



UNIVERSITÀ DEGLI STUDI DI BERGAMO  
DIPARTIMENTO DI INGEGNERIA  
QUADERNI DEL DIPARTIMENTO

**Department of Engineering**

**Working Paper**

**Series “*Mathematics and Statistics*”**

n. 05/MS – 2014

**Z-PROCESS METHOD FOR CHANGE POINT PROBLEMS**

by

I. Negri and Y. Nishiyama

## **COMITATO DI REDAZIONE<sup>§</sup>**

Series Mathematics and Statistics (MS): Luca Brandolini, Alessandro Fassò, Christian Vergara

---

<sup>§</sup> L'accesso alle *Series* è approvato dal Comitato di Redazione. I *Working Papers* della Collana dei Quaderni del Dipartimento di Ingegneria dell'Informazione e Metodi Matematici costituiscono un servizio atto a fornire la tempestiva divulgazione dei risultati dell'attività di ricerca, siano essi in forma provvisoria o definitiva.

# Z-process method for change point problems

Ilia Negri and Yoichi Nishiyama\*

*Department of Information Technology and Mathematical Methods, University of Bergamo*

*Viale Marconi, 5, 24044, Dalmine (BG), Italy*

`ilia.negri@unibg.it`

and

*The Institute of Statistical Mathematics*

*10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan*

`nishiyama@ism.ac.jp`

April 23, 2014

## Abstract

The aim of this paper is to develop a general, unified approach, based on some partial estimation functions which we call “Z-process”, to the change point problems. The method proposed can be applied not only to ergodic models but also to some models where the Fisher information matrix is random. Applications to some concrete models, including especially a parametric model for volatilities of diffusion processes are presented. Simulations for randomly time-transformed Brownian bridge process appearing as the limit of the proposed test statistics are performed with computer intensive use.

## 1 Introduction

The aim of this paper is to develop a general and unified approach to change point problems using the partial sum process of estimating equations, which may be called *Z-process method*.

Let us give an illustration by the example of independent data. Let  $(\mathcal{X}, \mathcal{A}, \mu)$  be a measure space, and let us be given a parametric family of probability densities  $f(\cdot; \theta)$  with respect to  $\mu$ , where  $\theta \in \Theta \subset \mathbb{R}^d$ . Let  $X_1, X_2, \dots$  be an independent sequence of  $\mathcal{X}$ -valued random variables from this parametric model. There are at least two ways to define the “maximum likelihood estimator (MLE)” in statistics.

---

\*Corresponding author.

One way is to define it as the maximum point of the random function

$$\theta \mapsto \mathbb{M}_n(\theta) = \sum_{k=1}^n \log f(X_k; \theta),$$

while the other is to do it as the solution to the estimating equation

$$\mathbb{Z}_n(\theta) := \dot{\mathbb{M}}_n(\theta) = 0,$$

where  $\dot{\mathbb{M}}_n(\theta)$  is the gradient vector of  $\mathbb{M}_n(\theta)$ . The former is a special case of “ $M$ -estimators”, and the latter is that of “ $Z$ -estimators”; see van der Vaart and Wellner (1996) for these terminologies.

To deal with change point problem let us introduce the partial sum process

$$\mathbb{M}_n(u, \theta) = \sum_{k=1}^{[un]} \log f(X_k; \theta), \quad \forall u \in [0, 1],$$

and denote its gradient vectors by  $\mathbb{Z}_n(u, \theta) = \dot{\mathbb{M}}_n(u, \theta)$ . Let  $\hat{\theta}_n$  be the MLE for the full data  $X_1, \dots, X_n$  as a special case of  $Z$ -estimators, that is,  $\hat{\theta}_n$  is the solution to the estimating equation

$$\mathbb{Z}_n(1, \theta) = \dot{\mathbb{M}}_n(1, \theta) = 0.$$

The fact that the random process

$$u \rightsquigarrow n^{-1/2} \mathbb{Z}_n(u, \theta_0) \quad \text{converges weakly to} \quad u \rightsquigarrow I(\theta_0)^{1/2} B(u)$$

in the Skorohod space  $D[0, 1]$ , where  $u \rightsquigarrow B(u)$  is a vector of independent standard Brownian motions, is immediate from Donsker’s theorem. However, it does not seem so well known that the random process

$$u \rightsquigarrow n^{-1/2} \mathbb{Z}_n(u, \hat{\theta}_n) \quad \text{converges weakly to} \quad u \rightsquigarrow I(\theta_0)^{1/2} B^\circ(u) \quad (1)$$

in  $D[0, 1]$ , where  $u \rightsquigarrow B^\circ(u)$  is a vector of independent standard Brownian bridges. Horváth and Parzen (1994) are apparently the firsts to introduce the statistic

$$\mathcal{T}_n = n^{-1} \sup_{u \in [0, 1]} \mathbb{Z}_n(u, \hat{\theta}_n)^\top \hat{I}_n^{-1} \mathbb{Z}_n(u, \hat{\theta}_n),$$

where  $\hat{I}_n$  is a consistent estimator for the Fisher Information matrix  $I(\theta_0)$ . It is straightforward from the fact (1) and the continuous mapping theorem that

$$\mathcal{T}_n \rightarrow^d \sup_{u \in [0, 1]} \|B^\circ(u)\|^2.$$

See also Tsukuda and Nishiyama (2014) who took an approach based on weak convergence in  $L^2$ -spaces to this problem.

Although Horváth and Parzen (1994) didn’t discuss the asymptotic behavior of the test under the alternative, Negri and Nishiyama (2012) who took the same

approach to the change point problem for an ergodic diffusion process model based on the continuous observation proved also the consistency of the test under an alternative which has sufficient generality. The argument for alternatives in the latter paper works also for the case of independent data. In Section 2 of this paper, we will present a generalized version of this methods which works also for some more general models. We prove that the limit of the test statistic under the null hypothesis is a standard Brownian bridge in ergodic statistical models, and that it is a randomly time-transformed Brownian bridge in non-ergodic ones. Moreover, we develop the method for the consistency under alternatives further. Some new examples will be given.

## 1.1 Notations and examples

In the rest of this section, we shall list up some examples which fit in our theories. In what follows, the parametric space  $\Theta$  is a bounded, open, convex subset of  $\mathbb{R}^d$ , where  $d$  is a fixed, positive integer. The word “vector” always means “ $d$ -dimensional real column vector”, and the word “matrix” does “ $d \times d$  real matrix”. The Euclidean norm is denoted by  $\|v\| := \sqrt{\sum_{i=1}^d |v^{(i)}|^2}$  for a vector  $v$  where  $v^{(i)}$  denotes the  $i$ -th component of  $v$ , and by  $\|A\| := \sqrt{\sum_{i,j=1}^d |A^{(i,j)}|^2}$  for a matrix  $A$  where  $A^{(i,j)}$  denotes the  $(i, j)$ -component of  $A$ . Note that  $\|Av\| \leq \|A\| \cdot \|v\|$  and  $\|AB\| \leq \|A\| \cdot \|B\|$  for vector  $v$  and matrices  $A, B$ . The notations  $v^\top$  and  $A^\top$  denote the transpose. We use also the notation  $A \circ B$  defined by  $(A \circ B)^{(i,j)} := A^{(i,j)} B^{(i,j)}$  for two matrices  $A, B$  (the Hadamard product). We denote by  $I_d$  the identity matrix. The notations  $\rightarrow^p$  and  $\rightarrow^d$  mean the convergence in probability and the convergence in distribution, as  $n \rightarrow \infty$ , respectively.

In the following examples, the rate matrices  $Q_n$  and  $R_n$ , which are diagonal matrices like  $\sqrt{n}I_d$ , are given in such a way that  $Q_n^{-1} \ddot{M}_n(u, \theta) R_n^{-1}$  converges to a limit, where  $\ddot{M}_n(u, \theta)$  is something like a Hessian matrix of the partial sum process  $\mathbb{M}_n(u, \theta)$  of a contrast function like a loglikelihood function.

