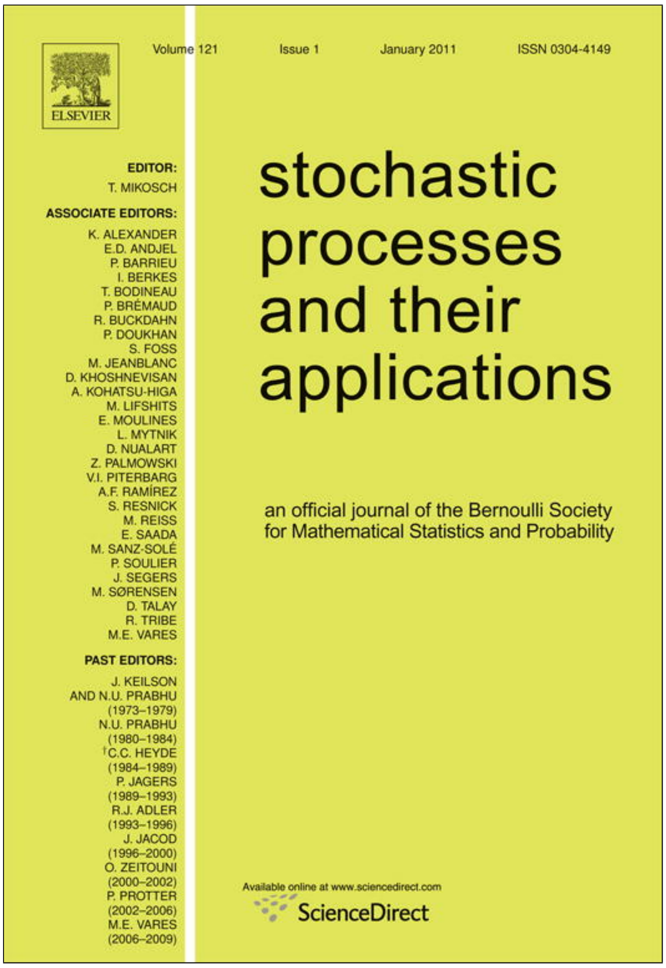
Volume 121
Issue 1
January 2011
ISSN 0304-4149
ELSEVIER
EDITOR:
T. MIKOSCH
ASSOCIATE EDITORS:
K. ALEXANDER
E.D. ANDJEL
P. BARRIEU
I. BERKES
T. BODINEAU
P. BRÉMAUD
R. BUCKDAHN
P. DOUKHAN
S. FOSS
M. JEANBLANC
D. KHOSHNEVISAN
A. KOHATSU-HIGA
M. LIFSHITS
E. MOULINES
L. MYTNIK
D. NUALART
Z. PALMOWSKI
V.I. PITERBARG
A.F. RAMÍREZ
S. RESNICK
M. REISS
E. SAADA
M. SANZ-SOLÉ
P. SOULIER
J. SEGERS
M. SØRENSEN
D. TALAY
R. TRIBE
M.E. VARES
PAST EDITORS:
J. KEILSON
AND N.U. PRABHU
(1973–1979)
N.U. PRABHU
(1980–1984)
†C.C. HEYDE
(1984–1989)
P. JAGERS
(1989–1993)
R.J. ADLER
(1993–1996)
J. JACOD
(1996–2000)
O. ZEITOUNI
(2000–2002)
P. PROTTER
(2002–2006)
M.E. VARES
(2006–2009)
stochastic processes and their applications
an official journal of the Bernoulli Society for Mathematical Statistics and Probability
Available online at www.sciencedirect.com
ScienceDirect

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Stochastic Processes and their Applications 122 (2012) 1068–1092

stochastic processes and their applications

www.elsevier.com/locate/spa

Estimation for the change point of volatility in a stochastic differential equation[☆]

Stefano M. Iacus[a,*], Nakahiro Yoshida[b,c]

[a] Department of Economics, Business and Statistics, University of Milan, Via Conservatorio 7, 20122 Milan, Italy

[b] University of Tokyo, Japan

[c] Japan Science and Technology Agency Graduate School of Mathematical Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan

Received 27 December 2010; received in revised form 15 November 2011; accepted 15 November 2011

Available online 1 December 2011

Abstract

We consider a multidimensional Itô process $Y = (Y_t)_{t\in[0,T]}$ with some unknown drift coefficient process b_t and volatility coefficient $\sigma(X_t, \theta)$ with covariate process $X = (X_t)_{t\in[0,T]}$, the function $\sigma(x, \theta)$ being known up to $\theta \in \Theta$. For this model, we consider a change point problem for the parameter θ in the volatility component. The change is supposed to occur at some point $t^* \in (0, T)$. Given discrete time observations from the process (X, Y), we propose quasi-maximum likelihood estimation of the change point. We present the rate of convergence of the change point estimator and the limit theorems of the asymptotically mixed type.

Keywords: Itô processes; Discrete time observations; Change point estimation; Volatility

1. Introduction

The problem of change point has been considered initially in the framework of independent and identically distributed data by many authors, see e.g. [9,7,5,13]. Recently, it naturally moved to context of time series analysis, see for example, [15,18,3] and the papers cited therein.

☆ This work was in part supported by the JST Basic Research Programs PRESTO, Grants-in-Aid for Scientific Research No. 19340021, the Global COE program "The research and training center for new development in mathematics" of Graduate School of Mathematical Sciences, University of Tokyo, and by the Cooperative Research Program of the Institute of Statistical Mathematics.

* Corresponding author.

E-mail addresses: stefano.iacus@unimi.it (S.M. Iacus), nakahiro@ms.u-tokyo.ac.jp (N. Yoshida).

0304-4149/$ - see front matter © 2011 Elsevier B.V. All rights reserved.

doi:10.1016/j.spa.2011.11.005

In fact, change point problems have originally arisen in the context of quality control, but the problem of abrupt changes in general arises in many contexts like epidemiology, rhythm analysis in electrocardiograms, seismic signal processing, study of archaeological sites and financial markets. In particular, in the analysis of financial time series, the knowledge of the change in the volatility structure of the process under consideration is of a certain interest.

In this paper we deal with a change-point problem for the volatility of a process solution to a stochastic differential equation, when observations are collected at discrete times. The instant of the change in volatility regime is identified retrospectively by maximum likelihood method on the approximated likelihood. For continuous time observations of diffusion processes [19] considered the change point estimation problem for the drift. In the present work we only assume regularity conditions on the drift process. De Gregorio and Iacus [6] considered a least squares approach following the lines of [1,2] of a simplified model also under discrete sampling while [22] considered a CUSUM approach. Finally it should be noted that the problems of the change-point for the drift function of ergodic diffusion processes have been treated by Kutoyants [16,17], however the asymptotic results and the sampling schemes are different from this paper. Notice also that, as usual in change point problems, due to non smoothness of the statistical model with respect to the parameter to be estimated (the change point instant), the rate of convergence of the change point estimator is faster than usual rate of estimators in regular models [12].

The paper is organized as follows. Section 2 introduces the model of observation, the regularity conditions and some notation. We shall treat two asymptotic settings: there are two models before and after the change point. In one case (A) the two models remain distinct in the limit (fixed alternatives), in the second case (B) they get closer and closer (contiguous alternatives, see [20]). Section 3 studies consistency and the rate of convergence of estimator of the change while asymptotic distributions are considered in Section 4. A mixture of certain Wiener functionals appears as the limit of the likelihood ratio random field, and it characterizes the limit distribution of the change-point estimator. Those sections assume that consistent estimators of the volatility parameters are available. Section 5 contains a preliminary interesting inequality which is used to study the asymptotic distribution of the change point estimator in case (A). Section 6 presents some practical considerations and a proposal to obtain first stage estimators of the volatility parameters which allow to obtain all asymptotic properties stated in the previous sections. Finally, Section 7 presents some numerical analysis to assess the performance of the estimators.

2. Estimator for the change-point of the volatility

Consider a d-dimensional Itô process $Y = (Y_t)_{t\in[0,T]}$ satisfying the stochastic differential equation

$$dY_t = b_t dt + \sigma(X_t, \theta) dW_t, \quad t \in [0, T]$$

on a probability space, where W_t is an r-dimensional standard Wiener process, on a stochastic basis, b_t and X_t are vector valued progressively measurable processes, and $\sigma(x,\theta) : \mathcal{X} \times \Theta \to \mathbb{R}^d \otimes \mathbb{R}^r$ is a matrix valued function.

We assume that there is the time t^* across which the diffusion coefficient changes from $\sigma(x, \theta_0)$ to $\sigma(x, \theta_1)$. The change point $t^* \in (0, T)$ is unknown and we want to estimate t^* based on the observations sampled from the path of (X, Y). The coefficient $\sigma(x, \theta)$ is assumed to be known up to the parameter θ, while b_t is completely unknown and unobservable, therefore possibly depending on θ and t^*.

The sample consists of (X_{t_i}, Y_{t_i}), $i = 0, 1, \dots, n$, where $t_i = ih$ for $h = h_n = T/n$. The parameter space Θ of θ is a bounded domain in $\mathbb{R}^{d_0}$, $d_0 \geq 1$, and the parameter θ is a nuisance in estimation of t^*. Denote by θ_i^* the true value of θ_i for $i = 0, 1$.

Let $\vartheta_n = |\theta_1^* - \theta_0^*|$. We will consider the following two different situations.

(A) θ_0^* and θ_1^* are fixed and do not depend on n.
(B) θ_0^* and θ_1^* depend on n, and as $n \to \infty$, $\theta_0^* \to \theta^* \in \Theta$, $\vartheta_n \to 0$ and $n\vartheta_n^2 \to \infty$.

In Case (A), ϑ_n is a constant ϑ_0 independent of n.

We shall formulate the problem more precisely. It will be assumed that the process Y generating the data is an Itô process realized on a stochastic basis $\mathcal{B} = (\Omega, \mathcal{F}, \mathbf{F}, P)$ with filtration $\mathbf{F} = (\mathcal{F}_t)_{t\in[0,T]}$, and satisfies the stochastic integral equation

$$Y_t = \begin{cases} Y_0 + \displaystyle\int_0^t b_s ds + \int_0^t \sigma(X_s, \theta_0^*) dW_s & \text{for } t \in [0, t^*) \\ Y_{t^*} + \displaystyle\int_{t^*}^t b_s ds + \int_{t^*}^t \sigma(X_s, \theta_1^*) dW_s & \text{for } t \in [t^*, T]. \end{cases}$$

Here W_t is an r-dimensional $\mathbf{F}$-Wiener process on $\mathcal{B}$, and b_t, X_t and $\sigma(x, \theta)$ satisfy the conditions below. Let $\mathcal{X}$ be a set in $\mathbb{R}^{d_1}$ (possibly $\mathcal{X} = \mathbb{R}^{d_1}$) and denote the modulus of continuity of a function $f : I \to \mathbb{R}^{d_1}$ by

$$w_I(\delta, f) = \sup_{s,t\in I, |s-t|\leq\delta} |f(s) - f(t)|.$$

For matrices $M = (m_{ij})$ and $N = (n_{ij})$ of the same size, we write $M^{\otimes 2} = M\,{}^{\mathrm{t}}M$, $M[N] = \sum_{ij} m_{ij}n_{ij} = \mathrm{Tr}(M\,{}^{\mathrm{t}}N)$, and the Euclidean norm of M by $|M| = (M[M])^{1/2}$. Set $S(x, \theta) = \sigma(x, \theta)^{\otimes 2}$. $\lambda_1(M)$ denotes the minimum eigenvalue of a symmetric matrix M. Let us denote by $\partial_\theta^\ell f$ the partial derivative of order ℓ of function f with respect to θ. Let α be a positive number.

$[\mathbf{H}]_j$ (i) $\sigma(x, \theta)$ is a measurable function defined on $\mathcal{X} \times \Theta$ satisfying
(a) $\inf_{(x,\theta)\in\mathcal{X}\times\Theta} \lambda_1(S(x, \theta)) > 0$,
(b) derivatives $\partial_\theta^\ell \sigma$ $(0 \leq \ell \leq j + [d_0/2])$ exist and those functions are continuous on $\mathcal{X} \times \Theta$,
(c) there exists a locally bounded function $L : \mathcal{X} \times \mathcal{X} \times \Theta \to \mathbb{R}_+$ such that
$$|\sigma(x, \theta) - \sigma(x', \theta)| \leq L(x, x', \theta)|x - x'|^\alpha \quad (x, x' \in \mathcal{X}, \theta \in \Theta).$$
(ii) $(X_t)_{t\in[0,T]}$ is a progressively measurable process taking values in $\mathcal{X}$ such that
$$w_{[0,T]}\left(\frac{1}{n}, X\right) = o_p(\vartheta_n^{1/\alpha})$$
as $n \to \infty$.
(iii) $(b_t)_{t\in[0,T]}$ is a progressively measurable process taking values in $\mathbb{R}^d$ such that $(b_t - b_0)_{t\in[0,T]}$ is locally bounded.

