \arg\max_\theta \mathcal{L}^{(b)}_{t_0,T}(\theta)$. Note that $\hat{\theta}_{t_0,T}$ is regarded as an estimator of $\theta(t_0/T)$, or of $\theta(u_0)$ for any u_0 such that $|u_0 - t_0/T| < 1/T$. Alternatively, the estimator is defined through the weighted loglikelihood score

$$S^{(b)}_{t_0,T}(\theta) = (bT)^{-1} \sum_{t=1}^{T} K\{(t_0 - t)/(bT)\} \nabla_{\theta^{\mathrm{T}}} \ell_t(\theta), \tag{5}$$

where $\hat{\theta}_{t_0,T}$ is the estimator solving the p-dimensional estimating equation $S^{(b)}_{t_0,T}(\theta) = 0$. This segment M-estimator is the solution to an estimating equation at t_0, where the kernel in time controls the length of the segment, assigning more weight to the observations closer to t_0.

The problem relating to the use of $S^{(b)}_{t_0,T}(\theta)$ is that for many diffusions the transition density is not known in closed form, so $\nabla_{\theta^{\mathrm{T}}} \ell_t(\theta)$ is also unavailable. For such cases we propose an extension of (5). A stepping stone for our construction is the introduction of M-estimators defined by equations in which $\nabla_{\theta^{\mathrm{T}}} \ell_t(\theta)$ is replaced by an alternative function. We set

$$G^{(b)}_{t_0,T}(\theta) = (bT)^{-1} \sum_{t=1}^{T} K\{(t_0 - t)/(bT)\} \psi(X_t, X_{t-1}, \theta), \tag{6}$$

where $\psi : \mathcal{S}^2 \times \Theta \to \mathbb{R}^p$ is a suitably chosen function, as explained in the next subsection. The estimator $\hat{\theta}_{t_0,T}$ is then the solution to $G^{(b)}_{t_0,T}(\theta) = 0$. For different specifications of ψ, (6) defines a general class of semiparametric segment M-estimators at t_0.

3.2. *Construction*

Considering the kernel and the bandwidth parameter as fixed, $G^{(b)}_{t_0,T}(\theta)$ is uniquely characterized by the estimating function ψ. The key idea of our construction is to specify ψ using, at each time-point, the stationary approximation to the original process.

To define a semiparametric segment M-estimator at t_0, we start by devising ψ as if we were estimating the constant parameters associated with the stationary approximating process $\{\tilde{X}_t(u_0),\ t \geqslant 0\}$ with $|u_0 - t_0/T| < 1/T$. Then, estimation of the time-varying parameters of $\{X_t,\ 0 \leqslant t \leqslant T\}$ is conducted, defining the estimating function in (6) by means of that ψ. If $\nabla_{\theta^{\mathrm{T}}} \ell_t(\theta)$ is not known in closed form, an estimator of θ is obtained using ψ, which mimics the behaviour of the likelihood score. To this end, various types of estimating functions for discretely observed diffusions have been proposed in the stationary case; see Kessler et al. (2012). Among the available techniques, martingale estimating functions define inferential procedures with a well-established theoretical underpinning and moderate computational cost; see Bibby et al. (2010). We propose using them to conduct inference in our setting.

Let $u_0 \in [0, 1]$, and assume that we are given a discrete sample from the stationary process $\{\tilde{X}_t(u_0),\ t \geqslant 0\}$ satisfying a stochastic differential equation whose drift and dispersion depend on a constant parameter $\theta = \theta(u_0)$. To define a function ψ, we consider k square-integrable functions $h_j : \mathcal{S}^2 \times \Theta \to \mathbb{R}$ $(j = 1, \ldots, k)$ such that for each j, $E[h_j\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta\} \mid \tilde{X}_{t-1}(u_0)] = 0$. Each function h_j defines a relationship between an observation $\tilde{X}_t(u_0)$, the history represented by $\tilde{X}_{t-1}(u_0)$, and the parameter θ. Now, denoting by $\mathcal{M}_{p,k}$ the set of $p \times k$ matrices with real entries, let $a : \mathcal{S} \times \Theta \to \mathcal{M}_{p,k}$ be a function that is differentiable in its second argument. For $h = (h_1, \ldots, h_k)^{\mathrm{T}}$ we define

$$\psi\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta\} = a\{\tilde{X}_{t-1}(u_0), \theta\} h\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta\}, \quad (7)$$

i.e., ψ is a martingale difference, with a determining the weight of each h_j. The resulting

$$\tilde{G}_T(u_0, \theta) = \sum_{t=1}^{T} a\{\tilde{X}_{t-1}(u_0), \theta\} h\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta\} \quad (8)$$

is a p-dimensional martingale estimating function. The function h can be constructed using, for example, the first few conditional moments, which are often available in closed form for diffusions. For instance, using the first conditional moment, one would set $h\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta\} = \tilde{X}_t(u_0) - E\{\tilde{X}_t(u_0) \mid \tilde{X}_{t-1}(u_0)\}$. Martingale estimating functions using higher conditional moments are defined analogously; see the Supplementary Material.

For a given h, we consider the class $\tilde{\mathcal{G}}_T(u_0, \theta)$ of martingale estimating functions of the form (8) obtained from different choices of the weight a. Clearly, each estimating function in $\tilde{\mathcal{G}}_T(u_0, \theta)$ yields a Fisher-consistent M-estimator, and any choice of a can be used in (7). Among all the available options, one may choose a to tune the efficiency. The optimal estimating function $\tilde{G}^*_T(u_0, \theta)$ is defined by $\psi^*\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta\} = a^*\{\tilde{X}_{t-1}(u_0), \theta\} h\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta\}$, where

$$a^*\{\tilde{X}_{t-1}(u_0), \theta\} = -E\big[\nabla_{\theta^{\mathrm{T}}} h\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta\} \mid \tilde{X}_{t-1}(u_0)\big]^{\mathrm{T}} V_h^{-1}\{\tilde{X}_{t-1}(u_0), \theta\}$$

with

$$V_h\{\tilde{X}_{t-1}(u_0), \theta\} = E\big[h\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta\} h\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta\}^{\mathrm{T}} \mid \tilde{X}_{t-1}(u_0)\big].$$

The resulting M-estimator is efficient in the semiparametric model which assumes that $\{\tilde{X}_t(u_0),\ t \geqslant 0\}$ is a Markov process and specifies conditional moments up to the appropriate order; see Heyde (1997) and Bibby et al. (2010) for a related discussion.

Examples of M-estimators obtained via martingale estimating functions are given in Gouriéroux et al. (1984), Kessler & Sørensen (1999), Aït-Sahalia (2002, 2008), Bibby et al. (2010) and La Vecchia & Trojani (2010). In particular, the approach based on the normal pseudo maximum likelihood method is a special case of a suboptimal martingale estimating function obtained by taking ψ to be the logarithmic derivative with respect to θ of the Gaussian pseudo transition density.

4. Asymptotic theory

In this section, we discuss the asymptotic properties of our estimators. The main idea underlying an asymptotic theory in the locally stationary setting is the rescaling of time. The key to studying an estimator at a particular time t_0 is to keep the ratio t_0/T fixed and let T diverge. This implies that as $T \to \infty$, more and more information on the local characteristics of the process becomes available. We do not assume that the time between two consecutive discrete-time observations decreases in the original time scale. To keep the theory general, in Propositions 2 and 3 we derive the asymptotics without specifying any functional form for μ and σ in (4). In § 5 we show that when suitable functional forms are defined, the assumptions can be relaxed.

