Estimation for the change point of volatility in a stochastic differential equation✩

Stefano M. Iacus\textsuperscript{a,∗}, Nakahiro Yoshida\textsuperscript{b,c}

\textsuperscript{a}Department of Economics, Business and Statistics, University of Milan, Via Conservatorio 7, 20122 Milan, Italy
\textsuperscript{b}University of Tokyo, Japan
\textsuperscript{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 \( \theta \) 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.

© 2011 Elsevier B.V. All rights reserved.

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 $\theta$, while $b_t$ is completely unknown and unobservable, therefore possibly depending on $\theta$ and $t^*$.

The sample consists of $(X_t, Y_t)$, $i = 0, 1, \ldots, n$, where $t_i = ih$ for $h = h_n = T/n$. The parameter space $\Theta$ of $\theta$ is a bounded domain in $\mathbb{R}^{d_0}$, $d_0 \geq 1$, and the parameter $\theta$ is a nuisance in estimation of $t^*$. Denote by $\theta_i^*$ the true value of $\theta_i$ for $i = 0, 1$.

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

(A) $\theta_0^*$ and $\theta_1^*$ are fixed and do not depend on $n$.

(B) $\theta_0^*$ and $\theta_1^*$ depend on $n$, and as $n \to \infty$, $\theta_0^* \to \theta^* \in \Theta$, $\partial_n \to 0$ and $n\partial_n^2 \to \infty$.

In Case (A), $\partial_n$ is a constant $\partial_0$ 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 $\mathcal{B} = (\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in [0, T]}$, and satisfies the stochastic integral equation

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

Here $W_t$ is an $r$-dimensional $\mathbb{F}$-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 $M^{\otimes 2} = M^\dagger M$, $M[N] = \sum_{ij} m_{ij}n_{ij} = \text{Tr}(M^\dagger N)$, 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 $\ell$ of function $f$ with respect to $\theta$. Let $\alpha$ be a positive number.

[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 $\partial_\theta^\ell \sigma$ ($0 \leq \ell \leq [d_0/2]$) 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 $|\sigma(x, \theta) - \sigma(x', \theta)| \leq L(x, x', \theta)|x - x'|^\alpha$ ($x, x' \in \mathcal{X}, \theta \in \Theta$).

(ii) $(X_t)_{t \in [0, T]}$ is a progressively measurable process taking values in $\mathcal{X}$ such that $w_{[0, T]} \left( \frac{1}{n}, X \right) = o_p(\partial_n^{1/\alpha})$ 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 [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 $b_{s_n}$ 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 $\sigma$ 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 [H] (ii), $t \mapsto X_t$ is continuous a.s. Also, [H] (ii) imposes a restriction on the rate $\theta_n$. For example, when $\alpha = 1$, for a Brownian motion $X$, it suffices that $n^{\frac{\alpha}{d}}/\log n \to \infty$, 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–Rényi type estimate before making use of inequality (4) below, then it is possible to limit the range of $\ell$ from “$0 \leq \ell \leq j + [d_0/2]$” to “$0 \leq \ell \leq j$”.

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 $\theta^*_k$, $k = 0, 1$. Each estimator is based on $(X_{t_i}, Y_{t_i})_{i=0,1,...,n}$ 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 $\theta^*_k$ are known, we define $\hat{\theta}_k$ just as $\theta^*_k$. This article proposes

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

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

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

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

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

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

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

[C] $|\hat{\theta}_k - \theta^*_k| = o_p(\theta_n)$ as $n \to \infty$ for $k = 0, 1$. 

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

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

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

(a) \([H]_1\), \([A]\) and \([C]\) hold in Case (A).

(b) \([H]_2\), \([B]\) and \([C]\) hold in Case (B).

In both Case (A) and (B) this result gives consistency of \( \hat{\theta}_n \), 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

\[
\tau(K) = \inf \{ t \mid |X_t| + |b_t| > K \} \land T
\]

for \( K > 0 \). \( X^\tau \) denotes the process \( X \) stopped at \( \tau \). Write \( S_i(\theta) = S(X^\tau_{i-1}, \theta) \), and \( \Delta_i Y^\tau = Y^\tau_i - Y^\tau_{i-1} \). Let

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

where

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

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

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

\[
A_n(t; \theta_0, \theta_1) = \mathbf{1}_{\{\tau > 0\}} \sum_{i=\lfloor nt^*/T \rfloor + 1}^{\lfloor nt/T \rfloor} \left\{ \text{Tr} \left( S_{i-1}(\theta_0)^{-1} S_{i-1}(\theta_1) - I_d \right) \right\} \times \log \det \left( S_{i-1}(\theta_0)^{-1} S_{i-1}(\theta_1) \right),
\]

\[
\rho_n(t; \theta_0, \theta_1) = \mathbf{1}_{\{\tau > 0\}} \sum_{i=\lfloor nt^*/T \rfloor + 1}^{\lfloor nt/T \rfloor} \text{Tr} \left\{ \left( S_{i-1}(\theta_1)^{-1} - S_{i-1}(\theta_0)^{-1} \right) \cdot \left( S_{i-1}(\theta_1) - h^{-1} E_{i-1}^{\theta^*} [(\Delta_i Y^\tau)^{\otimes 2}] \right) \right\}.
\]

The proof of Lemma 1 is straightforward and omitted.
Remark 3. Later we will consider substitution of estimators \( \hat{\theta}_k \) to \( \theta_k \), \( k = 0, 1 \). Then the expectation \( E_{i-1}^\theta [g_i(\theta_0) - g_i(\hat{\theta}_1)] \) is taken before the substitution, and so
\[
M_n(t; \hat{\theta}_0, \hat{\theta}_1) = \sum_{i=[nt/T]+1}^{[nt/T]} \left\{ g_i(\hat{\theta}_0) - g_i(\hat{\theta}_1) \right\} - E_{i-1}^\theta [g_i(\theta_0) - g_i(\hat{\theta}_1)] - E_{i-1}^\theta [g_i(\theta_0) - g_i(\hat{\theta}_1)]|_{\theta_0=\hat{\theta}_0, \hat{\theta}_1=\hat{\theta}_1}.
\]
In particular, the second term in the braces is not necessarily \( \mathcal{F}_{t-1} \)-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
\[
\|f\|_{s,p} = \left\{ \sum_{i=0}^{s} \|\partial^i f\|_{L^p(D)}^{p} \right\}^{1/p}
\]
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
\[
\sup_{\theta \in D} |f(\theta)| \leq C \|f\|_{s,p} \quad (f \in W^{s,p}(D))
\]
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 \( \epsilon_0 \) such that
\[
\inf_{\theta \in D} v(B_\epsilon(\theta)) \geq a \epsilon^d \quad \text{for} \ \epsilon \in (0, \epsilon_0),
\]
where \( v \) is the Lebesgue measure on \( D \) and \( B_\epsilon(\theta) \) is the \( \epsilon \)-ball centered at \( \theta \), 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 \((\theta, i) \in D \times \{0, 1, \ldots, s\}, (\partial^i \theta X_j(\theta))_{j \in \mathbb{Z}_+}\) is a zero-mean \( L^p \)-martingale with respect to \( \mathbf{F} \).