### Example A: Ergodic diffusion process

Let  $I = (l, r)$ , where  $-\infty \leq l < r \leq \infty$ , be given. Let us consider an  $I$ -valued diffusion process  $t \rightsquigarrow X_t$  which is the unique strong solution to the stochastic differential equation (SDE)

$$X_t = X_0 + \int_0^t S(X_s; \alpha) ds + \int_0^t \sigma(X_s; \beta) dW_s,$$

where  $s \rightsquigarrow W_s$  is a standard Wiener process. The parameters come from  $\alpha \in \Theta_A \subset \mathbb{R}^{d_A}$  and  $\beta \in \Theta_B \subset \mathbb{R}^{d_B}$ , and we denote  $\theta = (\alpha^\top, \beta^\top)^\top$ . We are supposed to be able to observe the process  $X$  at discrete time grids  $0 = t_0^n < t_1^n < \dots < t_n^n$ , and we shall consider the asymptotic scheme  $n\Delta_n^2 \rightarrow 0$  and  $t_n^n \rightarrow \infty$  as  $n \rightarrow \infty$ , where

$$\Delta_n = \max_{1 \leq k \leq n} |t_k^n - t_{k-1}^n|,$$

and

$$\sum_{k=1}^n \left| \frac{|t_k^n - t_{k-1}^n|}{t_n^n} - \frac{1}{n} \right| \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (2)$$

We introduce

$$\mathbb{Z}_n(u, \theta) = \dot{\mathbb{M}}_n(u, \theta) = \dot{\gamma}_n(u, \theta),$$

where

$$\begin{aligned} & \gamma_n(u, \theta) \\ &= - \sum_{k:t_{k-1}^n \leq ut_n^n} \left\{ \log \sigma(X_{t_{k-1}^n}; \beta) + \frac{|X_{t_k^n} - X_{t_{k-1}^n} - S(X_{t_{k-1}^n}; \alpha)|t_k^n - t_{k-1}^n|^2}{2\sigma(X_{t_{k-1}^n}; \beta)^2 |t_k^n - t_{k-1}^n|} \right\} \end{aligned}$$

and  $R_n = Q_n$  are the diagonal matrices such that  $R_n^{(i,i)}$  and  $Q_n^{(i,i)}$  are  $\sqrt{t_n^n}$  for  $i = 1, \dots, d_A$  and  $\sqrt{n}$  for  $i = d_A + 1, \dots, d$  with  $d = d_A + d_B$ .

Song and Lee (2009) proposed a statistic for testing the existence of a change point of the parameter  $\beta$ , but the problem to test it for both parameters was left as an open problem in their paper (see their Section 5). We will give an answer to this problem with the new method that can treat both parameters simultaneously even if, as in this problem, the parameters  $\alpha$  and  $\beta$ , have different rates of convergence.

### Example B: Volatility of diffusion process

Let  $I = (l, r)$ , where  $-\infty \leq l < r \leq \infty$ , be given. Let us consider an  $I$ -valued diffusion process  $t \rightsquigarrow X_t$  which is the unique strong solution to the SDE

$$X_t = X_0 + \int_0^t S(X_s) ds + \int_0^t \sigma(X_s; \theta) dW_s,$$

where  $s \rightsquigarrow W_s$  is a standard Wiener process. Here, the drift coefficient  $S(\cdot)$  is treated as an unknown nuisance function. We are supposed to be able to observe the process  $X$  at discrete time grids  $0 = t_0^n < t_1^n < \dots < t_n^n = T < \infty$ , and we shall consider the asymptotic scheme (2).

We introduce

$$\mathbb{Z}_n(u, \theta) = \dot{\mathbb{M}}_n(u, \theta) = \dot{\gamma}_n(u, \theta),$$

where

$$\gamma_n(u, \theta) = - \sum_{k:t_{k-1}^n \leq ut_n^n} \left\{ \log \sigma(X_{t_{k-1}^n}; \theta) + \frac{|X_{t_k^n} - X_{t_{k-1}^n}|^2}{2\sigma(X_{t_{k-1}^n}; \theta)^2 |t_k^n - t_{k-1}^n|} \right\}.$$

The rate matrices are given by  $R_n = Q_n = \sqrt{n}I_d$ .

As we already mentioned, Song and Lee (2009) proposed a statistic for testing the existence of a change point in the volatility of an ergodic diffusion process under the asymptotic scheme  $n\Delta_n^a \rightarrow 0$  and  $n\Delta_n^b \rightarrow \infty$  for some  $a > b > 4$ , by an approach which is different from ours.

A new aspect of interest is that in the current situation of parametric models for the volatility of diffusion process based on discrete time observation in a compact time interval where the Fisher information matrix is random and depending on  $t$ , some computer intensive method is used to perform computer simulations for randomly time-transformed Brownian bridges appearing as the limit of the proposed test statistics.

### Example C: Cox's regression model

Let a sequence of counting processes  $t \rightsquigarrow N_t^k$ ,  $k = 1, 2, \dots$ , which do not have simultaneous jumps, be observed the time interval  $[0, T]$ . Suppose that  $t \rightsquigarrow N_t^k$  has the intensity

$$\lambda_t^k(\theta) = \alpha(t)e^{\theta^\top Z_t^k} Y_t^k,$$

where the baseline hazard function  $\alpha$  which is common for all  $k$ 's is non-negative and satisfies that  $\int_0^T \alpha(t)dt < \infty$ , the random process  $t \rightsquigarrow Z_t^k$  is a vector valued covariate for the individual  $k$ , and the random process  $t \rightsquigarrow Y_t^k$  is given by

$$Y_t^k = \begin{cases} 1, & \text{if the individual } k \text{ is observed at time } t, \\ 0, & \text{otherwise.} \end{cases}$$

This model was introduced by Cox (1972), and its asymptotic theory was developed by Andersen and Gill (1982).

We introduce

$$\mathbb{Z}_n(u, \theta) = \dot{\mathbb{M}}_n(u, \theta) = \dot{\gamma}_n(u, \theta),$$

where

$$\gamma_n(u, \theta) = \sum_{k=1}^n \int_0^{uT} (\theta^\top Z_t^k - \log S_t^{n,0}(\theta)) dN_t^k$$

with

$$S_t^{n,0}(\theta) = \sum_{k=1}^n e^{\theta^\top Z_t^k} Y_t^k.$$

The rate matres are  $R_n = Q_n = \sqrt{n}I_d$ .

### Example D: Counting process models

Let  $t \rightsquigarrow N_t$  be a counting process with the intensity  $t \rightsquigarrow \lambda_t(\theta)$ . We suppose that we can observe the processes on the compact time interval  $[0, T_n]$ , and consider the asymptotic scheme  $T_n \rightarrow \infty$ . Our method may be typically applied to

$$\mathbb{Z}_n(u, \theta) = \dot{\mathbb{M}}_n(u, \theta) = \frac{1}{T_n} \dot{\gamma}_n(u, \theta),$$

where

$$\gamma_n(u, \theta) = \int_0^{uT_n} \log \lambda_t(\theta) dN_t - \int_0^{uT_n} \lambda_t(\theta) dt.$$

The rate matrix is typically  $R_n = \sqrt{T_n}I_d$ . An example of this model is the stress release process introduced by Isham and Westcott (1979):

$$\lambda_t(\alpha, \beta, \theta) = \phi(\alpha t - \beta N_{t-}; \theta).$$

It is known that the process  $t \rightsquigarrow X_t := \alpha t - \beta N_{t-}$  is ergodic under some mild conditions. A test statistic for the change point problem in this model, which is different from ours, has been proposed by Fujii and Nishiyama (2011).

### Example E: Non-linear time series models

Our method may be applied to the time series models of the form

$$X_k = a(X_{k-1}, X_{k-2}, \dots; \theta) + b(X_{k-1}, X_{k-2}, \dots; \theta)\varepsilon_k, \quad k = 1, 2, \dots$$

Here,  $\{\varepsilon_k\}$  is an i.i.d. sequence with  $E[\varepsilon_1] = 0$  or, more generally, a martingale difference sequence with respect to the filtration  $(\mathcal{F}_k)_{k \geq 0}$  where  $\mathcal{F}_k = \sigma(X_k, X_{k-1}, \dots)$ . A possible way to define the estimating function is

$$\mathbb{Z}_n(u, \theta) = \dot{\mathbb{M}}_n(u, \theta) = \dot{\gamma}_n(u, \theta),$$

where

$$\gamma_n(u, \theta) = - \sum_{k=1}^{[un]} \left\{ \log b(X_{k-1}, X_{k-2}, \dots; \theta) + \frac{|X_k - a(X_{k-1}, X_{k-2}, \dots; \theta)|^2}{2b(X_{k-1}, X_{k-2}, \dots; \theta)^2} \right\}.$$

The rate matrices are typically given by  $R_n = Q_n = \sqrt{n}I_d$ .

Some detailed discussions for Examples A, B, and C will be given in Sections 3, 3.1 and 3.2, respectively, while those for Examples D and E are left for readers.