For $|u_0 - t_0/T| < 1/T$ and $\theta = \theta(u_0)$, consider $G^{(b)}_{t_0,T}(\theta)$ in (6) and the estimating function

$$\tilde{G}^{(b)}_T(u_0,\theta) = (bT)^{-1}\sum_{t=1}^{T} K\{(t_0 - t)/(bT)\}\psi\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta\}.$$

A key step is to show that $G^{(b)}_{t_0,T}(\theta)$ and $\tilde{G}^{(b)}_T(u_0,\theta)$ converge to the same nonstochastic limit $G(u_0,\theta) = E[\psi\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta\}]$, which is equal to the null vector when θ is the true vector of parameters $\theta_0 \equiv \theta_0(u_0)$. We prove that the bias due to deviation from stationarity, defined as $B_{t_0,T}(\theta) = G^{(b)}_{t_0,T}(\theta) - \tilde{G}^{(b)}_T(u_0,\theta)$, vanishes as $T \to \infty$. Consistency and asymptotic normality then follow, respectively, from the law of large numbers for dependent data and the martingale central limit theorem; see Sørensen (1999).

PROPOSITION 2 (Consistency). *Suppose that* $\{X_t,\ 0 \leqslant t \leqslant T\}$ *is a locally stationary diffusion given by* (4). *Let* $\theta(u_0) \in \Theta \subset \mathbb{R}^p$ *with* $p \geqslant 1$. *Let* $\psi : \mathcal{S}^2 \times \Theta \to \mathbb{R}^p$ *be a function satisfying conditions* (a)–(e) *of Assumption* A2, *defining an estimator* $\hat{\theta}_{t_0,T}$ *through the equation* $G^{(b)}_{t_0,T}(\hat{\theta}_{t_0,T}) = 0$. *Then, for* $|u_0 - t_0/T| < 1/T$, $\hat{\theta}_{t_0,T}$ *converges in probability to* $\theta_0(u_0)$ *whenever* $b \to 0$ *and* $bT \to \infty$ *as* $T \to \infty$.

PROPOSITION 3 (Asymptotic normality). *Suppose that* $\{X_t,\ 0 \leqslant t \leqslant T\}$ *is a locally stationary diffusion given by* (4). *Let* $\theta(u_0) \in \Theta \subset \mathbb{R}^p$ *with* $p \geqslant 1$. *Let* $\psi : \mathcal{S}^2 \times \Theta \to \mathbb{R}^p$ *be a function satisfying conditions* (a)–(h) *of Assumption* A2, *defining an estimator* $\hat{\theta}_{t_0,T}$ *through the equation* $G^{(b)}_{t_0,T}(\hat{\theta}_{t_0,T}) = 0$. *Then, for* $|u_0 - t_0/T| < 1/T$, *the following statements hold:*

(i) *If* $b^3 \ll T^{-1}$, *then* $(bT)^{1/2}\{\hat{\theta}_{t_0,T} - \theta_0(u_0)\}$ *converges in distribution to* $N\{0, \Sigma(u_0)\}$ *where*

$$\Sigma(u_0) = \left(\int_{-1/2}^{1/2} K(s)^2\,\mathrm{d}s\right)\Omega(u_0)^{-1}\Upsilon(u_0)\Omega(u_0)^{-1\,\mathrm{T}}$$
$$\Omega(u_0) = E\big[\nabla_{\theta^{\mathrm{T}}}\psi\big\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta_0(u_0)\big\}\big]^{\mathrm{T}},$$
$$\Upsilon(u_0) = E\big[\psi\big\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta_0(u_0)\big\}\psi\big\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta_0(u_0)\big\}^{\mathrm{T}}\big].$$

(ii) *If* $b^{13} \ll T^{-1}$ *and condition* (i) *of Assumption* A2 *holds, then*

$$(bT)^{1/2}\left\{\hat{\theta}_{t_0,T} - \theta_0(u_0) + \frac{b^2}{2}\left(\int_{-1/2}^{1/2} K(s)s^2\,\mathrm{d}s\right)\Omega(u_0)^{-1}\Lambda(u_0)\right\}$$

converges in distribution to $N\{0, \Sigma(u_0)\}$, *where*

$$\Lambda(u_0) = E\big[\partial_u^2\psi\{\tilde{X}_t(u), \tilde{X}_{t-1}(u), \theta_0(u_0)\}\big|_{u=u_0}\big].$$

The term $\Lambda(u_0)$ in Proposition 3 shows that the bias depends on the degree of smoothness of the family of approximating processes, which enters via ψ and its derivatives in time. In the Supplementary Material, we use Proposition 3 to determine an optimal bandwidth.

5. Illustrative example

5.1. *Local stationarity*

Time-homogeneous polynomial diffusions are characterized by a linear drift vector and a diffusion matrix which can be quadratic (Filipović & Larsson, 2016; Larsson & Pulido, 2017). Thanks to these features, martingale estimating functions for the model parameters in (4) are typically available in closed form; see the Supplementary Material. Examples of polynomial diffusions include Brownian motion, geometric Brownian motion, Ornstein–Uhlenbeck processes, squared Bessel processes such as the Cox–Ingersoll–Ross diffusion, and Jacobi processes. The analytical tractability of polynomial diffusions makes them good candidates for the application of our inferential method. However, the polynomial nature of the infinitesimal coefficients entails that some of the assumptions in Propositions 1–3 may be not satisfied. Nevertheless, the violated conditions are not always necessary: the results in § § 2 and 4 can be proved by imposing suitable restrictions on $\theta(\cdot)$ and following the steps in the Supplementary Material. This suggests that the applicability of our method goes beyond the framework used for the derivation of our theory. We illustrate this by sketching the proofs of local stationarity and consistency, see § 5.2, in the benchmark case of a linear time-inhomogeneous diffusion.

Let $\{W_t,\ t \geqslant 0\}$ be a d-dimensional Brownian motion and let $\gamma(t/T) = \{\gamma_1(t/T), \ldots, \gamma_d(t/T)\}^{\mathrm{T}}$. We consider the time-inhomogeneous diffusion with $\mathcal{S} = \mathbb{R}$ solving the equation $\mathrm{d}X_{t,T} = -\kappa(t/T)X_{t,T}\,\mathrm{d}t + \gamma(t/T)\,\mathrm{d}W_t$. At the rescaled time-point $u_0 = t_0/T$, we consider the process solution to $\mathrm{d}\tilde{X}_t(u_0) = -\kappa(u_0)\tilde{X}_t(u_0)\,\mathrm{d}t + \gamma(u_0)\,\mathrm{d}W_t$. If $\kappa(u_0) > 0$, the process $\{\tilde{X}_t(u_0)\}$ is stationary (Kessler & Rahbek, 2004). To show the local stationarity of $\{X_t\}$, we cannot readily apply Proposition 1, since Assumption A1(b) is not satisfied, for instance. However, we can still prove the result under very reasonable conditions. Specifically, for $m = 1, 2, 3$ and constants C_κ and C_γ, we assume that $\inf_{u\in[0,1]} \kappa(u) > 0$, $\inf_{u\in[0,1]} \gamma_i(u) > 0$, $\sup_{u\in[0,1]} |\partial_u^m \kappa(u)| < C_\kappa$ and $\sup_{u\in[0,1]} |\partial_u^m \gamma_i(u)| < C_\gamma$. The conditions on the time derivatives prevent the time-varying parameters from changing abruptly. Next, we illustrate how local stationarity can be proved under these assumptions. For a fixed u_0, an application of the triangle inequality yields $\|X_t - \tilde{X}_t(u_0)\| \leqslant \|X_t - \tilde{X}_t(t/T)\| + \|\tilde{X}_t(t/T) - \tilde{X}_t(u_0)\|$. To prove local stationarity, we should show that $\|\tilde{X}_t(t/T) - \tilde{X}_t(u_0)\| = O_{\mathrm{p}}(|t/T - u_0|)$ while $\|X_t - \tilde{X}_t(t/T)\| = O_{\mathrm{p}}(1/T)$. A similar approach applied to univariate time-inhomogeneous mean-reverting diffusions is presented in Koo & Linton (2012). However, we remark that here we may exploit the available functional form of the drift and diffusion matrix and the smoothness of the time-varying parameters. Thus, proving $\|X_t - \tilde{X}_t(u_0)\| = O_{\mathrm{p}}(1/T + |t/T - u_0|)$ becomes more straightforward.