Remark 1. The term "locally bounded" in $[\mathbf{H}]_j$ (i) (c) means, as usual, being bounded on every compact set, i.e. b_t is locally bounded if there exists a sequence of increasing stopping times s_n such that $b_{s_n\wedge t}$ is bounded. The case where the drift b_t changes its structure at time t^*, or any time in force, is included in our context because b_t admits jumps. The case of time dependent σ is included by making X_t have argument t, i.e. taking X_t as a non-homogeneous process. Needless to say, if we set X or a part of X as Y, then our model can express a system with feedback, in particular, a diffusion process. By $[\mathbf{H}]_j$ (ii), $t \mapsto X_t$ is continuous a.s. Also, $[\mathbf{H}]_j$ (ii) imposes a restriction on the rate ϑ_n. For example, when $\alpha = 1$, for a Brownian motion X, it suffices that $n\vartheta_n^2/\log n \to \infty$, due to Lévy property. The additional $[d_0/2]$ time differentiability to j is used only in Step (iii) of the proof of Theorem 1. If one can introduce a different set of conditions that ensures the Hájek–Renyi type estimate before making use of inequality (4) below, then it is possible to limit the range of ℓ from "$0 \le \ell \le j + [d_0/2]$" to "$0 \le \ell \le j$".

Write $\Delta_i Y = Y_{t_i} - Y_{t_{i-1}}$ and let

$$\Phi_n(t;\theta_0,\theta_1) = \sum_{i=1}^{[nt/T]} G_i(\theta_0) + \sum_{i=[nt/T]+1}^{n} G_i(\theta_1),$$

where

$$G_i(\theta) = \log\det S(X_{t_{i-1}},\theta) + h^{-1} S(X_{t_{i-1}},\theta)^{-1}[(\Delta_i Y)^{\otimes 2}].$$

Suppose that there exists an estimator $\hat\theta_k$ for each θ_k^*, $k = 0, 1$. Each estimator is based on $(X_{t_i}, Y_{t_i})_{i=0,1,\ldots,n}$ and so depends on n. See Section 6 for some discussion on how to obtain consistent estimators $\hat\theta_k$, $k = 0, 1$. To make our discussion complete, in case θ_k^* are known, we define $\hat\theta_k$ just as $\hat\theta_k = \theta_k^*$. This article proposes

$$\hat t_n = \operatorname*{argmin}_{t\in[0,T]} \Phi_n(t;\hat\theta_0,\hat\theta_1)$$

for the estimation of t^*. More precisely, $\hat t_n$ is any measurable function of $(X_{t_i})_{i=0,1,\ldots,n}$ satisfying

$$\Phi_n(\hat t_n;\hat\theta_0,\hat\theta_1) = \min_{t\in[0,T]} \Phi_n(t;\hat\theta_0,\hat\theta_1).$$

Remark that our quasi-likelihood approach generalizes previously proposed methods, see e.g. [8], to the case of stochastic regression models with coordinates (Y_t, X_t).

3. Rate of convergence

We introduce identifiability conditions in order to ensure consistent estimation. In Case (A) we assume

[A] $P\left[S(X_{t^*};\theta_0^*) \neq S(X_{t^*};\theta_1^*)\right] = 1$;

In Case (B) we assume

[B] $\Xi(X_{t^*},\theta^*)$ is positive-definite a.s., where

$$\Xi(x,\theta) = \left(\mathrm{Tr}((\partial_{\theta^{(i_1)}}S)S^{-1}(\partial_{\theta^{(i_2)}}S)S^{-1})(x,\theta)\right)_{i_1,i_2=1}^{d_0}, \quad \theta = (\theta^{(i)}).$$

Remark 2. Since $\Xi(x,\theta^*)$ is the Hessian matrix of the nonnegative function

$$Q(x,\theta^*,\theta) := \mathrm{Tr}\left(S(x,\theta^*)^{-1}S(x,\theta) - I_d\right) - \log\det\left(S(x,\theta^*)^{-1}S(x,\theta)\right)$$

of θ at θ^*, $\Xi(x,\theta^*)$ is nonnegative-definite.

The following property will be necessary to validate our estimating procedure.

[C] $|\hat\theta_k - \theta_k^*| = o_p(\vartheta_n)$ as $n \to \infty$ for $k = 0, 1$.

In case the parameters are known, $\hat{\theta}_k$ should read θ_k^*, and then Condition [C] requires nothing. Section 6 presents an example of estimator for θ_k which satisfies Condition [C].

Here we state the result on the rate of convergence of our change-point estimator.

Theorem 1. *The family* $\{n\vartheta_n^2(\hat{t}_n - t^*)\}_{n\in\mathbb{N}}$ *is tight under any one of the following conditions.*

(a) $[\mathbf{H}]_1$, $[A]$ *and* $[C]$ *hold in Case* (A).
(b) $[\mathbf{H}]_2$, $[B]$ *and* $[C]$ *hold in Case* (B).

In both Case (A) and (B) this result gives consistency of $\hat{t}_n$, since $n\vartheta_n^2 \to \infty$ by assumption, which is true also in Case (B).

The rest of this section will be devoted to the proof of Theorem 1. Define a stopping time $\tau = \tau(K)$ by

$$\tau(K) = \inf\{t;\, |X_t| + |b_t| > K\} \wedge T$$

for $K > 0$. X^τ denotes the process X stopped at τ. Write $S_i(\theta) = S(X_{t_i}^\tau, \theta)$, and $\Delta_i Y^\tau = Y_{t_i}^\tau - Y_{t_{i-1}}^\tau$. Let

$$\Psi_n(t;\theta_0,\theta_1) = \sum_{i=1}^{[nt/T]} g_i(\theta_0) + \sum_{i=[nt/T]+1}^{n} g_i(\theta_1),$$

where

$$\begin{aligned} g_i(\theta) &= 1_{\{\tau>0\}}\left\{\log\det S_{i-1}(\theta) + h^{-1}S_{i-1}(\theta)^{-1}[(\Delta_i Y^\tau)^{\otimes 2}]\right\} \\ &= 1_{\{\tau>0\}}\log\det S_{i-1}(\theta) + h^{-1}S_{i-1}(\theta)^{-1}[(\Delta_i Y^\tau)^{\otimes 2}]. \end{aligned}$$

Then $\sup_{\theta\in\mathcal{K}}|g_i(\theta)| \in L^{\infty-} = \cap_{p>1}L^p$ for any compact set $\mathcal{K}$ in Θ under $[\mathrm{H}]_1$. Denote by $E_{i-1}^{\theta_1^*}$ the conditional expectation with respect to $\mathcal{F}_{t_{i-1}}$ under the true distribution for $t_{i-1} \geq t^*$.

Lemma 1. *For* $t > t^*$,

$$\Psi_n(t;\theta_0,\theta_1) - \Psi_n(t^*;\theta_0,\theta_1) = M_n(t;\theta_0,\theta_1) + A_n(t;\theta_0,\theta_1) + \rho_n(t;\theta_0,\theta_1),$$

where

$$M_n(t;\theta_0,\theta_1) = \sum_{i=[nt^*/T]+1}^{[nt/T]} \left\{[g_i(\theta_0) - g_i(\theta_1)] - E_{i-1}^{\theta_1^*}[g_i(\theta_0) - g_i(\theta_1)]\right\},$$

$$\begin{aligned} A_n(t;\theta_0,\theta_1) &= 1_{\{\tau>0\}} \sum_{i=[nt^*/T]+1}^{[nt/T]} \left\{\mathrm{Tr}\left(S_{i-1}(\theta_0)^{-1}S_{i-1}(\theta_1) - I_d\right)\right. \\ &\qquad \left. - \log\det\left(S_{i-1}(\theta_0)^{-1}S_{i-1}(\theta_1)\right)\right\}, \end{aligned}$$

$$\begin{aligned} \rho_n(t;\theta_0,\theta_1) &= 1_{\{\tau>0\}} \sum_{i=[nt^*/T]+1}^{[nt/T]} \mathrm{Tr}\left\{\left(S_{i-1}(\theta_1)^{-1} - S_{i-1}(\theta_0)^{-1}\right)\right. \\ &\qquad \left. \cdot \left(S_{i-1}(\theta_1) - h^{-1}E_{i-1}^{\theta_1^*}[(\Delta_i Y^\tau)^{\otimes 2}]\right)\right\}. \end{aligned}$$

The proof of Lemma 1 is straightforward and omitted.

Remark 3. Later we will consider substitution of estimators $\hat{\theta}_k$ to θ_k, $k = 0, 1$. Then the expectation $E^{\theta_1^*}_{i-1}[g_i(\theta_0) - g_i(\theta_1)]$ is taken before the substitution, and so

$$M_n(t; \hat{\theta}_0, \hat{\theta}_1) = \sum_{i=[nt^*/T]+1}^{[nt/T]} \left\{ [g_i(\hat{\theta}_0) - g_i(\hat{\theta}_1)] - E^{\theta_1^*}_{i-1}[g_i(\theta_0) - g_i(\theta_1)]|_{\theta_0=\hat{\theta}_0, \theta_1=\hat{\theta}_1} \right\}.$$

In particular, the second term in the braces is not necessarily $\mathcal{F}_{t_{i-1}}$-measurable.

We will need a uniform Hájek–Renyi inequality. Let D be a bounded open set in $\mathbb{R}^d$. The Sobolev norm is denoted by

$$\|f\|_{s,p} = \left\{ \sum_{i=0}^{s} \|\partial_\theta^i f\|^p_{L^p(D)} \right\}^{1/p}$$

for $f \in W^{s,p}(D)$, the Sobolev space with indices (s, p). Suppose that $p > 1$ and $s > d/p$. The embedding inequality is the following

$$\sup_{\theta \in D} |f(\theta)| \leq C\|f\|_{s,p} \quad (f \in W^{s,p}(D)) \tag{1}$$

where C is a constant depending only on s, p and D. We will apply this inequality for $f \in C^s(D)$, and the validity of such an inequality depends on the regularity of the boundary of D; the Garsia–Rodemich–Rumsey (GRR) inequality validates it if there exist positive numbers a and ϵ_0 such that

$$\inf_{\theta \in D} \nu(B_\epsilon(\theta)) \geq a\,\epsilon^d \quad \text{for } \epsilon \in (0, \epsilon_0), \tag{2}$$

where ν is the Lebesgue measure on D and $B_\epsilon(\theta)$ is the ϵ-ball centered at θ, see e.g. [24] for details.

Lemma 2. *Let $(\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_j)_{j\in\mathbb{Z}_+}, P)$ be a stochastic basis. Let D be a bounded domain in $\mathbb{R}^d$ admitting Sobolev's inequality (1) for some $p \in (1, 2]$ and $s \in \mathbb{Z}_+$ such that $s > d/p$. Let $(c_j)_{j\in\mathbb{Z}_+}$ be a nondecreasing sequence of positive numbers. Let $X = (X_j)_{j\in\mathbb{Z}_+}$ be a sequence of random fields on D for $j \in \mathbb{Z}_+$ satisfying the following conditions:*

(i) *For each $(w, j) \in \Omega \times \mathbb{Z}_+$, $X_j \in C^s(D)$;*
(ii) *For each $(\theta, i) \in D \times \{0, 1, \ldots, s\}$, $(\partial_\theta^i X_j(\theta))_{j\in\mathbb{Z}_+}$ is a zero-mean L^p-martingale with respect to* **F**.

Then there exists a constant C' depending only on s, p and D, not depending on X, such that

$$P\left[\max_{j\leq n} \frac{1}{c_j} \sup_D |X_j(\theta)| \geq a\right] \leq \frac{C'}{a^p} \sum_{j=0}^{n} \frac{1}{c_j^p} E\left[\|X_j - X_{j-1}\|^p_{s,p}\right]$$

for all $a > 0$ and $n \in \mathbb{Z}_+$.

Proof. Let $B = L^p(D)$, then B is p-uniformly smooth; see Definition 2.2 of [23, pp. 245–246, and Example 2.2, p. 247]. We apply Theorem in [21, p. 245], to conclude

$$P\left[\max_{j\leq n} \frac{1}{c_j} \|\partial_\theta^i X_j\|_B \geq a\right] \leq \frac{C_1}{a^p} \sum_{j=0}^{n} \frac{1}{c_j^p} E\left[\|\partial_\theta^i X_j - \partial_\theta^i X_{j-1}\|^p_B\right]$$

for $i \in \{0, 1, \ldots, s\}$ for some constant C_1. Therefore (1) yields the result. □

Proof of Theorem 1. For the proof, we may assume $T = 1$ for notational simplicity without loss of generality.

(i) Let ϵ be an arbitrary positive number. Set

$$H(x) = 4Q(x, \theta_0^*, \theta_1^*)\vartheta_0^{-2}$$

in Case (A), and set $H(x) = \lambda_1(\Xi(x, \theta^*))$ in Case (B). We denote $\sigma(t; \theta) = \sigma(X_t^\tau, \theta)$ and $h(t) = H(X_t^\tau)$ in what follows. Those processes depend on K by definition while it is suppressed from the symbols. Set $B_K = \{\tau = 1\}$ and fix a sufficiently large K so that $P[B_K^c] < \epsilon/4$.