Then there exists a constant \( C' \) depending only on \( s, p \) and \( D \), not depending on \( X \), such that
\[
P \left[ \max_{j \leq n} \frac{1}{c_j} \sup_{D} \|X_j(\theta)| \geq a \right] \leq \frac{C'}{a^p} \sum_{j=0}^{n} \frac{1}{c_j^p} E \left[ \|X_j - X_{j-1}\|_{s,p}^p \right]
\]
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
\[
P \left[ \max_{j \leq n} \frac{1}{c_j} \|\partial^i \theta X_j\|_B \geq a \right] \leq \frac{C_1}{a^p} \sum_{j=0}^{n} \frac{1}{c_j^p} E \left[ \|\partial^i \theta X_j - \partial^i \theta X_{j-1}\|_B^p \right]
\]
for \( i \in \{0, 1, \ldots, s\} \) for some constant \( C_1 \). Therefore (1) yields the result. \( \square \)
Proof of Theorem 1. For the proof, we may assume $T = 1$ for notational simplicity without loss of generality.

(i) Let $\epsilon$ be an arbitrary positive number. Set

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

in Case (A), and set $H(x) = \lambda_1(\Xi(x, \theta^*))$ in Case (B). We denote $\sigma(t; \theta) = \sigma(X_t^*, \theta)$ and $h(t) = H(X_t^*)$ 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 $X_{t^*} = X_{t^*}$ on $B_K$. We will show that there exists a positive constant $c_\epsilon$ such that

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

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 $t \in [t^*+\delta, 1]$,

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

Choose a $\delta$ 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

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

Then

$$P\left[ \inf_{t \in [t^*, 1] \setminus t^*} \frac{1}{t-t^*} \int_{t^*}^t h(s) \, ds \leq 5c_\epsilon \right] \leq P[\mathcal{A}_\delta^c]$$

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

(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

$$P[n \vartheta_n(\hat{t}_n - t^*) > M] \leq P\left[ \inf_{t: n \vartheta_n^2(t - t^*) > M} \Phi_n(t; \hat{\theta}_0, \hat{\theta}_1) \leq \Phi_n(t^*; \hat{\theta}_0, \hat{\theta}_1) \right]$$

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

where

$$P_{1,n} = P\left[ \sup_{t: n \vartheta_n^2(t - t^*) > M} \frac{1}{nt} \left| M_n(t; \hat{\theta}_0, \hat{\theta}_1) \right| \geq \frac{c \vartheta_n}{3} \right]$$

$$P_{2,n} = P\left[ \inf_{t: n \vartheta_n^2(t - t^*) > M} \frac{1}{nt} A_n(t; \hat{\theta}_0, \hat{\theta}_1) \leq c \vartheta_n \right]$$

$$P_{3,n} = P\left[ \sup_{t: n \vartheta_n^2(t - t^*) > M} \frac{1}{nt} \left| \rho_n(t; \hat{\theta}_0, \hat{\theta}_1) \right| \geq \frac{c \vartheta_n}{3} \right].$$

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

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

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

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

one has

$$P_{1,n} \leq P\left[ \sup_{t: n \vartheta_n^2(t - t^*) > M} \frac{1}{nt} \left| M_n(t; \hat{\theta}_0, \hat{\theta}_1) \right| |\hat{\theta}_0 - \hat{\theta}_1|^{-1} \geq \frac{c \vartheta_n}{6}, \hat{\theta}_0, \hat{\theta}_1 \in \hat{\Theta} \right]$$

$$+ P[|\hat{\theta}_0 - \hat{\theta}_1| \geq 2 \vartheta_n] + P[\hat{\theta}_0 \notin \hat{\Theta}] + P[\hat{\theta}_1 \notin \hat{\Theta}]$$

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

$$+ P[|\hat{\theta}_0 - \hat{\theta}_1| \geq 2 \vartheta_n] + P[\hat{\theta}_0 \notin \hat{\Theta}] + P[\hat{\theta}_1 \notin \hat{\Theta}].$$

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

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

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

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

(4)
Thus

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

for large \(n\).

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

\[
P_{1,n} \leq P \left[ \sup_{t:nt^2(t-t^*) > M} \frac{1}{nt} - \frac{1}{nt^*} \sup_{\theta_0} \sup_{\theta_1} |A_n(t; \theta_0, \theta_1)| \right] + P[\hat{\theta}_0 \notin \Theta_0] + P[\hat{\theta}_1 \notin \Theta_1].
\]

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

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

\[
\text{Tr} \left( S_{i-1}(\hat{\theta}_0)^{-1}S_{i-1}(\hat{\theta}_1) - I_d \right) - \log \det \left( S_{i-1}(\hat{\theta}_0)^{-1}S_{i-1}(\hat{\theta}_1) \right) \\
\geq \Xi(X^*_{i-1}, \theta^*)[(\hat{\theta}_1 - \hat{\theta}_0)^\otimes 2] + r_{n,i-1}|\hat{\theta}_1 - \hat{\theta}_0|^2 \\
\geq \{\lambda_1(\Xi(X^*_{i-1}, \theta^*)) + r_{n,i-1}\}|\hat{\theta}_1 - \hat{\theta}_0|^2
\]

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

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

Thus

\[
P_{2,n} \leq P \left[ \sup_{t:nt^2(t-t^*) > M} \frac{1}{nt} - \frac{1}{nt^*} \inf_{A_n(t; \theta_0, \theta_1)} |\hat{\theta}_1 - \hat{\theta}_0|^{-2} \leq 4c_\epsilon, B_{K,n} \right] \\
+ P \left[ |\hat{\theta}_1 - \hat{\theta}_0| \leq \frac{1}{2}\vartheta_n \right] + P[B_{K,n}^c] \\
\leq P \left[ \sup_{t:nt^2(t-t^*) > M} \frac{1}{nt} - \frac{1}{nt^*} \sum_{i=nt^*}^{nt} (\lambda_1(\Xi(X^*_{i-1}, \theta^*)) + r_{n,i-1}) \leq 4c_\epsilon \right] + \epsilon
\]

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

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

for large \(n\).