To see this, first notice that Corollary 8.2.5 in Arnold (1974) implies that the exact discrete-time version of the continuous-time process is a time-varying autoregressive process of order 1 with Gaussian innovations. Specifically, for $\varepsilon_t \sim N(0,1)$ we have $X_{t,T} = A(t/T)X_{t-1,T} + \Sigma^{1/2}(t/T)\varepsilon_t$ where $A(t/T) = \exp\{-\varkappa(t,T)\}$, with $\varkappa(t,T) = \int_{t-1}^{t} \kappa(s/T)\,\mathrm{d}s$, and

$$\Sigma(t/T) = \int_{t-1}^{t} \exp\left\{-2\int_{s}^{t} \kappa(r/T)\,\mathrm{d}r\right\} \sum_{i=1}^{d} \gamma_i^2(s/T)\,\mathrm{d}s.$$

Similar arguments hold for the approximating process $\tilde{X}_t(u_0) = A(u_0)\tilde{X}_{t-1}(u_0) + \Sigma^{1/2}(u_0)\varepsilon_t$. Second, observe that $\tilde{X}_t(t/T) = \sum_{j=0}^{\infty} A^j(t/T)\Sigma^{1/2}(t/T)\varepsilon_{t-j}$. Repeated substitutions and Theorem 2.3(ii) in Dahlhaus (1996) yield

$$X_{t,T} = \sum_{j=0}^{\infty}\left[\prod_{k=0}^{j-1} A\{(t-k)/T\}\right]\Sigma^{1/2}\{(t-j)/T\}\varepsilon_{t-j}$$
$$= \sum_{j=0}^{\infty} A^j(t/T)\Sigma^{1/2}(t/T)\varepsilon_{t-j} + O_{\rm p}(1/T)$$
$$= \tilde{X}_t(t/T) + O_{\rm p}(1/T). \tag{9}$$

Thus $X_{t,T} = \tilde{X}_t(t/T) + O_{\rm p}(1/T)$, which implies that $\|X_t - \tilde{X}_t(t/T)\| = O_{\rm p}(1/T)$. Using the assumption on the third derivatives of the time-varying parameters, Taylor expansion yields

$$\tilde{X}_t(t/T) - \tilde{X}_t(u_0) = (t/T - u_0)\partial_u \tilde{X}_t(u)\big|_{u=u_0} + \frac{1}{2}(t/T-u_0)^2\partial_u^2\tilde{X}_t(u)\big|_{u=u_0} + O_{\rm p}(t/T-u_0)^3. \tag{10}$$

We remark that $\partial_u\tilde{X}_t(u)$ depends on both $\partial_u A^j(u)$ and $\partial_u\Sigma^{1/2}(u)$. Leibniz's rule and the uniform boundedness of the first derivative of the time-varying parameters imply that $\|\partial_u A(u)\| = O(1)$ and $\|\partial_u\Sigma(u)\| = O(1)$. Moreover, $\rho_1 = \|A(t/T)\| < 1$. Then, for $C^* > 0$, we have

$$\left\|\partial_u\tilde{X}_t(u)\big|_{u=u_0}\right\|$$
$$\leqslant \left\|\sum_{j=0}^{\infty}\left\{jA^{j-1}(u_0)\partial_u A(u)\big|_{u=u_0}\Sigma^{1/2}(u_0) + A^j(u_0)\partial_u\Sigma^{1/2}(u)\big|_{u=u_0}\right\}\varepsilon_{t-j}\right\|$$
$$\leqslant C^*\sum_{j=0}^{\infty}(j\rho_1^{-1}+1)\rho_1^j\,\|\varepsilon_{t-j}\|, \tag{11}$$

implying that $\|\partial_u\tilde{X}_t(u)|_{u=u_0}\| = O_{\rm p}(1)$. A similar calculation involving the second derivative of $A(u)$ and $\Sigma(u)$ implies that $\|\partial_u^2\tilde{X}_t(u)|_{u=u_0}\| = O_{\rm p}(1)$. These results yield $\|X_t - \tilde{X}_t(u_0)\| = O_{\rm p}(1/T + |t/T - u_0|)$, even when Assumption A1(b) does not hold.

5.2. *Estimation and consistency*

Given a discrete sample, an estimate of $\theta = \{\kappa(t_0/T), \gamma_1^2(t_0/T) + \cdots + \gamma_d^2(t_0/T)\}^{\rm T}$ at a fixed time-point t_0 can be obtained via the segment maximum likelihood estimator, defined as follows. The transition density of $X_{t,T}$ is $p_{t,t-1}(X_{t,T}; X_{t-1,T}, \theta)$, representing the probability density function of $N\{A(t/T)X_{t-1,T}, \Sigma(t/T)\}$, with $A(t/T)$ and $\Sigma(t/T)$ as in the previous subsection. Hence, let $\ell_t(\theta) = \log p_{t,t-1}(X_{t,T}; X_{t-1,T}, \theta)$. For $|u_0 - t_0/T| < 1/T$, we approximate $\ell_t(\theta)$ by $\tilde{\ell}_t(u_0, \theta) = \log p_{t,t-1}\{\tilde{X}_t(u_0); \tilde{X}_{t-1}(u_0), \theta\}$, which has the same functional form as $\ell_t(\theta)$, but with $N\{A(t/T)X_{t-1,T}, \Sigma(t/T)\}$ replaced by $N\{A(u_0)\tilde{X}_{t-1}(u_0), \Sigma(u_0)\}$. The semiparametric segment M-estimator is the solution to (6), and it has a bias, due to the nonstationarity, expressed as $B_{t_0,T}(\theta) = (bT)^{-1}\sum_{t=1}^{T} K\{(t_0-t)/(bT)\}\nabla_{\theta^{\rm T}}\{\ell_t(\theta) - \tilde{\ell}_t(u_0,\theta)\}$.