We notice that $h(s) \geq 0$ and that $h(t^*) > 0$ a.s. on B_K from the identifiability condition [A]/[B] since $X_{t^*}^\tau = X_{t^*}$ on B_K. We will show that there exists a positive constant c_ϵ such that

$$P\left[\inf_{t\in[t^*,1]} \frac{1}{t-t^*}\int_{t^*}^t h(s)\,ds \leq 5c_\epsilon\right] < \epsilon.$$

Define the event $\mathcal{A}_\delta$ by

$$\mathcal{A}_\delta = \left\{\inf_{t\in[t^*,t^*+\delta]} h(s) \geq \frac{1}{2}h(t^*)\right\}$$

for $\delta \in (0, 1-t^*)$. On $\mathcal{A}_\delta$, it holds that

$$\inf_{t\in[t^*,t^*+\delta]} \frac{1}{t-t^*}\int_{t^*}^t h(s)\,ds \geq \frac{1}{2}h(t^*) \geq \frac{\delta}{2(1-t^*)}h(t^*)$$

and also that, for $t \in [t^*+\delta, 1]$,

$$\begin{aligned}\frac{1}{t-t^*}\int_{t^*}^t h(s)\,ds &\geq \frac{1}{1-t^*}\int_{t^*}^t h(s)\,ds\\ &\geq \frac{1}{1-t^*}\int_{t^*}^{t^*+\delta} h(s)\,ds\\ &\geq \frac{\delta}{2(1-t^*)}h(t^*).\end{aligned}$$

Choose a δ so that $P[\mathcal{A}_\delta] > 1-\epsilon/2$ by the continuity of h, and next choose a positive number $c_\epsilon = c(\epsilon, \delta)$ such that

$$P\left[\frac{\delta}{2(1-t^*)}h(t^*) > 5c_\epsilon\right] \geq P\left[\left\{\frac{\delta}{2(1-t^*)}h(t^*) > 5c_\epsilon\right\}\bigcap B_K\right] > 1 - \frac{\epsilon}{2}.$$

Then

$$\begin{aligned}P\left[\inf_{t\in[t^*,1]} \frac{1}{t-t^*}\int_{t^*}^t h(s)\,ds \leq 5c_\epsilon\right] &\leq P[\mathcal{A}_\delta^c]\\ &\quad + P\left[\mathcal{A}_\delta, \inf_{t\in[t^*,1]} \frac{1}{t-t^*}\int_{t^*}^t h(s)\,ds \leq 5c_\epsilon\right]\\ &< \epsilon.\end{aligned}$$

(ii) With Lemma 1, we decompose $\Psi_n(t; \hat\theta_0, \hat\theta_1) - \Psi_n(t^*; \hat\theta_0, \hat\theta_1)$ as follows:

$$\Psi_n(t; \hat\theta_0, \hat\theta_1) - \Psi_n(t^*; \hat\theta_0, \hat\theta_1) = M_n(t; \hat\theta_0, \hat\theta_1) + A_n(t; \hat\theta_0, \hat\theta_1) + \rho_n(t; \hat\theta_0, \hat\theta_1).$$

Let $M \geq 1$. We have

$$\begin{aligned} P[n\vartheta_n^2(\hat{t}_n - t^*) > M] &\leq P\left[\inf_{t:n\vartheta_n^2(t-t^*)>M} \Phi_n(t;\hat{\theta}_0,\hat{\theta}_1) \leq \Phi_n(t^*;\hat{\theta}_0,\hat{\theta}_1)\right] \\ &\leq P\left[\inf_{t:n\vartheta_n^2(t-t^*)>M} \Psi_n(t;\hat{\theta}_0,\hat{\theta}_1) \leq \Psi_n(t^*;\hat{\theta}_0,\hat{\theta}_1)\right] \\ &\quad + P[B_K^c] < P_{1,n} + P_{2,n} + P_{3,n} + \epsilon, \end{aligned} \tag{3}$$

where

$$\begin{aligned} P_{1,n} &= P\left[\sup_{t:n\vartheta_n^2(t-t^*)>M} \frac{1}{[nt]-[nt^*]}\left|M_n(t;\hat{\theta}_0,\hat{\theta}_1)\right| \geq \frac{c_\epsilon\vartheta_n^2}{3}\right] \\ P_{2,n} &= P\left[\inf_{t:n\vartheta_n^2(t-t^*)>M} \frac{1}{[nt]-[nt^*]}A_n(t;\hat{\theta}_0,\hat{\theta}_1) \leq c_\epsilon\vartheta_n^2\right] \\ P_{3,n} &= P\left[\sup_{t:n\vartheta_n^2(t-t^*)>M} \frac{1}{[nt]-[nt^*]}\left|\rho_n(t;\hat{\theta}_0,\hat{\theta}_1)\right| \geq \frac{c_\epsilon\vartheta_n^2}{3}\right]. \end{aligned}$$

Here we read $\inf\emptyset = \infty$ and $\sup\emptyset = -\infty$. We will estimate these terms.
(iii) Estimate of $P_{1,n}$. In Case (B), let

$$\mathcal{M}_n(t;\theta) = \sum_{i=[nt^*]+1}^{[nt]} \left\{\partial_\theta g_i(\theta) - E_{i-1}^{\theta_1^*}[\partial_\theta g_i(\theta)]\right\}.$$

Let $\dot{\Theta}$ be an open ball such that $\theta^* \in \dot{\Theta}$ and $\overline{\dot{\Theta}} \subset \Theta$. Since

$$\sup_{\theta_0,\theta_1\in\dot{\Theta}} |M_n(t;\theta_0,\theta_1)|\,|\theta_0-\theta_1|^{-1} \leq \sup_{\theta\in\dot{\Theta}} |\mathcal{M}_n(t;\theta)|,$$

one has

$$\begin{aligned} P_{1,n} &\leq P\left[\sup_{t:n\vartheta_n^2(t-t^*)>M} \frac{1}{[nt]-[nt^*]}\left|M_n(t;\hat{\theta}_0,\hat{\theta}_1)\right|\,|\hat{\theta}_0-\hat{\theta}_1|^{-1} \geq \frac{c_\epsilon\vartheta_n}{6}, \hat{\theta}_0,\hat{\theta}_1 \in \dot{\Theta}\right] \\ &\quad + P[|\hat{\theta}_0-\hat{\theta}_1| \geq 2\vartheta_n] + P[\hat{\theta}_0 \notin \dot{\Theta}] + P[\hat{\theta}_1 \notin \dot{\Theta}] \\ &\leq P\left[\sup_{t:n\vartheta_n^2(t-t^*)>M} \frac{1}{[nt]-[nt^*]} \sup_{\theta\in\dot{\Theta}} |\mathcal{M}_n(t;\theta)| \geq \frac{c_\epsilon\vartheta_n}{6}\right] \\ &\quad + P[|\hat{\theta}_0-\hat{\theta}_1| \geq 2\vartheta_n] + P[\hat{\theta}_0 \notin \dot{\Theta}] + P[\hat{\theta}_1 \notin \dot{\Theta}]. \end{aligned}$$

By the uniform version of the Hájek–Renyi inequality in Lemma 2 applied to the case $p = 2$, $s = 2 + [d_0/2]$ and $D = \dot{\Theta}$, we see under $[\mathbf{H}]_2$ that

$$P\left[\sup_{t:n\vartheta_n^2(t-t^*)>M} \frac{1}{[nt]-[nt^*]} \sup_{\theta\in\dot{\Theta}} |\mathcal{M}_n(t;\theta)| \geq \frac{c_\epsilon\vartheta_n}{6}\right] \leq \frac{\mathsf{C}}{c_\epsilon^2 M} =: \rho_\epsilon(M),$$

where C denotes a generic constant independent of n and M. Therefore

$$\overline{\lim_{n\to\infty}}\, P_{1,n} \leq \rho_\epsilon(M) \tag{4}$$

thanks to

$$P\left[|\hat{\theta}_0 - \hat{\theta}_1| \geq 2\vartheta_n\right] \leq P\left[|\hat{\theta}_0 - \theta_0^*| \geq \frac{1}{3}\vartheta_n\right] + P\left[|\hat{\theta}_1 - \theta_1^*| \geq \frac{1}{3}\vartheta_n\right]$$

for large n.

In Case (A), Let $\dot{\Theta}_k$ be an open ball such that $\overline{\dot{\Theta}_k} \subset \Theta$ and $\theta_k^* \in \dot{\Theta}_k$ for each $k = 0, 1$.

$$P_{1,n} \leq P\left[\sup_{t:n\vartheta_0^2(t-t^*)>M} \frac{1}{[nt]-[nt^*]} \sup_{\substack{\theta_0\in\dot{\Theta}_0 \\ \theta_1\in\dot{\Theta}_1}} |M_n(t;\theta_0,\theta_1)| \geq \frac{c_\epsilon \vartheta_0^2}{3}\right] + P[\hat{\theta}_0 \notin \dot{\Theta}_0] + P[\hat{\theta}_1 \notin \dot{\Theta}_1].$$

We apply the Hájek–Renyi inequality for $M_n(t;\theta_0,\theta_1)$, which is a difference of two random fields on $\dot{\Theta}_k$ to be done with one by one, in order to obtain (4) under $[\mathbf{H}]_1$.

(iv) Estimation of $P_{2,n}$. First we consider Case (B). There is a positive constant c_2 independent of n such that

$$\begin{aligned}
&\mathrm{Tr}\left(S_{i-1}(\hat{\theta}_0)^{-1}S_{i-1}(\hat{\theta}_1) - I_d\right) - \log\det\left(S_{i-1}(\hat{\theta}_0)^{-1}S_{i-1}(\hat{\theta}_1)\right) \\
&\quad\geq \Xi(X^\tau_{t_{i-1}},\theta^*)[(\hat{\theta}_1-\hat{\theta}_0)^{\otimes 2}] + r_{n,i-1}|\hat{\theta}_1-\hat{\theta}_0|^2 \\
&\quad\geq \{\lambda_1(\Xi(X^\tau_{t_{i-1}},\theta^*)) + r_{n,i-1}\}|\hat{\theta}_1-\hat{\theta}_0|^2
\end{aligned}$$

for all i, where $\max_i |r_{n,i-1}| \leq c_2\vartheta_n$, on the event

$$B_{K,n} = B_K \cap \{\hat{\theta}_0, \hat{\theta}_1 \in \dot{\Theta}, |\hat{\theta}_k - \theta^*| \leq \vartheta_n \ (k=0,1)\}.$$

Thus

$$\begin{aligned}
P_{2,n} &\leq P\left[\inf_{t:n\vartheta_n^2(t-t^*)>M} \frac{1}{[nt]-[nt^*]} A_n(t;\hat{\theta}_0,\hat{\theta}_1)\,|\hat{\theta}_1-\hat{\theta}_0|^{-2} \leq 4c_\epsilon, B_{K,n}\right] \\
&\quad + P\left[|\hat{\theta}_1-\hat{\theta}_0| \leq \frac{1}{2}\vartheta_n\right] + P[B^c_{K,n}] \\
&\leq P\left[\inf_{t:n\vartheta_n^2(t-t^*)>M} \frac{1}{[nt]-[nt^*]} \sum_{i=[nt^*]+1}^{[nt]} \{\lambda_1(\Xi(X^\tau_{t_{i-1}},\theta^*)) + r_{n,i-1}\} \leq 4c_\epsilon\right] + \epsilon
\end{aligned}$$

for large n. The scaled summation converges to the corresponding scaled integral uniformly in t a.s., hence from Step (i) we have

$$\overline{\lim_{n\to\infty}} P_{2,n} \leq P\left[\inf_{t\in[t^*,1]} \frac{1}{t-t^*}\int_{t^*}^t h(s)\,ds \leq 5c_\epsilon\right] + \epsilon < 2\epsilon$$

for large n.

We will consider Case (A). There is a positive constant c_2 independent of n such that

$$\begin{aligned}
&\mathrm{Tr}\left(S_{i-1}(\hat{\theta}_0)^{-1}S_{i-1}(\hat{\theta}_1) - I_d\right) - \log\det\left(S_{i-1}(\hat{\theta}_0)^{-1}S_{i-1}(\hat{\theta}_1)\right) \\
&\quad\geq \mathrm{Tr}\left(S_{i-1}(\theta_0^*)^{-1}S_{i-1}(\theta_1^*) - I_d\right) - \log\det\left(S_{i-1}(\theta_0^*)^{-1}S_{i-1}(\theta_1^*)\right) \\
&\qquad - c_2(|\hat{\theta}_1-\theta_1^*| + |\hat{\theta}_0-\theta_0^*|)
\end{aligned}$$

for all i on the event $B'_{K,n} = B_K \cap \{\hat{\theta}_0 \in \dot{\Theta}_0, \hat{\theta}_1 \in \dot{\Theta}_1\}$ because there exists a continuous derivative $\partial_\theta \sigma$ by $[\mathbf{H}]_1$. In this way,

$$P_{2,n} \le P\left[\inf_{t:n(t-t^*)>M} \frac{1}{[nt]-[nt^*]} A_n(t;\hat{\theta}_0,\hat{\theta}_1) \le c_\epsilon \vartheta_0^2, B'_{K,n}\right] + P\left[B'^{c}_{K,n}\right].$$

Therefore,

$$\overline{\lim_{n\to\infty}}\, P_{2,n} \le P\left[\inf_{t\in[t^*,1]} \frac{1}{t-t^*}\int_{t^*}^t h(s)\,ds \le 5c_\epsilon\right] + \epsilon < 2\epsilon$$

by Step (i).