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

\[
\text{Tr} \left( S_{i-1}(\hat{\theta}_0)^{-1}S_{i-1}(\hat{\theta}_1) - I_d \right) - \log \det \left( S_{i-1}(\hat{\theta}_0)^{-1}S_{i-1}(\hat{\theta}_1) \right) \\
\geq \text{Tr} \left( S_{i-1}(\theta_0^*)^{-1}S_{i-1}(\theta_1^*) - I_d \right) - \log \det \left( S_{i-1}(\theta_0^*)^{-1}S_{i-1}(\theta_1^*) \right) \\
- c_2(|\hat{\theta}_1 - \theta_1^*| + |\hat{\theta}_0 - \theta_0^*|)
\]
for all \( i \) on the event \( B'_{K,n} = B_K \cap \{ \hat{\theta}_0 \in \hat{\theta}_0, \hat{\theta}_1 \in \hat{\theta}_1 \} \) because there exists a continuous derivative \( \partial_0 \sigma \) by \([H]_1\). In this way,

\[
P_{2,n} \leq P \left[ \inf_{t: n(t-t^*) > M} \frac{1}{|nt| - |nt^*|} A_n(t; \hat{\theta}_0, \hat{\theta}_1) \leq c_\epsilon \partial_0^2, B'_{K,n} \right] + P \left[ B'_{K,n} \right].
\]

Therefore,

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

by Step (i).

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

\[
\sup_{t: t-t^* \in [t^*, t]} \left| S(X_t, \hat{\theta}_k) - S(X_t, \theta_k^*) \right| 1_{|\hat{\theta}_k - \theta_k^*| < 2\partial_n} \leq C \partial_n \quad (k = 0, 1),
\]

\[
\sup_{t: t-t^* \in [t^*, t]} \left| S(X_t, \hat{\theta}_k)^{-1} - S(X_t, \theta_k^*)^{-1} \right| 1_{|\hat{\theta}_k - \theta_k^*| < 2\partial_n} \leq C \partial_n \quad (k = 0, 1)
\]

and

\[
\sup_{i: i \geq n(t^*)+2} \left| S_{i-1}(\theta_k^*)^\alpha E_{i-1}[(\Delta_i Y) \otimes 2] \right| 1_{B_K} \leq C w([0, 1]) \left( X, \frac{1}{n} \right)^\alpha.
\]

In the last estimate, the local \( \alpha \)-Hölder continuity of \( \sigma \) was used. Then on \( B_K \cap \{ |\hat{\theta}_k - \theta_k^*| \leq 2\partial_n \ (k = 0, 1) \} ,

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

because of \([H]_j\) (ii). Consequently, we see \( \lim_{n \to \infty} P_{3,n} \leq \epsilon \) due to \([C]\) and the localization by \( B_K \).

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

\[
\lim_{n \to \infty} \frac{1}{|nt| - |nt^*|} \left| \rho_n(t; \hat{\theta}_0, \hat{\theta}_1) \right| \partial_n^{-2} = o_p(1)
\]

for any \( M \geq 1 \) and \( \epsilon > 0 \). Therefore,

\[
\lim_{M \to \infty} \lim_{n \to \infty} P \left[ n\partial_n^2(\hat{t}_n - t^*) > M \right] \leq \rho_\epsilon(M) + 5\epsilon
\]

which shows the tightness of \( \{ n\partial_n^2(\hat{t}_n - t^*)_+ \}_n \). In a quite similar way, we can show that \( \{ n\partial_n^2(\hat{t}_n - t^*)_- \}_n \) is tight, and hence the family \( \{ n\partial_n^2(\hat{t}_n - t^*) \}_n \) 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 \((\Omega, \mathcal{F}, \mathbb{Q})\) and a Markov kernel \( \hat{P} \) from \((\Omega, \mathcal{F})\) to a measurable space \((\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})\), the extension \((\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})\) is defined as \( \hat{\Omega} = \Omega \times \hat{\Omega}, \hat{\mathcal{F}} = \mathcal{F} \times \hat{\mathcal{F}} \) (product \( \sigma \)-field) and \( \hat{\mathbb{P}}(A \times B) = \int_A P(d\omega) \hat{\mathbb{P}}(\omega, B) \) for \( A \in \mathcal{F} \) and \( B \in \hat{\mathcal{F}} \). Let \( \mathcal{G} \) be a sub \( \sigma \)-field of \( \mathcal{F} \) and \( \mathbb{E} \) a Polish space. Let \( \tilde{Z} \) be an \( \mathbb{E} \)-valued random variable defined on \((\hat{\Omega}, \hat{\mathcal{F}})\). It is said that a sequence \((\mathbb{Z}_n)_{n \in \mathbb{N}}\) of \( \mathbb{E} \)-valued random variables defined on \((\Omega, \mathcal{F})\) stably
converges (in distribution) to $Z$ if $E[f(Z_n)G] \to \tilde{E}[f(Z)G]$ as $n \to \infty$ for all $f \in C_B(\mathbb{E})$ and all bounded $\mathcal{G}$-measurable random variables $G$, where $\tilde{E}$ stands for the expectation with respect to $\tilde{P}$ and $G$ is extended in a natural way to $\tilde{\Omega}$. 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

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

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

**Theorem 2.** Suppose that the limit $\eta = \lim_{n \to \infty} \theta_n^{-1}(\theta_t^* - \theta_k^*)$ exists. Suppose that $[\mathbf{H}]_2$, $[C]$ and $[B]$ are fulfilled in Case (B). Then $n \theta_n^2(\hat{t}_n - t^*) \to d_s \arg\min_{v \in \mathbb{R}} H(v)$ 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 $t = t^*_v := t^* + v(n\theta_n^2)^{-1}$. Let

$$D_n(v) = \left\{ \Psi_n(t^*_v; \hat{\theta}_0, \hat{\theta}_1) - \Psi_n(t^*; \hat{\theta}_0, \hat{\theta}_1) \right\} - \left\{ \Psi_n(t^*_v; \theta_0^*, \theta_1^*) - \Psi_n(t^*; \theta_0^*, \theta_1^*) \right\}$$

$$= \left\{ M_n(t^*_v; \hat{\theta}_0, \hat{\theta}_1) - M_n(t^*_v; \theta_0^*, \theta_1^*) \right\} + \left\{ A_n(t^*_v; \hat{\theta}_0, \hat{\theta}_1) - A_n(t^*_v; \theta_0^*, \theta_1^*) \right\}$$

$$+ \{ \rho_n(t^*_v; \hat{\theta}_0, \hat{\theta}_1) - \rho_n(t^*_v; \theta_0^*, \theta_1^*) \}.$$

**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

$$M_n(t^*_v; \hat{\theta}_0, \hat{\theta}_1) - M_n(t^*_v; \theta_0^*, \theta_1^*)$$

$$= \int_0^1 \frac{\partial}{\partial \theta} M_n(t^*_v; \theta_0^* + u(\hat{\theta}_0 - \theta_0^*), \theta_1^* + u(\hat{\theta}_1 - \theta_1^*)) du [\partial^{-1}(\hat{\theta}_0 - \theta_0^*, \hat{\theta}_1 - \theta_1^*)].$$