## 2 Z-process method for change point problems

Let  $D[0, 1]$  be the space of functions defined on  $[0, 1]$  taking values in a finite-dimensional Euclidean space, which are right continuous and have left hand limits; we equip this space with the Skorohod metric. Throughout this section, all random processes, denoted as like  $u \rightsquigarrow X(u)$ , are assumed to take values in  $D[0, 1]$ .

Let  $\Theta$  be a bounded, open, convex subset of  $\mathbb{R}^d$ . For every  $n \in \mathbb{N}$ , let  $u \rightsquigarrow \mathbb{Z}_n(u, \theta)$  be an  $\mathbb{R}^d$ -valued random process indexed by  $\theta \in \Theta$ , defined on a probability space  $(\Omega, \mathcal{F}, P)$  that is common for all  $n \in \mathbb{N}$ . (However, it will be clear from our proofs that the underlying probability spaces do not have to be common for  $n \in \mathbb{N}$  if the objects  $\mathcal{Z}(u, \theta)$  and  $V(u, \theta_0)$  appearing in the limit below are non-random.)

This framework includes the case where  $\mathbb{Z}_n(u, \theta)$  is given as the gradient vector  $\dot{\mathbb{M}}_n(u, \theta)$  of  $\theta \mapsto \mathbb{M}_n(u, \theta)$  which is assumed to be two times continuously differentiable with the Hessian matrix  $\ddot{\mathbb{M}}_n(u, \theta)$ .

We consider the following testing problem:



$H_0$ : the true value  $\theta_0 \in \Theta$  does not change during  $u \in [0, 1]$ ;  
 $H_1$ : “not  $H_0$ ”.

The meaning of “not  $H_0$ ” will be precisely specified in the condition [A] below. Let us describe some properties which the “limits”  $Z_{\theta_0}(u, \theta)$  and  $\mathcal{Z}(u, \theta)$  of the random vectors  $\mathbb{Z}_n(u, \theta)$  under  $H_0$  and under  $H_1$ , respectively, have to satisfy.

[N] Under  $H_0$ , there exists a sequence of diagonal matrices  $Q_n$  such that

$$\sup_{\theta \in \Theta} \|Q_n^{-2} \mathbb{Z}_n(1, \theta) - Z_{\theta_0}(1, \theta)\| \xrightarrow{p} 0, \quad (3)$$

where the limits  $Z_{\theta_0}(1, \theta)$ 's satisfy

$$\inf_{\theta: \|\theta - \theta_0\| > \varepsilon} \|Z_{\theta_0}(1, \theta)\| > 0 = \|Z_{\theta_0}(1, \theta_0)\|, \quad \text{almost surely, } \forall \varepsilon > 0. \quad (4)$$

[A] Under  $H_1$ , there exists a sequence of diagonal matrices  $Q_n$  such that

$$\sup_{u \in [0, 1]} \sup_{\theta \in \Theta} \|Q_n^{-2} \mathbb{Z}_n(u, \theta) - \mathcal{Z}(u, \theta)\| \xrightarrow{p} 0, \quad (5)$$

and there exists a  $\Theta$ -valued random vector  $\theta_*$  such that

$$\inf_{\theta: \|\theta - \theta_*\| > \varepsilon} \|\mathcal{Z}(1, \theta)\| > 0 = \|\mathcal{Z}(1, \theta_*)\|, \quad \text{almost surely, } \forall \varepsilon > 0, \quad (6)$$

and such that

$$\sup_{u \in (0, 1)} \|\mathcal{Z}(u, \theta_*)\| > 0, \quad \text{almost surely.} \quad (7)$$

Assuming the conditions (3), (4), (5) and (6) is natural; see e.g. Theorems 5.7 and 5.9 of van der Vaart (1998). Let us explain how to check (7) in the most typical form of alternatives in the change problems:

$H'_1$ : there exists a constant  $u_* \in (0, 1)$  such that the true value is  $\theta_0 \in \Theta$  for  $u \in [0, u_*]$ , and  $\theta_1 \in \Theta$  for  $u \in (u_*, 1]$ , where  $\theta_0 \neq \theta_1$ .

In many cases of “ergodic models”, under  $H'_1$ , the condition (5) is satisfied with  $\mathcal{Z}(u, \theta)$  such that

$$\mathcal{Z}(u_*, \theta) = u_* Z_{\theta_0}(1, \theta) \quad \text{and} \quad \mathcal{Z}(1, \theta) = u_* Z_{\theta_0}(1, \theta) + (1 - u_*) Z_{\theta_1}(1, \theta),$$

where not only  $Z_{\theta_0}(1, \theta)$ 's but also  $Z_{\theta_1}(1, \theta)$ 's are assumed to satisfy (4) with trivial change of notation. To see that the condition (7) is satisfied, notice that

$$\mathcal{Z}(u_*, \theta_*) = \mathcal{Z}(u_*, \theta_*) - u_* \mathcal{Z}(1, \theta_*) = u_*(1 - u_*)(Z_{\theta_0}(1, \theta_*) - Z_{\theta_1}(1, \theta_*));$$

if this were zero with positive probability, then it should follow from  $\mathcal{Z}(1, \theta_*) = 0$  that  $Z_{\theta_0}(1, \theta_*) = Z_{\theta_1}(1, \theta_*) = 0$  with positive probability, and this contradicts with (4) and the assumption that  $\theta_0 \neq \theta_1$ . Therefore, we have

$$\begin{aligned} \sup_{u \in (0, 1)} \|\mathcal{Z}(u, \theta_*)\| &\geq \|\mathcal{Z}(u_*, \theta_*)\| \\ &= u_*(1 - u_*) \|Z_{\theta_0}(\theta_*) - Z_{\theta_1}(\theta_*)\| > 0, \quad \text{almost surely.} \end{aligned}$$

This positive value is closely related to the power of our test under  $H'_1$ .

Now, we prepare a lemma to prove the consistency of a sequence of  $Z$ -estimators. This lemma can be proved exactly in the same way as Theorems 5.7 and 5.9 of van der Vaart (1998), so the proof is omitted.

**Lemma 2.1** (i) Under  $[N]$ , for any sequence of  $\Theta$ -valued random vectors  $\widehat{\theta}_n$  such that  $\|\mathbb{Z}_n(1, \widehat{\theta}_n)\| = o_P(1)$ , it holds that  $\widehat{\theta}_n \rightarrow^P \theta_0$ .

(ii) Under  $[A]$ , for any sequence of  $\Theta$ -valued random vectors  $\widehat{\theta}_n$  such that  $\|\mathbb{Z}_n(1, \widehat{\theta}_n)\| = o_P(1)$ , it holds that  $\widehat{\theta}_n \rightarrow^P \theta_*$ .

Let us turn to stating our main results of this paper. Let  $\widehat{\theta}_n$  be any sequence of  $\Theta$ -valued random vectors such that  $\|Q_n^{-1}\mathbb{Z}_n(1, \widehat{\theta}_n)\| = o_P(1)$  under  $H_0$  and  $H_1$ . Let  $u \rightsquigarrow V(u, \theta_0)$  be a non-negative definite matrix valued random process such that  $V(1, \theta_0)$  is positive definite almost surely, and let  $u \rightsquigarrow B(u)$  be a vector of independent standard Brownian motions; (In general we do not assume that  $u \rightsquigarrow V(u, \theta_0)$  and  $u \rightsquigarrow B(u)$  are independent). Let  $u \rightsquigarrow \widehat{V}_n(u)$  be any sequence of matrix valued random processes, which are regular except for  $u = 0$  almost surely, and it should be a uniformly consistent sequence of estimators for the non-negative definite matrix valued random process  $u \rightsquigarrow V(u, \theta_0)$ . Introduce the test statistic

$$\mathcal{T}_n = \sup_{u \in (0,1]} (Q_n^{-1}\mathbb{Z}_n(u, \widehat{\theta}_n))^\top (u\widehat{V}_n(u)^{-1})Q_n^{-1}\mathbb{Z}_n(u, \widehat{\theta}_n).$$

**Theorem 2.2** Under  $[N]$  with a sequence of diagonal matrices  $Q_n$  in the above set up, suppose that there exist a sequence of diagonal matrices  $R_n$  and a sequence of matrix valued random processes  $u \rightsquigarrow V_n(u, \theta_0)$  such that  $V_n(1, \theta_0)$ 's are regular almost surely and that for any sequence of  $\Theta$ -valued random vectors  $\widetilde{\theta}_n(u)$  indexed by  $u \in [0, 1]$  satisfying  $\sup_{u \in [0,1]} \|\widetilde{\theta}_n(u) - \theta_0\| \rightarrow^P 0$ ,

$$\sup_{u \in [0,1]} \|Q_n^{-1}\dot{\mathbb{Z}}_n(u, \widetilde{\theta}_n(u))R_n^{-1} - (-V_n(u, \theta_0))\| \rightarrow^P 0. \quad (8)$$

Suppose also that

$$(Q_n^{-1}\mathbb{Z}_n(u, \theta_0), V_n(u, \theta_0)) \rightarrow^d ((u^{-1}V(u, \theta_0))^{1/2}B(u), V(u, \theta_0)), \quad \text{in } D[0, 1], \quad (9)$$

where  $u \rightsquigarrow B(u)$  is a  $d$ -dimensional standard Brownian motion, and the value of the first vector of the limit in (9) at  $u = 0$  should be read as zero.