To prove consistency of the resulting estimator, we have to show that $B_{t_0,T}(\theta) = O_{\rm p}(b)$; see Proposition 2. In the general case treated in the Supplementary Material, we explain how this can be proved by imposing the Lipschitz condition in Assumption A2(d). This condition is not satisfied by the score function of the process considered in this section, since $\nabla_{\theta^{\rm T}}\ell_t(\theta)$ is polynomial in $X_{t,T}$ and in $X_{t-1,T}$. Nevertheless, $B_{t_0,T}(\theta) = O_{\rm p}(b)$ can be proved using the restrictions on the derivatives of the time-varying parameters. To see this, first we rewrite

$$B_{t_0,T}(\theta) = (bT)^{-1} \sum_{t=1}^{T} K\{(t_0 - t)/(bT)\} \nabla_{\theta^{\mathrm{T}}} \left\{\tilde{\ell}_t(t/T, \theta) - \tilde{\ell}_t(u_0, \theta)\right\} + R_T,$$

where $R_T = (bT)^{-1} \sum_{t=1}^{T} K\{(t_0 - t)/(bT)\} \nabla_{\theta^{\mathrm{T}}}\{\ell_t(\theta) - \tilde{\ell}_t(t/T, \theta)\}$. The polynomial structure of the likelihood score and (9) imply that $\|R_T\| = O_{\mathrm{p}}(1/T)$. Thus, in the bias, we replace $\nabla_{\theta^{\mathrm{T}}}\{\ell_t(\theta)\}$ with $\nabla_{\theta^{\mathrm{T}}}\{\tilde{\ell}_t(t/T, \theta)\}$ and obtain

$$\begin{aligned}\nabla_{\theta^{\mathrm{T}}}\{\tilde{\ell}_t(t/T, \theta) - \tilde{\ell}_t(u_0, \theta)\} = {} & \partial_u \nabla_{\theta^{\mathrm{T}}}\{\tilde{\ell}_t(u, \theta)\}\big|_{u=u_0} (t/T - u_0) \\ & + \frac{1}{2} \partial_u^2 \nabla_{\theta^{\mathrm{T}}}\{\tilde{\ell}_t(u, \theta)\}\big|_{u=u_0} (t/T - u_0)^2 \\ & + \frac{1}{6} \partial_u^3 \nabla_{\theta^{\mathrm{T}}}\{\tilde{\ell}_t(u, \theta)\}\big|_{u=\tilde{U}_t^*} (t/T - u_0)^3, \end{aligned} \qquad (12)$$

with the random variable $\tilde{U}_t^* \in (0, 1]$. The chain rule and the total derivative formula yield

$$\partial_u \nabla_{\theta^{\mathrm{T}}}\{\tilde{\ell}_t(u_0, \theta)\}\big|_{u=u_0} = \sum_{i=1}^{2} \left[\partial\{\partial_{\theta_i} \tilde{\ell}_t(u, \theta)\}/\partial \tilde{X}_t(u)\right] \partial_u \tilde{X}_t(u)\big|_{u=u_0}. \qquad (13)$$

Since $\partial_u \tilde{X}_t(u)$ is as in (11), the process in (13) is stationary, with constant expected value. Similar considerations hold for the terms in (12) involving the second and third derivatives. Now we observe that the domain of the kernel gives the inequality $|t/T - u_0| \leqslant 1/T + b/2$ whenever $|u_0 - t_0/T| < 1/T$. Plugging this expression into (12) and using the boundedness of the derivatives of the time-varying parameters, as in (10), we obtain

$$\begin{aligned}\|B_{t_0,T}(\theta)\| \leqslant \Bigg\| (1/T + b/2)(bT)^{-1} \sum_{t=1}^{T} K\{(t_0 - t)/(bT)\} \Big[& \partial_u \nabla_{\theta^{\mathrm{T}}}\{\tilde{\ell}_t(u, \theta)\}\big|_{u=u_0} \\ & + \frac{1}{2} \partial_u^2 \nabla_{\theta^{\mathrm{T}}}\{\tilde{\ell}_t(u, \theta)\}\big|_{u=u_0} (1/T + b/2) + O_{\mathrm{p}}(b^2) \Big] \Bigg\|.\end{aligned}$$

An application of Lemma A.2 in Dahlhaus & Subba Rao (2006) gives that the term $(bT)^{-1} \sum_{t=1}^{T} K\{(t_0 - t)/(bT)\} \partial_u \nabla_{\theta^{\mathrm{T}}}\{\tilde{\ell}_t(u, \theta)\}|_{u=u_0}$ is $O_{\mathrm{p}}(1)$. A similar argument applies to the term involving $\partial_u^2 \nabla_{\theta^{\mathrm{T}}}\{\tilde{\ell}_t(u, \theta)\}|_{u=u_0}$. So, finally, we conclude that $B_{t_0,T}(\theta) = O_{\mathrm{p}}(b)$ as required, even though Assumption A2(d) is not satisfied.

6. Monte Carlo simulations

6.1. *Time-inhomogeneous Cox–Ingersoll–Ross diffusion*

We consider a time-inhomogeneous version of the process introduced by Cox et al. (1985) which is the solution to $\mathrm{d}X_t = \kappa(t/T)\{\alpha(t/T) - X_t\}\,\mathrm{d}t + \gamma(t/T) {X_t}^{1/2}\,\mathrm{d}W_t$. The approximating process, indexed by u_0, is the solution to

$$\mathrm{d}\tilde{X}_t(u_0) = \kappa(u_0)\{\alpha(u_0) - \tilde{X}_t(u_0)\}\,\mathrm{d}t + \gamma(u_0) {\tilde{X}_t(u_0)}^{1/2}\,\mathrm{d}W_t. \qquad (14)$$

We set $\theta(u_0) \equiv \theta = (\kappa, \alpha, \gamma^2)^{\mathrm{T}}$ and specify the parameter curves as $\kappa(t/T) = 0.4 - 0.1(t/T) + 0.05(t/T)^2$, $\alpha(t/T) = 0.1 + 0.3(t/T) + 0.1(t/T)^2$ and $\gamma^2(t/T) = 0.075[\cos\{0.7\pi(t/T)\} +$

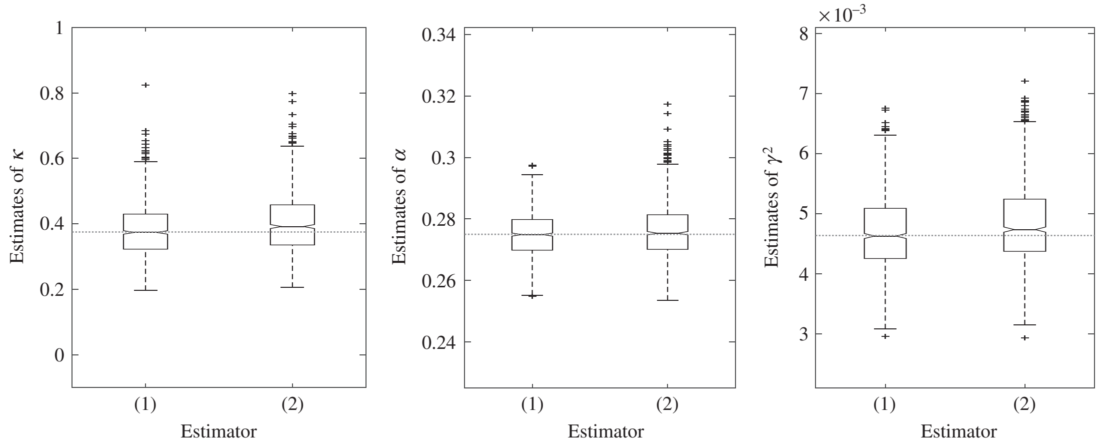

Fig. 1. Boxplots for $\hat{\theta}(0.5)$ with $T = 2000$; in each panel the dotted central line represents the true parameter value, (1) corresponds to the semiparametric segment M-estimator based on ψ^*, and (2) corresponds to the segment maximum likelihood estimator.

$\sin\{0.3\pi(t/T)\}]^2$. For this mean-reverting diffusion, local stationarity follows from the smoothness of the time-varying parameters and can be established as in § 5; see also Theorem 1 and the numerical examples in Koo & Linton (2012).