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

$$E \left[ \sup_{t \in [t^*, t^* + L(n\theta_n^2)^{-1}]} |\partial_{\theta_k}^j M_n(t; \theta_0, \theta_1)|^2 \right] \leq 8E \left[ |\partial_{\theta_k}^j M_n(t^* + L(n\theta_n^2)^{-1}; \theta_0, \theta_1)|^2 \right]$$

$$+ O(1)$$

$$\leq 8L \theta_n^{-2} \sup_{i \geq 1} E \left[ |\partial_{\theta_k}^j g_i(\theta_k)|^2 \right]$$

$$+ O(1)$$

$$\leq CL \theta_n^{-2}.$$
Then Sobolev’s inequality implies
\[ \vartheta_n \sup_{t \in [t^*, t^* + L(n \theta_1^* n^{-1})^{-1}]} |\partial_\theta M_n (t; \theta_0, \theta_1)| = O_p(1). \]

As a result,
\[ \sup_{v \in [0, L]} |M_n (t_v^{\uparrow}; \hat{\theta}_0, \hat{\theta}_1) - M_n (t_v^{\uparrow}; \theta_0^*, \theta_1^*)| \rightarrow P 0 \]
as \( n \rightarrow \infty \).

Set \( r_n = |\hat{\theta}_0 - \theta_0^*| + |\hat{\theta}_1 - \theta_1^*| \). Simple calculus yields
\[ |(\text{Tr} y - \log \det(I_d + y)) - (\text{Tr} x - \log \det(I_d + x))| \leq c_3 |y - x|(|x| + |y - x|) \]
for \( d \times d \)-symmetric matrices \( x \) and \( y \) whenever \( |x|, |y| \leq c_4 \), where \( c_3 \) and \( c_4 \) are some positive constants independent of \( x, y \). Indeed, the formula \( \int \exp(-2^{-1}(I_d + \epsilon x)(z \otimes z))^T \frac{dz}{(2\pi)^{d/2} \det(I_d + \epsilon x)^{-1/2}} \) is convenient for explicit computation.

Applying this inequality to \( y = S_{i-1}(\hat{\theta}_0) - \frac{1}{2} S_{i-1}(\hat{\theta}_0) S_{i-1}(\hat{\theta}_0)^{-1/2} - I_d \) and \( x = S_{i-1}(\theta_0^*) - \frac{1}{2} S_{i-1}(\theta_0^*) S_{i-1}(\theta_0^*)^{-1/2} - I_d \), we see that there exists a constant \( c_4 \) such that for large \( n \), on \( B_K \cap \{ |\hat{\theta}_k - \theta_0^*| < \vartheta_n (k = 0, 1) \} \),
\[ |A_n (t; \hat{\theta}_0, \hat{\theta}_1) - A_n (t; \theta_0^*, \theta_1^*)| \leq c_4 \sum_{i=\lceil nt^* \rceil + 1}^{[nt]} r_n (\vartheta_n + r_n). \]

Therefore, for any \( \epsilon > 0 \), if we take sufficiently large \( K \), then
\[ \lim_{n \rightarrow \infty} P \left[ \sup_{t \in [t^*, t^* + L(n \theta_1^*)^{-1}]} |A_n (t; \hat{\theta}_0, \hat{\theta}_1) - A_n (t; \theta_0^*, \theta_1^*)| \geq \epsilon \right] \leq \epsilon. \]

This implies
\[ \sup_{v \in [0, L]} |A_n (t_v^{\uparrow}; \hat{\theta}_0, \hat{\theta}_1) - A_n (t_v^{\uparrow}; \theta_0^*, \theta_1^*)| \rightarrow P 0 \]
as \( n \rightarrow \infty \). The convergence
\[ \sup_{v \in [0, L]} |\rho_n (t_v^{\uparrow}; \hat{\theta}_0, \hat{\theta}_1) - \rho_n (t_v^{\uparrow}; \theta_0^*, \theta_1^*)| \rightarrow P 0 \]
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. \( \square \)

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

Thus we can focus only on \( \Psi_n (t_v^{\uparrow}; \theta_0^*, \theta_1^*) = \Psi_n (t^*; \theta_0^*, \theta_1^*) \). For simplicity, we write \( \Psi_n^*(t) \) for \( \Psi_n (t; \theta_0^*, \theta_1^*) \). By assumption, there exists a limit \( \eta = \lim_{n \rightarrow \infty} \partial_n^{-1}(\theta_1^* - \theta_0^*) \). \( \mathbb{D}(I) \) denotes the \( D \)-space on an interval \( I \) of \( t \), i.e. the space of càdlàg functions on \( I \). Let
\[ \mathbb{H}_n (v) = \Psi_n^* \left( t^* + v(n \theta_1^*)^{-1} \right) - \Psi_n^* (t^*). \]
and
\[
\mathbb{H}^\tau(v) = -2 \left( \Gamma_{\eta,\tau}^{1/2} \mathcal{W}(v) - \frac{1}{2} \Gamma_{\eta,\tau} |v| \right)
\]
for \( \Gamma_{\eta,\tau} = 1_{\{\tau > 0\}} (2T)^{-1} \Xi(X_{i^*}, \delta^*) [\eta^{\otimes 2}] \).