(v) Estimation of $P_{3,n}$. We have

$$\sup_{t\in[t^*,1]} \left|S(X_t,\hat{\theta}_k) - S(X_t,\theta_k^*)\right| 1_{\{|\hat{\theta}_k-\theta_k^*|<2\vartheta_n\}\cap B_K} \le \mathsf{C}\,\vartheta_n \quad (k=0,1),$$

$$\sup_{t\in[t^*,1]} \left|S(X_t,\hat{\theta}_k)^{-1} - S(X_t,\theta_k^*)^{-1}\right| 1_{\{|\hat{\theta}_k-\theta_k^*|<2\vartheta_n\}\cap B_K} \le \mathsf{C}\,\vartheta_n \quad (k=0,1)$$

and

$$\sup_{i:\ge[nt^*]+2} \left|S_{i-1}(\theta_1^*) - h^{-1}E_{i-1}^{\theta_1^*}[(\Delta_i Y)^{\otimes 2}]\right| 1_{B_K} \le \mathsf{C}\, w_{[0,T]}\left(X, \frac{1}{n}\right)^\alpha.$$

In the last estimate, the local α-Hölder continuity of σ was used. Then on $B_K \cap \{|\hat{\theta}_k - \theta_k^*| \le 2\vartheta_n\ (k=0,1)\}$,

$$\sup_{t:n\vartheta_n^2(t-t^*)>M} \frac{1}{[nt]-[nt^*]} \left|\rho_n(t;\hat{\theta}_0,\hat{\theta}_1)\right| \vartheta_n^{-2} = o_p(1) \tag{5}$$

because of $[\mathbf{H}]_j$ (ii). Consequently, we see $\overline{\lim}_{n\to\infty} P_{3,n} \le \epsilon$ due to [C] and the localization by B_K.

(vi) From the estimates in Steps (ii)–(iv) and making K sufficiently large, we have

$$\overline{\lim_{n\to\infty}}\, P[n\vartheta_n^2(\hat{t}_n - t^*) > M] \le \rho_\epsilon(M) + 5\epsilon$$

for any $M \ge 1$ and $\epsilon > 0$. Therefore,

$$\overline{\lim_{M\to\infty}}\, \overline{\lim_{n\to\infty}}\, P[n\vartheta_n^2(\hat{t}_n - t^*) > M] \le 5\epsilon,$$

which shows the tightness of $\{n\vartheta_n^2(\hat{t}_n - t^*)_+\}_n$. In a quite similar way, we can show that $\{n\vartheta_n^2(\hat{t}_n - t^*)_-\}_n$ is tight, and hence the family $\{n\vartheta_n^2(\hat{t}_n - t^*)\}_n$ is tight. $\square$

4. Asymptotic distribution of the change point estimator

This section discusses limit theorems for the distributions of the estimators. The notion of the stable convergence will be necessary. Given a probability space $(\Omega, \mathcal{F}, P)$ and a Markov kernel $\hat{P}$ from $(\Omega, \mathcal{F})$ to a measurable space $(\hat{\Omega}, \hat{\mathcal{F}})$, the extension $(\check{\Omega}, \check{\mathcal{F}}, \check{P})$ is defined as $\check{\Omega} = \Omega \times \hat{\Omega}$, $\check{\mathcal{F}} = \mathcal{F} \times \hat{\mathcal{F}}$ (product σ-field) and $\check{P}(A\times B) = \int_A P(d\omega)\hat{P}(\omega, B)$ for $A \in \mathcal{F}$ and $B \in \hat{\mathcal{F}}$. Let $\mathcal{G}$ be a sub σ-field of $\mathcal{F}$ and $\mathbb{E}$ a Polish space. Let Z be an $\mathbb{E}$-valued random variable defined on $(\check{\Omega}, \check{\mathcal{F}})$. It is said that a sequence $(\mathbf{Z}_n)_{n\in\mathbb{N}}$ of $\mathbb{E}$-valued random variables defined on $(\Omega, \mathcal{F})$ stably converges (in distribution) to Z if $E[f(Z_n)G] \to \check{E}[f(Z)G]$ as $n \to \infty$ for all $f \in C_B(\mathbb{E})$[1] and all bounded $\mathcal{G}$-measurable random variables G, where $\check{E}$ stands for the expectation with respect to $\check{P}$ and G is extended in a natural way to $\check{\Omega}$. The stable convergence is denoted by $Z_n \to^{d_s(\mathcal{G})} Z$. We simply write d_s for $d_s(\mathcal{F})$.

First we consider Case (B). Let

$$\mathbb{H}(v) = -2\left(\Gamma_\eta^{\frac{1}{2}}\,\mathcal{W}(v) - \frac{1}{2}\Gamma_\eta|v|\right)$$

for $\Gamma_\eta = (2T)^{-1}\Xi(X_{t^*}, \theta^*)[\eta^{\otimes 2}]$. Here $\mathcal{W}$ is a two-sided standard Wiener process defined on a probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$. For $(\check{\Omega}, \check{\mathcal{F}}, \check{P})$, we consider the extension of $(\Omega, \mathcal{F}, P)$ by the product of those spaces, i.e., $\check{P} = P \times \hat{P}$.

Theorem 2. *Suppose that the limit $\eta = \lim_{n\to\infty} \vartheta_n^{-1}(\theta_1^* - \theta_0^*)$ exists. Suppose that $[\mathbf{H}]_2$, [C] and [B] are fulfilled in Case (B). Then $n\vartheta_n^2(\hat{t}_n - t^*) \to^{d_s} \operatorname{argmin}_{v\in\mathbb{R}} \mathbb{H}(v)$ as $n \to \infty$.*

We will prove Theorem 2 and assume for a while that $T = 1$ to simplify the notation. Introduce a new parameter v as $t = t_v^\dagger := t^* + v(n\vartheta_n^2)^{-1}$. Let

$$\begin{aligned} D_n(v) &= \left\{\Psi_n(t_v^\dagger; \hat{\theta}_0, \hat{\theta}_1) - \Psi_n(t^*; \hat{\theta}_0, \hat{\theta}_1)\right\} - \left\{\Psi_n(t_v^\dagger; \theta_0^*, \theta_1^*) - \Psi_n(t^*; \theta_0^*, \theta_1^*)\right\} \\ &= \{M_n(t_v^\dagger; \hat{\theta}_0, \hat{\theta}_1) - M_n(t_v^\dagger; \theta_0^*, \theta_1^*)\} + \{A_n(t_v^\dagger; \hat{\theta}_0, \hat{\theta}_1) - A_n(t_v^\dagger; \theta_0^*, \theta_1^*)\} \\ &\quad + \{\rho_n(t_v^\dagger; \hat{\theta}_0, \hat{\theta}_1) - \rho_n(t_v^\dagger; \theta_0^*, \theta_1^*)\}. \end{aligned}$$

Lemma 3. *For every $L > 0$,*

$$\sup_{v\in[-L,L]} |D_n(v)| \to^p 0$$

as $n \to \infty$.

Proof. We assume that $v > 0$. We have

$$\begin{aligned} &M_n(t_v^\dagger; \hat{\theta}_0, \hat{\theta}_1) - M_n(t_v^\dagger; \theta_0^*, \theta_1^*) \\ &\quad = \int_0^1 \vartheta_n \partial_\theta M_n(t_v^\dagger; \theta_0^* + u(\hat{\theta}_0 - \theta_0^*), \theta_1^* + u(\hat{\theta}_1 - \theta_1^*))\, du\, [\vartheta_n^{-1}(\hat{\theta}_0 - \theta_0^*, \hat{\theta}_1 - \theta_1^*)]. \end{aligned}$$

For $k = 0, 1$ and $j = 1, 2$,

$$\begin{aligned} E\left[\sup_{t\in[t^*, t^*+L(n\vartheta_n^2)^{-1}]} |\partial_{\theta_k}^j M_n(t; \theta_0, \theta_1)|^2\right] &\le 8E\left[|\partial_{\theta_k}^j M_n(t^* + L(n\vartheta_n^2)^{-1}; \theta_0, \theta_1)|^2\right] \\ &\quad + O(1) \\ &\le 8L\vartheta_n^{-2} \sup_{i\ge 1} E\left[|\partial_{\theta_k}^j g_i(\theta_k)|^2\right] \\ &\quad + O(1) \\ &\le \mathsf{C}L\vartheta_n^{-2}. \end{aligned}$$

[1] The set of all bounded continuous functions on $\mathbb{E}$.

Then Sobolev's inequality implies

$$\vartheta_n \sup_{\substack{t\in[t^*,t^*+L(n\vartheta_n^2)^{-1}],\\ \theta_0,\theta_1\in\dot{\Theta}}} |\partial_\theta M_n(t;\theta_0,\theta_1)| = O_p(1).$$

As a result,

$$\sup_{v\in[0,L]} |M_n(t_v^\dagger;\hat{\theta}_0,\hat{\theta}_1) - M_n(t_v^\dagger;\theta_0^*,\theta_1^*)| \to^p 0$$

as $n \to \infty$.

Set $r_n = |\hat{\theta}_0 - \theta_0^*| + |\hat{\theta}_1 - \theta_1^*|$. Simple calculus yields

$$|\{\mathrm{Tr}\, y - \log\det(I_d + y)\} - \{\mathrm{Tr}\, x - \log\det(I_d + x)\}| \le c_3|y - x|(|x| + |y - x|)$$

for $d \times d$-symmetric matrices x and y whenever $|x|, |y| \le c_3'$, where c_3' and c_3 are some positive constants independent of x, y. Indeed, the formula $\int \exp(-2^{-1}(I_d + \epsilon x)[z^{\otimes 2}])dz = (2\pi)^{d/2}\det(I_d + \epsilon x)^{-1/2}$ is convenient for explicit computation.

Applying this inequality to $y = S_{i-1}(\hat{\theta}_0)^{-1/2}S_{i-1}(\hat{\theta}_1)S_{i-1}(\hat{\theta}_0)^{-1/2} - I_d$ and $x = S_{i-1}(\theta_0^*)^{-1/2}S_{i-1}(\theta_1^*)S_{i-1}(\theta_0^*)^{-1/2} - I_d$, we see that there exists a constant c_4 such that for large n, on $B_K \cap \{|\hat{\theta}_k - \theta^*| < \vartheta_n\ (k = 0, 1)\}$,

$$|A_n(t;\hat{\theta}_0,\hat{\theta}_1) - A_n(t;\theta_0^*,\theta_1^*)| \le c_4 \sum_{i=[nt^*]+1}^{[nt]} r_n(\vartheta_n + r_n).$$

Therefore, for any $\epsilon > 0$, if we take sufficiently large K, then

$$\varlimsup_{n\to\infty} P\left[\sup_{t\in[t^*,t^*+L(n\vartheta_n^2)^{-1}]} |A_n(t;\hat{\theta}_0,\hat{\theta}_1) - A_n(t;\theta_0^*,\theta_1^*)| \ge \epsilon\right] \le \epsilon.$$

This implies

$$\sup_{v\in[0,L]} |A_n(t_v^\dagger;\hat{\theta}_0,\hat{\theta}_1) - A_n(t_v^\dagger;\theta_0^*,\theta_1^*)| \to^p 0$$

as $n \to \infty$. The convergence

$$\sup_{v\in[0,L]} |\rho_n(t_v^\dagger;\hat{\theta}_0,\hat{\theta}_1) - \rho_n(t_v^\dagger;\theta_0^*,\theta_1^*)| \to^p 0$$

can be shown in the same way as (5).

A similar proof of the uniform convergence on $[-L, 0]$ is possible. After all, we obtained the desired result. □

Remark 4. When θ_k^* $(k = 0, 1)$ are known, we do not need Lemma 3.

Thus we can focus only on $\Psi_n(t_v^\dagger;\theta_0^*,\theta_1^*) - \Psi_n(t^*;\theta_0^*,\theta_1^*)$. For simplicity, we write $\Psi_n^*(t)$ for $\Psi_n(t;\theta_0^*,\theta_1^*)$. By assumption, there exists a limit $\eta = \lim_{n\to\infty} \vartheta_n^{-1}(\theta_1^* - \theta_0^*)$. $\mathbb{D}(I)$ denotes the D-space on an interval I of t, i.e. the space of càdlàg functions on I. Let

$$\mathbb{H}_n(v) = \Psi_n^*\left(t^* + v(n\vartheta_n^2)^{-1}\right) - \Psi_n^*(t^*).$$

and

$$\mathbb{H}^\tau(v) = -2\left(\Gamma_{\eta,\tau}^{\frac{1}{2}}\,\mathcal{W}(v) - \frac{1}{2}\Gamma_{\eta,\tau}|v|\right)$$

for $\Gamma_{\eta,\tau} = 1_{\{\tau>0\}}(2T)^{-1}\Xi(X^\tau_{t^*},\theta^*)[\eta^{\otimes 2}]$.