The transition probability density of the process characterized by (14) is noncentral χ^2 (Shreve, 2004). Inference on θ can be conducted using the segment maximum likelihood estimator, as obtained via the loglikelihood score $\nabla_{\theta^{\mathrm{T}}}\ell_t(\theta)$. However, the solution of the corresponding estimating equations can be unstable. To overcome this problem in the stationary setting, an alternative estimator proposed by Bibby et al. (2010) is based on an optimal martingale estimating function. The method proposed in § 3 extends this construction to the locally stationary setting. The optimal semiparametric segment M-estimator obtained using ψ^* is defined by

$$\exp(-\hat{\kappa}) = \frac{\sum_{t=1}^{T} w_{t_0,t}X_t/X_{t-1} - \left(\sum_{t=1}^{T} w_{t_0,t}X_t\right)\left(\sum_{t=1}^{T} w_{t_0,t}/X_{t-1}\right)}{1 - \left(\sum_{t=1}^{T} w_{t_0,t}X_{t-1}\right)\left(\sum_{t=1}^{T} w_{t_0,t}/X_{t-1}\right)},$$

$$\hat{\alpha} = \{1 - \exp(-\hat{\kappa})\}^{-1}\sum_{t=1}^{T} w_{t_0,t}\left\{X_t - \exp(-\hat{\kappa})X_{t-1}\right\},$$

$$\hat{\gamma}^2 = \hat{\kappa}\frac{\sum_{t=1}^{T}(w_{t_0,t}/X_{t-1})\{X_t - \hat{\alpha} - \exp(-\hat{\kappa})(X_{t-1} - \hat{\alpha})\}^2}{\sum_{t=1}^{T}(w_{t_0,t}/X_{t-1})\{(\hat{\alpha}/2 - X_{t-1})\exp(-2\hat{\kappa}) - (\hat{\alpha} - X_{t-1})\exp(-\hat{\kappa}) + \hat{\alpha}/2\}},$$

with $w_{t_0,t} = K\{(t_0 - t)/(bT)\}/\sum_{s=1}^{T} K\{(t_0 - s)/(bT)\}$.

The availability of these expressions is a perk of our method, which is simpler and faster to implement than maximum likelihood. One might wonder if the computational advantages come with a cost in terms of accuracy. Therefore, we study numerically the effects of replacing $\nabla_{\theta^{\mathrm{T}}}\ell_t(\theta)$ by ψ^*. In Fig. 1 we display boxplots for the segment estimator based on ψ^* and for the segment maximum likelihood estimator. We consider $\theta(u_0)$ at $u_0 = 0.5$, a bandwidth of $b = 0.05$ and a sample size of $T = 2000$. The plots provide numerical evidence that the two estimators perform similarly in terms of bias and variance, with the segment maximum likelihood estimator having more outliers due to its numerical instability. Thus, in this example, replacing $\nabla_{\theta^{\mathrm{T}}}\ell_t(\theta)$ with ψ^* yields numerical advantages, with little cost in terms of efficiency.

6.2. *Time-inhomogeneous Jacobi diffusion*

We consider a time-inhomogeneous Jacobi diffusion over the standard simplex in $\mathbb{R}^3$. The time-homogeneous version of this process was used by Gouriéroux & Jasiak (2006) to model the dynamics of a discrete probability distribution. The process is a polynomial diffusion on the compact state space $\mathcal{S} = \{(x_1, x_2, x_3)^{\mathrm{T}} \in \mathbb{R}^3_+ : x_1 + x_2 + x_3 = 1\}$. Because of the constraint on the coordinates, the process is characterized by

$$\mu\{\theta(t/T), x_1, x_2\} = \kappa(t/T) \begin{Bmatrix} \alpha_1(t/T) - x_1 \\ \alpha_2(t/T) - x_2 \end{Bmatrix}$$

and

$$\sigma\{\theta(t/T), x_1, x_2\} = \gamma(t/T) \begin{Bmatrix} (1-x_1)x_1^{1/2} & -x_1 x_2^{1/2} & -x_1(1-x_1-x_2)^{1/2} \\ -x_2 x_1^{1/2} & (1-x_2)x_2^{1/2} & -x_2(1-x_1-x_2)^{1/2} \end{Bmatrix}.$$

We specify the following time-varying parameters: $\kappa(t/T) = 0.2 + 0.1(t/T) + 0.05(t/T)^2$, $\alpha_1(t/T) = 0.2 + 0.05(t/T) + 0.02(t/T)^2$, $\alpha_2(t/T) = 0.3 + 0.05(t/T) + 0.05(t/T)^2$ and $\gamma(t/T) = 0.1 + 0.05(t/T) + 0.05(t/T)^2$. These specifications guarantee the existence of the stationary approximations, where the drift and diffusion are defined analogously (Gouriéroux & Jasiak, 2006). Formal verification of the conditions needed to apply Propositions 1–3 follows the same kind of steps as in § 5, but is much more cumbersome. Our simulation study demonstrates that our procedure yields reliable results.

We evaluate the performance of two semiparametric segment M-estimators, defined using two choices of ψ. The first is based on the optimal martingale estimating function obtained from the first two eigenfunctions of the infinitesimal generators of the approximating family of stationary processes. The second estimating function is based on the Gaussian pseudo maximum likelihood method. Both martingale estimating functions are polynomial in x_1 and x_2; see the Supplementary Material. The proposed comparison is of interest since the bias due to nonstationarity depends on ψ. It may be advantageous to use a ψ that entails a smaller bias, even if it leads to a higher variance; this trade-off depends on the bandwidth.

To study these aspects, we simulate 8000 sample paths with an Euler scheme; we use a total of 100 000 points and obtain the data by sampling $T = 5000$ equidistant points. We use an Epanechnikov kernel with two values of the bandwidth: a small value $b = 0.1$ and a large value $b = 0.5$. For comparison, we consider the mean squared error at the rescaled time-point $u_0 = 0.5$. Table 1 shows that the two estimators exhibit comparable performance. This is an interesting finding, since pseudo maximum likelihood estimation is easy to implement.

To investigate further the behaviour of the proposed estimators, we simulate 100 trajectories with a moderate sample size of $T = 1000$. Figure 2 shows the functional boxplots for the estimator defined by ψ^*, for $b = 0.1$. We display the results for the estimates of α_1 and γ^2. Each plot summarizes the 100 estimated parameter curves, treated as functional data. Displayed are the median function, the 50% central region, and the functional limits based on the 25th and 75th functional percentiles. The centred outward ordering is induced by band depth for functional data; see Sun & Genton (2011). The plots illustrate the location and scatter of the estimated parameter curves; in each panel, the true parameter curve lies within the central region.