**Lemma 4.** Let \( \eta = \lim_{n \to \infty} \delta_n^{-1} (\theta_1^* - \theta_0^*) \). Suppose that \([H]_2, [C] \) and \([B] \) are fulfilled in Case (B). Then \( \mathbb{H}_n \to^{d_s} \mathbb{H}^\tau \) in \( \mathcal{D}([-L, L]) \) 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
\[
M_n^\Delta(v) = M_n(t^* + v(n \delta_n^2)^{-1}; \theta_0^*, \theta_1^*),
A_n^\Delta(v) = A_n(t^* + v(n \delta_n^2)^{-1}; \theta_0^*, \theta_1^*),
\rho_n^\Delta(v) = \rho_n(t^* + v(n \delta_n^2)^{-1}; \theta_0^*, \theta_1^*).
\]
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
\[
M_n^\Delta(v) = 1_{\{\tau > 0\}} \sum_{i = [nt^* + \delta_n^{-2}v]}^{[nt^* + \delta_n^{-2}v]} \text{Tr} \left( S_{i - 1}^{-1} - S_{i - 1}^{-1} \right) \beta_i \]
\[
\cdot h^{-1} \left( \left( \int_{h_{i-1}}^{h_i} \sigma(X_{i}^\tau, \theta_1^*) dW_t \right)^{\otimes 2} - E_{i - 1}^{\theta_1^*} \left[ \int_{h_{i-1}}^{h_i} S(X_{i}^\tau, \theta_1^*) dt \right] \right) + \tilde{\sigma}_p(1) \quad (6)
\]
where \( U_n(v) = \tilde{\sigma}_p(1) \) means that \( \sup_{v \in [0, L]} |U_n(v)| \to^p 0 \), and we used the hypothesis \( n \delta_n^2 \to \infty \) and the fact that \( |S_{i - 1}^{-1} - S_{i - 1}^{-1}| \leq C \delta_n \) with the localization. To obtain \( \tilde{\sigma}_p(1) \), \( L^1 \)-estimate helps. It follows from \([H] \_j \) (i)(c) and (ii) that
\[
\left| h^{-1} \left( \int_{h_{i-1}}^{h_i} S(X_{i}^\tau, \theta_1^*) dt - E_{i - 1}^{\theta_1^*} \left[ \int_{h_{i-1}}^{h_i} S(X_{i}^\tau, \theta_1^*) dt \right] \right) \right|
\]
\[
= \left| h^{-1} \left( \int_{h_{i-1}}^{h_i} [S(X_{i}^\tau, \theta_1^*) - S(X_{h_{i-1}}, \theta_1^*]) dt \right) \right|
\]
\[
- E_{i - 1}^{\theta_1^*} \left[ \int_{h_{i-1}}^{h_i} [S(X_{i}^\tau, \theta_1^*) - S(X_{h_{i-1}}, \theta_1^*)] dt \right] \right| \leq C w_{[0, T]}(n^{-1}, X)^\alpha
\]
\[
= o_p(\delta_n).
\]
Moreover, with the Burkholder–Davis–Gundy inequality, the first terms on the right-hand side of (6) equals \( \tilde{M}_n^\Delta(v) + \tilde{\sigma}_p(1) \) with \( \tilde{M}_n^\Delta(v) = \sum_{i = [nt^* + \delta_n^{-2}v]} \xi_n, i \), where
\[ \xi_{n,i} = 1_{\{\tau > 0\}} \text{Tr} \left[ t \sigma_{i-1}(\theta^*_n) \left( S_i^{-1}(\theta^*_n)^{-1} - S_i^{-1}(\theta^*_0)^{-1} \right) \sigma_{i-1}(\theta^*_n) \right] \times \left( h^{-1}(\Delta_t W) \otimes 2 - I_r \right) \]

and \( \sigma_{i-1}(\theta) = \sigma(X^*_i, \theta) \).

Now we introduce the backward approximation

\[ \tilde{\xi}_{n,i} = 1_{\{\tau > 0\}} \text{Tr} \left[ t \sigma(X^*_i - \epsilon_n, \theta^*_n) \left( S(X^*_i - \epsilon_n, \theta^*_0)^{-1} - S(X^*_i - \epsilon_n, \theta^*_1)^{-1} \right) \sigma(X^*_i - \epsilon_n, \theta^*_n) \right] \times \left( h^{-1}(\Delta_t W) \otimes 2 - I_r \right) \]

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

\[ M_n^\Delta(v) = \tilde{M}_n^\Delta(v) + \delta_p(1). \] (8)

Since

\[ 1_{\{\tau > 0\}} \vartheta_n^{-2} v \left| \sigma(X^*_i - \epsilon_n, \theta^*_n) \left( S(X^*_i - \epsilon_n, \theta^*_0)^{-1} - S(X^*_i - \epsilon_n, \theta^*_1)^{-1} \right) \sigma(X^*_i - \epsilon_n, \theta^*_n) \right|^2 \]
\[ = 1_{\{\tau > 0\}} \vartheta_n^{-2} v \Xi(X^*_i - \epsilon_n, \theta^*)[(\sigma^*_n - \theta^*_0)^\otimes 2] + \delta_p(1) \]
\[ \to p \quad 1_{\{\tau > 0\}} \Xi(X^*_i, \theta^*)[\eta^\otimes 2] v, \]

the central limit theorem ensures the convergence \( \tilde{M}_n^\Delta \to d - 2I_{\eta, \tau}^\frac{1}{2} W \) in \( \mathbb{D}([0, L]) \). In the same fashion, we can show \( \tilde{M}_n^\Delta \to d - 2I_{\eta, \tau}^\frac{1}{2} W \) 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 \( A_n^\Delta \) and \( \rho_n^\Delta \) to \([-L, L] \), which completes the proof. \( \Box \)

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

\[ \lim_{n \to \infty} P[n \vartheta_n^2(\hat{v}_n - t^*) \leq x] = \epsilon \]
\[ \leq \lim_{n \to \infty} P[ \inf_{v \in [-L, x]} \mathbb{H}_n^\Delta(v) \leq \inf_{v \in [-L, x]} \mathbb{H}_n^\Delta(v) ] + \sup_n P[n \vartheta_n^2(\hat{v}_n - t^*) \not\in [-L, L]] \]
\[ = P[ \inf_{v \in [-L, x]} \mathbb{H}_n^\Delta(v) \leq \inf_{v \in [-L, x]} \mathbb{H}_n^\Delta(v) ] + \sup_n P[n \vartheta_n^2(\hat{v}_n - t^*) \not\in [-L, L]] \]
\[ \leq \epsilon + P[\hat{v} \leq x] + P[\hat{v} \not\in [-L, L]] + \sup_n P[n \vartheta_n^2(\hat{v}_n - t^*) \not\in [-L, L]]. \]

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

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

The estimate of \( P[n \vartheta_n^2(\hat{v}_n - t^*) \leq x] \) from below can be done in a similar manner. It is easy to see the stable convergence if we replace \( P \) by \( G \cdot \hat{P} \) in expectations for bounded \( G \)-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

\[
n\vartheta_n^2(\hat{u}_n - u^*) \rightarrow_d \arg\min_{\tilde{v} \in \mathbb{R}} \tilde{\mathbb{H}}(\tilde{v}), \tag{9}
\]

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

\[
T \arg\min_{\tilde{v} \in \mathbb{R}} \tilde{\mathbb{H}}(\tilde{v}) = \arg\min_{v \in \mathbb{R}} \tilde{\mathbb{H}}\left( \frac{v}{T} \right) \equiv^d \arg\min_{v \in \mathbb{R}} \mathbb{H}(v)
\]

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

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 \( \hat{u}_n \) so that it takes values in \( \mathcal{G}_n \), and set \( \hat{k}_n = n\hat{u}_n/T \). Let

\[
\mathbb{K}(v) = \sum_{i=1}^v \left\{ \text{Tr} \left[ 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}^*) \xi_{i}^{2} \right] \right.