Lemma 4. *Let $\eta = \lim_{n\to\infty}\vartheta_n^{-1}(\theta_1^* - \theta_0^*)$. Suppose that $[\mathbf{H}]_2$, [C] and [B] are fulfilled in Case (B). Then $\mathbb{H}_n \to^{d_s} \mathbb{H}^\tau$ in $\mathbb{D}([-L,L])$ as $n\to\infty$ for every $L>0$.*

Proof. We will only consider positive v since the argument is essentially the same for negative v. Let $T=1$ as before. It follows from Lemma 1 that

$$\mathbb{H}_n(v) = M_n^\Delta(v) + A_n^\Delta(v) + \rho_n^\Delta(v),$$

where

$$\begin{aligned}
M_n^\Delta(v) &= M_n(t^* + v(n\vartheta_n^2)^{-1};\theta_0^*,\theta_1^*),\\
A_n^\Delta(v) &= A_n(t^* + v(n\vartheta_n^2)^{-1};\theta_0^*,\theta_1^*),\\
\rho_n^\Delta(v) &= \rho_n(t^* + v(n\vartheta_n^2)^{-1};\theta_0^*,\theta_1^*).
\end{aligned}$$

The evaluation of these terms will be done in the following. As repeated previously, we may proceed discussion on the event B_K hereafter. First

$$\begin{aligned}
M_n^\Delta(v) = 1_{\{\tau>0\}}\sum_{i=[nt^*]+1}^{[nt^*+\vartheta_n^{-2}v]}\mathrm{Tr}\Bigg[&\left(S_{i-1}(\theta_0^*)^{-1} - S_{i-1}(\theta_1^*)^{-1}\right)\\
&\cdot h^{-1}\left(\left(\int_{t_{i-1}}^{t_i}\sigma(X_t^\tau,\theta_1^*)dW_t\right)^{\otimes 2} - E_{i-1}^{\theta_1^*}\left[\int_{t_{i-1}}^{t_i}S(X_t^\tau,\theta_1^*)dt\right]\right)\Bigg] + \bar{o}_p(1) \qquad (6)
\end{aligned}$$

where $U_n(v) = \bar{o}_p(1)$ means that $\sup_{v\in[0,L]}|U_n(v)| \to^p 0$, and we used the hypothesis $n\vartheta_n^2 \to \infty$ and the fact that $|S_{i-1}(\theta_0^*)^{-1} - S_{i-1}(\theta_1^*)^{-1}| \le \mathsf{C}\,\vartheta_n$ with the localization. To obtain $\bar{o}_p(1)$, L^1-estimate helps. It follows from $[\mathbf{H}]_j$ (i)(c) and (ii) that

$$\begin{aligned}
&\left|h^{-1}\left(\int_{t_{i-1}}^{t_i}S(X_t^\tau,\theta_1^*)dt - E_{i-1}^{\theta_1^*}\left[\int_{t_{i-1}}^{t_i}S(X_t^\tau,\theta_1^*)dt\right]\right)\right|\\
&\quad= \Bigg|h^{-1}\Bigg(\int_{t_{i-1}}^{t_i}[S(X_t^\tau,\theta_1^*) - S(X_{t_{i-1}}^\tau,\theta_1^*)]dt\\
&\qquad - E_{i-1}^{\theta_1^*}\left[\int_{t_{i-1}}^{t_i}[S(X_t^\tau,\theta_1^*) - S(X_{t_{i-1}}^\tau,\theta_1^*)]dt\right]\Bigg)\Bigg|\\
&\quad\le \mathsf{C}\,w_{[0,T]}(n^{-1},X)^\alpha\\
&\quad= o_p(\vartheta_n).
\end{aligned}$$

Moreover, with the Burkholder–Davis–Gundy inequality, the first terms on the right-hand side of (6) equals $\bar{M}_n^\Delta(v) + \bar{o}_p(1)$ with $\bar{M}_n^\Delta(v) = \sum_{i=[nt^*]+1}^{[nt^*+\vartheta_n^{-2}v]}\xi_{n,i}$, where

$$\xi_{n,i} = 1_{\{\tau>0\}}\mathrm{Tr}\Big[{}^{\mathrm{t}}\sigma_{i-1}(\theta_1^*)\left(S_{i-1}(\theta_0^*)^{-1} - S_{i-1}(\theta_1^*)^{-1}\right)\sigma_{i-1}(\theta_1^*) \times \left(h^{-1}(\Delta_i W)^{\otimes 2} - I_r\right)\Big] \tag{7}$$

and $\sigma_{i-1}(\theta) = \sigma(X^\tau_{t_{i-1}}, \theta)$.

Now we introduce the backward approximation

$$\tilde{\xi}_{n,i} = 1_{\{\tau>0\}}\mathrm{Tr}\Big[{}^{\mathrm{t}}\sigma(X^\tau_{t^*-\epsilon_n}, \theta_1^*)\left(S(X^\tau_{t^*-\epsilon_n}, \theta_0^*)^{-1} - S(X^\tau_{t^*-\epsilon_n}, \theta_1^*)^{-1}\right)\sigma(X^\tau_{t^*-\epsilon_n}, \theta_1^*) \times \left(h^{-1}(\Delta_i W)^{\otimes 2} - I_r\right)\Big]$$

to $\xi_{n,i}$ for $\epsilon_n = 2Ln^{-1}\vartheta_n^{-2}$. After all, by $\tilde{M}_n^\Delta(v) = \sum_{i=[nt^*]+1}^{[nt^*+\vartheta_n^{-2}v]} \tilde{\xi}_{n,i}$, we have

$$M_n^\Delta(v) = \tilde{M}_n^\Delta(v) + \bar{o}_p(1). \tag{8}$$

Since

$$\begin{aligned}
&1_{\{\tau>0\}}\vartheta_n^{-2}v\left|{}^{\mathrm{t}}\sigma(X^\tau_{t^*-\epsilon_n}, \theta_1^*)\left(S(X^\tau_{t^*-\epsilon_n}, \theta_0^*)^{-1} - S(X^\tau_{t^*-\epsilon_n}, \theta_1^*)^{-1}\right)\sigma(X^\tau_{t^*-\epsilon_n}, \theta_1^*)\right|^2 \\
&\quad = 1_{\{\tau>0\}}\vartheta_n^{-2}v\,\Xi(X^\tau_{t^*-\epsilon_n}, \theta^*)[(\theta_1^* - \theta_0^*)^{\otimes 2}] + \bar{o}_p(1) \\
&\quad \to^p 1_{\{\tau>0\}}\Xi(X^\tau_{t^*}, \theta^*)[\eta^{\otimes 2}]\, v,
\end{aligned}$$

the central limit theorem ensures the convergence $\tilde{M}_n^\Delta \to^d -2\Gamma_{\eta,\tau}^{\frac{1}{2}}\mathcal{W}$ in $\mathbb{D}([0, L])$. In the same fashion, we can show $\tilde{M}_n^\Delta \to^d -2\Gamma_{\eta,\tau}^{\frac{1}{2}}\mathcal{W}$ in $\mathbb{D}([-L, 0])$ if $\tilde{M}_n^\Delta$ is defined in a natural way for negative v. Those convergence take place jointly and stably; apply Theorem 3-2 of [14] to the triangular array with many zeros.

Easy calculations yield $\sup_{v\in[-L,L]}|A_n^\Delta(v) - \Gamma_{\eta,\tau}|v|| \to^p 0$ and $\sup_{v\in[-L,L]}|\rho_n^\Delta(v)| \to^p 0$ for extended A_n^Δ and ρ_n^Δ to $[-L, L]$, which completes the proof. □

Proof of Theorem 2. We have supposed that $T = 1$ to state the lemmas, and we start with this case. Write $\hat{v} = \mathrm{argmin}_{v\in\mathbb{R}}\,\mathbb{H}(v)$. For $\epsilon > 0$, take large K so that $P[\tau = T] > 1 - \epsilon$. It follows from Lemma 4 that for every $x \in \mathbb{R}$,

$$\begin{aligned}
&\overline{\lim_{n\to\infty}}\, P[n\vartheta_n^2(\hat{t}_n - t^*) \le x] - \epsilon \\
&\quad \le \overline{\lim_{n\to\infty}}\, P[\inf_{v\in[-L,x]}\mathbb{H}_n^\tau(v) \le \inf_{v\in[x,L]}\mathbb{H}_n^\tau(v)] + \sup_n P[n\vartheta_n^2(\hat{t}_n - t^*) \notin [-L, L]\} \\
&\quad = P[\inf_{v\in[-L,x]}\mathbb{H}^\tau(v) \le \inf_{v\in[x,L]}\mathbb{H}^\tau(v)] + \sup_n P[n\vartheta_n^2(\hat{t}_n - t^*) \notin [-L, L]\} \\
&\quad \le \epsilon + P[\hat{v} \le x] + P[\hat{v} \notin [-L, L]] + \sup_n P[n\vartheta_n^2(\hat{t}_n - t^*) \notin [-L, L]].
\end{aligned}$$

As $L \to \infty$, the last two terms of the right-hand side of the above inequality tend to 0 thanks to Theorem 1 (b). So we have obtained

$$\overline{\lim_{n\to\infty}}\, P[n\vartheta_n^2(\hat{t}_n - t^*) \le x] \le P[\hat{v} \le x].$$

The estimate of $P[n\vartheta_n^2(\hat{t}_n - t^*) \le x]$ from below can be done in a similar manner. It is easy to see the stable convergence if we replace P by $G \cdot \check{P}$ in expectations for bounded $\mathcal{G}$-measurable variables G, which completes the proof in case $T = 1$.

For general T, we introduce a stochastic basis $\tilde{\mathcal{B}} = (\Omega, \mathcal{F}, \tilde{\mathbf{F}}, P)$ with $\tilde{\mathbf{F}} = (\mathcal{F}_{Tu})_{u\in[0,1]}$, and the processes $\tilde{b}_u = b_{Tu}$, $\tilde{X}_u = X_{Tu}$ and $\tilde{Y}_u = Y_{Tu}$, $u \in [0,1]$, to scale the time as $t = Tu$. Those stochastic processes satisfy the stochastic integral equation

$$\tilde{Y}_u = \tilde{Y}_0 + \int_0^u \tilde{b}_r dr + \int_0^u \tilde{\sigma}(\tilde{X}_r, \theta) d\tilde{W}_r,$$

where $\tilde{\sigma}(x,\theta) = \sqrt{T}\sigma(x,\theta)$ and $\tilde{W}$ is an r-dimensional $\tilde{\mathbf{F}}$-Wiener process. The sampling times $(iT/n)_{i=0}^n$ now change to $(i/n)_{i=0}^n$ in the new setting after scaling time. For the change point estimator $\hat{u}_n$ for $u^* = T^{-1}t^*$, we know

$$n\vartheta_n^2(\hat{u}_n - u^*) \to^{d_s} \underset{\tilde{v}\in\mathbb{R}}{\operatorname{argmin}}\ \tilde{\mathbb{H}}(\tilde{v}), \tag{9}$$

where $\tilde{\mathbb{H}}(\tilde{v}) = -2\left(\tilde{\Gamma}_\eta \tilde{\mathcal{W}}(\tilde{v}) - 2^{-1}\tilde{\Gamma}_\eta|\tilde{v}|\right)$, $\tilde{\Gamma}_\eta = 2^{-1}\Xi(\tilde{X}_{u^*}, \theta^*)[\eta^{\otimes 2}]$ and $\tilde{\mathcal{W}}$ is a two-sided Wiener process independent of $\tilde{\sigma}(\tilde{X}_{u^*}, \theta^*) = \sqrt{T}\sigma(X_{t^*}, \theta^*)$. Since

$$\begin{aligned} T \underset{\tilde{v}\in\mathbb{R}}{\operatorname{argmin}}\ \tilde{\mathbb{H}}(\tilde{v}) &= \underset{v\in\mathbb{R}}{\operatorname{argmin}}\ \tilde{\mathbb{H}}\left(\frac{v}{T}\right) \\ &=^d \underset{v\in\mathbb{R}}{\operatorname{argmin}}\ \mathbb{H}(v) \end{aligned}$$

thanks to $\mathcal{W}(\cdot) =^d T^{1/2}\tilde{\mathcal{W}}(\cdot/T)$. Thus (9) gives the desired convergence of $\hat{t}_n$ since $\hat{t}_n = T\hat{u}_n$. $\square$

Let us investigate the limit distribution of the estimator in Case (A). Proposition 1 of Section 5 gives a result to obtain the asymptotic distribution of the change point presented in Theorem 3 below. By nature of the sampling scheme, only the set $\mathcal{G}_n = \{kT/n; k \in \mathbb{Z}\}$ has essential meaning for the optimization with respect to the parameter t. Without loss of generality, we modify $\hat{t}_n$ so that it takes values in $\mathcal{G}_n$, and set $\hat{k}_n = n\hat{t}_n/T$. Let

$$\begin{aligned} \mathbb{K}(v) = \sum_{i=1}^{v} \Big\{ &\mathrm{Tr}\left[{}^{\mathrm{t}}\sigma(X_{t^*}, \theta_1^*)\left(S(X_{t^*}, \theta_0^*)^{-1} - S(X_{t^*}, \theta_1^*)^{-1}\right)\sigma(X_{t^*}, \theta_1^*)\zeta_i^{\otimes 2}\right] \\ &- \log\det\left(S(X_{t^*}, \theta_0^*)^{-1} S(X_{t^*}, \theta_1^*)\right)\Big\}, \end{aligned}$$

where ζ_i are independent r-dimensional standard normal variables independent of $\mathcal{F}_T$.

Theorem 3. *Suppose that* $[\mathbf{H}]_1$, *[C] and [A] are fulfilled in Case (A). Then* $\hat{k}_n - [\frac{nt^*}{T}] \to^{d_s} \operatorname{argmin}_{v\in\mathbb{Z}} \mathbb{K}(v)$ *as* $n \to \infty$.

Proof. We change the definition of $t_v^\dagger$ and newly set $t_v^\dagger = [\frac{nt^*}{T}]\frac{T}{n} + \frac{Tv}{n}$. Lemma 3 is still valid by essentially the same proof and hence we may only consider $\Psi_n(t_v^\dagger; \theta_0^*, \theta_1^*) - \Psi_n(t^*; \theta_0^*, \theta_1^*)$. Writing $\Psi_n^*(t)$ for $\Psi_n(t; \theta_0^*, \theta_1^*)$, we will investigate the behavior of the random field

$$\mathbb{K}_n(v) = \Psi_n^*(t_v^\dagger) - \Psi_n^*(t^*)$$

on $v \in \mathbb{Z}$. For a while, we consider nonnegative v. The argument is similar for negative v. According to Lemma 1, we have the decomposition

$$\mathbb{K}_n(v) = \mathbb{M}_n(v) + \mathbb{A}_n(v) + \varrho_n(v),$$

where $\mathbb{M}_n(v) = M_n(t_v^\dagger; \theta_0^*, \theta_1^*)$, $\mathbb{A}_n(v) = A_n(t_v^\dagger; \theta_0^*, \theta_1^*)$ and $\varrho_n(v) = \rho_n(t_v^\dagger; \theta_0^*, \theta_1^*)$.