7. Electroencephalogram data analysis

Electroencephalograms, EEGs, are noninvasive measures of electrical activity in the brain, recorded by electrodes on the scalp of a patient. We analyse EEG signals from a patient with

Table 1. *Estimation of the time-varying parameters for the time-inhomogeneous Jacobi diffusion at $u_0 = 0.5$: different values of the bandwidth b are used; when a value has its first nonzero digit six places after the decimal point, it is written as $<10^{-5}$*

	Optimal estimating function		Pseudo maximum likelihood		Relative
	Bias	Standard error	Bias	Standard error	mean squared error
	(10^{-3})	(10^{-3})	(10^{-3})	(10^{-3})	
$\hat{\kappa}$ $(b = 0.1)$	5.59	19.9	5.24	20.2	0.98
$\hat{\alpha}_1$ $(b = 0.1)$	$<10^{-5}$	7.27	0.05	7.41	0.96
$\hat{\alpha}_2$ $(b = 0.1)$	0.07	8.23	0.10	8.32	0.98
$\hat{\gamma}^2$ $(b = 0.1)$	0.33	0.75	0.32	0.75	1.00
$\hat{\kappa}$ $(b = 0.5)$	−13.0	8.50	−12.5	8.70	1.04
$\hat{\alpha}_1$ $(b = 0.5)$	2.70	3.32	1.95	3.33	1.19
$\hat{\alpha}_2$ $(b = 0.5)$	3.50	3.37	2.70	3.38	1.21
$\hat{\gamma}^2$ $(b = 0.5)$	1.25	0.39	1.25	0.40	0.99

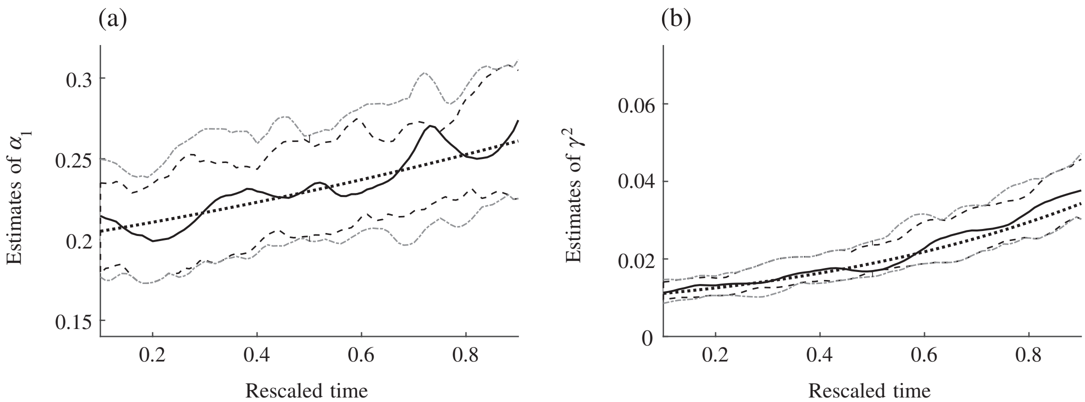

Fig. 2. Functional boxplots for the estimates of (a) α_1 and (b) γ^2: median function (solid line), the 50% central region (area between dashed lines), the whiskers (dash-dotted lines), and the true time-varying parameter (dotted line).

epilepsy; there are three main statistical challenges. First, brain activity happens in continuous time, but only discrete-time observations are available. Second, people with epilepsy suffer from seizures, i.e., transient aberrations in the brain's electrical activity. Third, brain regions do not act in isolation, and synchronous activations have to be modelled. We demonstrate how our method can be used to model EEG signals. Viewing the data as discrete-time observations of a time-inhomogeneous multivariate diffusion, we obtain estimates of the model parameters characterizing the continuous-time dynamics of the EEG signal in different brain areas, while using a parsimonious and interpretable model. Data are taken from the CHB–MIT Scalp EEG database; see Goldberger et al. (2000) and the description given in the 2009 Massachusetts Institute of Technology PhD thesis of A. H. Shoeb. We consider $T = 3600$ seconds of EEG recording from Patient 1, an 11-year-old female, and focus on the channels FP2-F4 and T8-P8. According to a specialist, the patient had a seizure for about 40 seconds. Figure 3 shows the signals and the seizure time frame: the signals oscillate about zero, with co-movements and variance that seem to vary over rescaled time.

We model the signals with the simple, yet flexible, time-inhomogeneous bivariate Ornstein–Uhlenbeck process solution to

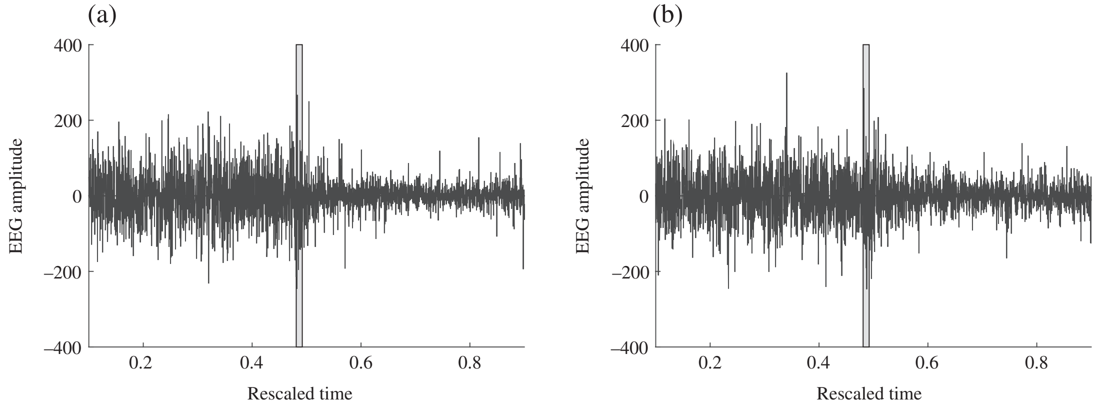

Fig. 3. EEG for channels (a) FP2-F4 and (b) T8-P8; the shaded areas indicate the seizure time frame.

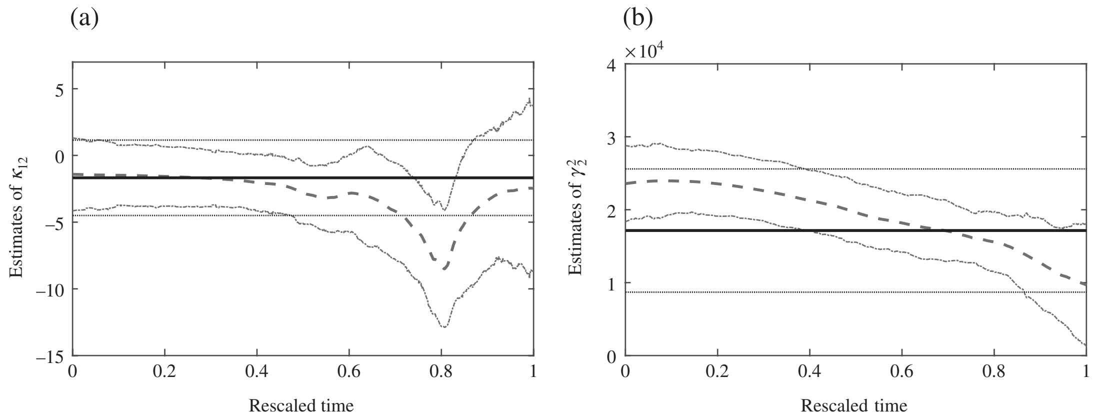

Fig. 4. Time-varying (dashed) and constant (solid) estimates of (a) κ_{12} and (b) γ_2^2, along with their 95% confidence intervals (bounded by dashed lines for time-varying and dotted lines for constant estimates).

$$\mathrm{d}X_t = \begin{Bmatrix} \kappa_{11}(t/T) & \kappa_{12}(t/T) \\ 0 & \kappa_{22}(t/T) \end{Bmatrix} \begin{Bmatrix} \alpha_1(t/T) - X_{1,t} \\ \alpha_2(t/T) - X_{2,t} \end{Bmatrix} \mathrm{d}t + \begin{Bmatrix} \gamma_1(t/T) & 0 \\ 0 & \gamma_2(t/T) \end{Bmatrix} \mathrm{d}W_t, \tag{15}$$

where $\kappa_{12}(\cdot)$ represents the time-varying dependence between the series. Our semiparametric segment M-estimator is available in closed form; see the Supplementary Material.