- \log \det \left( S(X_{t^*}, \theta_{0}^*)^{-1} S(X_{t^*}, \theta_{1}^*) \right) \right\},
\]

where \( \xi_i \) are independent \( r \)-dimensional standard normal variables independent of \( \mathcal{F}_T \).

**Theorem 3.** Suppose that \([H]_1\), \([C]\) and \([A]\) are fulfilled in Case (A). Then \( \hat{k}_n - [n^{1/2}] \rightarrow_d \arg\min_{v \in \mathbb{Z}} \mathbb{K}(v) \) as \( n \rightarrow \infty \).

**Proof.** We change the definition of \( t^*_v \) and newly set \( t^*_v = [n^{1/2}] T/n + T/ n \). Lemma 3 is still valid by essentially the same proof and hence we may only consider \( \Psi_n(t^*_v; \theta_{0}^*, \theta_{1}^*) - \Psi_n(t^*; \theta_{0}^*, \theta_{1}^*) \). Writing \( \Psi_n(t; \theta_{0}^*, \theta_{1}^*) \) for \( \Psi_n(t^*_v; \theta_{0}^*, \theta_{1}^*) \), we will investigate the behavior of the random field

\[
\mathbb{K}_n(v) = \Psi_n(t^*_v) - \Psi_n(t^*)
\]

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

\[
\mathbb{K}_n(v) = \mathbb{M}_n(v) + \mathbb{A}_n(v) + \mathbb{G}_n(v),
\]

where \( \mathbb{M}_n(v) = M_n(t^*_v; \theta_{0}^*, \theta_{1}^*), \mathbb{A}_n(v) = A_n(t^*_v; \theta_{0}^*, \theta_{1}^*) \) and \( \mathbb{G}_n(v) = \rho_n(t^*_v; \theta_{0}^*, \theta_{1}^*) \).
Now, $\mathbb{M}_n(v)$ admits a similar expansion as before:

$$\mathbb{M}_n(v) = \sum_{i=\lfloor nt/T \rfloor+1}^{\lfloor (n+1)t/T \rfloor+1} \xi_{n,i} + \tilde{\sigma}(1)$$

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,

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

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 $\sigma$ and because $W$ is an $\mathcal{F}$-Wiener process, Proposition 1 of Section 5 gives

$$(G, X_{t^*}^T - \epsilon_n, (h^{-1}(\Delta_i W)^{\otimes 2})_{i=\lfloor nt/T \rfloor+L_0}^{\lfloor nt/T \rfloor+L_0}) \rightarrow^d (G, X_{t^*}^T, (\zeta_i)^{\otimes 2})_{i=-L_0}^{L_0},$$

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

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

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

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

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

$$\mathbb{A}_\infty(v) = 1_{\{r > 0\}} v \left\{ \text{Tr} \left( S(X_{t^*}^T, \theta_0^*)^{-1} S(X_{t^*}^T, \theta_1^*) - I_d \right) \right. \right.$$  

$$\left. - \log \det \left( S(X_{t^*}^T, \theta_0^*)^{-1} S(X_{t^*}^T, \theta_1^*) \right) \right\}.$$

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

$$(\mathbb{K}_n(v))_{v=-L_0}^{L_0} \rightarrow^d (\mathbb{K}_\infty(v))_{v=-L_0}^{L_0},$$

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

**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 $|v| \rightarrow \infty$. 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 \( X^n = (\xi_{n,i})_{i = [nt^*/T] - L_0} \) with
\[
\xi_{n,i} = h^{-1}(\Delta_i W) \otimes^2 - I_r.
\]
Let \( F \) denote any bounded \( \mathcal{F} \)-measurable function and let \( G = E[F|\mathcal{F}_T] \). Since \( L[X^n] \) does not depend on \( n \), \((X^n, G)_{n \in \mathbb{N}} \) is tight. By the subsequence argument, we may assume that \((X^n, G) \rightarrow^d (X^\infty, G^\infty) \) as \( n \rightarrow \infty \) for the canonical process \((X^\infty, G^\infty) \) on \( \mathbb{R}^{2L_0+2} \) equipped with a probability measure \( P^\infty \); the uniqueness of the limit will be found later. Thus we have \( E[f(X^n)G] \rightarrow E_\infty[f(X^\infty)G^\infty] \) for every \( f \in C_B(\mathbb{R}^{2L_0+2}) \), where \( E^\infty \) denotes the expectation with respect to \( P^\infty \). Let \( \mathcal{W} = (\xi_i \otimes^2 - I_r)_{i = -L_0} \). If the equality
\[
E_\infty[f(X^\infty)G^\infty] = E_\infty[f(X^\infty)]E_\infty[G^\infty]
\]
is obtained for \( \mathcal{F}_T \)-measurable functions \( G \), we also have
\[
E[f(X^n)F] = E[f(X^n)G] \\
= E_\infty[f(X^\infty)G^\infty] \\
= E_\infty[f(X^\infty)]E_\infty[G^\infty] \\
= E_\mathcal{W}[f(\mathcal{W})]E[f] \\
= E_\mathcal{W}[f(\mathcal{W})F].
\]
This characterizes the possible limit of the sequence \((X^n, F)_{n \in \mathbb{N}} \) uniquely and implies the stable convergence of \( X^n \). 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
\[
J(h, \alpha)_t = \int_0^t h^1_{s_1} dW_{s_1}^\alpha_1 \int_0^{s_1} h^2_{s_2} dW_{s_2}^\alpha_2 \cdots \int_0^{s_{m-1}} h^m_{s_m} dW_{s_m}^\alpha_m.
\]
Let
\[
\mathcal{E}(h_1, \alpha_1, p_1; \ldots; h_k, \alpha_k, p_k; G)_t = E[J(h_1, \alpha_1)^{p_1} \cdots J(h_k, \alpha_k)^{p_k} G],
\]
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 \( G \),
\[
\sup_{t \in [0,1]} |\mathcal{E}(h_1, \alpha_1, p_1; \ldots; h_k, \alpha_k, p_k; G)_t - \mathcal{E}(h_1, \alpha_1, p_1; \ldots; h_k, \alpha_k, p_k; 1)_t E[G]| \\
\leq C(p_1, \ldots, p_k) \prod_{i=1}^k \left( \prod_{j=1}^{m_j} \|h^j_i\|_2 \right)^{p_i} \max_{i,j} \left( E \left[ \int_0^1 1_{\{h_i^j \neq 0\}} d(E[G|\mathcal{F}_i])_t \right] \right)^{1/2},
\]
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 $G = E[G|\mathcal{F}_1] = E[G|\mathcal{F}_0] + M^c_t + M^d_t$ for some continuous $L^2$-martingale $M^c$ with $M^c_0 = 0$ and some purely discontinuous $L^2$-martingale $M^d$ with $M^d_0 = 0$.