If  $\sup_{u \in [0,1]} \|\widehat{V}_n(u) - V(u, \theta_0)\| \rightarrow^P 0$ , then it holds that

$$\mathcal{T}_n \rightarrow^d \sup_{u \in [0,1]} \|B(u) - u^{1/2}V(u, \theta_0)^{1/2}V(1, \theta_0)^{-1/2}B(1)\|^2. \quad (10)$$

**Remark.** If  $V(u, \theta_0) = uV(1, \theta_0)$  for every  $u \in [0, 1]$ , then the test is asymptotically distribution free. In this case the limit is reduced to  $\sup_{u \in [0,1]} \|B^\circ(u)\|^2$  where  $u \rightsquigarrow B^\circ(u) = B(u) - uB(1)$  is a vector of independent standard Brownian bridges.

In the general case, if  $u \rightsquigarrow V(u, \theta_0)$  and  $u \rightsquigarrow B(u)$  are independent, then the limit in (10) is approximated by

$$\sup_{u \in [0,1]} \|B(u) - u^{1/2} \widehat{V}_n(u)^{1/2} \widehat{V}_n(1)^{-1/2} B(1)\|^2,$$

whose approximate distribution can be computed by some computer simulations for the standard Brownian motions  $u \rightsquigarrow B(u)$ .

The following theorem establishes the consistency of the test.

**Theorem 2.3** *Under [A], it holds for any random point  $\check{u}$  in  $(0, 1)$  that*

$$\mathcal{T}_n \geq \lambda(\check{u} Q_n \widehat{V}_n(\check{u})^{-1} Q_n) \{ \|\mathcal{Z}(\check{u}, \theta_*)\|^2 + o_P(1) \},$$

where  $\lambda(A)$  denotes the smallest eigenvalue of the random matrix  $A$ . Hence, if there exists a random point  $\check{u}$  in  $(0, 1)$  such that  $\|\mathcal{Z}(\check{u}, \theta_*)\| > 0$  almost surely and that  $\lambda(Q_n(\widehat{V}_n(\check{u})^{-1} Q_n))$  tends to  $\infty$  in probability, then the test is consistent.

*Proof of Theorem 2.2.* By Lemma 2.1 (i) we know that  $\widehat{\theta}_n$  is a consistent estimator for  $\theta_0$  under  $H_0$ . It follows from the Taylor expansion that

$$Q_n^{-1} \mathbb{Z}_n(1, \widehat{\theta}_n) = Q_n^{-1} \mathbb{Z}_n(1, \theta_0) + (Q_n^{-1} \dot{\mathbb{Z}}_n(1, \widetilde{\theta}_n(1)) R_n^{-1}) R_n (\widehat{\theta}_n - \theta_0). \quad (11)$$

So we have

$$R_n (\widehat{\theta}_n - \theta_0) = A_n + B_n R_n (\widehat{\theta}_n - \theta_0), \quad (12)$$

where

$$\begin{aligned} A_n &= V_n(1, \theta_0)^{-1} Q_n^{-1} (\mathbb{Z}_n(1, \theta_0) - \mathbb{Z}_n(1, \widehat{\theta}_n)), \\ B_n &= V_n(1, \theta_0)^{-1} (Q_n^{-1} \dot{\mathbb{Z}}_n(1, \widetilde{\theta}_n(1)) R_n^{-1} + V_n(1, \theta_0)), \end{aligned}$$

and  $\widetilde{\theta}_n$  is a random vector on the segment connecting  $\theta_0$  and  $\widehat{\theta}_n$ . It follows from the extended continuous mapping theorem (e.g., Theorem 1.11.1 of van der Vaart and Wellner (1996)) that  $V_n(u, \theta_0)^{-1} \rightarrow^p V(u, \theta_0)^{-1}$ , thus we have  $\|A_n\| = O_P(1)$  and  $\|B_n\| = o_P(1)$ . It therefore holds that

$$\|R_n (\widehat{\theta}_n - \theta_0)\| \leq O_P(1) + o_P(1) \cdot \|R_n (\widehat{\theta}_n - \theta_0)\|,$$

which implies that  $\|R_n (\widehat{\theta}_n - \theta_0)\| = O_P(1)$ . Hence, going back to (12) we obtain

$$R_n (\widehat{\theta}_n - \theta_0) = A_n + o_P(1) = V_n(1, \theta_0)^{-1} Q_n^{-1} \mathbb{Z}_n(1, \theta_0) + o_P(1).$$

Now we can write

$$\begin{aligned} & Q_n^{-1} \mathbb{Z}_n(u, \widehat{\theta}_n) \\ &= Q_n^{-1} \mathbb{Z}_n(u, \theta_0) + (Q_n^{-1} \dot{\mathbb{Z}}_n(u, \widetilde{\theta}_n(u)) R_n^{-1}) R_n (\widehat{\theta}_n - \theta_0) \\ &= Q_n^{-1} \mathbb{Z}_n(u, \theta_0) - V_n(u, \theta_0) V_n(1, \theta_0)^{-1} Q_n^{-1} \mathbb{Z}_n(1, \theta_0) + e_n(u) \\ &\rightarrow^d (u^{-1} V(u, \theta_0))^{1/2} (B(u) - u^{1/2} V(u, \theta_0)^{1/2} V(1, \theta_0)^{-1/2} B(1)), \quad \text{in } D[0, 1], \end{aligned}$$

where  $\tilde{\theta}_n(u)$  is a random vector on the segment connecting  $\theta_0$  and  $\hat{\theta}_n$ , and the reminder terms  $e_n(u)$  appearing above satisfy that  $\sup_{u \in [0,1]} \|e_n(u)\| \rightarrow^P 0$ . Hence the claim of the theorem follows from the continuous mapping theorem.  $\square$

*Proof of Theorem 2.3.* The inequality is proved as follows:

$$\begin{aligned} \mathcal{T}_n &\geq (Q_n^{-1} \mathbb{Z}_n(\check{u}, \hat{\theta}_n))^\top (\check{u} \hat{V}_n(\check{u})^{-1}) Q_n^{-1} \mathbb{Z}_n(\check{u}, \hat{\theta}_n) \\ &= Q_n^{-2} \mathbb{Z}_n(\check{u}, \hat{\theta}_n)^\top (\check{u} Q_n \hat{V}_n(\check{u})^{-1} Q_n) Q_n^{-2} \mathbb{Z}_n(\check{u}, \hat{\theta}_n) \\ &\geq \lambda(\check{u} Q_n \hat{V}_n(\check{u})^{-1} Q_n) \|Q_n^{-2} \mathbb{Z}_n(\check{u}, \hat{\theta}_n)\|^2 \\ &= \lambda(\check{u} Q_n (\hat{V}_n(\check{u})^{-1}) Q_n) \{ \|\mathcal{Z}(\check{u}, \theta_*)\|^2 + o_P(1) \}. \end{aligned}$$

The proof is finished.  $\square$

### 3 Examples

In this section we show how the results of the previous section can be applied to some ergodic and non ergodic models. Applications of the  $Z$ -process methods to test if there is a parameter change in some diffusion models and in Cox's regression model are presented and some numerical studies are also discussed.