Now, $\mathbb{M}_n(v)$ admits a similar expansion as before:

$$\mathbb{M}_n(v) = \sum_{i=[nt^*/T]+1}^{[nt^*/T]+v} \xi_{n,i} + \bar{o}_p(1)$$

with $\xi_{n,i}$ given by (7). Moreover, for $\epsilon_n = n^{-1/2}$ this time, we consider the backward approximation of $\xi_{n,i}$, that is,

$$\xi_{n,i} = \tilde{\xi}_{n,i} + \bar{o}_p(1).$$

Here $v \in [0, L] \cap \mathbb{Z}$, however this approximation is available when we consider $v \in [-L, 0]$. let L_0 be the maximum integer in $[0, L]$. By continuity of σ and because W is an $\mathbf{F}$-Wiener process, Proposition 1 of Section 5 gives

$$\left(G, X^\tau_{t^*-\epsilon_n}, (h^{-1}(\Delta_i W)^{\otimes 2})_{i=[nt^*/T]-L_0}^{[nt^*/T]+L_0}\right) \to^d \left(G, X^\tau_{t^*}, (\zeta_i^{\otimes 2})_{i=-L_0}^{L_0}\right),$$

where G is any $\mathcal{F}$-measurable function and ζ_i are independent r-dimensional standard normal variables independent of $\mathcal{F}_T$; we use the same symbol ζ_i as in the statement. Consequently,

$$(G, \mathbb{M}_n(v))_{v=-L_0}^{L_0} \to^d (G, \mathbb{M}_\infty(v))_{v=-L_0}^{L_0}$$

for all $\mathcal{F}$-measurable random variables G, where

$$\mathbb{M}_\infty(v) = \sum_{i=1}^{v} \xi_{\infty,i}$$

and $\xi_{\infty,i}$ is given by

$$\xi_{\infty,i} = 1_{\{\tau>0\}} \mathrm{Tr}\left[{}^{\mathrm{t}}\sigma(X^\tau_{t^*}, \theta_1^*)\left(S(X^\tau_{t^*}, \theta_0^*)^{-1} - S(X^\tau_{t^*}, \theta_1^*)^{-1}\right)\sigma(X^\tau_{t^*}, \theta_1^*) \cdot \left(\zeta_i^{\otimes 2} - I_r\right)\right].$$

For $\mathbb{A}_n$, we have $\mathbb{A}_n(v) \to \mathbb{A}_\infty(v)$ with

$$\mathbb{A}_\infty(v) = 1_{\{\tau>0\}} v \left\{\mathrm{Tr}\left(S(X^\tau_{t^*}, \theta_0^*)^{-1} S(X^\tau_{t^*}, \theta_1^*) - I_d\right) - \log\det\left(S(X^\tau_{t^*}, \theta_0^*)^{-1} S(X^\tau_{t^*}, \theta_1^*)\right)\right\}.$$

On the other hand, $\varrho_n(v)$ tends to 0 uniformly in v. Therefore,

$$(\mathbb{K}_n(v))_{v=-L_0}^{L_0} \to^{d_s} (\mathbb{K}^\tau(v))_{v=-L_0}^{L_0},$$

where $\mathbb{K}^\tau(v) = \mathbb{M}_\infty(v) + \mathbb{A}_\infty(v)$. Removing τ by letting $K \to \infty$, and using Theorem 1, we obtain the limit distribution of $\hat{t}_n$. □

Remark 5. If we compensate $\zeta_i^{\otimes 2}$ in the representation of $\mathbb{K}(v)$, it can be observed, with the identifiability, that $\mathbb{K}(v)$ diverges a.s. as $|v| \to \infty$. It is also clear that $\mathbb{K}$ has no tie a.s. Therefore, the argmin-operation is well defined.

The results of this section include as a case previous results like, for example, the ones in [7] if we take $Y = X$ and one dimensional with $b_t = 0$ and $\sigma(x, \theta) = \theta$.

5. Mixing inequality and stable convergence

This section contains a result used in the proof of Theorem 3 of Section 4 but we present it here to make it self-contained. Let $\mathbb{X}^n = (\dot{\xi}_{n,i})_{i=[nt^*/T]-L_0}^{[nt^*/T]+L_0}$ with

$$\dot{\xi}_{n,i} = h^{-1}(\Delta_i W)^{\otimes 2} - I_r.$$

Let F denote any bounded $\mathcal{F}$-measurable function and let $G = E[F|\mathcal{F}_T]$. Since $\mathcal{L}\{\mathbb{X}^n\}$ does not depend on n, $\{(\mathbb{X}^n, G)\}_{n\in\mathbb{N}}$ is tight. By the subsequence argument, we may assume that $(\mathbb{X}^n, G) \to^d (\mathbb{X}^\infty, G_\infty)$ as $n \to \infty$ for the canonical process $(\mathbb{X}^\infty, G_\infty)$ on $\mathbb{R}^{2L_0+2}$ equipped with a probability measure P_∞; the uniqueness of the limit will be found later. Thus we have $E[f(\mathbb{X}^n)G] \to E_\infty[f(\mathbb{X}^\infty)G_\infty]$ for every $f \in C_B(\mathbb{R}^{2L_0+2})$, where E^∞ denotes the expectation with respect to P^∞. Let $\mathcal{W} = (\zeta_i^{\otimes 2} - I_r)_{i=-L_0}^{L_0}$. If the equality

$$E_\infty[f(\mathbb{X}^\infty)G_\infty] = E_\infty[f(\mathbb{X}^\infty)]E_\infty[G_\infty] \tag{10}$$

is obtained for $\mathcal{F}_T$-measurable functions G, we also have

$$\begin{aligned} E[f(\mathbb{X}^n)F] &= E[f(\mathbb{X}^n)G] \\ &\to E_\infty[f(\mathbb{X}^\infty)G_\infty] \\ &= E_\infty[f(\mathbb{X}^\infty)]E_\infty[G_\infty] \\ &= \check{E}[f(\mathcal{W})]E[G] \\ &= \check{E}[f(\mathcal{W})]E[F] \\ &= \check{E}[f(\mathcal{W})F]. \end{aligned}$$

This characterizes the possible limit of the sequence $\{(\mathbb{X}^n, F)\}_{n\in\mathbb{N}}$ uniquely and implies the stable convergence of $\mathbb{X}^n$. In order to show (10), it is sufficient to establish it only for monomials. Then the mixing property below serves to do this. Let us give the result in a slightly general setting. In the following, we assume $T = 1$ without loss of generality.

For $h = (h^1, \ldots, h^m) \in L^2([0,1])^m$ and $\alpha = (\alpha^1, \ldots, \alpha^m) \in \{1, \ldots, r\}^m$, let

$$J(h,\alpha)_t = \int_0^t h^1_{s_1} dW^{\alpha^1}_{s_1} \int_0^{s_1} h^2_{s_2} dW^{\alpha^2}_{s_2} \cdots \int_0^{s_{m-1}} h^m_{s_m} dW^{\alpha^m}_{s_m}.$$

Let

$$\mathcal{E}(h_1, \alpha_1, p_1; \cdots; h_k, \alpha_k, p_k; G)_t = E[J(h_1,\alpha_1)_t^{p_1} \cdots J(h_k,\alpha_k)_t^{p_k} G],$$

where $h_i \in L^2([0,1])^{m_i}$, $\alpha_i \in \{1, \ldots, r\}^{m_i}$, $m_i \in \mathbb{N}$ and $p_i \in \mathbb{N}$ for $i = 1, \ldots, k$ and $k \in \mathbb{N}$. Let $\mathcal{E}(\emptyset) = 1$. The following shows a mixing property.

Proposition 1. *For an $\mathcal{F}_1$-measurable bounded function G,*

$$\sup_{t\in[0,1]} |\mathcal{E}(h_1, \alpha_1, p_1; \cdots; h_k, \alpha_k, p_k; G)_t - \mathcal{E}(h_1, \alpha_1, p_1; \cdots; h_k, \alpha_k, p_k; 1)_t E[G]|$$

$$\le C(p_1, \ldots, p_k) \prod_{i=1}^k \left(\prod_{j=1}^{m_i} \|h_i^j\|_2 \right)^{p_i} \max_{i,j} \left(E\left[\int_0^1 1_{\{(h_i^j)_s \neq 0\}} d\langle E[G|\mathcal{F}_t]^c \rangle_s \right] \right)^{\frac{1}{2}},$$

where $E[G|\mathcal{F}_t]^c$ is the continuous martingale part of the L^2-martingale $(E[G|\mathcal{F}_t])_{t\in[0,1]}$, and $C(p_1, \ldots, p_k)$ is a constant independent of $h_1, \ldots, h_k$ and G.

Proof. For G, we have $G = E[G|\mathcal{F}_1] = E[G|\mathcal{F}_0] + M_1^c + M_1^d$ for some continuous L^2-martingale M^c with $M_0^c = 0$ and some purely discontinuous L^2-martingale M^d with $M_0^d = 0$.

First,

$$\mathcal{E}(h_1, \alpha_1, p_1; \cdots; h_k, \alpha_k, p_k; E[G|\mathcal{F}_0])_t = \mathcal{E}(h_1, \alpha_1, p_1; \cdots; h_k, \alpha_k, p_k; 1)_t E[G] \quad (11)$$

by the independence of $\mathcal{F}_0$ and the Wiener processes.

Next, thanks to Itô's formula,

$$\begin{aligned}
&\mathcal{E}(h_1, \alpha_1, p_1; \cdots; h_k, \alpha_k, p_k; G)_t \\
&\quad = \sum_{i=1}^k E\left[\int_0^t p_i J(h_1, \alpha_1)_s^{p_1} \cdots J(h_i, \alpha_i)_s^{p_i - 1}\right. \\
&\qquad \times \left. \cdots J(h_k, \alpha_1)_s^{p_k}\; J(h_i^{(-1)}, \alpha_i^{(-1)})_s (h_i^1)_s dW_s^{\alpha_i^1} G\right] \\
&\qquad + \sum_{i,j:i<j} p_i p_j \int_0^t \mathcal{E}\Big(\cdots; h_i^{(-1)}, \alpha_i^{(-1)}, p_i - 1; \cdots; h_j^{(-1)}, \alpha_j^{(-1)}, p_j - 1; \cdots; \\
&\qquad \times\; h_i^{(-1)}, \alpha_i^{(-1)}, 1; h_j^{(-1)}, \alpha_j^{(-1)}, 1; G\Big)_s (h_i^1)_s (h_j^1)_s\, d\langle W^{\alpha_i^1}, W^{\alpha_j^1}\rangle_s \\
&\qquad + \sum_i p_i(p_i - 1) \int_0^t \mathcal{E}\Big(\cdots; h_i^{(-1)}, \alpha_i^{(-1)}, p_i - 2; \cdots; h_i^{(-1)}, \alpha_i^{(-1)}, 2; G\Big)_s (h_i^1)_s^2\, ds
\end{aligned} \quad (12)$$

where $h_i^{(-1)} = (h_i^2, \ldots, h_i^{m_i})$ and for $h_i = (h_i^1, h_i^2, \ldots, h_i^{m_i})$ and $\alpha_i^{(-1)} = (\alpha_i^2, \ldots, \alpha_i^{m_i})$ for $\alpha_i^{(-1)}$. We read $x^0 = 1$ and $x^{-n} = 0$ for $n \in \mathbb{N}$.

We have

$$\begin{aligned}
&\left|E\left[\int_0^t J(h_1, \alpha_1)_s^{p_1} \cdots J(h_i, \alpha_i)_s^{p_i - 1} \cdots J(h_k, \alpha_1)_s^{p_k}\; J(h_i^{(-1)}, \alpha_i^{(-1)})_s (h_i^1)_s dW_s^{\alpha_i^1} M_1^c\right]\right| \\
&\quad \le \left|E\Big[\int_0^t J(h_1, \alpha_1)_s^{p_1} \cdots J(h_i, \alpha_i)_s^{p_i - 1}\right. \\
&\qquad \left. \cdots J(h_k, \alpha_1)_s^{p_k}\; J(h_i^{(-1)}, \alpha_i^{(-1)})_s (h_i^1)_s d\langle W^{\alpha_i^1}, M^c\rangle_s\Big]\right| \\
&\quad \le \left(E\left[\int_0^t \left\{J(h_1, \alpha_1)_s^{p_1} \cdots J(h_i, \alpha_i)_s^{p_i - 1} \cdots J(h_k, \alpha_1)_s^{p_k}\; J(h_i^{(-1)}, \alpha_i^{(-1)})_s (h_i^1)_s\right\}^2 ds\right]\right)^{\frac{1}{2}} \\
&\qquad \times \left(E\left[\int_0^t 1_{\{(h_i^1)_s \neq 0\}} d\langle M^c\rangle_s\right]\right)^{\frac{1}{2}}
\end{aligned} \quad (13)$$

and

$$\begin{aligned}
&E\left[\int_0^t \left\{J(h_1, \alpha_1)_s^{p_1} \cdots J(h_i, \alpha_i)_s^{p_i - 1} \cdots J(h_k, \alpha_1)_s^{p_k}\; J(h_i^{(-1)}, \alpha_i^{(-1)})_s (h_i^1)_s\right\}^2 ds\right] \\
&\quad \le E\left[\sup_{s\in[0,t]} \left\{J(h_1, \alpha_1)_s^{p_1} \cdots J(h_i, \alpha_i)_s^{p_i - 1} \cdots J(h_k, \alpha_1)_s^{p_k}\; J(h_i^{(-1)}, \alpha_i^{(-1)})_s\right\}^2\right] \|h_i^1\|_2^2 \\
&\quad \le C(p_1, \ldots, p_k) \prod_{i=1}^k \left(\prod_{j=1}^{m_i} \|h_i^j\|_2\right)^{2p_i}
\end{aligned} \quad (14)$$

for all $t \in [0,1]$, where $C(p_1,\ldots,p_k)$ denotes a generic constant depending only on $p_1,\ldots,p_k$ and varying from line to line. The last inequality used the Hölder inequality and the Burkholder–Davis–Gundy inequality.