First,

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

(11)

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

Next, thanks to Itô’s formula,

$$E(h_1, \alpha_1, p_1; \ldots; h_k, \alpha_k, p_k; G)_t$$

$$= \sum_{i=1}^k E \left[ \int_0^t p_i J(h_1, \alpha_1)_s \cdots J(h_i, \alpha_i)_s \cdots J(h_k, \alpha_k)_s dW^a_i G \right]$$

$$\times \cdots J(h_k, \alpha_k)_s J(h_i^{(-1)}, \alpha_i^{(-1)})_s (h_i^1)_s dW^a_i$$

$$+ \sum_{i,j,i\neq j} p_i p_j \int_0^t E \left( \cdots; h_i^{(-1)}, \alpha_i^{(-1)}, p_i - 1; \cdots; h_j^{(-1)}, \alpha_j^{(-1)}, p_j - 1; \cdots; \right.$$}

$$\times h_i^{(-1)}, \alpha_i^{(-1)}, 1; h_j^{(-1)}, \alpha_j^{(-1)}, 1; G) \right) (h_i^1)_s (h_j^1)_s d(W^a_i, W^a_j)_s$$

$$\left( h_i^{(-1)}, \alpha_i^{(-1)}, p_i - 2; \cdots; h_i^{(-1)}, \alpha_i^{(-1)}, 2; G \right)_s (h_i^1)_s ds$$

(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

$$\left| E \left[ \int_0^t J(h_1, \alpha_1)_s \cdots J(h_i, \alpha_i)_s \cdots J(h_k, \alpha_k)_s dW^a_i M^c_t \right] \right|$$

$$\leq \left| E \left[ \int_0^t J(h_1, \alpha_1)_s \cdots J(h_i, \alpha_i)_s \cdots J(h_k, \alpha_k)_s dW^a_i M^c_t \right] \right|$$

(13)

and

$$E \left[ \int_0^t \left\{ J(h_1, \alpha_1)_s \cdots J(h_i, \alpha_i)_s \cdots J(h_k, \alpha_k)_s \right\}^2 ds \right]$$

$$\leq C(p_1, \ldots, p_k) \prod_{i=1}^k \left( \prod_{j=1}^{m_i} \|h_i^j\|_2 \right)^{2p_i}$$

(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 \( G = M^c_1 \) and using (13) and (14), we obtain

\[
\sup_{t \in [0,1]} \left| \mathcal{E}(h_1, \alpha_1, p_1; \cdots; h_k, \alpha_k, p_k; M^c_1)_t \right| \\
\leq C(p_1, \ldots, p_k) \prod_{i=1}^{k} \left( \prod_{j=1}^{m_i} \|h^j_i\|_2 \right)^{p_i} \max_{i,j} \left( E \left[ \int_0^1 1_{\{h^j_i\}_\tau \neq 0} d(M^c)_\tau \right] \right)^{1/2} \\
+ C(p_1, \ldots, p_k) \\
\times \sup_{t \in [0,1]} \left| \mathcal{E} \left( \cdots; h^{(-1)}_i, \alpha^{(-1)}_i, p_i - 1; \cdots; h^{(-1)}_j, \alpha^{(-1)}_j, p_j - 1; \cdots; h^{(-1)}_i, \alpha^{(-1)}_i, 1; h^{(-1)}_j, \alpha^{(-1)}_j, 1; M^c_1 \right)_t \right| \|h^1_i\|_2 \|h^1_j\|_2 \\
+ C(p_1, \ldots, p_m) \\
\times \sup_{t \in [0,1]} \left| \mathcal{E} \left( \cdots; h^{(-1)}_i, \alpha^{(-1)}_i, p_i - 2; \cdots; h^{(-1)}_i, \alpha^{(-1)}_i, 2; M^c_1 \right)_t \right| \|h^1_i\|_2^2. \tag{15}\]

We apply (15) repeatedly together with \( E[M^c_1] = 0 \) to obtain

\[
\sup_{t \in [0,1]} \left| \mathcal{E}(h_1, \alpha_1, p_1; \cdots; h_k, \alpha_k, p_k; M^c_1)_t \right| \\
\leq C(p_1, \ldots, p_k) \prod_{i=1}^{k} \left( \prod_{j=1}^{m_i} \|h^j_i\|_2 \right)^{p_i} \max_{i,j} \left( E \left[ \int_0^1 1_{\{h^j_i\}_\tau \neq 0} d(M^c)_\tau \right] \right)^{1/2}. \tag{16}\]

If apply (12) for \( G = M^d_1 \) instead of \( M^c_1 \), the first term on the right-hand side of (12) vanishes, and we obtain

\[
\mathcal{E}(h_1, \alpha_1, p_1; \cdots; h_k, \alpha_k, p_k; M^d_1)_t = 0 \tag{17}\]

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 \( \theta_k \)

When the values of the parameters \( \theta_k \) \( (k = 0, 1) \) are unknown, our estimating function for the change point requires certain estimators for \( \theta_k \). In this section, we will briefly discuss estimation of the nuisance parameters \( \theta_k \). 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 \( \mathbb{T} = [t_0, t_1] \subset (0, T) \) for given numbers \( t_0 \) and \( t_1 \). In this case, the structural change occurs in neither interval \([0, t_0)\) nor \((t_1, T]\), and we can use this knowledge to construct estimators for \( \theta_k \). 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 \( \theta_k \) by shrinkage of the sampling time intervals.
Let
\[
\phi_0^k(t; \theta_0) = \sum_{i=1}^{[nt/T]} G_j(\theta_0) \quad \text{and} \quad \phi_1^k(t; \theta_1) = \sum_{i=[nt/T]+1}^{n} G_i(\theta_1).
\]

Suppose that \(t_0\) and \(t_1\) are known. Let us consider estimators \(\hat{\theta}_k\) for \(\theta_k\) such that
\[
\phi_n^k(t_k; \hat{\theta}_k) = \min_{\theta_k} \phi_n^k(t_k; \theta_k)
\]
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
\[
\int_0^{b_n} Q(X_t, \theta_0, \theta) \, dt > 0 \quad \text{a.s. and} \quad \int_0^{T} Q(X_t, \theta_0, \theta) \, dt > 0 \quad \text{a.s.}
\]
(18) for every \(\theta \neq \theta_0\), 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 \(\tilde{\theta}_n\) are defined. According to the previous sections, \(\tilde{\theta}_n\) possesses \(\theta_0^2\)-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 \(n \theta_0^2 b_n \to \infty\) as \(n \to \infty\). Construct \(\tilde{\theta}_k\) so that
\[
\phi_n^k(\tilde{\theta}_n + (-1)^{k+1} b_n; \tilde{\theta}_k) = \min_{\theta_k} \phi_n^k(\tilde{\theta}_n + (-1)^{k+1} b_n; \theta_k)
\]
for \(k = 0, 1\).\(^3\) The new estimators \(\tilde{\theta}_k\) 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 \(\tilde{\theta}_k\), we define \(\tilde{\theta}_n\) for \(t^*\) as
\[
\tilde{\theta}_n = \arg\min_{t \in [0, T]} \phi_n(t; \tilde{\theta}_0, \tilde{\theta}_1).
\]

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

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 \(\theta_0\) 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 \(\theta_1\). This consideration suggests the use of estimators \(\tilde{\theta}_k\) based on the data over time interval \([0, a_n]\) for \(\theta_0\) and the data over \([T - a_n, T]\) for \(\theta_1\), 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 \(\theta_k\), as they are necessary for the discussions here, in a fairly general setting even for possibly moving targets \(\theta_k\).