#### Example A: Ergodic diffusion process

Recall the description of Example A in Section 1.1, where the first  $d_A$ -components  $\alpha$  of the parameter  $\theta = (\alpha^\top, \beta^\top)^\top$  is involved in the drift coefficient, and the latter  $d_B$ -components  $\beta$  is in the diffusion coefficient. Recalling also the definition of the rate matrix  $R_n$  there, let us consider the  $(d_A + d_B)$ -dimensional random vectors  $\mathbb{Z}_n(u, \theta) = \dot{\mathbb{M}}_n(u, \theta)$  and the  $(d_A + d_B) \times (d_A + d_B)$ -random matrices  $\dot{\mathbb{Z}}_n(u, \theta) = \ddot{\mathbb{M}}_n(u, \theta)$  given by the trivial notations as follows:

$$\begin{aligned} \dot{\mathbb{M}}_n(u, \theta) &= (\dot{\mathbb{M}}_n^A(u, \theta)^\top, \dot{\mathbb{M}}_n^B(u, \theta)^\top)^\top, \\ \ddot{\mathbb{M}}_n(u, \theta) &= \begin{pmatrix} \ddot{\mathbb{M}}_n^A(u, \theta) & \ddot{\mathbb{M}}_n^C(u, \theta) \\ \ddot{\mathbb{M}}_n^C(u, \theta)^\top & \ddot{\mathbb{M}}_n^B(u, \theta) \end{pmatrix}. \end{aligned}$$

Under some regularity conditions which are usually assumed in the asymptotic theory for ergodic diffusion process models, it is standard to show the following facts (see e.g. the appendix of Kessler (1997) for some techniques needed to prove them; or Nishiyama (2011), in Japanese, for the detailed proofs that are omitted in Kessler's (1997) appendix):

$$\begin{aligned} \sup_{u \in [0,1]} \left\| \frac{1}{t_n} \dot{\mathbb{M}}_n^A(u, \theta_0) - \frac{1}{t_n} \sum_{k: t_{k-1}^n \leq ut_n} \frac{\dot{S}(X_{t_{k-1}^n}; \alpha_0)}{\sigma(X_{t_{k-1}^n}; \beta_0)} (W_{t_k^n} - W_{t_{k-1}^n}) \right\| &= o_P((t_n)^{-1/2}), \\ \sup_{u \in [0,1]} \left\| \frac{1}{n} \ddot{\mathbb{M}}_n^B(u, \theta_0) - \frac{1}{n} \sum_{k: t_{k-1}^n \leq ut_n} \frac{\dot{\sigma}(X_{t_{k-1}^n}; \beta_0)}{\sigma(X_{t_{k-1}^n}; \beta_0)} \left\{ \frac{|W_{t_k^n} - W_{t_{k-1}^n}|^2}{|t_k^n - t_{k-1}^n|} - 1 \right\} \right\| &= o_P(n^{-1/2}), \end{aligned}$$

$$\begin{aligned} \sup_{u \in [0,1]} \sup_{\theta \in \Theta} \left\| \frac{1}{t_n^n} \ddot{M}_n^A(u, \theta) - \frac{1}{t_n^n} \sum_{k: t_{k-1}^n \leq ut_n^n} H^A(X_{t_{k-1}^n}; \theta_0, \theta) |t_k^n - t_{k-1}^n| \right\| &= o_P((t_n^n)^{-1/2}), \\ \sup_{u \in [0,1]} \sup_{\theta \in \Theta} \left\| \frac{1}{n} \ddot{M}_n^B(u, \theta) - \frac{1}{n} \sum_{k: t_{k-1}^n \leq ut_n^n} H^B(X_{t_{k-1}^n}; \theta_0, \theta) \right\| &= o_P(n^{-1/2}), \\ \sup_{u \in [0,1]} \sup_{\theta \in \Theta} \left\| \frac{1}{\sqrt{t_n^n n}} \ddot{M}_n^C(u, \theta) \right\| &= o_P(n^{-1/4}), \end{aligned}$$

where

$$\begin{aligned} H^A(x; \theta_0, \theta) &= \frac{\ddot{S}(x; \alpha)(S(x; \alpha_0) - S(x; \alpha)) - \dot{S}(x; \alpha)\dot{S}(x; \alpha)^\top}{\sigma(x; \beta)^2}, \\ H^B(x; \theta_0, \theta) &= \left\{ \frac{\ddot{\sigma}(x; \beta)}{\sigma(x; \beta)^3} - 3 \frac{\dot{\sigma}(x; \beta)\dot{\sigma}(x; \beta)^\top}{\sigma(x; \beta)^4} \right\} (\sigma(x; \beta_0)^2 - \sigma(x; \beta)^2) \\ &\quad - 2 \frac{\dot{\sigma}(x; \beta)\dot{\sigma}(x; \beta)^\top}{\sigma(x; \beta)^2}. \end{aligned}$$

Under some standard conditions on the parametric family for the drift and diffusion coefficients in the context of ergodic diffusion processes, we can show that the condition (3) under  $H_0$  is satisfied with  $\dot{M}_{\theta_0}(1, \theta) = (\dot{M}_{\theta_0}^A(1, \theta)^\top, \dot{M}_{\theta_0}^B(1, \theta)^\top)^\top$ , where

$$\dot{M}_{\theta_0}^A(1, \theta) = \int_I \frac{\dot{S}(x; \alpha)}{\sigma(x; \beta)} (S(x; \alpha_0) - S(x; \alpha)) \mu_{\theta_0}(dx)$$

and

$$\dot{M}_{\theta_0}^B(1, \theta) = \int_I \frac{\dot{\sigma}(x; \beta)}{\sigma(x; \beta)^3} (\sigma(x; \beta_0)^2 - \sigma(x; \beta)^2) \mu_{\theta_0}(dx),$$

and that the condition (5) under  $H_1'$  is satisfied with

$$\mathcal{Z}(u, \theta) = (u \wedge u_*) \dot{M}_{\theta_0}(1, \theta) + ((u - u_*) \vee 0) \dot{M}_{\theta_1}(1, \theta).$$

As stated there, the condition (7) is automatically satisfied as soon as the natural conditions (4) and (6) are satisfied.

Using the facts which we saw at the beginning of this section and the usual martingale central limit theorem, we can see that the condition (8) and (9) hold for

$$V_n(u, \theta_0) = \begin{pmatrix} V_n^A(u, \theta_0) & 0 \\ 0 & V_n^B(u, \theta_0) \end{pmatrix},$$

where

$$\begin{aligned} V_n^A(u, \theta_0) &= \frac{1}{t_n^n} \sum_{k: t_{k-1}^n \leq ut_n^n} \frac{\dot{S}(X_{t_{k-1}^n}; \alpha_0)\dot{S}(X_{t_{k-1}^n}; \alpha_0)^\top}{\sigma(X_{t_{k-1}^n}; \beta_0)^2} |t_k^n - t_{k-1}^n|, \\ V_n^B(u, \theta_0) &= \frac{2}{n} \sum_{k: t_{k-1}^n \leq ut_n^n} \frac{\dot{\sigma}(X_{t_{k-1}^n}; \beta_0)\dot{\sigma}(X_{t_{k-1}^n}; \beta_0)^\top}{\sigma(X_{t_{k-1}^n}; \beta_0)^2}. \end{aligned}$$

The limit of  $V_n(u, \theta_0)$  is  $V(u, \theta_0) = uI_{\theta_0}(\theta_0)$ , where

$$I_{\theta_0}(\theta) = \begin{pmatrix} I_{\theta_0}^A(\theta) & 0 \\ 0 & I_{\theta_0}^B(\theta) \end{pmatrix}$$

with

$$I_{\theta_0}^A(\theta) = \int_I \frac{\dot{S}(x; \alpha)\dot{S}(x; \alpha)^\top}{\sigma(x; \beta)^2} \mu_{\theta_0}(dx), \quad I_{\theta_0}^B(\theta) = 2 \int_I \frac{\dot{\sigma}(x; \beta)\dot{\sigma}(x; \beta)^\top}{\sigma(x; \beta)^2} \mu_{\theta_0}(dx).$$

We suppose that  $I_{\theta_0}(\theta)$ 's are positive definite.

As a consistent estimator  $\widehat{V}_n(u)$  for  $V(u, \theta_0)$ , we introduce

$$\widehat{V}_n(u) = \begin{pmatrix} u\widehat{I}_n^A & 0 \\ 0 & u\widehat{I}_n^B \end{pmatrix},$$

where

$$\widehat{I}_n^A = \frac{1}{n} \sum_{k=1}^n \frac{\dot{S}(X_{t_{k-1}^n}; \widehat{\alpha}_n)\dot{S}(X_{t_{k-1}^n}; \widehat{\alpha}_n)^\top}{\sigma(X_{t_{k-1}^n}; \widehat{\beta}_n)^2}, \quad \widehat{I}_n^B = \frac{2}{n} \sum_{k=1}^n \frac{\dot{\sigma}(X_{t_{k-1}^n}; \widehat{\beta}_n)\dot{\sigma}(X_{t_{k-1}^n}; \widehat{\beta}_n)^\top}{\sigma(X_{t_{k-1}^n}; \widehat{\beta}_n)^2}.$$

Finally, it is clear that under  $H_1'$ ,  $\lambda(Q_n \widehat{V}_n(u_*) Q_n)$  tends to  $\infty$  in probability since the matrix  $u_* I_{\theta_0}(\theta_*) + (1 - u_*) I_{\theta_1}(\theta_*)$ , which is the limit of  $\widehat{V}_n(u_*)$ , is positive definite. Hence the test is consistent.