Applying (12) to $G = M_1^c$ and using (13) and (14), we obtain

$$\begin{aligned}
&\sup_{t\in[0,1]} \left|\mathcal{E}(h_1,\alpha_1,p_1;\cdots;h_k,\alpha_k,p_k;M_1^c)_t\right| \\
&\le C(p_1,\ldots,p_k)\prod_{i=1}^{k}\left(\prod_{j=1}^{m_i}\|h_i^j\|_2\right)^{p_i}\max_i\left(E\left[\int_0^1 1_{\{(h_i^1)_s\neq 0\}}d\langle M^c\rangle_s\right]\right)^{\frac{1}{2}} \\
&\quad + C(p_1,\ldots,p_k) \\
&\quad \times \sup_{t\in[0,1]}\Big|\mathcal{E}\Big(\cdots;h_i^{(-1)},\alpha_i^{(-1)},p_i-1;\cdots;h_j^{(-1)},\alpha_j^{(-1)},p_j-1;\cdots; \\
&\quad \times\ h_i^{(-1)},\alpha_i^{(-1)},1;h_j^{(-1)},\alpha_j^{(-1)},1;M_1^c\Big)_t\Big|\ \|h_i^1\|_2\|h_j^1\|_2 \\
&\quad + C(p_1,\ldots,p_m) \\
&\quad \times \sup_{t\in[0,1]}\left|\mathcal{E}\left(\cdots;h_i^{(-1)},\alpha_i^{(-1)},p_i-2;\cdots;h_i^{(-1)},\alpha_i^{(-1)},2;M_1^c\right)_t\right|\ \|h_i^1\|_2^2.
\end{aligned} \tag{15}$$

We apply (15) repeatedly together with $E[M_1^c]=0$ to obtain

$$\begin{aligned}
&\sup_{t\in[0,1]} \left|\mathcal{E}(h_1,\alpha_1,p_1;\cdots;h_k,\alpha_k,p_k;M_1^c)_t\right| \\
&\le C(p_1,\ldots,p_k)\prod_{i=1}^{k}\left(\prod_{j=1}^{m_i}\|h_i^j\|_2\right)^{p_i}\max_{i,j}\left(E\left[\int_0^1 1_{\{(h_i^j)_s\neq 0\}}d\langle M^c\rangle_s\right]\right)^{\frac{1}{2}}.
\end{aligned} \tag{16}$$

If apply (12) for $G = M_1^d$ instead of M_1^c, the first term on the right-hand side of (12) vanishes, and we obtain

$$\mathcal{E}(h_1,\alpha_1,p_1;\cdots;h_k,\alpha_k,p_k;M_1^d)_t = 0 \tag{17}$$

by the orthogonality between a continuous martingale and M^d. The desired inequality follows from (11), (16) to (17). □

6. Remarks on estimation of the nuisance parameters θ_k

When the values of the parameters θ_k ($k=0,1$) are unknown, our estimating function for the change point requires certain estimators for θ_k. In this section, we will briefly discuss estimation of the nuisance parameters θ_k. There are two situations according to the prior knowledge of the parameter space $\mathbb{T}$ of the change point. The first one is the case where $\mathbb{T} = [\mathsf{t}_0,\mathsf{t}_1] \subset (0,T)$ for given numbers t_0 and t_1. In this case, the structural change occurs in neither interval $[0,\mathsf{t}_0)$ nor $(\mathsf{t}_1,T]$, and we can use this knowledge to construct estimators for θ_k. Contrarily, in the second case, we do not assume any prior information about the location of the change point t^*. Due to lack of information, we cannot use data over any interval of fixed length. However, as we will see later, it is still possible to construct estimators for θ_k by shrinkage of the sampling time intervals.

Let

$$\Phi_n^0(t;\theta_0)=\sum_{i=1}^{[nt/T]}G_i(\theta_0)\quad\text{and}\quad\Phi_n^1(t;\theta_1)=\sum_{i=[nt/T]+1}^{n}G_i(\theta_1).$$

Suppose that t_0 and t_1 are known. Let us consider estimators $\hat{\theta}_k$ for θ_k such that

$$\Phi_n^k(\mathrm{t}_k;\hat{\theta}_k)=\min_{\theta_k}\Phi_n^k(\mathrm{t}_k;\theta_k)$$

for $k=0,1$. Obviously, $\hat{\theta}_0$ is a function of the data up to t_0, and $\hat{\theta}_1$ is that of after t_1; they take advantage of the prior knowledge about $\mathbb{T}$.[2] Under suitable regularity conditions as well as the identifiability conditions that

$$\int_0^{\mathrm{t}_0}Q(X_t,\theta^*,\theta)\,dt>0\quad\text{a.s.}\quad\text{and}\int_{\mathrm{t}_1}^{T}Q(X_t,\theta^*,\theta)\,dt>0\text{ a.s.}\tag{18}$$

for every $\theta\neq\theta^*$, it is possible in general to show that $\hat{\theta}_k-\theta_k^*=O_p(n^{-1/2})$, then Condition [C] is satisfied in both cases (A) and (B); see the references given in Introduction for estimation of diffusion coefficients. Based on $\hat{\theta}_k$, the estimator $\hat{t}_n$ are defined. According to the previous sections, $\hat{t}_n$ possesses $n\vartheta_n^2$-consistency and the asymptotic distribution in each case is already known.

We can also construct the second stage estimators. Let b_n be a sequence of positive numbers such that $n\vartheta_n^2 b_n\to\infty$ as $n\to\infty$. Construct $\check{\theta}_k$ so that

$$\Phi_n^k(\hat{t}_n+(-1)^{k+1}b_n;\check{\theta}_k)=\min_{\theta_k}\Phi_n^k(\hat{t}_n+(-1)^{k+1}b_n;\theta_k)$$

for $k=0,1$.[3] The new estimators $\check{\theta}_k$ are expected to improve $\hat{\theta}_k$ since they utilize data from each end point to a sampling time near t^*. Further, we can construct a new change-point estimator with those estimators. Based on $\check{\theta}_k$, we define $\check{t}_n$ for t^* as

$$\check{t}_n=\operatorname*{argmin}_{t\in[0,T]}\Phi_n(t;\check{\theta}_0,\check{\theta}_1).$$

Since it is usually easy to verify Condition [C] for $\check{\theta}_k$, we will be able to obtain the same asymptotic results for $\check{t}_n$ as $\hat{t}_n$.

Next, let us consider the second situation. The knowledge of the distance from t^* to each end point is not available and it means that any data set sampled over a fixed time interval $[0,a]$ is useless for estimating θ_0 since t^* may be less than a and then the data over $(t^*,a]$ causes bias in general. A similar note also applies to estimation of θ_1. This consideration suggests the use of estimators $\hat{\theta}_k$ based on the data over time interval $[0,a_n]$ for θ_0 and the data over $[T-a_n,T]$ for θ_1, respectively, for some sequence a_n tending to zero. We assume that there exists a constant

2 To validate asymptotic properties of the estimators, it is sufficient that these relations are satisfied asymptotically. It is possible and rather routine to prove asymptotic properties for the estimator of θ_k, as they are necessary for the discussions here, in a fairly general setting even for possibly moving targets θ_k.

3 The estimating functions depend on the unknown parameters in a mild way as sequence b_n satisfies the mentioned condition. However, such a situation is quite common in asymptotic statistics; recall, for example, the smoothness assumption for the true density in the kernel density estimation. The theory has a meaning only when the condition is satisfied even though it is impossible to verify from the data.

$\beta \in (0, 1/2)$ such that $a_n \geq 1/(n\vartheta_n^{1/\beta})$ and that $|\hat{\theta}_k - \theta_k^*| = o_p((na_n)^{-\beta})$ for $k = 0, 1$. When $\overline{\lim}_{n\to\infty}\vartheta_n > 0$, we also assume $na_n \to \infty$. In particular, the first condition implies $n\vartheta_n^2 \to \infty$. The second condition is natural because the number of the available data over each time interval is proportional to na_n. To obtain consistent $\hat{\theta}_k$, we may need at least an identifiability condition such that $\sigma(\theta, x) = \sigma(\theta', x)$ implies $\theta = \theta'$; it is a strong condition like monotonicity of $\sigma(\theta, x)$ in θ, however it is necessary because the sampling interval is shrinking to an end point, by lack of knowledge of $\mathbb{T}$.

Here we show how to obtain an initial estimator of θ^*, considering only the one-dimensional case for notational simplicity. Usual regularity of the functions involved are also assumed to hold true. In order to obtain desired rate of convergence, it is for example assumed that $\inf_\theta |\partial_\theta S(x, \theta)| \geq p(x)$ $(\theta \in \Theta)$ for some positive continuous function p in a neighborhood of the initial value $X_0 = x_0 \in \mathbb{R}$. Let $\hat{\theta}_0^m$ be a root to the estimating equation $\psi_n(\theta) = 0$, where

$$\psi_n(\theta) := \frac{1}{a_n} \sum_{j:t_j \leq a_n} (\Delta_j Y)^2 - \frac{T}{a_n n} \sum_{j:t_j \leq a_n} S(X_{t_{j-1}}, \theta).$$

Under the usual regularity conditions, it is straightforward to prove that

$$\psi_n(\theta_0^*) = O_p\left(\frac{1}{\sqrt{na_n}}\right)$$

as $n \to \infty$. Then the nondegeneracy of $\partial_\theta S(x, \theta)$ near x_0 with the Taylor formula yields the stochastic order $\hat{\theta}_0^m - \theta_0^* = O_p(na_n)^{-1/2}$. This is the case of a moment-type estimator, however, a similar argument applies to the quasi maximum likelihood estimator based on the data on $[0, a_n]$ as, intuitively, this method is locally a kind of moment-type method.

Under the assumptions, [C] holds and after that it is possible to construct $\hat{t}_n$, $\check{\theta}_k$ and $\check{t}_n$ in turn as mentioned above. The asymptotic properties of $\check{t}_n$ are the same as $\hat{t}_n$ because $\check{\theta}_k$'s satisfy Condition [C]. It is expected that the new estimator $\check{t}_n$ possesses equal or better precision than $\hat{t}_n$, as suggested by the numerical studies in Section 7. Some proposals of first stage explicit and consistent estimators are given in [11].

7. Numerical analysis

This section contains results of a Monte Carlo analysis performed to assess the quality of the estimator of the change point and of the unknown volatilities under two different models. As the process b_t does not play a relevant role in the framework of this paper, only models without drift are considered. Moreover, the observed process Y_t and the coordinate process X_t are the same, i.e. $Y_t = X_t$.

7.1. First experiment

The first model considered is a diffusion process solution to the following stochastic differential equation

$$X_t = \begin{cases} X_0 + \int_0^t (1 + X_s^2)^{\theta_0^*} dW_s & \text{for } t \in [0, t^*) \\ X_{t^*} + \int_{t^*}^t (1 + X_s^2)^{\theta_1^*} dW_s & \text{for } t \in [t^*, T] \end{cases} \tag{19}$$

Table 1
Monte Carlo estimates for model (19) over 10,000 replications. True values: $\theta_0^* = 0.2$, $\theta_1^* = 0.378$, 0.350, and 0.319 for different sample sizes $n = 1000$, 2000 and 5000. True change point $t^* = 0.6$.

n	a_n	$\tilde{t}_n$	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{t}_n$	$\check{\theta}_0$	$\check{\theta}_1$	$\check{t}_n$
5000	0.1189	0.601 (0.005)	0.200 (0.009)	0.319 (0.014)	0.601 (0.011)	0.200 (0.005)	0.319 (0.013)	0.601 (0.012)
2000	0.1495	0.601 (0.008)	0.200 (0.013)	0.349 (0.020)	0.601 (0.014)	0.200 (0.008)	0.349 (0.017)	0.601 (0.015)
1000	0.1778	0.601 (0.011)	0.199 (0.017)	0.377 (0.025)	0.601 (0.019)	0.200 (0.011)	0.377 (0.026)	0.602 (0.018)

where t^* is the true change point assumed to be $t^* = 0.6$. The true value of the parameters are $\theta_0^* = 0.2$ and $\theta_1^* = \theta_0^* + n^{-\gamma}$, with $\gamma = \frac{1}{4}$, n is the sample size and $T = nh = 1$. The initial value is X_0 assumed to be constant, in particular we take $X_0 = 5$. The sequences $a_n = b_n = \frac{1}{n\vartheta_n^\delta}$ with $\delta = 3$ so that they satisfy the properties required in Section 6. The first stage estimator of θ_0^* (resp. θ_1^*) is obtained using the first na_n observations from the left (resp. na_n from the right). We denote the first stage estimators with $\hat{\theta}_i$, $i = 0, 1$. Once the first stage estimators of θ_0^* and θ_1^* are available, the first stage estimator of t^*, i.e. $\hat{t}_n$ is obtained via

$$\Phi_n(\hat{t}_n; \hat{\theta}_0, \hat{\theta}_1) = \min_{t \in [0,T]} \Phi_n(t; \hat{\theta}_0, \hat{\theta}_1).$$

Then, with the first stage estimator of t^* in hands, we calculate the second stage estimator of θ_i^* using observations in the interval $[0, \hat{t}_n - b_n]$ for θ_0^* and observations in the interval $[\hat{t}_n + b_n, T]$ for θ_1^*. We denote the second stage estimators of θ_i^* by $\check{\theta}_i$. Finally, the second stage estimator of t^*, i.e. $\check{t}_n$, is obtained as

$$\Phi_n(\check{t}_n; \check{\theta}_0, \check{\theta}_1) = \min_{t \in [0,T]} \Phi_n(t; \check{\theta}_0, \check{\theta}_1).$$

For comparison, we also report the value of the estimator $\tilde{t}_n$ obtained plugging the true parameter values in the contrast function, i.e. when the volatilities are supposed to be known

$$\Phi_n(\tilde{t}_n; \theta_0^*, \theta_1^*) = \min_{t \in [0,T]} \Phi_n(t; \theta_0^*, \theta_1^*),$$

and this can be considered as a benchmark. For the Monte Carlo setup, we consider different sample sizes $n = 1000, 2000, 5000$ and for each sample size n, we run $M = 10,000$ Monte Carlo replications. Under this choice of n the value of $\theta_1^* = 0.378, 0.350$, and 0.319 respectively. The values of the sequences a_n and b_n are reported in Table 1.