\(^3\)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 $a_n \geq 1/(n\theta_n^{1/\beta})$ and that $|\hat{\theta}_k - \theta_k^*| = o_p((n\theta_n)^{-\beta})$ for $k = 0, 1$. When $\lim_{n \to \infty} \theta_n > 0$, we also assume $n\theta_n \to \infty$. In particular, the first condition implies $n\theta_n^2 \to \infty$. The second condition is natural because the number of the available data over each time interval is proportional to $n\theta_n$. To obtain consistent estimators, 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 $\theta$, however it is necessary because the sampling interval is shrinking to an end point, by lack of knowledge of $T$.

Here we show how to obtain an initial estimator of $\theta^*$, 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

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

as $n \to \infty$. Then the nondegeneracy of $\hat{\theta}_0 S(x, \theta)$ near $x_0$ with the Taylor formula yields the stochastic order $\hat{\theta}_0^m - \theta_0^* = O_p(n\theta_n)^{-1/2}$. 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 $\hat{i}_n$, $\hat{\theta}_k$ and $\hat{i}_n$ in turn as mentioned above. The asymptotic properties of $\hat{i}_n$ are the same as $\hat{i}_n$ because $\hat{\theta}_k$’s satisfy Condition $[C]$. It is expected that the new estimator $\hat{i}_n$ possesses equal or better precision than $\hat{i}_n$, 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 $y_t$ 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

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

(19)
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 $\gamma = 1/4$, $n$ 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 $a_n = b_n = n^{-\gamma}$ with $\delta = 3$ so that they satisfy the properties required in Section 6. The first stage estimator of $\theta_0^*$ (resp. $\theta_1^*$) 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 $\theta_0^*$ and $\theta_1^*$ 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 $\theta_i^*$ using observations in the interval $[0, \hat{t}_n - b_n]$ for $\theta_0^*$ and observations in the interval $[\hat{t}_n + b_n, T]$ for $\theta_1^*$. We denote the second stage estimators of $\theta_i^*$ by $\tilde{\theta}_i$. Finally, the second stage estimator of $t^*$, i.e. $\hat{t}_n$, is obtained as

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

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 $\theta_1^* = 0.378, 0.350, \text{ and } 0.319$ 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 $\theta_0$ and $\theta_1$ and the change point $t^*$. In parenthesis are the standard deviations of the Monte Carlo estimates.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$a_n$</th>
<th>$\hat{t}_n$</th>
<th>$\tilde{t}_n$</th>
<th>$\hat{\theta}_0$</th>
<th>$\tilde{\theta}_0$</th>
<th>$\hat{\theta}_1$</th>
<th>$\tilde{\theta}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5000</td>
<td>0.1189</td>
<td>0.601</td>
<td>0.319</td>
<td>0.601</td>
<td>0.319</td>
<td>0.601</td>
<td>0.319</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.005)</td>
<td>(0.014)</td>
<td>(0.011)</td>
<td>(0.005)</td>
<td>(0.013)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>2000</td>
<td>0.1495</td>
<td>0.601</td>
<td>0.349</td>
<td>0.601</td>
<td>0.349</td>
<td>0.601</td>
<td>0.349</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.008)</td>
<td>(0.020)</td>
<td>(0.014)</td>
<td>(0.008)</td>
<td>(0.017)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>1000</td>
<td>0.1778</td>
<td>0.601</td>
<td>0.377</td>
<td>0.601</td>
<td>0.377</td>
<td>0.601</td>
<td>0.377</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.011)</td>
<td>(0.025)</td>
<td>(0.019)</td>
<td>(0.026)</td>
<td>(0.018)</td>
<td></td>
</tr>
</tbody>
</table>
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 \( t^* = 0.7 \).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( a_n )</th>
<th>( \hat{a}_n )</th>
<th>( \hat{\theta}_0 )</th>
<th>( \hat{\theta}_1 )</th>
<th>( \hat{t}_n )</th>
<th>( \hat{\theta}_0 )</th>
<th>( \hat{\theta}_1 )</th>
<th>( \hat{t}_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5000</td>
<td>0.1189</td>
<td>0.701</td>
<td>0.200</td>
<td>0.319</td>
<td>0.701</td>
<td>0.200</td>
<td>0.319</td>
<td>0.701</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.012)</td>
<td>(0.018)</td>
<td>(0.011)</td>
<td>(0.018)</td>
<td>(0.012)</td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>0.1495</td>
<td>0.702</td>
<td>0.200</td>
<td>0.350</td>
<td>0.701</td>
<td>0.200</td>
<td>0.350</td>
<td>0.701</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.016)</td>
<td>(0.029)</td>
<td>(0.024)</td>
<td>(0.009)</td>
<td>(0.030)</td>
<td>(0.021)</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>0.1778</td>
<td>0.703</td>
<td>0.200</td>
<td>0.378</td>
<td>0.701</td>
<td>0.200</td>
<td>0.377</td>
<td>0.701</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.021)</td>
<td>(0.040)</td>
<td>(0.038)</td>
<td>(0.012)</td>
<td>(0.056)</td>
<td>(0.040)</td>
<td></td>
</tr>
</tbody>
</table>

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 \( \tau^* \) 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 \( \hat{\tau} \) 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 \( n\theta_n^2(\hat{t}_n - t^*) \) is required. Therefore, the limiting distribution of the following sequence is studied

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

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

\[
F(x) = \begin{cases} 
g(x), & x > 0, \\
1 - g(-x), & x \leq 0.
\end{cases}
\]
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

$$g(x) = 1 + \sqrt{\frac{x}{2\pi}} e^{-\frac{x}{2}} - \frac{1}{2} (x + 5) \Phi \left( \frac{-\sqrt{x}}{2} \right) + \frac{3}{2} e^{\frac{x}{4}} \Phi \left( \frac{-3}{2} \sqrt{x} \right)$$

(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