### 3.1 Numerical study for Example A

In this section we observe finite sample performance of our test statistic through numerical experiments. Here, we adopt the Ornstein-Uhlenbeck process starting from  $x_0 = 0$  for the true (data-generating) process:

$$X_t = x_0 - \int_0^t \alpha X_s ds + \beta W_t, \quad t \in [0, T]. \quad (13)$$

For simplicity, we shall treat the equidistant sampling case, that is,  $\Delta_n = |t_k^n - t_{k-1}^n|$  for every  $k = 1, \dots, n$ .

We are going to observe the trajectory of the process (13) for different time horizons  $t_n^n = T$ , and the number  $n$  of observations for each trajectory is such that  $t_n^n = n^{1/3}$ , so  $\Delta_n = n^{-2/3}$ . For this process (13) the estimators for the parameters  $\alpha$  and  $\beta$  and the estimator of the information matrix can be explicitly calculated, and thus the test statistic can be easily computed. For any fixed level  $\varepsilon > 0$  the critical value  $c_\varepsilon$  is given by

$$P \left( \sup_{u \in [0, 1]} \sum_{i=1}^{d_A + d_B} |B^{\circ, (i)}(u)|^2 > c_\varepsilon \right) = \varepsilon.$$

Table 1 of Lee *et al.* (2003) gives a table of the critical values for the significance levels  $\varepsilon = 0.01, 0.05, 0.10$  and for different values of the dimension  $d = d_A + d_B$  computed by Monte Carlo simulation for the limit distribution. Throughout we take the significance level to be  $\varepsilon = 0.05$ . For two parameters ( $d = 2$ ) the critical value is  $c_\varepsilon = 2.408$ . Regarding the null hypothesis we generate  $M = 10^4$  trajectory of (13) and we evaluate the empirical size. The results are reported in Table 1. We observe that: the empirical size gains along with increasing terminal time  $T = t_n^n$ , attaining at 0.05, but also for small terminal  $T$ . In the second example reported in Table 1, the values of the parameter are the maximum likelihood estimate for the mostly federal funds data 1963-1998 in Aït-Sahalia (1999).

$T$	5	10	15	20	25
$n$	125	1000	3375	8000	15625
$\alpha_0 = 1, \beta_0 = 1$	0.044	0.054	0.050	0.052	0.053
$\alpha_0 = 0.25, \beta_0 = 0.02$	0.047	0.061	0.058	0.064	0.054

Table 1: Empirical size based on  $M = 10^4$  independent statistics, for different time horizons.

Regarding the alternative hypothesis we study the behavior of the test statistic in three different situations and for different changing points  $u_*T$  of the parameters, as follows:

- The drift coefficient changes from  $\alpha_0$  to  $\alpha_1$ , but the diffusion coefficient does not change.
- The drift coefficient does not change, but the diffusion coefficient changes from  $\beta_0$  to  $\beta_1$ .
- Both coefficients change.

For each of the above scenarios we consider the following change points,  $u_* = \frac{1}{2}, \frac{3}{4}, \frac{9}{10}$ .

The first scenario is the worst case for the diffusion, and in order to detect a change in the drift we have to observe the process as long as possible. Table 2 shows empirical power for different terminal times  $T$  and different change points  $u_*T$ . The values of the parameters are  $\alpha_0 = 0.25, \alpha_1 = 0.50$  and  $\beta = 0.02$ . The last does not vary. We simulate  $10^4$  independent copies of a trajectory of (13) to obtain different values of  $\mathcal{T}_n$ . The power increase as  $T$  increase and the performance is better when we can observe the process after the change for long time (the case  $u_* = \frac{1}{2}$ ). In such a case the power of the test is reasonable. In the worst case  $u_* = \frac{9}{10}$ , the test is not able to detect the change in the drift coefficient.

$T$	5	10	15	20	25
$n$	125	1000	3375	8000	15625
$u_* = \frac{1}{2}$	0.31	0.52	0.73	0.79	0.88
$u_* = \frac{3}{4}$	0.12	0.17	0.23	0.26	0.35
$u_* = \frac{9}{10}$	0.05	0.07	0.08	0.08	0.09

Table 2: Empirical power based on  $M = 10^4$  independent statistics. Here the significance level is 0.05. The values of the parameter are  $\alpha_0 = 0.25$ ,  $\alpha_1 = 0.50$  and  $\beta = 0.02$  (it does not vary).

Table 3 reports the results for simulation when only the drift changes, but the change is bigger. With  $\alpha_0 = 0.25$ ,  $\alpha_1 = 1.25$  and  $\beta = 0.02$ , the power increases not only for  $u_* = \frac{1}{2}$  but also for  $u_* = \frac{3}{4}$ . This was expected, but as in the previous example the performance of the test is not good when the chance of the parameter occur at the end of the observation window.

$T$	5	10	15	20	25
$n$	125	1000	3375	8000	15625
$u_* = \frac{1}{2}$	0.35	0.60	0.78	0.88	0.94
$u_* = \frac{3}{4}$	0.13	0.20	0.28	0.31	0.38
$u_* = \frac{9}{10}$	0.06	0.08	0.09	0.11	0.11

Table 3: Empirical power based on  $M = 10^4$  independent statistics. Here the significance level is 0.05. The values of the parameter are  $\alpha_0 = 0.25$ ,  $\alpha_1 = 1.25$  and  $\beta = 0.02$  (it does not vary).

Table 4 shows the empirical power for different terminal times  $T$  and different change points  $u_*$  in the second scenario:  $\alpha = 0.25$  does not vary, but  $\beta_0 = 0.02$  changes and becomes  $\beta_1 = 0.03$ . The power of the test is very good, but this is not surprising because a change in the diffusion coefficient can be easily detected. The situation reported in Table 5 is the same: also for very small change in the



diffusion coefficient, the performance of the test is very good with the empirical power that reaches the value 1 also for small  $T$ . We do not report the results for the third scenario where the drift changes at the same instant of the diffusion, because the performance of the test is the same as in the second scenario. This is not surprising and is due to the different rates of convergence of the estimators of the two parameters.

$T$	5	10	15	20	25
$n$	125	1000	3375	8000	15625
$u_* = \frac{1}{2}$	0.99	1	1	1	1
$u_* = \frac{3}{4}$	0.86	1	1	1	1
$u_* = \frac{9}{10}$	0.36	0.99	1	1	1

Table 4: Empirical power based on  $M = 10^4$  independent statistics. Here the significance level is 0.05. The values of the parameter are  $\alpha = 0.25$  (it does not vary),  $\beta_0 = 0.02$  and  $\beta_1 = 0.03$ .

$T$	5	10	15	20	25
$n$	125	1000	3375	8000	15625
$u_* = \frac{1}{2}$	0.87	1	1	1	1
$u_* = \frac{3}{4}$	0.52	0.99	1	1	1
$u_* = \frac{9}{10}$	0.14	0.62	0.99	1	1

Table 5: Empirical power based on  $M = 10^4$  independent statistics. Here the significance level is 0.05. The values of the parameter are  $\alpha = 0.25$  (it does not vary),  $\beta_0 = 0.020$  and  $\beta_1 = 0.025$ .

## Example B: Volatility of diffusion process

Recall the description of Example B in Section 1.1. An interesting point of this example is that the limit of  $-\tilde{M}_n(u, \tilde{\theta}_n(u))$  is random and depend on  $u \in [0, 1]$  in

a complex way.