Data are generated according to the Euler–Maruyama scheme with predictor corrector method for stability (see, e.g. Chapter 2, [10]) using a double discretization approach: first the simulated paths are generated with a mesh of size 10^{-5} and then data are subsampled at rate $h = 1/n$ to produce statistical data which are used to construct the quasi-likelihood function and to perform the inference. The CIR model is simulated using the exact conditional distribution.

Table 1 also reports Monte Carlo estimates (i.e. average over the M replications) of the volatility parameters θ_0 and θ_1 and the change point t^*. In parenthesis are the standard deviations of the Monte Carlo estimates.

Table 2
Monte Carlo estimates for model (20) over 10,000 replications. True values: $\theta_0^* = 0.2$, $\theta_1^* = 0.378, 0.350$, and 0.319 for different sample sizes $n = 1000$, 2000 and 5000. True change point $t^* = 0.7$.

n	a_n	$\tilde{t}_n$	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{t}_n$	$\check{\theta}_0$	$\check{\theta}_1$	$\check{t}_n$
5000	0.1189	0.701 (0.010)	0.200 (0.012)	0.319 (0.018)	0.701 (0.011)	0.200 (0.018)	0.319 (0.012)	0.701 (0.010)
2000	0.1495	0.702 (0.016)	0.200 (0.016)	0.350 (0.029)	0.701 (0.024)	0.200 (0.009)	0.350 (0.030)	0.701 (0.021)
1000	0.1778	0.703 (0.025)	0.200 (0.021)	0.378 (0.040)	0.701 (0.038)	0.200 (0.012)	0.377 (0.056)	0.701 (0.040)

7.2. *Second experiment*

The second model considered is the [4] diffusion model solution to the following stochastic differential equation

$$X_t = \begin{cases} X_0 + \int_0^t \sqrt{\theta_0^* X_s} dW_s & \text{for } t \in [0, t^*) \\ X_{t^*} + \int_{t^*}^t \sqrt{\theta_1^* X_s} dW_s & \text{for } t \in [t^*, T] \end{cases} \tag{20}$$

with change point $t^* = 0.7$ and all remaining experimental conditions are the same as in previous experiment. The results are reported in Table 2. The difference in the two experiments is in the regularity of the diffusion coefficient term and in the fact that, in the second experiment, the true change point instant τ^* is closer to the right border of the interval $(0, T)$. Comparing the two simulation results, it is possible to see that the second stage estimators in the second experiment performs slightly better in term of the standard deviation.

7.3. *Asymptotic distribution of the change point estimator*

This section analyzes the empirical distribution of the second stage change point estimator $\check{\tau}$ and compares it with its theoretical counterpart. The empirical distribution is considered for a sample size of $n = 5000$ observations under the setup of the first experiment. Due to mixed-normal type limit, in order to obtain the limiting distribution, a preliminary studentization of the sequence $n\theta_n^2(\check{t}_n - t^*)$ is required. Therefore, the limiting distribution of the following sequence is studied

$$Z_n = n\theta_n^2(\check{t}_n - t^*)\hat{\Gamma}(X_{t^*}, \theta_0),$$

with $\hat{\Gamma}(X_{t^*}, \theta_0) = (\log(1 + X_{t^*}^2))^2$. Then Z_n converges in distribution to $\mathcal{W}(v) - \frac{1}{2}|v|$ with density

$$f(x) = \frac{3}{2}e^{|x|}\left(1 - \Phi\left(\frac{3}{2}\sqrt{|x|}\right)\right) - \frac{1}{2}\left(1 - \Phi\left(\frac{1}{2}\sqrt{|x|}\right)\right)$$

and distribution function

$$F(x) = \begin{cases} g(x), & x > 0, \\ 1 - g(-x), & x \le 0. \end{cases}$$

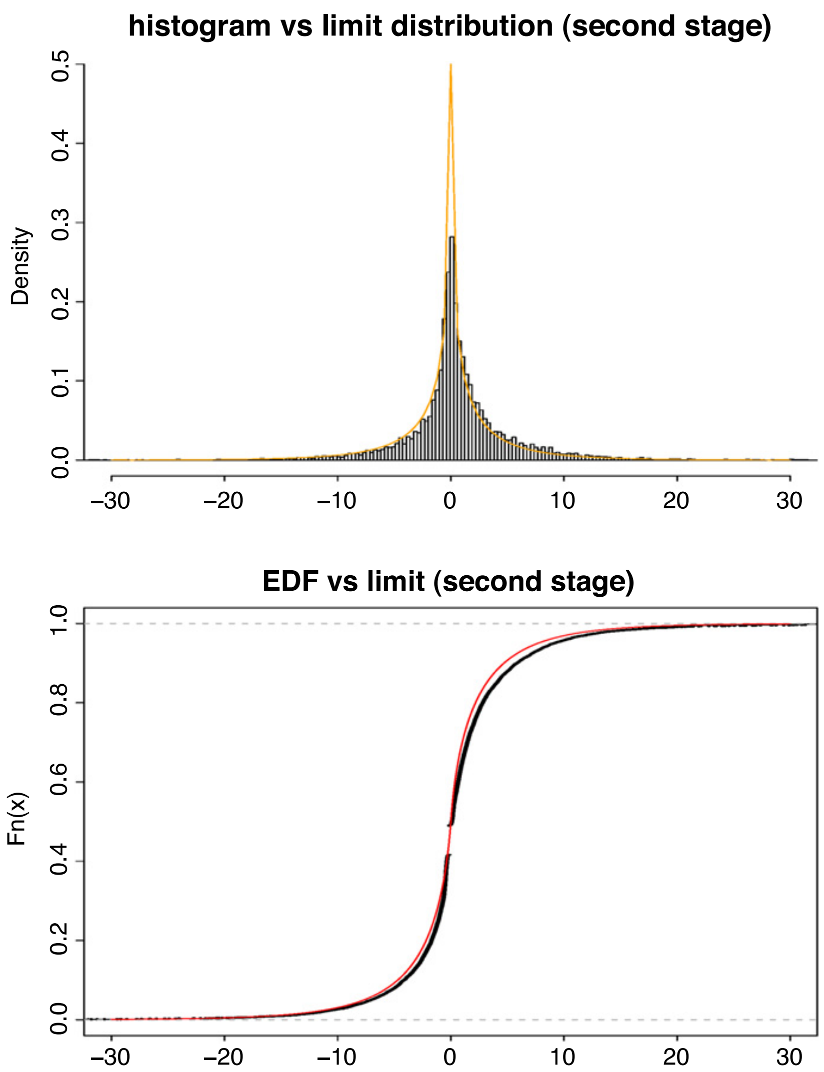

Fig. 1. Histogram versus theoretical density function (up) and empirical distribution function versus theoretical distribution function (bottom) for the second stage change point estimator. Results of 10, 000 Monte Carlo replications and sample size $n = 5000$ for the first model.

Here $\Phi(x)$ the distribution function of the Gaussian random variable, and

$$g(x) = 1 + \sqrt{\frac{x}{2\pi}} e^{-\frac{x}{8}} - \frac{1}{2}(x+5)\,\Phi\left(-\frac{\sqrt{x}}{2}\right) + \frac{3}{2} e^{x}\,\Phi\left(-\frac{3}{2}\sqrt{x}\right)$$

(see e.g. [5]). Fig. 1 reports the graphical representation of the histogram and empirical distribution function of Z (over 10,000 Monte Carlo replications) against their theoretical counterparts. The empirical results are inline with the expected theoretical quantities.

Acknowledgments

The authors wish to thank an anonymous referee for careful reading of the manuscript and fruitful remarks that led to an improved version of the paper.

References

[1] J. Bai, Least squares estimation of a shift in linear processes, Journal of Times Series Analysis 15 (1994) 453–472.
[2] J. Bai, Estimation of a change point in multiple regression models, The Review of Economics and Statistics 79 (1997) 551–563.

[3] G. Chen, Y.K. Choi, Y. Zhou, Nonparametric estimation of structural change points in volatility models for time series, Journal of Econometrics 126 (2005) 79–144.
[4] J.C. Cox, J.E. Ingersoll, S.A. Ross, A theory of the term structure of interest rates, Econometrica 53 (1985) 385–408.
[5] M. Csörgő, L. Horváth, Limit Theorems in Change-Point Analysis, Wiley, New York, 1997.
[6] A. De Gregorio, S.M. Iacus, Least squares volatility change point estimation for partially observed diffusion processes, Communications in Statistics, Theory and Methods 37 (15) (2008) 2342–2357.
[7] J. Deshayes, D. Picard, Lois asymptotiques des test et estimateurs de rupture dans un modèlle statistique classique, Annales Institute H. Poincaré, Section B 20 (4) (1984) 309–327.
[8] V. Genon-Catalot, J. Jacod, On the estimation of the diffusion coefficient for multi-dimensional diffusion processes, Annales Institute H. Poincaré Probability Statistics 29 (1993) 119–151.
[9] D.V. Hinkley, Inference about the change-point from cumulative sum tests, Biometrika 58 (1971) 509–523.
[10] S.M. Iacus, Simulation and Inference for Stochastic Differential Equations, With *R* Examples, Springer, NY, 2008.
[11] S.M. Iacus, N. Yoshida, Numerical analysis of volatility change point estimators for discretely sampled stochastic differential equations, Economic Notes 39 (1/2) (2010) 107–127.
[12] I.A. Ibragimov, R.Z. Hasminskii, Statistical Estimation: Asymptotic Theory, Springer, Berlin, 1981.
[13] C. Inclan, G.C. Tiao, Use of cumulative sums of squares for retrospective detection of change of variance, Journal of the American Statistical Association 89 (1994) 913–923.
[14] J. Jacod, On continuous conditional Gaussian martingales and stable convergence in law, in: Séminaire de Probabilitiés XXXI, in: Lecture Notes in Mathematics, vol. 1655, Springer, 1997, pp. 232–246.
[15] S. Kim, S. Cho, S. Lee, On the cusum test for parameter changes in GARCH(1,1) models, Communications of Statistics, Theory Methods 29 (2000) 445–462.
[16] Y. Kutoyants, Identification of Dynamical Systems with Small Noise, Kluwer, Dordrecht, 1994.
[17] Y. Kutoyants, Statistical Inference for Ergodic Diffusion Processes, Springer-Verlag, London, 2004.
[18] S. Lee, J. Ha, O. Na, S. Na, The Cusum test for parameter change in time series models, Scandinavian Journal of Statistics 30 (2003) 781–796.
[19] S. Lee, Y. Nishiyama, N. Yoshida, Test for parameter change in diffusion processes by cusum statistics based on one-step estimators, Annales Institute of Statististic and Mathematics 58 (2006) 211–222.
[20] G. Roussas, Contiguity of probability measures, in: Some Applications in Statistics, Cambridge University Press, Cambridge, 1972.
[21] G. Shixin, The Hájek–Renyi inequality for Banach space valued martingales and the p smoothness of Banach spaces, Statistics and Probability Letters 32 (1997) 245–248.
[22] J. Song, S. Lee, Test for parameter change in discretely observed diffusion processes, Statistical Inference for Stochastic Processes 12 (2) (2009) 165–183. doi:10.1007/s11203-008-9033-4.
[23] W.A. Woyczyński, Geometry and martingales in banach spaces, in: Winter School on Probability, Kapracz, in: Lecture notes in mathematics, vol. 472, Springer, 1975, pp. 235–275.
[24] N. Yoshida, Polynomial type large deviation inequality and its applications (reprint 2005), Annals of the Institute of Statistical Mathematics 63 (3) (2011) 431–479. doi:10.1007/s10463-009-0263-z.