Figure 4 shows the estimated parameter curves for $\kappa_{12}(\cdot)$ and $\gamma_2^2(\cdot)$, with the corresponding 95% confidence intervals implied by Proposition 3. For comparison, we superimpose the constant parameter estimates and their confidence intervals. The first plot illustrates that the intensity of the co-movements changes over time, becoming stronger around the seizure and then reverting to pre-seizure values. The constant parameter estimation procedure fails to reveal this feature: it yields an estimate not significantly different from zero, over the whole time span. Figure 4(b) confirms that the signal variability is higher before the seizure, so the variance of the signal is changing over time. The constant parameter estimation procedure does not capture this aspect: it yields a unique estimate, representing a kind of average of the time-varying estimates.

Figure 5 shows time-frequency plots of the spectrum for the second channel and for the coherence, obtained using the estimated parameters of the process solution to (15).

Figure 6 illustrates the results obtained with the smooth localized complex exponential transform method of Ombao et al. (2001, 2005), which represents a benchmark technique for the

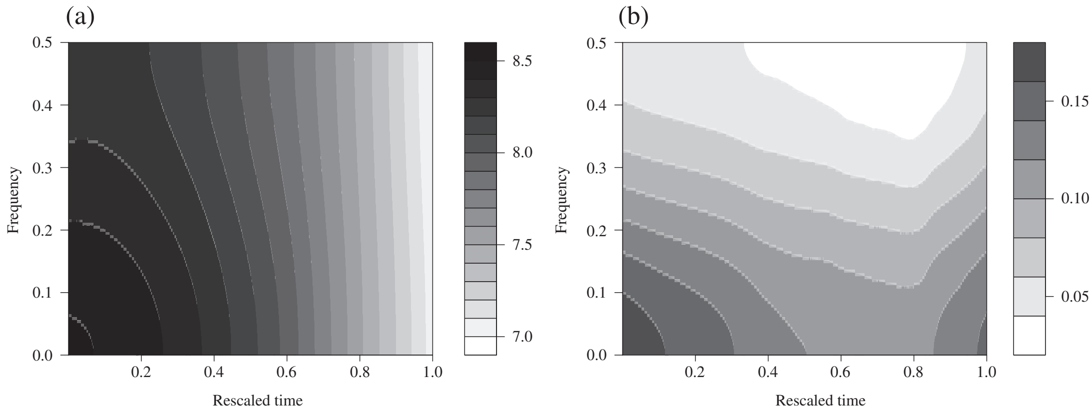

Fig. 5. Time-frequency plots of the spectrum, in log base, given by the Ornstein–Uhlenbeck model for (a) the second channel and (b) the coherence.

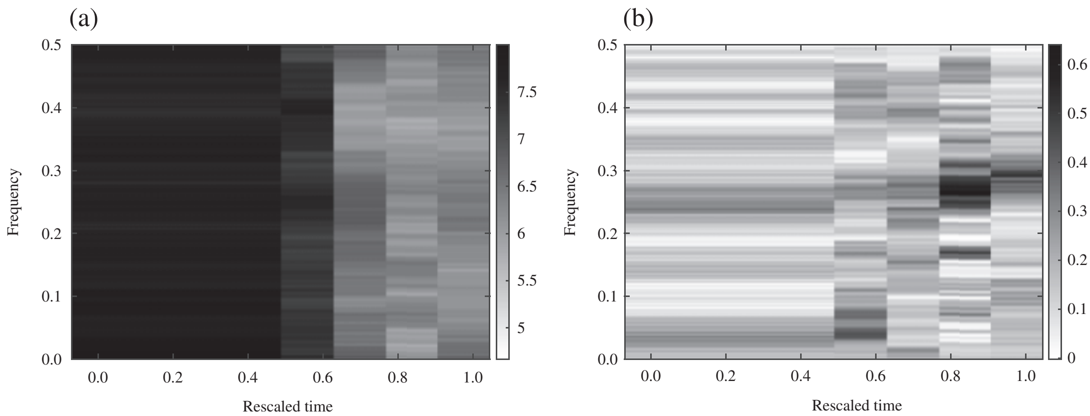

Fig. 6. Time-frequency plots of the spectrum, in log base, given by the method of Ombao et al. (2001, 2005) for (a) the second channel and (b) the coherence.

analysis of EEG data recorded from epileptic patients and is closely related to our definition of local stationarity, being based on a segmentation of the signals into stationary blocks. A comparison of Figs. 5 and 6 suggests that the plots convey similar information on the time evolution of the spectrum. However, our approach offers the possibility of conducting both time-domain and spectral analyses, while the method of Ombao et al. (2001, 2005) is designed specifically for the frequency domain. The Supplementary Material contains additional comparisons with the spectrum obtained using the discrete Fourier transform.

Acknowledgement

We thank Anthony C. Davison, the associate editor and the referees for comments that have led to significant improvements of the paper. We are grateful to Rainer Dahlhaus, Bonsoo Koo, Michael Sørensen, Howell Tong and Michael Vogt for stimulating discussions. The second author acknowledges financial support from the Swiss National Science Foundation.

Supplementary material

Supplementary material available at *Biometrika* online includes proofs of the theoretical results, further theoretical developments for the selection of the bandwidth, and additional simulations.

Appendix

Here we state the assumptions for our theoretical results.

Assumption A1.

(a) The functions μ and σ are Lipschitz continuous with respect to their first argument, i.e., there exist constants L^μ and L^σ such that for any $u_1, u_2 \in [0,1]$ and $x \in \mathcal{S}$, $\|\mu(u_1,x) - \mu(u_2,x)\| \leqslant L^\mu |u_1 - u_2|$ and $\|\sigma(u_1,x) - \sigma(u_2,x)\| \leqslant L^\sigma |u_1 - u_2|$.

(b) There exist positive constants C^μ, c^σ and C^σ such that for any $u \in [0,1]$ and $x \in \mathcal{S}$, $\|\mu(u,x)\|_\infty \leqslant C^\mu$ and $c^\sigma \leqslant \|\sigma(u,x)\|_\infty \leqslant C^\sigma$.

(c) The functions μ and σ are continuously differentiable with respect to their second argument. In addition, for any positive constant C, let $A_C = \{(u,x) : u \in [0,1], \|x\|_\infty > C\}$. Then there exists $\underline{\delta} = \underline{\delta}(C) < 1$ such that for all $p, q, r \in \{1, \ldots, d\}$, $-1 - \underline{\delta} \leqslant \sup_{(u,x)\in A_C} \partial_{x_p} \mu_p(u,x) \leqslant -1 + \underline{\delta}$, $\sup_{(u,x)\in A_C} |\partial_{x_q} \mu_p(u,x)| \leqslant \underline{\delta}$ for $p \neq q$, and $\sup_{(u,x)\in A_C} |\partial_{x_q} \sigma_{p,r}(u,x)| \leqslant \underline{\delta}$.

Assumption A2.

(a) The parameter space Θ is compact.

(b) The equation $G(u_0, \theta) = 0$ admits a unique solution $\theta_0 \equiv \theta_0(u_0) \in \text{int}(\Theta)$.

(c) The kernel $K(\cdot)$ has support $[-1/2, 1/2]$, is centred at zero, and is of bounded variation.

(d) The function ψ is twice continuously differentiable with respect to θ. Moreover, it is Lipschitz continuous in each of its first two arguments, i.e., there exist constants $C_1, C_2 < \infty$ such that $\|\psi(y_1, x, \theta) - \psi(y_2, x, \theta)\| \leqslant C_1 \|y_1 - y_2\|$ for all $y_1, y_2, x \in \mathcal{S}$ and $\theta \in \Theta$, and $\|\psi(y, x_1, \theta) - \psi(y, x_2, \theta)\| \leqslant C_2 \|x_1 - x_2\|$ for all $x_1, x_2, y \in \mathcal{S}$ and $\theta \in \Theta$.