Under some regularity conditions, it holds that

$$\sup_{u \in [0,1]} \left\| \dot{\mathbb{M}}_n(u, \theta_0) - \frac{1}{n} \sum_{k: t_{k-1}^n \leq ut_n^n} \frac{\dot{\sigma}(X_{t_{k-1}^n}; \theta_0)}{\sigma(X_{t_{k-1}^n}; \theta_0)} \left\{ \frac{|W_{t_k^n} - W_{t_{k-1}^n}|^2}{|t_k^n - t_{k-1}^n|} - 1 \right\} \right\| = o_P(n^{-1/2}),$$

$$\sup_{u \in [0,1]} \sup_{\theta \in \Theta} \left\| \ddot{\mathbb{M}}_n(u, \theta) - \frac{1}{n} \sum_{k: t_{k-1}^n \leq ut_n^n} H(X_{t_{k-1}^n}; \theta_0, \theta) \right\| = o_P(n^{-1/2}),$$

where

$$H(x; \theta_0, \theta) = \left\{ \frac{\ddot{\sigma}(x; \theta)}{\sigma(x; \theta)^3} - 3 \frac{\dot{\sigma}(x; \theta) \dot{\sigma}(x; \theta)^\top}{\sigma(x; \theta)^4} \right\} (\sigma(x; \theta_0)^2 - \sigma(x; \theta)^2) - 2 \frac{\dot{\sigma}(x; \theta) \dot{\sigma}(x; \theta)^\top}{\sigma(x; \theta)^2}.$$

Moreover, under some standard conditions, it holds that for any sequence of random vectors  $\tilde{\theta}_n(u)$  indexed by  $u \in [0, 1]$  such that  $\sup_{u \in [0,1]} \|\tilde{\theta}_n(u) - \theta_0\| \rightarrow^p 0$ ,

$$\sup_{u \in [0,1]} \left\| \frac{1}{n} \ddot{\mathbb{M}}_n(u, \tilde{\theta}_n(u)) - (-V_n(u, \theta_0)) \right\| \rightarrow^p 0,$$

where

$$V_n(u, \theta_0) = \frac{2}{n} \sum_{k: t_{k-1}^n \leq ut_n^n} \frac{\dot{\sigma}(X_{t_{k-1}^n}; \theta_0) \dot{\sigma}(X_{t_{k-1}^n}; \theta_0)^\top}{\sigma(X_{t_{k-1}^n}; \theta_0)^2}, \quad \forall u \in [0, 1].$$

Also, it follows from the well known theory of martingales that

$$(\sqrt{n} \dot{\mathbb{M}}_n(u, \theta_0), V_n(u, \theta_0)) \rightarrow^d ((u^{-1} V(u, \theta_0))^{1/2} B(u), V(u, \theta_0)) \quad \text{in } D[0, 1],$$

where  $u \rightsquigarrow B(u)$  is a vector of independent standard Brownian motions which is independent of the matrix valued random process  $u \rightsquigarrow V(u, \theta_0)$  given by

$$V(u, \theta_0) = 2 \int_0^{uT} \frac{\dot{\sigma}(X_s; \theta_0) \dot{\sigma}(X_s; \theta_0)^\top}{\sigma(X_s; \theta_0)^2} ds, \quad \forall u \in [0, 1].$$

Under some standard conditions on the parametric family for the diffusion coefficient, we can show that the condition (3) under  $H_0$  is satisfied with

$$\dot{M}_{\theta_0}(1, \theta) = \int_0^{uT} \frac{\dot{\sigma}(X_t; \theta)}{\sigma(X_t; \theta)^3} (\sigma(X_t; \theta_0)^2 - \sigma(X_t; \theta)^2) dt.$$

Under  $H'_1$ , we have

$$\sup_{u \in [0,1]} \left\| \frac{1}{n} \dot{\mathbb{M}}_n(u, \theta) - \mathcal{Z}(u, \theta) \right\| \rightarrow^p 0,$$

where

$$\begin{aligned}\mathcal{Z}(u, \theta) &= \int_0^{(u \wedge u_*)T} \frac{\dot{\sigma}(X_t; \theta)}{\sigma(X_t; \theta)^3} (\sigma(X_t; \theta_0)^2 - \sigma(X_t; \theta)^2) dt \\ &\quad + 1\{u > u_*\} \int_{u_*T}^{uT} \frac{\dot{\sigma}(X_t; \theta)}{\sigma(X_t; \theta)^3} (\sigma(X_t; \theta_1)^2 - \sigma(X_t; \theta)^2) dt.\end{aligned}$$

We can give a set of sufficient conditions for (7) as follows. Suppose that the Lebesgue measure of the random set  $T_0 = \{t \in [0, T] : \dot{\sigma}(X_t; \theta_*)/\sigma(X_t; \theta_*)^3 = 0\}$  is zero almost surely, which is true in many concrete models. In this case, replace the values  $\dot{\sigma}(X_t; \theta_*)/\sigma(X_t; \theta_*)^3$  in the definition of  $\mathcal{Z}(u, \theta_*)$  on the set  $T_0$  by 1 to construct  $\mathcal{Z}(u, \theta_*)^*$  which equals with the original  $\mathcal{Z}(u, \theta_*)$  for all  $u \in [0, 1]$ , almost surely. If we further assume that for any non-empty interval  $J \subset I$

$$\sigma(x; \theta) = \sigma(x; \theta'), \quad \forall x \in J \quad \iff \quad \theta = \theta'$$

and that the each of the random sets  $J_0 = \{X_t : t \in [0, u_*T]\}$  and  $J_1 = \{X_t : t \in (u_*T, T]\}$  includes a non-empty set almost surely, then it follows from the assumption  $\theta_0 \neq \theta_1$  that  $\frac{d}{du}\mathcal{Z}(u, \theta_*)^* \neq 0$  almost surely. Thus we have

$$\sup_{u \in (0,1)} \|\mathcal{Z}(u, \theta_*)\| = \sup_{u \in (0,1)} \|\mathcal{Z}(u, \theta_*)^*\| > 0, \quad \text{almost surely.}$$

Now, consider the matrices

$$V(u, \theta) = 2 \int_0^{uT} \frac{\dot{\sigma}(X_t; \theta)\dot{\sigma}(X_t; \theta)^\top}{\sigma(X_t; \theta)^2} dt, \quad \forall u \in [0, 1].$$

Let us assume that for every  $\theta \in \Theta$  there exists a set  $J_\theta \subset I$  such that the Lebesgue measure of  $J_\theta^c$  is zero and that the matrices  $\dot{\sigma}(x; \theta)\dot{\sigma}(x; \theta)^\top/\sigma(x; \theta)$  are positive definite for  $x \in J_\theta$ , which is a standard assumption. In this case, if the claim that the Lebesgue measure of the set  $\{t \in [0, T]; X_t(\omega) \in J\}$  is positive for any set  $J \subset I$  such that the Lebesgue measure of  $J^c$  is zero holds for almost all  $\omega$ , then  $V(u, \theta_0)$ 's and  $V(u, \theta_*)$ 's for  $u \in (0, 1]$  are positive definite almost surely under  $H_0$  and  $H_1'$ , respectively.

As we saw at the beginning of this section, the conditions (8) and (9) under  $H_0$  are satisfied. A consistent estimator  $\widehat{V}_n(u)$  for  $V(u, \theta_0)$  is given by

$$\widehat{V}_n(u) = \frac{2}{n} \sum_{k: t_{k-1}^n \leq ut_n} \frac{\dot{\sigma}(X_{t_{k-1}^n}; \widehat{\theta}_n)\dot{\sigma}(X_{t_{k-1}^n}; \widehat{\theta}_n)^\top}{\sigma(X_{t_{k-1}^n}; \widehat{\theta}_n)^2}, \quad \forall u \in [0, 1].$$

Finally, it is clear that under  $H_1'$ ,  $n\lambda(\check{u}(\widehat{V}_n(\check{u}))^{-1})$  tends to  $\infty$  in probability, because it follows from what we have assumed that  $\lambda(\check{u}(\widehat{V}_n(\check{u}))^{-1}) \xrightarrow{p} \lambda(\check{u}V(\check{u}, \theta_*)^{-1})$  and the limit is positive almost surely. Hence the test is consistent.

### 3.2 Numerical study for Example B

Also In this section we observe finite sample performance of our test statistic through numerical experiments. The data-generating process is the following:

$$X_t = 4 - \int_0^t (X_s - 4)ds + \int_0^t \exp\left(\theta \frac{X_s^2}{1 + X_s^2}\right) dW_s, \quad t \in [0, 1],$$

where the drift coefficient  $S(x) = -(x - 4)$  is treated as a nuisance function. We simulate  $M = 10^4$  independent copies of this process at the equidistant time grid  $t_k^n = \frac{k}{n}$ ,  $k = 0, 1, \dots, n$ . We can compute the test statistic, estimating

$$V(u, \theta_0) = 2 \int_0^u \left| \frac{X_s^2}{1 + X_s^2} \right|^2 ds$$

by the natural estimator

$$\widehat{V}_n(u) = \frac{2}{n} \sum_{k=1}^{[un]} \left| \frac{X_{t_{k-1}^n}^2}{1 + X_{t_{k-1}^n}^2} \right|^2.$$

The empirical size under  $H_0$  is reported in Table 6, where the true value of the parameter is set as  $\theta_0 = 1.0$  or  $1.5$ .

$n$	20	40	100	200
$\theta_0 = 1.0$	0.013	0.024	0.031	0.034
$\theta_0 = 1.5$	0.028	0.029	0.035	0.036

Table 6: Empirical size based on  $M = 10^4$  independent statistics. Here the significance level is 0.05. The value of the parameter is  $\theta_0 = 1.0$  or  $1.5$ .

The empirical power under  $H_1'$  is reported in Table 7, where the true values of the parameter change from  $\theta_0 = 1.0$  to  $\theta_1 = 1.5$  at time point  $u_* = \frac{1}{2}, \frac{3}{4}$  or  $\frac{9}{10}$ .

We can observe how in both situations the empirical size approaches the theoretical size increasing the numbers observations. We expect that increasing the number of observations (here stopped at  $n = 200$  because of time cost of simulation) the empirical size will stabilize around the theoretical value. Indeed simulations for the model with  $\theta_0 = 1.0$  seem to confirm the theoretical result: with  $M = 1000$  simulated trajectories, for  $n = 1000$  and  $n = 5000$  the empirical size is 0.039 and 0.049 respectively.

In Table 7 the performance of the test under the alternative is checked and we can observe how also in the worst case, where the change of the parameter occur at the end of the window of observation, ( $u_* = \frac{9}{10}$ ) the empirical power approach to 1 as  $n$  increase.

$n$	20	40	100	200
$u_* = \frac{1}{2}$	0.288	0.793	0.974	0.994
$u_* = \frac{3}{4}$	0.457	0.768	0.951	0.979
$u_* = \frac{9}{10}$	0.251	0.466	0.805	0.935

Table 7: Empirical power based on  $M = 10^4$  independent statistics. Here the significance level is 0.05. The values of the parameter change from  $\theta_0 = 1.0$  to  $\theta_1 = 1.5$  at time  $u_* = \frac{1}{2}, \frac{3}{4}$  or  $\frac{9}{10}$ .

### Example C: Cox's regression model

Recall the description of Example C in Section 1.1. Since all the arguments are similar to those in Section 3.1, we state only the key points in the discussion on the change point problem.

Introducing the notations

$$\begin{aligned}
S_t^{n,0}(\theta) &= \sum_{k=1}^n e^{\theta Z_t^k} Y_t^k, \\
S_t^{n,1}(\theta) &= \sum_{k=1}^n Z_t^k e^{\theta Z_t^k} Y_t^k, \\
S_t^{n,2}(\theta) &= \sum_{k=1}^n (Z_t^k)^\top Z_t^k e^{\theta Z_t^k} Y_t^k,
\end{aligned}$$

we suppose that

$$\sup_{\theta \in \Theta} \sup_{t \in [0, T]} \left\| \frac{1}{n} S_t^{n,l}(\theta) - \mathcal{S}_t^l(\theta) \right\| \rightarrow^p 0, \quad l = 0, 1, 2,$$

where the limits  $t \rightsquigarrow \mathcal{S}_t^l$  are some stochastic processes (c.f. Andersen and Gill (1982) who assumed that  $\mathcal{S}^l$ 's are not random).

Then, some arguments similar to Section 3.1 are possible for

$$\begin{aligned}
\dot{M}_{\theta_0}(1, \theta) &= \int_0^T \left( \frac{\mathcal{S}_t^1(\theta_0)}{\mathcal{S}_t^0(\theta_0)} - \frac{\mathcal{S}_t^1(\theta)}{\mathcal{S}_t^0(\theta)} \right) \mathcal{S}_t^0(\theta_0) \alpha(t) dt, \\
\mathcal{Z}(u, \theta) &= \int_0^{(u \wedge u_*)^T} \left( \frac{\mathcal{S}_t^1(\theta_0)}{\mathcal{S}_t^0(\theta_0)} - \frac{\mathcal{S}_t^1(\theta)}{\mathcal{S}_t^0(\theta)} \right) \mathcal{S}_t^0(\theta_0) \alpha(t) dt \\
&\quad + 1\{u > u_*\} \int_{u_*^T}^{u^T} \left( \frac{\mathcal{S}_t^1(\theta_1)}{\mathcal{S}_t^0(\theta_1)} - \frac{\mathcal{S}_t^1(\theta)}{\mathcal{S}_t^0(\theta)} \right) \mathcal{S}_t^0(\theta_1) \alpha(t) dt, \\
V_n(u, \theta_0) &= \frac{1}{n} \int_0^{u^T} \frac{S_t^{n,0}(\theta_0) S_t^{n,2}(\theta_0) - S_t^{n,1}(\theta_0) S_t^{n,1}(\theta_0)^\top}{S_t^{n,0}(\theta_0)} \alpha(t) dt, \\
V(u, \theta) &= \int_0^{(u \wedge u_*)^T} \frac{\mathcal{S}_t^0(\theta) \mathcal{S}_t^2(\theta) - \mathcal{S}_t^1(\theta) \mathcal{S}_t^1(\theta)^\top}{\mathcal{S}_t^0(\theta)^2} \mathcal{S}_t^0(\theta_0) \alpha(t) dt \\
&\quad + 1\{u > u_*\} \int_{u_*^T}^{u^T} \frac{\mathcal{S}_t^0(\theta) \mathcal{S}_t^2(\theta) - \mathcal{S}_t^1(\theta) \mathcal{S}_t^1(\theta)^\top}{\mathcal{S}_t^0(\theta)^2} \mathcal{S}_t^0(\theta_1) \alpha(t) dt, \\
\widehat{V}_n(u) &= \frac{1}{n} \sum_{k=1}^n \int_0^{u^T} \frac{S_t^{n,0}(\widehat{\theta}_n) S_t^{n,2}(\widehat{\theta}_n) - S_t^{n,1}(\widehat{\theta}_n) S_t^{n,1}(\widehat{\theta}_n)^\top}{S_t^{n,0}(\widehat{\theta}_n)^2} dN_t^k.
\end{aligned}$$

The details are omitted.

## Acknowledgements

This work was partially supported by by PRIN 2009JW2STY and University of Bergamo (I.N.) and by Grant-in-Aid for Scientific Research (C), 24540152, from Japan Society for the Promotion of Science (Y.N.).

## References

- [1] Ait-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusion. *J. Finance* **54**, 1361-1395.
- [2] Andersen, P.K. and Gill, R.D. (1982). Cox's regression models for counting processes: A large sample study. *Ann. Statist.* **10**, 1100-1120.
- [3] Cox, D.R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc. B* **34**, 187-220.
- [4] Fujii, T. and Nishiyama, Y. (2011). Testing for parameter change in a stress release process. *Preprint*.
- [5] Horváth, L. and Parzen, E. (1994). Limit theorems for Fisher-score change processes. In: *Change-point Problems*, (Edited by Carlstein, E., Müller H.-G. and Siegmund, D.) IMS Lecture Notes – Monograph Series **23**, 157-169.

- [6] Isham, V. and Westcott, M. (1979). A self-correcting point processes. *Stochastic Process. Appl.* **8**, 335-347.
- [7] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 211-229.
- [8] Lee, S., Ha, J., Na, O. and Na, S. (2003). The cusum test for parameter change in time series models. *Scand. J. Statist.* **30**, 781–796.
- [9] Negri, I. and Nishiyama, Y. (2012). Asymptotically distribution free test for parameter change in a diffusion process model. *Ann. Inst. Statist. Math.* **64**, 911-918.
- [10] Nishiyama, Y. (2011). *Statistical Analysis by the Theory of Martingales*. (In Japanese.) ISM Series **1**, Kindaikagakusha, Tokyo.
- [11] Song, J. and Lee, S. (2009). Test for parameter change in discretely observed diffusion processes. *Statist. Inference Stoch. Process.* **12**, 165-183.
- [12] Tsukuda, K. and Nishiyama, Y. (2014). On  $L^2$  space approach to change point problems. *To appear in J. Statist. Plann. Inference*.
- [13] van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.
- [14] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag, New York.