(e) The function $\nabla_{\theta^{\mathrm{T}}} \psi(\cdot)$ is Lipschitz continuous in each of its first two arguments. In addition, $E[\sup_{\theta\in\Theta} \|\nabla_{\theta^{\mathrm{T}}} \psi\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta\}\|] < \infty$.

(f) The matrix $\Omega(u_0) = E[\nabla_{\theta^{\mathrm{T}}} \psi\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta_0(u_0)\}^{\mathrm{T}}]$ is invertible.

(g) The matrix $\Upsilon(u_0) = E[\psi\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta_0(u_0)\}\psi\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta_0(u_0)\}^{\mathrm{T}}]$ exists.

(h) For every $\epsilon > 0$ and every $j \in \{1, \ldots, p\}$,

$$\frac{1}{bT}\sum_{t=1}^{T} K\left(\frac{t_0 - t}{bT}\right)^2 E\Big[\psi_j^2\{\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta_0(u_0)\}\mathbb{1}_{\{|\psi_j(\tilde{X}_t(u_0), \tilde{X}_{t-1}(u_0), \theta_0(u_0))| > \epsilon\}} \;\Big|\; \tilde{X}_{t-1}(u_0)\Big]$$

converges in probability to 0.

(i) There exist stationary processes $[\partial_u^i \psi\{\tilde{X}_t(u), \tilde{X}_{t-1}(u), \theta_0(u_0)\}|_{u=u_0}]$, $i \in \{1,2,3\}$, such that

$$\sum_{k=0}^{\infty} \Big\|\text{cov}\Big[\partial_u^i \psi\{\tilde{X}_t(u), \tilde{X}_{t-1}(u), \theta_0(u_0)\}\Big|_{u=u_0}, \partial_u^i \psi\{\tilde{X}_{t+k}(u), \tilde{X}_{t+k-1}(u), \theta_0(u_0)\}\Big|_{u=u_0}\Big]\Big\| < \infty$$

for $i \in \{1, 2\}$ and $E[\sup_{u\in[0,1]} \|\partial_u^3 \psi\{\tilde{X}_t(u), \tilde{X}_{t-1}(u), \theta_0(u_0)\}\|] < \infty$.

References

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica* **70**, 223–62.

Aït-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *Ann. Statist.* **36**, 906–37.

Aït-Sahalia, Y. & Mykland, P. (2004). Estimators of diffusions with randomly spaced discrete observations: A general theory. *Ann. Statist.* **32**, 2186–222.

Arnold, L. (1974). *Stochastic Differential Equations: Theory and Applications*. New York: Wiley.

Bandi, F. M. & Phillips, P. C. B. (2010). Nonstationary continuous-time processes. In *Handbook of Financial Econometrics*, Y. Aït-Sahalia & L. P. Hansen, eds., vol. 1. Amsterdam: North-Holland, pp. 139–201.

Bibby, B., Jacobsen, M. & Sørensen, M. (2010). Estimating functions for discretely sampled diffusion-type models. In *Handbook of Financial Econometrics*, Y. Aït-Sahalia & L. P. Hansen, eds., vol. 1. Amsterdam: North-Holland, pp. 203–68.

Cox, J. C., Ingersoll Jr, J. E. & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica* **53**, 385–407.

Dahlhaus, R. (1996). Maximum likelihood estimation and model selection for locally stationary processes. *J. Nonparam. Statist.* **6**, 171–91.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25**, 1–37.

Dahlhaus, R. (2012). Locally stationary processes. In *Handbook of Statistics*, T. Subba Rao, S. Subba Rao & C. R. Rao, eds., vol. 30. Amsterdam: North-Holland, pp. 351–412.

Dahlhaus, R. & Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist.* **34**, 1075–114.

Fan, J., Jiang, J., Zhang, C. & Zhou, Z. (2003). Time-dependent diffusion models for term structure dynamics. *Statist. Sinica* **13**, 965–92.

Filipović, D. & Larsson, M. (2016). Polynomial diffusions and applications in finance. *Finan. Stoch.* **20**, 931–72.

Fryzlewicz, P., Sapatinas, T. & Subba Rao, S. (2008). Normalized least-squares estimation in time-varying ARCH models. *Ann. Statist.* **36**, 742–86.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K. & Stanley, H. E. (2000). Physiobank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **101**, e215–20.

Gouriéroux, C. & Jasiak, J. (2006). Multivariate Jacobi process with application to smooth transitions. *J. Economet.* **131**, 475–505.

Gouriéroux, C., Monfort, A. & Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica* **52**, 681–700.

Hafner, C. M. & Linton, O. (2010). Efficient estimation of a multivariate multiplicative volatility model. *J. Economet.* **159**, 55–73.

Heyde, C. C. (1997). *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation*. New York: Springer.

Hull, J. & White, A. (1990). Pricing interest-rate-derivative securities. *Rev. Finan. Stud.* **3**, 573–92.

Kessler, M., Lindner, A. & Sørensen, M. (2012). *Statistical Methods for Stochastic Differential Equations*. Boca Raton, Florida: CRC Press.

Kessler, M. & Rahbek, A. (2004). Identification and inference for multivariate cointegrated and ergodic Gaussian diffusions. *Statist. Infer. Stoch. Proces.* **7**, 137–51.

Kessler, M. & Sørensen, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli* **5**, 299–314.

Koo, B. & Linton, O. (2012). Estimation of semiparametric locally stationary diffusion models. *J. Economet.* **170**, 210–33.

La Vecchia, D. & Trojani, F. (2010). Infinitesimal robustness for diffusions. *J. Am. Statist. Assoc.* **105**, 703–12.

Larsson, M. & Pulido, S. (2017). Polynomial diffusions on compact quadric sets. *Stoch. Proces. Appl.* **127**, 901–26.

Li, C. (2013). Maximum-likelihood estimation for diffusion processes via closed-form density expansions. *Ann. Statist.* **41**, 1350–80.

Ombao, H., Raz, J. A., von Sachs, R. & Malow, B. A. (2001). Automatic statistical analysis of bivariate nonstationary time series. *J. Am. Statist. Assoc.* **96**, 543–60.

Ombao, H., von Sachs, R. & Guo, W. (2005). SLEX analysis of multivariate nonstationary time series. *J. Am. Statist. Assoc.* **100**, 519–31.

Park, J. (2014). Nonstationary nonlinearity: A survey on Peter Phillips's contributions with a new perspective. *Economet. Theory* **30**, 894–922.

Shreve, S. E. (2004). *Stochastic Calculus for Finance II: Continuous-Time Models*. New York: Springer.

Sørensen, M. (1999). On asymptotics of estimating functions. *Braz. J. Prob. Statist.* **13**, 111–36.

Stroock, D. W. & Varadhan, S. S. (2006). *Multidimensional Diffusion Processes*. Berlin: Springer.

Sun, Y. & Genton, M. G. (2011). Functional boxplots. *J. Comp. Graph. Statist.* **20**, 316–34.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *J. Finan. Econ.* **5**, 177–88.

Vogt, M. (2012). Nonparametric regression for locally stationary time series. *Ann. Statist.* **40**, 2601–33.

[*Received on* 11 *July* 2017. *Editorial decision on* 28 *January* 2019]